

Identification of conserved skeletal enhancers associated with craniosynostosis risk genes

Xuan Anita He(何璇)^{1,2}, Anna Berenson^{3,4}, Michelle Bernard^{1,5}, Chris Weber⁶, Laura E. Cook⁷, Axel Visel^{7,8,9}, Juan I. Fuxman Bass³, Shannon Fisher^{1,*}

¹Department of Pharmacology, Physiology & Biophysics, Boston University, 700 Albany St, W607, Boston, MA 02118, United States

²Graduate Program in Biomolecular Medicine, Boston University, 72 East Concord St, Boston, MA 02118, United States

³Department of Biology, Boston University, 5 Cummington Mall, Boston, MA 02215, United States

⁴Program in Molecular Biology, Cell Biology, and Biochemistry, Boston University, 5 Cummington Mall, Boston, MA 02215, United States

⁵College of Arts and Sciences, Boston University, 5 Cummington Mall, Boston, MA 02215, United States

⁶Department of Cell and Developmental Biology, University of Pennsylvania, 421 Curie Boulevard, Philadelphia, PA 19104-6058, United States

⁷Environmental Genomics & System Biology Division, Lawrence Berkeley National Laboratory, 1 Cyclotron Road, Berkeley, CA 94720, United States

⁸U.S. Department of Energy Joint Genome Institute, 1 Cyclotron Road, Berkeley, CA 94720, United States

⁹School of Natural Sciences, 5200 Lake Road, University of California Merced, Merced, CA 95343, United States

*Corresponding author. Department of Pharmacology, Physiology & Biophysics, Boston University, 700 Albany St., W607, Boston, MA 02118, United States.

E-mail: shanfish@bu.edu

Abstract

Craniosynostosis, defined by premature fusion of one or multiple cranial sutures, is a common congenital defect affecting more than 1/2000 infants and results in restricted brain expansion. Single gene mutations account for 15%–20% of cases, largely as part of a syndrome, but the majority are nonsyndromic with complex underlying genetics. We hypothesized that the two noncoding genomic regions identified by a GWAS for craniosynostosis contain distal regulatory elements for the risk genes *BMPER* and *BMP2*. To identify such regulatory elements, we surveyed conserved noncoding sequences from both risk loci for enhancer activity in transgenic *Danio rerio*. We identified enhancers from both regions that direct expression to skeletal tissues, consistent with the endogenous expression of *bmp2* and *bmp2*. For each locus, we also found a skeletal enhancer that also contains a sequence variant associated with craniosynostosis risk. We examined the activity of each enhancer during craniofacial development and found that the *BMPER*-associated enhancer is active in the restricted region of cartilage closely associated with frontal bone initiation. The same enhancer is active in mouse skeletal tissues, demonstrating evolutionarily conserved activity. Using enhanced yeast one-hybrid assays, we identified transcription factors that bind each enhancer and observed differential binding between alleles, implicating multiple signaling pathways. Our findings help unveil the genetic mechanism of the two craniosynostosis risk loci. More broadly, our combined *in vivo* approach is applicable to many complex genetic diseases to build a link between association studies and specific genetic mechanisms.

Keywords: conserved regulatory elements; zebrafish transgenesis; craniosynostosis; enhanced yeast one-hybrid assay; BMP signaling

Introduction

The bone morphogenetic protein (BMP) pathway is subject to complex regulation through a network of extracellular factors [1]. Many of these are required in the development of multiple organ systems and are implicated in human disease, including several skeletal conditions and some cancers [2]. BMP signaling plays a particular role in the formation of the skeletal system, and recent findings have implicated dysregulation of the pathway in the craniofacial defect craniosynostosis (CS), in which one or more cranial sutures fuse prematurely. CS is one of the most common structural birth defects, affecting ~1/2000 infants. Mutations in single genes account for only 15%–20% of CS cases, which are part of syndromes of coincident defects and mostly affect the coronal sutures [3]. The genetic causes of the remaining cases, which largely affect the midline sagittal and metopic sutures, are more complex. An association study focused on nonsyndromic CS (NCS) identified two risk regions, one downstream of *BMP2* (hg38, chr20: 7112785–7245836, and the other in an intron of *BBS9* (hg38, chr7:

33179156–33384149) [4]. There were no coding sequence variants associated with either risk allele, so the causal mutations are presumed to affect noncoding sequences. However, in the absence of functional data, it is difficult to predict which of the several linked single nucleotide polymorphisms (SNPs) in each risk locus might contribute to disease risk or to explore the mechanism further.

BBS9 is one of 14 genes associated with Bardet Biedel syndrome (BBS) and encoding components of functional cilia [5]. While *BMP2* is one of the most osteogenic BMP ligands [6], *BBS9* has no known function in osteogenesis, and CS is not a consistent feature of BBS. However, the adjacent gene encodes BMP binding endothelial regulator (*BMPER*), an extracellular regulator of BMP signaling. Homologous to the fly *crossveinless2*, the vertebrate gene was originally described as a negative regulator of BMP signaling during early endothelial cell differentiation [7]. Later studies supported both pro- and anti-BMP activities [8, 9], suggesting the role of *BMPER* is context-dependent.

Several lines of evidence suggest that BMPER activity is largely pro-BMP in osteogenesis. Upregulation of BMPER promotes BMP2-induced osteogenic differentiation in human bone mesenchymal stem cells [10]. Homozygous null mutations of BMPER cause diaphanospondylodysostosis (DSD), a lethal perinatal skeletal condition characterized by reduced skeletal structures [11, 12]. BMPER mutations that result in a truncated protein cause a less severe form of DSD [13]. *Bmper* null mutant mice display similar skeletal defects [14]. Although the expression of the mammalian gene has not been described during skeletal development, zebrafish *bmper* (previously named *crossveinless 2*) is prominently expressed in cranial bones and cartilages in larvae [8, 15] and adult fish [16]. Taken together, the mammalian mutant phenotypes and the expression of the zebrafish gene in developing bone support a conserved role of BMPER in skeletal development across vertebrates. In addition, it is plausible that the two genes make up a conserved regulatory domain, given 1) the syntenic relationship of BBS9 and BMPER is conserved from mammals to zebrafish [17]; 2) the two genes are in the same topologically associated domains (TADs) [18].

We hypothesized that both regions implicated in CS risk harbored noncoding sequences important in regulating genes in the BMP signaling pathway, BMPER, and BMP2. Given that BBS9 and BMPER are in the same TAD, we also predicted and aimed to identify a broader regulatory domain in the intergenic region, to better understand the role of BMPER as an evolutionarily conserved BMP regulator during skull development. Taken together, our goal is to identify enhancers regulating the expression of both BMP2 and BMPER in skeletal tissues during craniofacial development to clarify the contribution of both genes to the genetic risk for CS. Using an assay for enhancer activity in transgenic zebrafish, we examined candidate sequences in the intergenic region between BBS9 and BMPER, and the identified CS risk loci, including the intronic regions of BBS9 and the region downstream of BMP2. We identified two sequences within the risk locus near BMP2 that regulates expression during craniofacial development. We similarly found multiple enhancers intronic to BBS9 and in the intergenic region with activity consistent with endogenous *bmper* expression, including two that are active in skeletal tissues, and one of which is specifically active near the site of frontal bone initiation. Importantly, from each identified risk locus, one of the active enhancers encompasses a SNP associated with CS risk [4]. Finally, we identified candidate interactions with transcription factors for the skeletal enhancers from both loci through an enhanced yeast one-hybrid (eY1H) assay. Therefore, these TF-enhancer interactions provide a basis for future examination in determining the mechanism underlying the genetic risk of CS associated with each regulatory domain.

Results

Identification of putative enhancers of BMPER and BMP2 and screen results

We first examined a 1.3 million base pair (bp) region (chr7:33,197,564-34,501,768) encompassing BMPER on GRCh38/hg38, to select sequences to test as regulatory elements. This region includes the risk locus encompassing several introns of BBS9 [4] and the intergenic region between BBS9 and BMPER (Fig. 1A–C). We also selected sequences in the risk locus downstream of BMP2 (Fig. 1D), including one containing rs18843302, which was previously suggested to act as an enhancer [19]. The primary criterion for selection was conservation across species, which was quantified by a log odds (LOD) score from PhastCons [20, 21] available on the

UCSC genome browser. We also examined cis-regulatory elements predicted from ENCODE data [22–24] to make the final selections. In total, we selected 69 sequences for transgenic analysis, 51 associated with BMPER and 18 with BMP2. In each region, the sequences are identified by the distance in kilobases relative to the annotated transcription start site of the respective gene, i.e. –117BMPER is 117 kb upstream of the BMPER start site.

For the risk locus within BBS9, only –707BMPER contains a significant SNP (rs10254116*), while –687BMPER, –684BMPER, and –655BMPER are each within 2 kb of a SNP. For the risk locus near BMP2, +421BMP2, i.e. the sequence beginning 421 kb downstream of BMP2 transcription start site, contains a significant SNP (rs6117669); +382BMP2, +402BMP2, +460BMP2, and +463BMP2 are each within 2 kb of a SNP. The previously reported enhancer by Justice et al. [19], +344BMP2 containing a significant SNP (rs18843302), was also included in our assay. Taken together, we hypothesized that, during skull development, one or more selected conserved sequences near BMPER regulate BMPER gene, and those near BMP2 regulate BMP2 gene.

We assessed each candidate element for tissue-specific enhancer activity using zebrafish transgenesis as previously described [25]. Each element was cloned in a Tol2-based vector, upstream of a minimal *cfos* promoter and *egfp* [25]; embryos injected with each construct were screened for mosaic transgene expression at 5dpf when the first cranial bone and cartilage elements have formed. To aid in screening for expression in cranial skeletal tissues, we injected into embryos transgenic for *sp7: mcherry*, expressed in early osteoblasts. Candidate elements that gave rise to tissue-specific activity are summarized in Supplementary Tables S1 and S2. Injected embryos with tissue-specific expression of interest were raised and bred to established transgenic lines. Several tested sequences near BMPER give rise to transgene expression consistent with the endogenous zebrafish *bmper* expression in the skeletal tissues, including pectoral fins, cranial cartilages, and bones [8, 15]. Some also showed transgene expression in the inner ear [8, 26] (Supplementary Table S1).

