Introduction

Bone morphogenetic protein 4 (Bmp4) is a member of the transforming growth factor-beta (Tgfl) superfamily of secreted signaling ligands. The vertebrate Bmp4 and Bmp2 genes are close paralogs that are highly similar to the fly dpp gene (Wozney et al., 1988). Given the high amino acid identity between the mature peptides of human Bmp4 and Bmp2 (92%) and Bmp4 and Dpp (76%) (Kingsley, 1994), it has been proposed that these proteins could function interchangeably. In fact, Dpp protein can induce subcutaneous bone formation in rats similarly to Bmp4 and Bmp2 (Sampath et al., 1993). Likewise, expression of the human Bmp4 mature signaling peptide in fly embryos, in place of Dpp, is sufficient to rescue dorsal–ventral patterning defects of dpp null embryos (Padgett et al., 1993). Therefore, despite approximately 990 million years of cumulative evolution (Ureta-Vidal et al., 2003) the signaling functions of the mature Bmp4 and Dpp ligands appear strongly maintained.

Due to the evolutionary history of dpp and Bmp4 and the ability of their protein products to function interchangeably, their transcriptional regulation may also share similarities. The fly dpp gene is expressed in distinct embryonic regions and imaginal discs. Mutational and transgenic reporter analysis of dpp has revealed multiple tissue-specific transcriptional enhancers distributed throughout the dpp locus, with some elements clearly residing greater than 30 kb from the promoter (Blackman et al., 1991; Huang et al., 1993; Jackson and Hoffmann, 1994; Masucci et al., 1990; Spencer et al., 1982; St Johnston et al., 1990). Several vertebrate BMP family genes, including Bmp5, Gdf6 and Bmp2, contain similar arrangements of modular enhancers spread over large distances (Chandler et al., 2007a; DiLeone et al., 2000; DiLeone et al., 1998; Mortlock et al., 2003).

Bmp4 regulates multiple developmental processes, including dorsoventral patterning, gastrulation, and organogenesis (Hogan, 1996; Kingsley, 1994) and it displays numerous precise spatiotemporal expression patterns throughout development (Jones et al., 1991). The majority of Bmp4-null mouse embryos die early in development, mostly at the onset of gastrulation. Interestingly, these embryos fail to form mesoderm (Winnier et al., 1995). Some Bmp4−/− embryos that persist beyond this stage exhibit defects in mesoderm development, including abnormalities in extraembryonic and embryonic mesoderm tissues such as blood islands, allantois, ventral–lateral mesoderm, and primordial germ cells (Lawson et al., 1999; Winnier et al., 1995).

Further evidence for the importance of BMP signaling in mesoderm was shown by germline deletion of the BMP receptors, Bmpr1a or Bmpr2, or deletion of several downstream Smad factors, which generally result in gastrulation failure, lack of mesoderm, and/or profound defects in mesoderm tissues (Beppu et al., 2000; Chang et al., 1999; Lechleider et al., 2001; Mishina et al., 1995; Nomura and Li, 1998; Tremblay et al., 2001; Waldrip et al., 1998; Weinstein et al., 1998). Taken together, these data clearly demonstrate that BMP signaling is critical for formation and
development of extraembryonic and embryonic mesoderm and that Bmp4 is a key ligand driving these events. Bmp4 is also important for many aspects of organogenesis. For example, homozygous knockout embryos that survive beyond gastrulation have delayed liver bud morphogenesis (Rossi et al., 2001). Bmp4 haploinsufficiency can result in abnormalities in skeletal structures, kidney, seminiferous tubules, the urogenital system, eyes, craniofacial tissues, and pulmonary vascular smooth muscle (Dunn et al., 1997; Frank et al., 2005; Miyazaki et al., 2003). Conditional inactivation of Bmp4 in the developing heart revealed its requirement for atrioventricular septation (Jiao et al., 2003). Likewise, conditional gene inactivation studies demonstrated roles for Bmp4 in outflow tract septation and branchial arch remodeling (Liu et al., 2004), digit patterning (Selever et al., 2004) and distal lung epithelium (Eblaghie et al., 2006).

Since Bmp4 is expressed in a dynamic, spatiotemporal-specific manner throughout development, it is necessary to assay Bmp4 cis-regulation in vivo to obtain a complete view of these events. A 2.4 kb fragment encompassing the major Bmp4 promoter has been tested in transgenic mice (Feng et al., 2002; Zhang et al., 2002) and its activity compared to expression of the Bmp4 locus-driven knock-in reporter mouse (Lawson et al., 1999). While this fragment drove expression similar to endogenous Bmp4 in tooth ameloblasts and developing hair follicle shafts and matrix, it failed to drive expression in many sites of normal Bmp4 which is likely related to the mesodermal defects of human or pufferfish (Lawson et al., 1999). While this fragment drove expression similar to Bmp4 genomic sequences from human, mouse and pufferfish. We focused our efforts on the transcriptional activity of large, partially overlapping segments of DNA in mice using BAC reporter transgenes. We focused our efforts on the transcriptional activity of large, partially overlapping segments of DNA in mice using BAC reporter transgenes. We focused our efforts on the transcriptional activity of large, partially overlapping segments of DNA in mice using BAC reporter transgenes. We focused our efforts on the transcriptional activity of large, partially overlapping segments of DNA in mice using BAC reporter transgenes. We focused our efforts on the transcriptional activity of large, partially overlapping segments of DNA in mice using BAC reporter transgenes.