For some of the tested sequences near BMP2, the enhancer activities consistent with the endogenous *bmp2* genes include pectoral fins, cleithrum [27, 28], otic vesicles [27], and gills [29]. The mosaic transgene expressions driven by +372BMP2 in the notochord and by +382BMP2 in the gills were particularly specific and strong (Supplementary Table S2). A sequence containing the tag reference SNP near BMP2 associated with CS risk, rs1884302, had previously been ascribed an allelic difference assessed for effect on regulatory function via a similar zebrafish assay [19]. We tested the same sequence encompassing the SNP in our assay, comparing the reference and alternate alleles, but we found no tissue-specific enhancer activity above the background for either allele. We further characterized the enhancers with activity primarily in cranial skeletal tissues, –117BMPER and –707BMPER regulating BMPER; +402BMP2 and +421BMP2 regulating BMP2.

–117BMPER is a conserved enhancer active in early osteoblasts

From the intergenic region between BMPER and BBS9, the sequence –117BMPER has a maximum LOD score of 488 in PhastCons and is conserved down to chicken. It also contains an enhancer element predicted by the enhancer-specific histone marks of ENCODE data [22–24] (Fig. 1B). It showed enhancer activity in the cranial bones and cartilage of mosaic fish at 5dpf (Fig. 2A and B). In 5dpf larvae carrying the –117BMPER construct, *egfp* is expressed widely in cranial cartilage and bone (Fig. 2C and D), which is consistent across lines from more

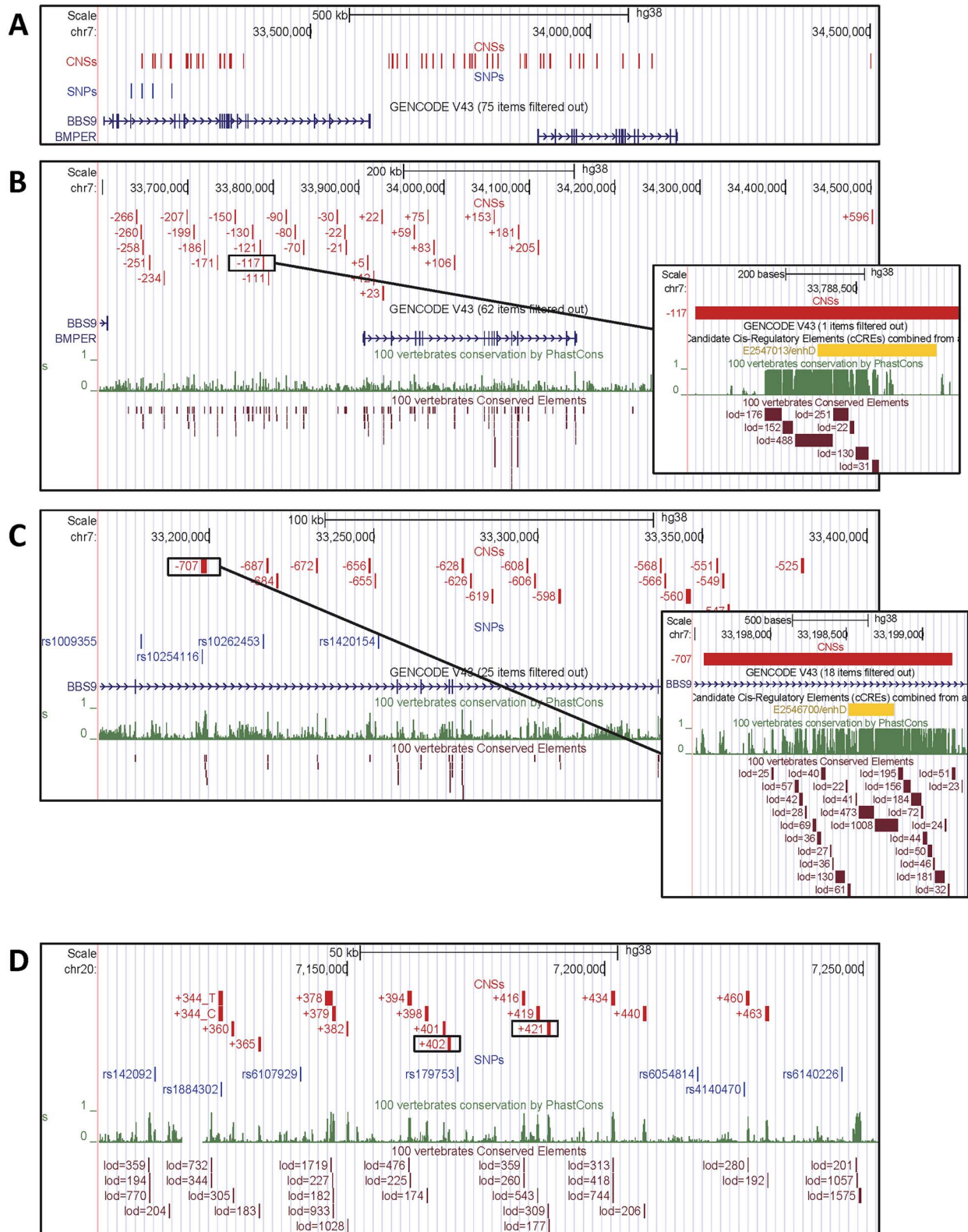


Figure 1. Selection of putative enhancers for *in vivo* testing. All panels are views from the UCSC genome browser, human genome assembly hg38; included are custom tracks indicating conserved sequences cloned for transgenesis (CNSs) and the SNPs linked to CS risk and confirmed after retesting (SNPs). (A) Overview of the region encompassing *BBS9* and *BMPER*, indicating all the tested sequences. (B) The region encompassing *BMPER*, indicating the 32 tested sequences; the inset is a closer view of -117 *BMPER*. (C) The risk locus within *BBS9* [4], showing the 19 tested sequences; the inset is a closer view of -707 *BMPER*. In B and C, the two lower tracks show two measures of multispecies conservation PhastCons and conserved elements, the primary criteria used to select sequences. The insets also include the track with predictions of regulatory elements from ENCODE data. (D) A similar view of the region downstream of *BMP2* (gene is not included in the browser window), indicating the 18 tested sequences and the SNPs associated with CS risk. Boxed sequences are those that tested positive for skeletal-specific enhancer activities in zebrafish transgenesis assay, $+402$ *BMP2* and $+421$ *BMP2*.

than three independent founders. Transgene expression remains prominent in Meckel's and ceratohyal cartilages throughout larval skull development. During this period of growth, the expression is also visible but less prominent at palatoquadrate, basibranchial, ethmoid plate, opercles, and subopercles. Most of these structures are homologous to human mandibles [30], which are areas clinically affected in some patients with DSD [13]. Notably, transgene expression aligns with the endogenous zebrafish *bmp2* expression [15], which strongly suggests the enhancer regulates *BMP2* expression. Interestingly, live confocal imaging revealed that -117BMP2 directs transgene expression at the osteogenic fronts when the frontal bones begin to form, and *egfp* expression precedes *sp7:mCherry* marking the osteoblasts (Fig. 2E–G). During the rapid growth of the frontal and parietal bones, the *egfp* expression is prominent at the osteogenic fronts in the region of osteoblast precursors. The expression is also visible at lateral ethmoids and supraorbital bones (Fig. 2H–J). This strong expression was transient and became less pronounced as the skull growth slowed and sutures formed at 8–9 mm standard length (SL). Therefore, -117BMP2 most likely regulates *BMP2* expression during early osteoblast differentiation. Notably, after the enhancer activity diminished at the osteogenic fronts, it became prominent at the supraorbital lateral line canals (data not shown), another site of rapid bone growth and remodeling.

The imaging data show the activity of -117BMP2 in subsets of cartilage and bone cells throughout skull development. To further characterize the specificity of enhancer activity, we performed single-cell RNA sequencing (scRNAseq) on dissected skulls of 2–3-week-old fish (~5.8 mm to 7 mm) transgenic for $-117\text{BMP2}:\text{egfp}$ and *sp7:mCherry*. At this stage, the frontal bones have just started to develop rapidly. There were a total of 23 cells that expressed both *egfp* and *bmp2*. We generated a heatmap of genes that were differentially expressed in the *egfp* and *bmp2* expressing cells compared to other cell clusters (Fig. 2K and L). The expression level for each differentially expressed gene is indicated by the log₂ fold changes of expression in either *egfp* or *bmp2* positive cells relative to other cell clusters. Consistent with -117BMP2 acting as an enhancer for *BMP2*, *egfp* is significantly upregulated in *bmp2*-expressing cells (Fig. 2L). Importantly, although the expression profiles of *egfp* and *bmp2* expressing cells do not completely overlap, they show the greatest overlap in osteoblasts indicated by the comparable expression level of *col10a1a* [15], *bgna* [31], and *sp7:mCherry* [32] (Supplementary Table S3).

-707BMP2 contains a risk-associated SNP and is active at the site of frontal bone initiation

-707BMP2 in the intronic region of *BBS9* is highly conserved, with a highest LOD score of 1008, and aligns with sequences down to *Xenopus tropicalis*. It also overlaps a predicted enhancer element based on ENCODE data [22–24] and contains the CS risk-associated SNP rs10254116 (Fig. 1C). In fish injected with the -707BMP2 construct at 5dpf, *egfp* is expressed mosaically in many major cranial cartilages, including ceratohyal and Meckel's cartilage (Fig. 3A), which was comparable to the transgene expression driven by -117BMP2 . We established transgenic lines from five independent founders that showed the enhancer activity of -707BMP2 containing the reference allele of rs10254116. Four of the lines have prominent *egfp* expression in cranial cartilage, similar to the mosaic pattern ($-707\text{BMP2}:\text{egfp}$ ©), so they were annotated as $-707\text{BMP2}:\text{egfp}$ (c1, c2, c3, c5), c for cartilage. The fifth line has *egfp* expression predominantly in perichondrium ($-707\text{BMP2}:\text{egfp}$ (pc)). At the larval stage, the enhancer activity

in the $-707\text{BMP2}:\text{egfp}$ (c) fish was very prominent at ceratohyal, Meckel's cartilage, and palatoquadrate (Fig. 3B and C), while the enhancer activity of the $-707\text{BMP2}:\text{egfp}$ (pc) fish was prominent in the perichondral cells around the Meckel's cartilage, posterior ceratohyal and palatoquadrate (Fig. 3D). These areas of expression remained strong throughout skull development. In addition, this enhancer was active in the forelimb and mandibular process of pharyngeal arch 1 in mouse embryos at E11.5, and in the humerus and structures associated with the inner ear at E13.5 (Supplementary Fig. S1), consistent with the mouse endogenous *Bmp2* expression at similar stage [33]. To investigate the enhancer activity of -707BMP2 at the later stage during skull formation, we performed confocal imaging on both expression patterns during the period of frontal bone initiation (~5–6 mm SL) and rapid planar expansion (~6–7 mm SL). During frontal bone initiation, cells of the zebrafish frontal bone were first detected near the juncture of the taenia marginalis and epiphyseal bar cartilages, and the bone grew on top of the cartilage toward the apex of the skull over the next several days [34]. Just prior to frontal bone initiation, $-707\text{BMP2}:\text{egfp}$ (c) fish expressed *egfp* in chondrocytes and perichondral cells in this region (Fig. 3E and F), and $-707\text{BMP2}:\text{egfp}$ (pc) fish expressed *egfp* in perichondral cells of the same region (Fig. 3I and J). Expression in both lines was maintained as the frontal bone expanded along the epiphyseal bar (Fig. 3G, H, K and L). Notably, $-707\text{BMP2}:\text{egfp}$ (c3) (Fig. 3G and H) and (c5) (Supplementary Fig. S2) fish also showed *egfp* expression in the osteoblasts of growing frontal bones overlapping with *mCherry* expression. Despite some variation among the five lines, they all demonstrated enhancer activity in the taenia marginalis region adjacent to the frontal bone, pointing to a role for *BMP2* in positively regulating the early growth of the frontal bone.

Because -707BMP2 contains a CS risk-associated SNP, we also examined whether the minor allele of rs10254116 had an effect on the enhancer activity via zebrafish transgenesis. We obtained transgenic lines from two independent founders, annotated as $-707\text{BMP2}_{\text{minor}}$ (c1) & (c2): *egfp*. Similar to the major allele lines, fish of the minor allele lines showed enhancer activities in early cranial cartilage and bones, as well as inner ears at 5dpf (Supplementary Fig. S2). During frontal bone development, the enhancer with the minor allele was also active in cartilage and perichondrium of taenia marginalis (Fig. 3M–P, Supplementary Fig. S3). The minor allele did not lead to a consistent qualitative difference in transgene expression, and the variability in expression among lines made quantitative comparisons impractical.

$+402\text{BMP2}$ directs expression in cranial bones

In mosaic fish at 5dpf, $+402\text{BMP2}$ directed prominent expression in cranial cartilage (data not shown). After germline transmission, transgenic progeny displayed broad expression in developing bone. At 5dpf, *egfp* expression completely overlapped with *sp7:mCherry* in branchiostegal ray 1, opercles, maxilla, and cleithrum (Fig. 4A–D). As the fish grow, *egfp* expression remained primarily on the edges of most cranial bones (Fig. 4E–H) as well as vertebrae. Despite the early specific bone expression, no enhancer activity was detected in the cranial vault suggesting $+402\text{BMP2}$ plays a role primarily in the development of the face and vertebrae, which are tissues responsive to *BMP2*-driven signaling.

$+421\text{BMP2}$ directs expression in cranial cartilage

A second sequence within the *BMP2* risk locus, $+421\text{BMP2}$, has a maximum LOD score of 509 and contains the CS risk-associated

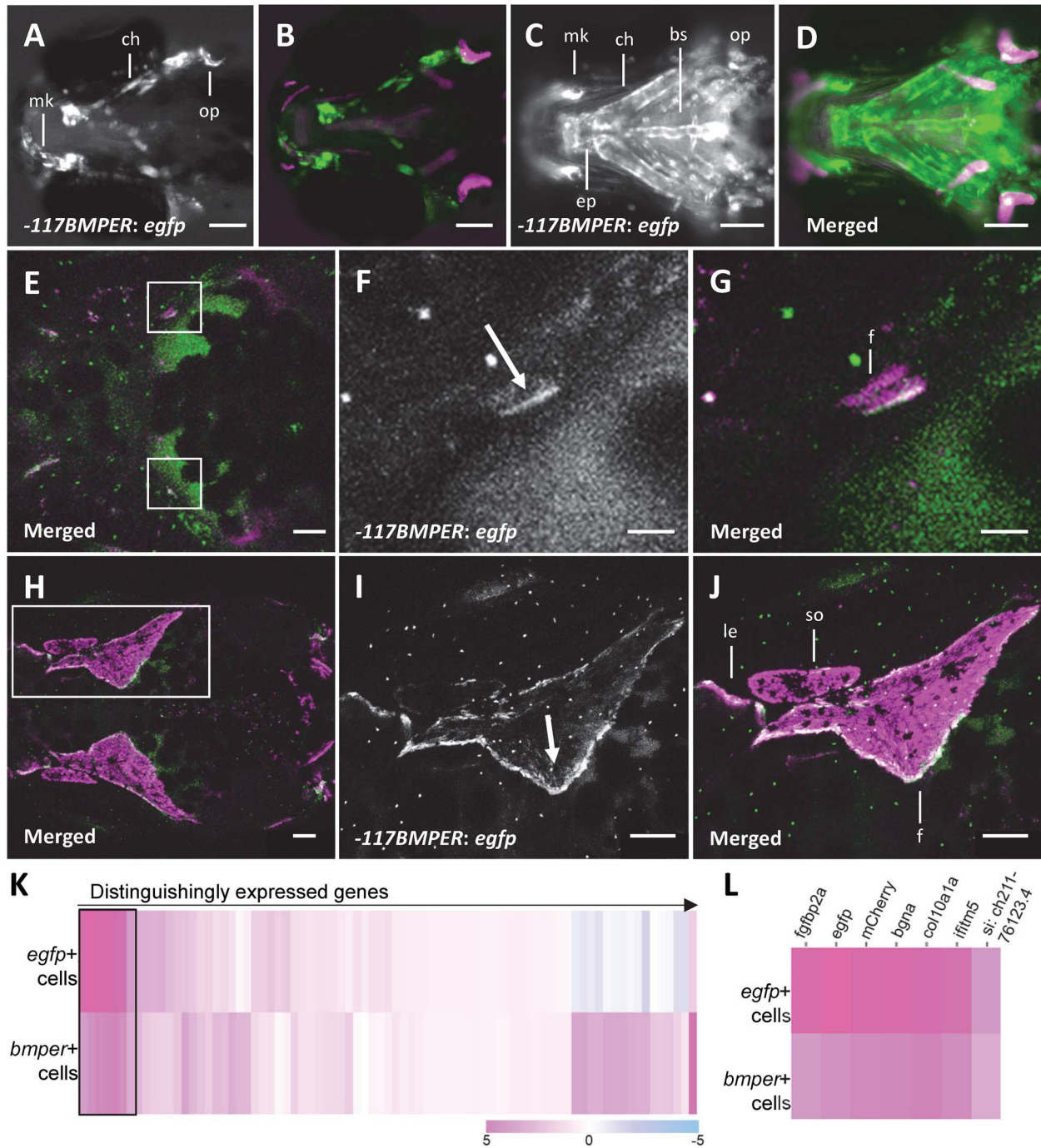


Figure 2. *-117BMPER* directs *egfp* expression to developing craniofacial bones and cartilage in zebrafish. (A, B) Ventral view of 5dpf zebrafish mosaic for *-117BMPER*. (A) Enhancer is active in early chondrocytes and osteoblasts, shown by *egfp* expression. (B) Overlap is shown with osteoblast marker *sp7: mcherry*. (C–D) Ventral view of 5dpf stable transgenic Tg (*-117BMPER: egfp; sp7: mcherry*) zebrafish. Enhancer is broadly active in craniofacial bones and cartilage. (E–G) compressed Z-stacks from confocal imaging of a transgenic Tg (*-117BMPER: egfp; sp7: mcherry*) animal at 5.2 mm SL (19dpf). Enhancer activity is prominent at the osteogenic front of the developing frontal bone (arrow). (H–J) Dorsal view of enhancer activity in the same transgenic animal at 6.62 mm SL (26dpf); each image is a maximum intensity projection of confocal slices. Enhancer is active in early differentiated osteoblasts, arrow points to the osteogenic front and the direction of growth. (K and L) Single cell RNAseq was performed on cells isolated from transgenic zebrafish skulls at 2 and 3 weeks. The enhancer, indicated by *egfp* expression, reflects the part of *bmpert* expression that is in osteoblast precursor cells. (K) Heatmap of genes (in columns) that are distinguishingly expressed in *egfp+* (first row) and *bmpert* expressing osteoblasts (second row). The quantification represents the Log₂ fold changes in expression of each gene in each column relative to the entire dataset. (L) Significantly upregulated genes based on adjusted P-value ($P < 0.05$), in either *egfp+* cells or *bmpert+* cells, boxed in (K). The transgene mCherry directed by *sp7* marks osteoblasts. bs: basibranchial; ch: ceratohyal; ep: ethmoid plate; f: frontal bone; le: lateral ethmoid; mk: Meckel's cartilage; op: opercle; so: supraorbital. Scale bars are 100 μm in A–E, H–J, and 25 μm in F and G.

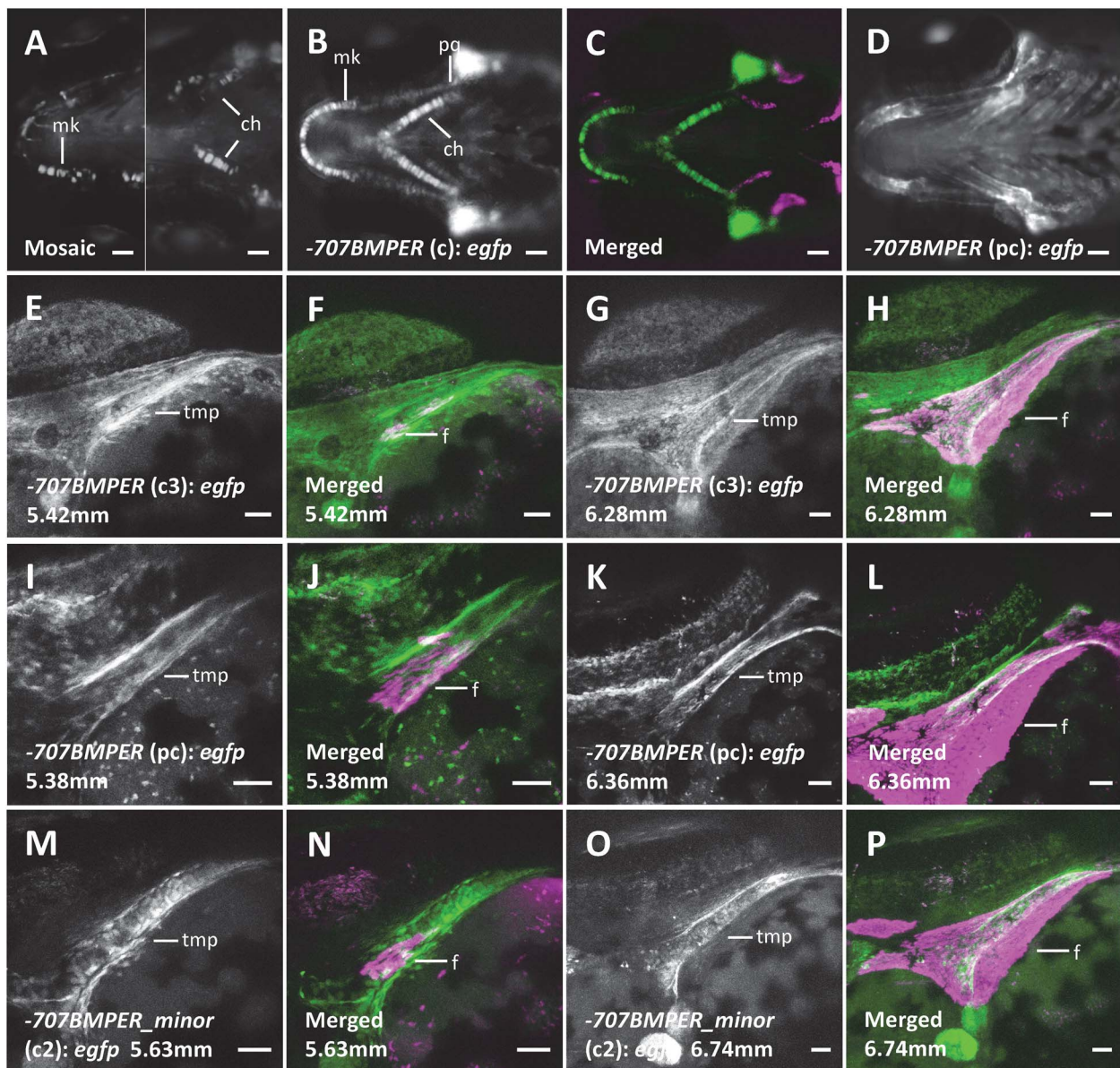


Figure 3. $-707BMPER$ directs *egfp* expression to developing craniofacial bones and cartilage in zebrafish. (A) Ventral view of two 5dpf zebrafish mosaic for $-707BMPER: egfp$ expression, showing enhancer activity in early chondrocytes of the skulls. The $-707BMPER$ contains the reference allele of CS-associated SNP rs10254116. (B–D) Ventral view of 5dpf larvae from two independent transgenic lines of Tg ($-707BMPER: egfp; sp7: mcherry$) zebrafish showing different expression patterns of the transgene. (B–C) In the $-707BMPER(c): egfp$ line, the enhancer is primarily active in early chondrocytes, with *sp7:mcherry* expressed in osteoblasts; (D) In the $-707BMPER(pc): egfp$ line, the expression is prominent in perichondral cells. (E–L) Compressed Z stacks from confocal imaging of two $-707BMPER$ transgenic lines during frontal bone initiation (5–6 mm SL) and rapid expansion (6–7 mm SL). (E–H) Dorsal view of the enhancer activity of a transgenic animal Tg ($-707BMPER(c3): egfp; sp7: mCherry$). The enhancer activity was prominent in the chondrocytes of tmp but also visible in osteoblasts of the frontal bone. (I–L) Dorsal view of the enhancer activity of a transgenic fish Tg ($-707BMPER(pc): egfp; sp7: mCherry$). The enhancer was active at the perichondrium of taenia marginalis posterior (tmp) next to the initiation site of the frontal bone. (M–O) Compressed Z stacks from confocal imaging of a transgenic line expressing $-707BMPER$ containing the minor allele of rs10254116, Tg ($-707BMPER_minor(c2): egfp; sp7: mCherry$). The enhancer was active at the chondrocytes of tmp during frontal bone development. ch: ceratohyal; f: frontal bone; mk: Meckel's cartilage; pg: palatoquadrate; tmp: taenia marginalis posterior. Scale bars are 50 μm .

SNP rs6117669. There is no predicted enhancer element within +421BMP2, according to ENCODE. In mosaic larvae at 5dpf, +421BMP2 primarily directed transgene expression in cranial cartilage. After germline transmission, we isolated a line with broad cartilage expression, including in the ceratohyal, Meckel's cartilage, and pectoral fins. At later stages, the enhancer was active in the otic capsule and anterior ceratohyal (Supplementary Fig. S4). This enhancer also lacks activity in the developing cranial vault.

Identification of transcription factor binding interactions

Enhancer activity is dictated by interactions with transcription factors (TFs), which recruit cofactors and transcriptional machinery. In addition, SNPs in noncoding genomic regions can disrupt these TF-DNA interactions, contributing to the dysregulation of target genes and disease pathology. Therefore, we aimed to identify potential transcription factor interactions with each enhancer. To do so, we screened these enhancers via an enhanced

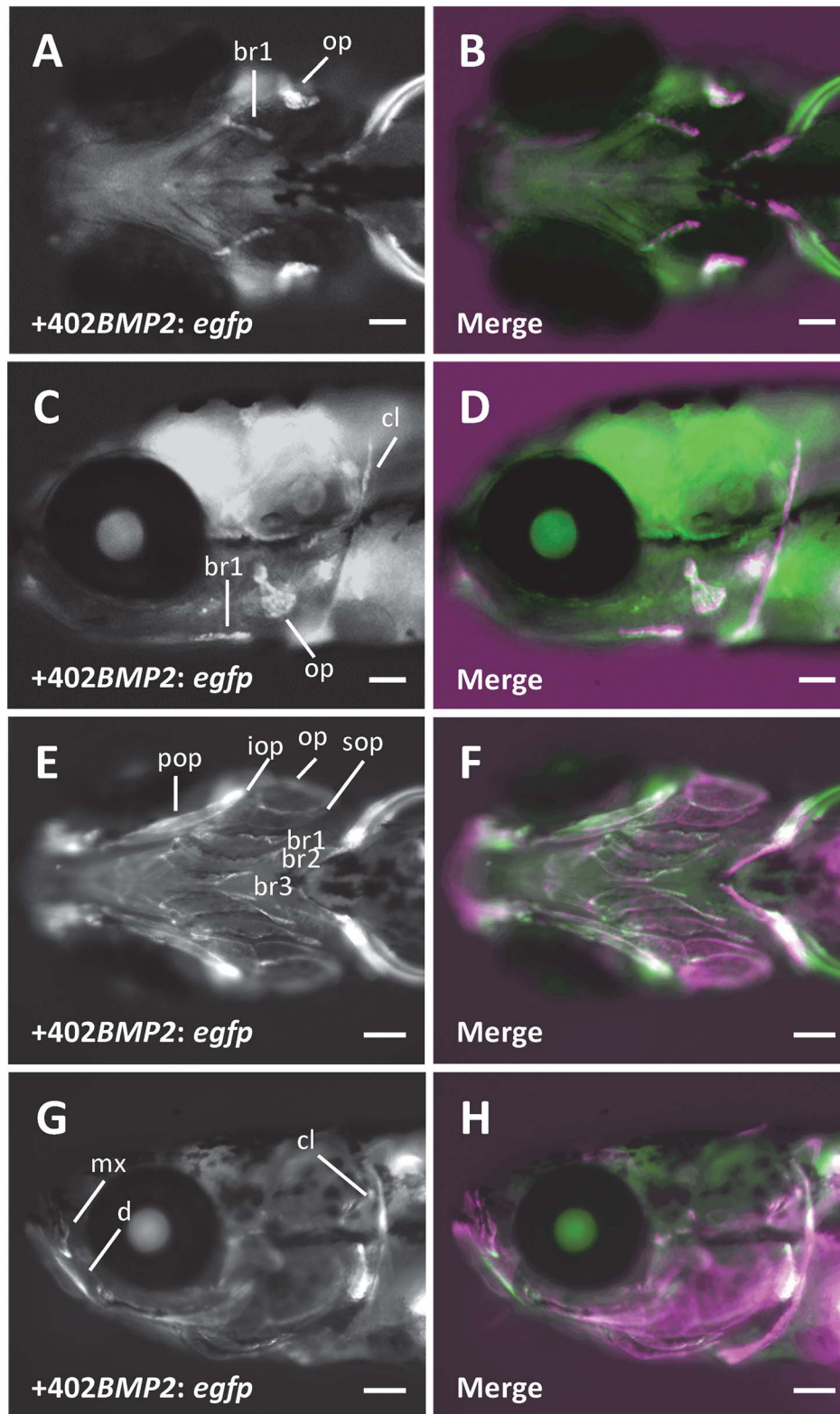


Figure 4. +402BMP2 directs *egfp* expression to developing craniofacial bones. (A–D) Ventral and lateral view of a 5dpf transgenic Tg (+402BMP2: *egfp*; *sp7*: *mcherry*) zebrafish. (B, D) Overlap is shown with osteoblast marker *sp7*: *mcherry*. Enhancer is active in early craniofacial osteoblasts. (E–H) Ventral and lateral view of Tg (+402BMP2: *egfp*; *sp7*: *mcherry*) animal at 6.27 mm SL (14dpf). Enhancer is active around the edge of many developing cranial bones, consistent with early osteoblasts. br 1–3: branchiostegal ray 1–3; cl: cleithrum; d: dentary; iop: interopercle; mx: maxilla; op: opercle; pop: preopercle; sop: subopercle. Scale bars are 100 μ m in A–D and 300 μ m in E–H.

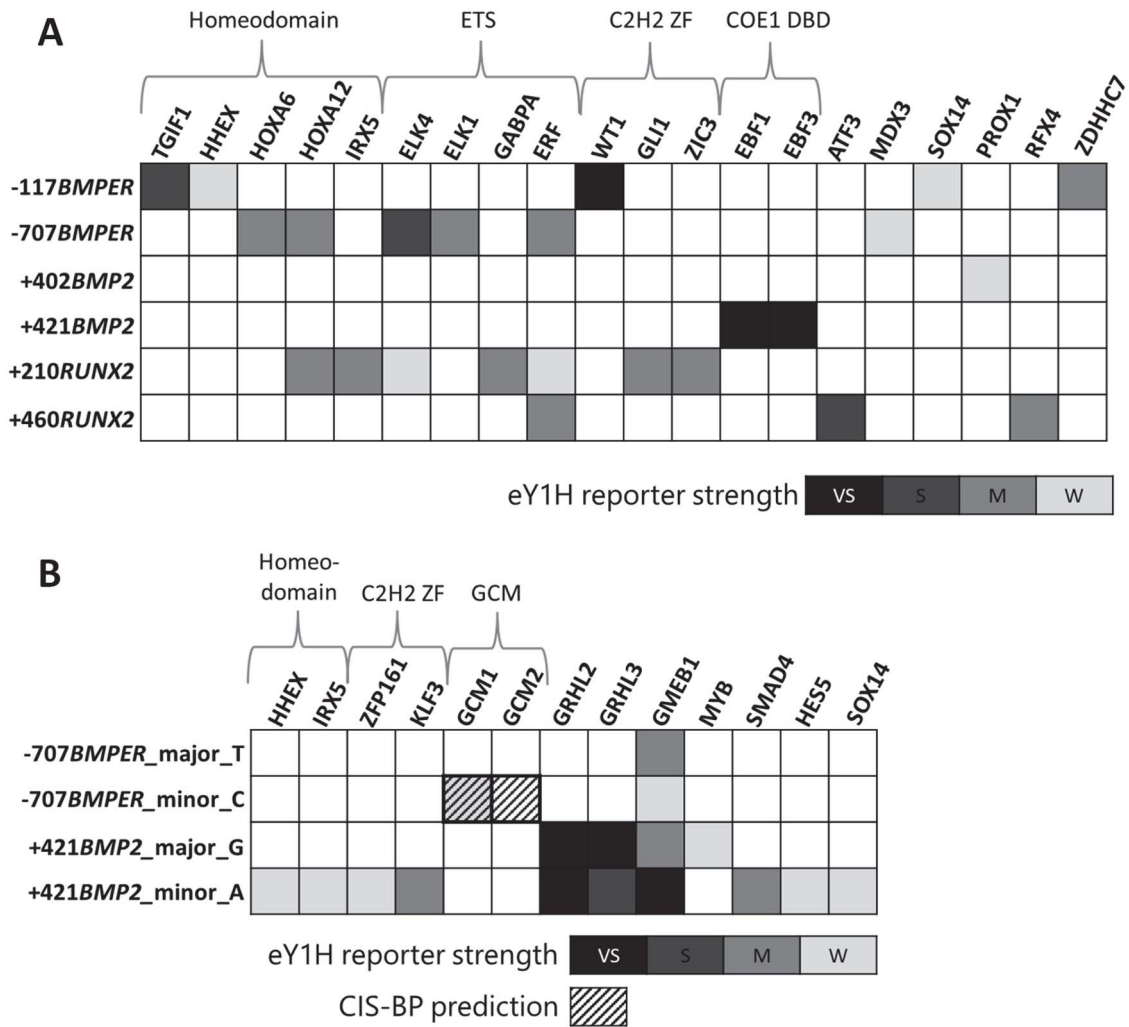


Figure 5. Identification of interacting TFs via eY1H assays and differential enhancer activities between reference and alternate alleles. The scale indicates the summed strength of TF-bait interactions from two strains of yeasts: very strong (VS), strong (S), medium (M), weak (W). Only interactions positive in both strains of yeasts are demonstrated in the matrices. Quantification of strengths and interactions positive in one strain of yeast are included in [Supplementary Table S5](#). (A) TFs that interacted with full-length enhancers in eY1H assays. (B) TFs that interacted with 40 bp-sequences around CS-associating SNPs. GCM1 showed up positive in both eY1H assay and CIS-BP differential binding prediction, whereas GCM2 from the same family only did by CIS-BP.

yeast one-hybrid (eY1H) assay [35]. The library represents 1086 out of ~1600 annotated human TFs which were tested for interactions with each enhancer. For comparison, we included two previously identified RUNX2 enhancers with specific activity in osteoblasts [36]. Each screen was performed in quadruplicate in two independent yeast strains, and only interactions detected by both yeast strains in at least 3/4 replicates were considered positive interactions (Fig. 5A). Notably, three out of five enhancers, -707BMPER and the two RUNX2 enhancers, interact with ERF, an inhibitory ETS transcription factor that directly binds to ERK1/2 [37]. ERK1/2 signaling is activated in CS [38], and haploinsufficiency of ERF causes CS [39].

To test for changes in TF interactions associated with risk alleles, we carried out a targeted screen of the two enhancers containing risk-associated SNPs, -707BMPER (rs10254116, C/T) and +421BMP2 (rs6117669, A/G). For each enhancer, two versions of the 40-base pair sequence centered on the SNP, corresponding to the reference and alternate alleles, were screened as above. In parallel, we used the CIS-BP database [40] to predict differential TF binding at each pair of alleles (Fig. 5B).

The alternate allele of -707BMPER had a novel interaction with GCM1, one of the two mammalian homologs of the *Drosophila* transcription factor glia cell missing. The interaction was not detected with the reference allele and agreed with the output of CIS-BP, which predicted the binding of GCM1 only with the alternate allele. For +421BMP2, there were multiple interactions gained and lost for the alternate allele compared to the reference.

Discussion

BBS9 and BMPER constitute a regulatory domain important in craniofacial development

The identification of a risk locus for CS within a *BBS9* intron implicated dysregulation of *BBS9* expression as a risk factor for CS. However, several lines of evidence suggested an alternative hypothesis that changes in the regulation of the adjacent gene *BMPER* underlie the risk. First, CS is not a consistent feature of BBS, and *BBS9* has no known role in osteogenesis. In contrast, *BMPER* has a known role in bone formation, supported by the phenotype of human patients and mouse mutants. Although a

null mutation in zebrafish *bmp2* has not been described, it is expressed prominently in developing skeletal tissues, suggesting a highly conserved function in skeletal development.

Subsequent to the publication of the GWAS that identified the CS risk locus within BBS9, additional support for the existence of a broad conserved regulatory domain has been reported. Through comparative genomic analysis of human and chimpanzee cultured cranial neural crest cells, Prescott *et al.* predicted common and differentially active enhancers in both species. Two of the 5000 most active common enhancers were in introns of BBS9, and the entire intergenic region between BBS9 and BMPER is a coordinated regulatory region [41]. More recently, a pig strain selectively bred for rapid weight gain was shown to carry a 112 kb deletion encompassing the 3' end of *Bbs9* and a portion of the intergenic region [42]. Obesity is a consistent feature of BBS in humans, and heterozygous loss of *Bbs9* is thought to account for the increased weight gain. However, pigs homozygous for the deletion die in utero. Although the deletion is not in the *Bmp2* gene, *Bmp2* expression is greatly decreased in the homozygous fetuses, suggesting a loss of critical enhancers that resulted in an intolerable level of *Bmp2* for proper development and contributed to the lethality. Our identification of two skeletal enhancers, from the BBS9/BMPER intergenic region and from a BBS9 intron, are also consistent with a large regulatory domain with critical and conserved function in craniofacial development. Notably, neither of our enhancers was identified by Prescott *et al.* [41], and the addition of -707BMPER extends the upstream limit of the regulatory domain by almost 300 kb. Finally, BBS9 and BMPER are part of a single topological associated domain (Supplementary Fig. S5), further supporting a conserved regulatory domain encompassing both genes. Of the enhancers we characterized, -707BMPER containing the SNP rs10254116 is the most plausible candidate to explain the risk of CS. Although statistical fine mapping was not performed to confirm the causal association, this SNP was again identified in a separate cohort through targeted sequencing [43], strengthening our proposed model.

Presence of skeletal-tissue-specific enhancers within the risk locus near BMP2

Justice *et al.* tested a sequence encompassing the tag SNP rs1884302 for enhancer activity using a similar zebrafish transgenesis assay, comparing the reference and alternate alleles. They reported a higher level of expression from the alternate allele relative to the reference allele, although the activity was not in skeletal tissues and was only observed up to 72 h post-fertilization, prior to the development of any bones of the cranial vault [19]. In our assay, we failed to observe any expression above the background driven by either sequence (+344BMP2 T and C). Differences in the transgenic vectors may account for our failure to detect any specific expression, but it was not possible to verify the published results.

We identified two BMP2 enhancers that are active in skeletal tissues, +402BMP2 in bones, and +421BMP2 in cranial cartilage. However, neither has enhancer activity in the developing cranial vault, suggesting the genetic risk of CS near BMP2 is carried by undiscovered sequences or mechanisms involving multiple regulatory elements. A subsequent study by Justice *et al.* also suggested a cumulative effect of more than one sequence variant could underly the increased CS risk [43]. They performed targeted sequencing of CS patients carrying the BMP2-linked risk allele and found a linked haplotype of >100kb with 116 significant sequence variants located either in the intergenic region between BMP2 and the noncoding RNA gene LINC01428 or in the

intronic regions of the nearby gene LOC101929265. Of interest, an additional enhancer identified in our study with activity in craniofacial cartilage contains one of the newly identified SNPs (+365BMP2, Supplementary Table S2). The sequence gave rise to mosaic expression in skeletal tissues; however, germline transmission of this expression was not found. In the eY1H assay, +365BMP2 had a very weak interaction with ZBTB39. Since ZBTB39 is ubiquitously expressed, we considered the sequence unlikely to contribute to the risk of CS and did not carry out a targeted screen centered on the SNP. Our work provides a paradigm to functionally test putative enhancer elements in this linked haplotype, which will help define the causal sequence variants.

Enhancer-TF interactions suggest underlying regulatory mechanisms

To find regulatory interactions with the identified enhancers and potentially reveal the mechanisms of increased CS risk, we took an unbiased approach to identify TF-DNA interactions through an eY1H assay. In contrast to *in silico* predictions, which have a high rate of false positives, TF interactions identified through eY1H show a subsequent verification rate of as high as 40%–70% in reporter assays in mammalian cells [44–46]. We analyzed the four identified skeletal enhancers, -117BMPER, -707BMPER, +402BMP2, and +421BMP2, and for comparison, included previously characterized RUNX2 enhancers with specific expression in developing bone. We identified robust interactions implicating multiple signaling pathways, including some already known to play important roles in craniofacial development. These interactions generate specific predictions about developmental regulation that can be tested directly in future experiments to verify their *in vivo* function in regulating skeletal gene expression.

Among the interactions detected, we found that ETS2 repressor factor (ERF) bound to the +460RUNX2, +210RUNX2, and -707BMPER enhancers. Significantly, haploinsufficiency for ERF is associated with cases of both non-syndromic and syndromic CS [39]. ERF is an atypical member of the ETS2 TF family that acts as a transcriptional repressor rather than an activator and is thought to prevent the binding of activating ETS factors at the same site. *Erf* and *Runx2* share similar expression patterns at the osteogenic fronts of parietal bones in mice and were shown to act antagonistically in function [39]. In addition, our scRNAseq data from skulls of 2- and 3-week-old fish suggests that zebrafish *erf* and *bmp2* are likely to be co-expressed in cartilage cells but not in osteoblasts, consistent with the enhancer activity of -707BMPER (Supplementary Fig. S6). Our results support a specific mechanism in which loss of ERF leads directly to the upregulation of BMPER and RUNX2 through decreased binding to their enhancers. More broadly, our results point to the -707BMPER enhancer as a critical convergence point for the regulation of craniofacial development, where either sequence variants in the enhancer itself or alterations in the level of the interacting factor ERF leads to changes in BMPER expression, misregulation of BMP signaling, and ultimately CS.

Through the eY1H assay, we also found TGF β -induced factor homeobox1 (TGIF1) interacted with -117BMPER. TGIF1 belongs to a subgroup of homeobox proteins that are highly conserved transcription regulators. Mutations in TGIF1 cause holoprosencephaly (HPE), a common human congenital disease that manifests structural brain and craniofacial defects [47]. TGIF1 is a corepressor of the SMAD2-dependent signaling pathway [48], which is proposed as a potential mechanism of TGIF1-related HPE. In mouse models, loss of *Tgif* protein, either *Tgif1* or *Tgif2* or both, disturbed

the Sonic Hedgehog (Shh) activities via dysregulation of SMAD2-Nodal signaling [49]. Another disease model includes dysfunction in retinoic acid metabolism, demonstrated in the zebrafish morphants [50], which is also a risk factor for craniosynostosis [51]. Although no specific role of TGIF1 has been described in osteogenesis, our scRNAseq data showed that *tgif1* and *egfp* driven by -117BMPER are expressed in the same subsets of osteoblasts and cartilage (Supplementary Fig. S7), consistent with the role for TGIF1 in regulating BMP signaling during craniofacial development.

We used two strategies to probe the functional significance of the putative causal SNPs in our identified enhancers. To look for differences in the *in vivo* activity of -707BMPER , we made an equivalent transgenic construct carrying the alternative allele and examined both mosaic expression and patterns of activity in established lines. Both the major and minor allele lines of -707BMPER showed consistent enhancer activities in the cartilage closely associated with developing frontal bones. The specific role of this cartilage in suture formation is yet to be determined. Because our Tol2-based approach mediates random integration of transgenic construct, direct quantitative comparison between the effect of the variants is not practical. A system that employs a single integration site [52] would facilitate future experiments to quantify allele-specific effects in transgenic zebrafish.

We also carried out targeted eY1H screens on short sequences centered around the putative causal SNPs for both loci, in parallel with *in silico* predictions of TFBSs using CIS-BP [40]. For $+421\text{BMP2}$, CIS-BP predicted several differences in binding sites for the minor allele, but these did not correspond to the eY1H results. Instead, there were a number of binding interactions gained by the alternate allele and one lost. For -707BMPER , the only difference detected in the eY1H assay was the gain of binding by GCM1 to the minor allele, also predicted *in silico* by CIS-BP.

The two mammalian genes *GCM1* and *GCM2* are homologs of the *Drosophila* gene *glial cells missing* (*gcm*), with well-studied roles in trophoblast and parathyroid cell differentiation, respectively. *GCM1* is expressed in a subset of trophoblast cells and is required for normal placenta development [53]. Haploinsufficiency for *Gcm1* in mice causes phenotypes resembling preeclampsia [54], while homozygous mutants fail to form functional placenta [55, 56]. Because of the early lethality of *Gcm1* mouse mutant, there is limited information on other possible roles. A conditional mouse allele was used to selectively delete *Gcm1* in the kidney, demonstrating an important role in recovery from ischemic injury [57]. Mammalian *GCM2* is expressed prominently in cells of the parathyroid, and mutations have demonstrated its requirement for parathyroid development in both mice [58, 59] and humans [60–62]. Interestingly, mice lacking *Gcm2* still have almost normal circulating levels of parathyroid hormone (PTH), apparently due to the upregulation of *Gcm1* and secretion of PTH by cells of the thyroid [58]. This suggests significant functional redundancy for the two mammalian genes, which could also mask their roles in other contexts. The zebrafish genome apparently has no *gcm1* orthologue; zebrafish *gcm2* is 82% homologous in amino acids sequences to human *GCM2* and 73% to human *GCM1* [63]. Zebrafish *gcm2* is expressed in parathyroid cells, ionocytes [63], and chondrocytes of the pharyngeal arches [64]. Antisense morpholino suppression of *gcm2* confirmed a conserved requirement for the zebrafish gene in parathyroid development and further suggested a role in craniofacial cartilage and bone formation [64]. More recently, a *gcm2* mutant revealed its important role in maintaining pH in lateral line neuromasts, resulting in hair cell dysfunction

[65]. However, the craniofacial development of the mutant was not described, and the mutants failed to inflate their swim bladders and did not survive past larval stages. Online expression atlases indicate other areas of expression for both *Gcm* genes in mice, including during craniofacial development, suggesting they may function in other contexts as well. Interestingly, a genome-wide screen for *Gcm* binding sites during *Drosophila* development uncovered previously undescribed target genes and roles in new tissues [66]. They extended the functional assays to mammalian cells and uncovered potentially conserved roles for the mammalian *Gcm* proteins. Future studies will focus on confirming *in vivo* roles for specific binding interactions and defining the importance of the SNPs in enhancer function and CS risk.

Although the eY1H assay did not uncover other differential bindings between enhancer alleles, the library used in our experiments contained only transcription factors that bind as monomers or homodimers. Similar screens with combinatorial libraries, yeast libraries expressing TF pairs, could uncover additional differences between alleles. It is also possible that the risk-associated SNP in the enhancer is tightly linked to a rare SNP with a greater effect on enhancer function and disease risk. Ongoing projects to acquire genomic sequence data from CS patients may reveal the existence of additional SNPs that can be functionally tested in future studies.

We have identified enhancers from the risk-associated loci near *BMP2* and *BMPER* that regulate expression in the developing craniofacial skeleton and propose that changes in their activities may lead to an increase in BMP signaling in skeletogenesis, which contributes to CS risk in conjunction with other risk factors, genetic or environmental. This hypothesis is consistent with the analysis of the existing zebrafish and mouse models of CS that implicate accelerated growth of the skull bones as the underlying biological mechanism [67]. We also find that -707BMPER binds ERF, implicating it as part of the mechanism in CS due to ERF haploinsufficiency. Overall, our findings provide further insights into the disease mechanism of CS. Importantly, our paradigm is broadly applicable to other complex genetic diseases, potentially illuminating many connections between genome variation and disease risk.

Materials and methods

Zebrafish lines

Zebrafish stocks were maintained in the central aquatic facility at Boston University. All animal experiments were approved by the Institutional Animal Care and Use Committee of Boston University (Approval # TR20220000043). Previously reported zebrafish lines used in this study include *sp7: mcherry* and *sp7-/-* [32]. We established additional new transgenic lines in the course of these studies: -117BMPER : *egfp*, -707BMPER : *egfp*, $+402\text{BMP2}$: *egfp*, $+421\text{BMP2}$: *egfp*, as described in the text [25].

Selection of candidate regulatory elements

Within the genomic regions of interest, the selection of sequences for analysis as potential regulatory elements was based primarily on multispecies conservation. We selected candidates based on PhastCons [20, 21], included in the 100 Vertebrate Conserved Elements track on the UCSC genome browser. Sequences containing an element with a conservation LOD score of at least 90 were included in our analysis. We also referenced ENCODE Candidate Cis-Regulatory Elements (cCREs) database [22–24] to check if selected sequences contain any predicted enhancer elements. The

names of the enhancers are indicative of their location relative to the transcription start site of either *BMPER* or *BMP2*, with a negative sign indicating upstream and a positive sign indicating downstream of the gene.

Zebrafish transgenesis assay

Primers were designed to amplify candidate sequences from human genomic DNA using Primer3Plus, with primers on either end located ~200 bp outside conserved elements. Adjacent conserved elements were grouped together when convenient, and amplicons ranged from 401 to 1631 bp. Primers used for amplification and the genome coordinates of all amplicons are listed in [Supplementary Table S4](#). Purified PCR products were cloned into a Gateway™ entry vector using Invitrogen pCR™8/GW/TOPO™ TA Cloning (Invitrogen, catalog # K252002) followed by LR clonase™ reaction (Invitrogen, catalog #11791100) into the pGW-cFos-EGFP vectors for analysis [25]. All entry vectors were verified by sequencing prior to LR recombination. Tol2 transposase mRNA was transcribed from the plasmid pCS-TP with the mMessage mMachine kit (Qiagen).

We tested the regulatory potential of each candidate sequence through mosaic transgenesis analysis in zebrafish larvae, as previously described [25]. For each sequence, at least 150 embryos were injected at the 1-cell stage with 1 nl of injection mix containing Tol2 mRNA (final concentration of 35 ng/ul), the purified plasmid (final concentration of 15 ng/ul), and 1% phenol red for visualization. We examined injected larvae for tissue-specific expression of *egfp* at 4–5 days post fertilization (dpf). For sequences with expression patterns of interest, injected larvae were raised and bred to generate transgenic lines.

Microscopic imaging

For screening and low-magnification imaging, we sedated fish of interest with 4 mg/ml Tricaine/MS-222 and mounted them in 3% methylcellulose. We screened injected and transgenic fish for expression by epifluorescence on Olympus MVX10 and captured images using a high-resolution camera. Images were processed using Image J.

We performed serial confocal imaging of live fish as previously described [34]. Briefly, each fish of interest is sedated using Tricaine/MS-222 and placed in 3% methylcellulose to measure the standard length. The fish is mounted in 2% low-melt agarose, space cleared around the gills for respiration, and the fish is covered with fresh Tricaine solution. Stabilized fish are placed onto the stage of the confocal microscope (Leica TCS LSI-III) for image capture, which typically takes less than ten min., and recovered in fresh tank water. Imaging may be conducted multiple times on the same individual from 15dpf (~5–6 mm) to 35dpf (9–11 mm). Image stacks were exported and processed using ImageJ.

[Figure 2E–L](#) were selected from images of the six Tg (–117BMPER: *egfp*; *sp7*: *mcherry*) fish, which were followed individually over the course between 13dpf (average 4.4 mmSL) to 35dpf (average 10.3 mmSL). For all other figures, one to three fish were randomly taken from the tank of over 60 fish with the appropriate transgenes to conduct the imaging experiment.

Single cell sequencing

For single-cell profiling of the developing zebrafish skull, we isolated skull tissue from WT and *sp7* mutant fish at ~5.8 mm (2 weeks post-fertilization) and ~7 mm (3wpf). Five fish were included in each sample. To aid in dissection and transcript analysis, fish carried two transgenes, –117BMPER: *egfp* and *sp7*: *mcherry*. To enrich skeletal tissues, brains and eyes were removed during

dissection. Each sample was then dissociated as described [68]. Dissociated cells were immediately loaded onto a 10× Genomics microfluidic chip processed for single cell RNAseq at the BUMC Single-Cell Sequencing Core. Sequencing on Illumina NextSeq was performed to a depth of at least 1 000 000 reads/ library. Cell Ranger software (v.3.1.0, 10× Genomics) was used for barcode recovery, genome alignment (Ensembl GRCz11), and to generate gene-by-cell count matrices with default parameters for each library. Cell clustering and differential gene expression analysis were performed using Loupe Browser 5.0. The raw and processed data files have been uploaded to the Gene Expression Omnibus (GEO GSE246579).

Mouse LacZ transgenesis assay

To verify the enhancer activity of –707BMPER in a mammalian model, we performed site-directed LacZ transgenesis in mice as previously described [69].

Briefly, –707BMPER was cloned upstream of a minimal *Shh* promoter and LacZ reporter gene (Addgene #139098) using Gibson Assembly. This design allows the integration of enhancer-reporter constructs into the safe harbor locus, *H11*. The sequence of the cloned constructs was verified with Primordium whole-plasmid sequencing. Transgenic mice were generated using pronuclear injection, as previously described [69]. Embryos were collected at E11.5 and E13.5 and stained with X-gal as previously described [69]. The embryos were genotyped for the presence of the transgenic construct, and embryos positive for transgene integration into the *H11* locus were cleared in 100% glycerol and imaged on an Olympus MVX10.

eY1H assay

To explore possible transcription factors that interact with identified enhancers, we performed a gene-centered yeast one-hybrid (eY1H) assay as previously described [35, 70]. Briefly, each enhancer, serving as a DNA bait, was cloned into two reporter constructs, expressing either *HIS3* or *LacZ*, and both constructs were integrated into the yeast genome. The yeast with the DNA bait was then mated to a collection of 1086 yeast strains expressing a TF fused to the *GAL4* activation domain. Binding to a TF activates expression of the reporter genes, allowing the yeast to grow in the absence of histidine and turning blue in the presence of X-gal. Each interaction was tested in quadruplicate, and the strength of the interaction is determined by the intensity of blue product. For the targeted screening that aimed to identify the differential TF binding at the CS-associated SNP, a 40 bp-sequence with the SNP in the middle was used as a DNA bait. The results of targeted screening were also compared with the differential TF binding prediction using publicly available CIS-BP database [40].

Acknowledgements

We thank BUMC Single-Cell Sequencing Core for their support of the single cell sequencing experiments and the staff of the BUMC Aquatics Facility for their expert care of our zebrafish colony. We thank Alexandra M Scalici and Kenta Kawasaki for their guidance and assistance with major experimental techniques. We also thank Kelly Ziyi Miao for her support with experiments as well as for offering constructive discussions, and we thank Joan Richtsmeier and Kazuhiko Kawasaki for helpful discussions on mouse developmental anatomy.

Supplementary data

Supplementary data is available at HMG Journal online.

Conflict of interest statement: The authors have declared that no competing interests exist.

Funding

This work was funded by National Institutes of Health grants R01 HG005039, U01 DE024434 and R21 DE029916 to SF, R35 GM128625 to JIFB, and R01DE028599 and U01DE024427 to AV. Mouse transgenic research was conducted at the E.O. Lawrence Berkeley National Laboratory and performed under U.S. Department of Energy Contract DE-AC02-05CH11231, University of California (UC).

References

- Chen G, Deng C, Li Y-P. TGF- β and BMP signaling in osteoblast differentiation and bone formation. *Int J Biol Sci* 2012;**8**:272–88.
- Bandyopadhyay A, Yadav PS, Prashar P. BMP signaling in development and diseases: a pharmacological perspective. *Biochem Pharmacol* 2013;**85**:857–64.
- Flaherty K, Singh N, Richtsmeier JT. Understanding craniosynostosis as a growth disorder. *Wiley Interdiscip Rev Dev Biol* 2016;**5**:429–59.
- Justice CM, Yagnik G, Kim Y. et al. A genome-wide association study identifies susceptibility loci for nonsyndromic sagittal craniosynostosis near BMP2 and within BBS9. *Nat Genet* 2012;**44**:1360–4.
- Veleri S, Bishop K, Nogare DED. et al. Knockdown of Bardet-Biedl syndrome gene BBS9/PTHB1 leads to cilia defects. *PLoS One* 2012;**7**:e34389.
- Wu M, Chen G, Li Y-P. TGF- β and BMP signaling in osteoblast, skeletal development, and bone formation, homeostasis and disease. *Bone Res* 2016;**4**:1–21.
- Moser M, Binder O, Wu Y. et al. BMPER, a novel endothelial cell precursor-derived protein, antagonizes bone morphogenetic protein signaling and endothelial cell differentiation. *Mol Cell Biol* 2003;**23**:5664–79.
- Rentzsch F, Zhang J, Kramer C. et al. Crossveinless 2 is an essential positive feedback regulator of bmp signaling during zebrafish gastrulation. *Development* 2006;**133**:801–11.
- Serpe M, Umulis D, Ralston A. et al. The BMP binding protein crossveinless 2 is a short-range, concentration-dependent, biphasic modulator of BMP signaling in Drosophila. *Dev Cell* 2008;**14**:940–53.
- Xiao F, Wang C, Wang C. et al. BMPER enhances bone formation by promoting the osteogenesis-angiogenesis coupling process in mesenchymal stem cells. *Cell Physiol Biochem* 2018;**45**:1927–39.
- Funari VA, Krakow D, Nevarez L. et al. BMPER mutation in diaphanospondylodysostosis identified by ancestral autozygosity mapping and targeted high-throughput sequencing. *Am J Hum Genet* 2010;**87**:532–7.
- Ben-Neriah Z, Michaelson-Cohen R, Inbar-Feigenberg M. et al. A deleterious founder mutation in the BMPER gene causes diaphanospondylodysostosis (DSD). *Am J Med Genet A* 2011;**155**:2801–6.
- Zong Z, Tees S, Miyajima F. et al. BMPER variants associated with a novel, attenuated subtype of diaphanospondylodysostosis. *J Hum Genet* 2015;**60**:743–7.
- Ikeya M, Kawada M, Kiyonari H. et al. Essential pro-bmp roles of crossveinless 2 in mouse organogenesis. *Development* 2006;**133**:4463–73.
- Li N, Felber K, Elks P. et al. Tracking gene expression during zebrafish osteoblast differentiation. *Dev Dyn* 2009;**238**:459–66.
- Kessels MY, Huitema LFA, Boeren S. et al. Proteomics analysis of the zebrafish skeletal extracellular matrix. *PLoS One* 2014;**9**:e90568.
- Ragvin A, Moro E, Fredman D. et al. Long-range gene regulation links genomic type 2 diabetes and obesity risk regions to HHX, SOX4, and IRX3. *Proc Natl Acad Sci U S A* 2010;**107**:775–80.
- Dixon JR, Jung I, Selvaraj S. et al. Chromatin architecture reorganization during stem cell differentiation. *Nature* 2015;**518**:331–6.
- Justice CM, Kim J, Kim S-D. et al. A variant associated with sagittal nonsyndromic craniosynostosis alters the regulatory function of a non-coding element. *Am J Med Genet A* 2017;**173**:2893–7.
- Siepel A, Haussler D. Combining phylogenetic and hidden Markov models in biosequence analysis. *J Comput Biol* 2004;**11**:413–28.
- Siepel A, Bejerano G, Pedersen JS. et al. Evolutionarily conserved elements in vertebrate, insect, worm, and yeast genomes. *Genome Res* 2005;**15**:1034–50.
- Consortium, T.E.P. A user's guide to the encyclopedia of DNA elements (ENCODE). *PLoS Biol* 2011;**9**:e1001046.
- Dunham I, Kundaje A, Aldred SF. et al. An integrated encyclopedia of DNA elements in the human genome. *Nature* 2012;**489**:57–74.
- Moore JE, Purcaro MJ, Pratt HE. et al. Expanded encyclopaedias of DNA elements in the human and mouse genomes. *Nature* 2020;**583**:699–710.
- Fisher S, Grice EA, Vinton RM. et al. Evaluating the biological relevance of putative enhancers using Tol2 transposon-mediated transgenesis in zebrafish. *Nat Protoc* 2006;**1**:1297–305.
- Kudoh T, Tsang M, Hukriede N. et al. A gene expression screen in zebrafish embryogenesis. *Genome Res* 2001;**11**:1979–87.
- Thisse B, Pflumio S, Fürthauer M. et al. Expression of the zebrafish genome during embryogenesis (NIH R01 RR15402). 2001. <http://zfin.org>.
- Thisse B, Thisse C. High throughput expression analysis of ZF-models consortium clones. 2005. <http://zfin.org>.
- Holzschuh J, Wada N, Wada C. et al. Requirements for endoderm and BMP signaling in sensory neurogenesis in zebrafish. *Development* 2005;**132**:3731–42.
- Mork L, Crump G. Zebrafish craniofacial development: a window into early patterning. *Curr Top Dev Biol* 2015;**115**:235–69.
- Thisse B, Thisse C, Wright GJ. Embryonic and larval expression patterns from a large scale screening for novel low affinity extracellular protein interactions. 2008. <http://zfin.org>.
- Kague E, Roy P, Asselin G. et al. Osterix/Sp7 limits cranial bone initiation sites and is required for formation of sutures. *Dev Biol* 2016;**413**:160–72.
- Richardson L, Venkataraman S, Stevenson P. et al. EMAGE mouse embryo spatial gene expression database: 2014 update. *Nucleic Acids Res* 2014;**42**:D835–44.
- Kanther M, Scalici A, Rashid A. et al. Initiation and early growth of the skull vault in zebrafish. *Mech Dev* 2019;**160**:103578.
- Fuxman Bass JI, Sahni N, Shrestha S. et al. Human gene-centered transcription factor networks for enhancers and disease variants. *Cell* 2015;**161**:661–73.
- Knopf F, Hammond C, Chekuru A. et al. Bone regenerates via dedifferentiation of osteoblasts in the zebrafish fin. *Dev Cell* 2011;**20**:713–24.

37. Katsianou MA, Adamopoulos C, Vastardis H. *et al.* Signaling mechanisms implicated in cranial sutures pathophysiology: craniosynostosis. *BBA Clin* 2016;**6**:165–76.
38. Kim H-J, Lee M-H, Park H-S. *et al.* Erk pathway and activator protein 1 play crucial roles in FGF2-stimulated premature cranial suture closure. *Dev Dyn* 2003;**227**:335–46.
39. Twigg SRF, Vorgia E, McGowan SJ. *et al.* Reduced dosage of ERF causes complex craniosynostosis in humans and mice and links ERK1/2 signaling to regulation of osteogenesis. *Nat Genet* 2013;**45**:308–13.
40. Weirauch MT, Yang A, Albu M. *et al.* Determination and inference of eukaryotic transcription factor sequence specificity. *Cell* 2014;**158**:1431–43.
41. Prescott SL, Srinivasan R, Carolina Marchetto M. *et al.* Enhancer divergence and cis-regulatory evolution in the human and chimp neural crest. *Cell* 2015;**163**:68–83.
42. Derks MFL, Lopes MS, Bosse M. *et al.* Balancing selection on a recessive lethal deletion with pleiotropic effects on two neighboring genes in the porcine genome. *PLoS Genet* 2018;**14**:e1007661.
43. Justice CM, Musolf AM, Cuellar A. *et al.* Targeted sequencing of candidate regions associated with sagittal and metopic nonsyndromic craniosynostosis. *Genes (Basel)* 2022;**13**:816.
44. Fuxman Bass JI, Pons C, Kozłowski L. *et al.* A gene-centered *C. elegans* protein–DNA interaction network provides a framework for functional predictions. *Mol Syst Biol* 2016;**12**:884.
45. Santoso CS, Li Z, Lal S. *et al.* Comprehensive mapping of the human cytokine gene regulatory network. *Nucleic Acids Res* 2020;**48**:12055–73.
46. Shrestha S, Liu X, Santoso CS. *et al.* Enhanced yeast one-hybrid screens to identify transcription factor binding to human DNA sequences. *J Vis Exp* 2019;**144**:10.3791/59192.
47. Gripp KW, Wotton D, Edwards MC. *et al.* Mutations in TGIF cause holoprosencephaly and link NODAL signalling to human neural axis determination. *Nat Genet* 2000;**25**:205–8.
48. Wotton D, Lo RS, Lee S. *et al.* A Smad transcriptional corepressor. *Cell* 1999;**97**:29–39.
49. Taniguchi K, Anderson AE, Sutherland AE. *et al.* Loss of Tgif function causes holoprosencephaly by disrupting the Shh signaling pathway. *PLoS Genet* 2012;**8**:e1002524.
50. Gongal PA, Waskiewicz AJ. Zebrafish model of holoprosencephaly demonstrates a key role for TGIF in regulating retinoic acid metabolism. *Hum Mol Genet* 2008;**17**:525–38.
51. Laue K, Pogoda H-M, Daniel PB. *et al.* Craniosynostosis and multiple skeletal anomalies in humans and zebrafish result from a defect in the localized degradation of retinoic acid. *Am J Hum Genet* 2011;**89**:595–606.
52. Bhatia S, Kleinjan DJ, Uttley K. *et al.* Quantitative spatial and temporal assessment of regulatory element activity in zebrafish. *eLife* 2021;**10**:e65601.
53. Baczyk D, Drewlo S, Proctor L. *et al.* Glial cell missing-1 transcription factor is required for the differentiation of the human trophoblast. *Cell Death Differ* 2009;**16**:719–27.
54. Bainbridge SA, Minhas A, Whiteley KJ. *et al.* Effects of reduced Gcm1 expression on trophoblast morphology, fetoplacental vascularity, and pregnancy outcomes in mice. *Hypertension* 2012;**59**:732–9.
55. Anson-Cartwright L, Dawson K, Holmyard D. *et al.* The glial cells missing-1 protein is essential for branching morphogenesis in the chorioallantoic placenta. *Nat Genet* 2000;**25**:311–4.
56. Schreiber J, Riethmacher-Sonnenberg E, Riethmacher D. *et al.* Placental failure in mice lacking the mammalian homolog of glial cells missing, GCMa. *Mol Cell Biol* 2000;**20**:2466–74.
57. Kamejima S, Tatsumi N, Anraku A. *et al.* Gcm1 is involved in cell proliferation and fibrosis during kidney regeneration after ischemia–reperfusion injury. *Sci Rep* 2019;**9**:7883.
58. Günther T, Chen Z-F, Kim J. *et al.* Genetic ablation of parathyroid glands reveals another source of parathyroid hormone. *Nature* 2000;**406**:199–203.
59. Yamada T, Tatsumi N, Anraku A. *et al.* Gcm2 regulates the maintenance of parathyroid cells in adult mice. *PLoS One* 2019;**14**:e0210662.
60. Guan B, Welch JM, Sapp JC. *et al.* GCM2-activating mutations in familial isolated hyperparathyroidism. *Am J Hum Genet* 2016;**99**:1034–44.
61. García-Castaño A, Madariaga L, Gómez-Conde S. *et al.* Five patients with disorders of calcium metabolism presented with GCM2 gene variants. *Sci Rep* 2021;**11**:2968.
62. Canaff L, Guarnieri V, Kim Y. *et al.* Novel Glial Cells Missing-2 (GCM2) variants in parathyroid disorders. *Eur J Endocrinol* 2022;**186**:351–66.
63. Hogan BM, Hunter MP, Oates AC. *et al.* Zebrafish gcm2 is required for gill filament budding from pharyngeal ectoderm. *Dev Biol* 2004;**276**:508–22.
64. Hanaoka R, Ohmori Y, Uyemura K. *et al.* Zebrafish gcm2 is required for pharyngeal cartilage formation. *Mech Dev* 2004;**121**:1235–47.
65. Stawicki TM, Owens KN, Linbo T. *et al.* The zebrafish merovingian mutant reveals a role for pH regulation in hair cell toxicity and function. *Dis Model Mech* 2014;**7**:847–56.
66. Cattenoz PB, Popkova A, Southall TD. *et al.* Functional conservation of the glide/Gcm regulatory network controlling glia, hemocyte, and tendon cell differentiation in drosophila. *Genetics* 2016;**202**:191–219.
67. Teng CS, Ting M, Farmer DT. *et al.* Altered bone growth dynamics prefigure craniosynostosis in a zebrafish model of Saethre-Chotzen syndrome. *eLife* 2018;**7**:e37024.
68. Bresciani E, Broadbridge E, Liu PP. An efficient dissociation protocol for generation of single cell suspension from zebrafish embryos and larvae. *MethodsX* 2018;**5**:1287–90.
69. Osterwalder M, Tran S, Hunter RD. *et al.* Characterization of mammalian in vivo enhancers using mouse transgenesis and CRISPR genome editing. *Methods Mol Biol* 2022;**2403**:147–86.
70. Reece-Hoyes JS, Diallo A, Lajoie B. *et al.* Enhanced yeast one-hybrid assays for high-throughput gene-centered regulatory network mapping. *Nat Methods* 2011;**8**:1059–64.