To perform comparative analyses, genomic sequences containing human or pufferfish Bmp4 and extending to adjacent 5′ and 3′ genes were obtained from the UCSC Genome Browser (http://genome.ucsc.edu) (Kent et al., 2002) May 2004 (hg17) human assembly and October 2004 pufferfish assembly. Genomic sequences corresponding to mouse Bmp4 BACs RP23-26C16 and RP23-145J23 were obtained from the UCSC Genome Browser May 2004 mouse (mm5) assembly.

To detect conserved elements, we first used PipMaker (Schwartz et al., 2000) to generate BLASTZ alignments with the mouse BAC sequences and human and pufferfish genomic sequences. Repetitive elements were pre-masked using RepeatMasker (Smit et al., 1996–2004). VISTA analysis (Mayor et al., 2000) was also used to confirm that the identified ECRs were in the same order and orientation in each species. ECR coordinates were defined from the BLASTZ alignments, as generated by PipMaker.

**Multi-sequence alignment and binding motif identification**

The orthologous sequences to mouse ECR2 in the pufferfish, zebrafish, chicken and human genomes were identified using the UCSC Genome Browser BLAT homology search tool (Kent, 2002). 1–1.5 kb of sequence spanning the ECR2 homology were obtained from each species. Coordinates of the sequences were: pufferfish (T. rubripes), chrUn:53534849–53536267, March 2006 (hg18) assembly. The MULAN multiple-sequence alignment tool (http://mulan.dcode.org/) (Ovcharenko et al., 2005a) was used to align these to a 688-nucleotide mouse genomic sequence containing ECR2 (corresponding to the longest sequence tested for enhancer activity; chr14:47056018–47056685, July 2007 (mm9) assembly). To find predicted transcription factor binding sites, the weight matrix-based MATCH tool and the TRANSFAC® Professional database (release 11.4) of transcription factors were utilized (Kel et al., 2003; Matys et al., 2003) via the BOCBASE online portal (https://portal.biobase-international.com). The matrix profile used was “vertebrate non-redundant minSUM” and the cutoff selection for the profile used was “minimize false positives” (minFP).

**BAC reporter transgenes**

Mouse Bmp4 BACs RPCI23-26C16 (227,097 bp) and RPCI23-145J23 (227,220 bp) were obtained from Children’s Hospital Oakland Research Institute (CHORI) (http://bapac.chori.org/) and verified using PCR and restriction enzyme digestion using standard procedures.

**Insertion of GFP-IRES-β-geo cassette into Bmp4 BACs**

BAC vectors were modified using homologous recombination in E. coli EL250 cells essentially as described (Lee et al., 2001; Mortlock et al., 2003) to contain a GFP-IRES-β-geo-SV40pA cassette inserted into the ATG start codon of the Bmp4 transcription unit (IRES = internal ribosome entry site; β-geo = lacZNeo fusion cassette). To generate the GFP-IRES-β-geo cassette, a PacI site was inserted was inserted upstream of the Kozak consensus of EGFP in pEGFP-C1 (kind gift of David Piston) and a PacI/Xhol fragment containing the EGFP open reading frame was inserted into pIBG-FTET (Chandler et al., 2007b) to create pIBGFTET. For simplicity, BAC RP23-145J23 and BAC RP23-26C16 were renamed 5′ BAC and 3′ BAC respectively. To generate the recombination cassette, 50 bp homology arms were designed to flank the Bmp4 start codon. Homology arm sequences (relative to Bmp4 coding strand) were as follows: for the 5′ arm, 5′-GTGT...ACACCA...CATC...-3′; for the 3′ arm, 3′-ATGAGTC...CATC...GT...-3′. Homology arm oligos were inserted into pGBGFET, and the final targeting cassette was gel purified. 250 ng of cassette were used for electroporation into recombination-competent EL250 cells containing...
Table 1

<table>
<thead>
<tr>
<th>Oligos used for BAC engineering</th>
</tr>
</thead>
<tbody>
<tr>
<td>galk replacement oligos</td>
</tr>
<tr>
<td>DelmECR1-F</td>
</tr>
<tr>
<td>DelmECR1-R</td>
</tr>
<tr>
<td>DelmECR2-F</td>
</tr>
<tr>
<td>DelmECR3-R</td>
</tr>
<tr>
<td>DelmECR2-R</td>
</tr>
<tr>
<td>RepmECR3-R</td>
</tr>
<tr>
<td>RepmECR3-F</td>
</tr>
<tr>
<td>RepmECR2-R</td>
</tr>
<tr>
<td>RepmECR1-R</td>
</tr>
<tr>
<td>galK replacement oligos</td>
</tr>
<tr>
<td>DelmECR3-F</td>
</tr>
<tr>
<td>DelmECR1-R</td>
</tr>
<tr>
<td>galK deletion oligos</td>
</tr>
<tr>
<td>DelmECR1-F</td>
</tr>
<tr>
<td>DelmECR1-R</td>
</tr>
<tr>
<td>DelmECR2-F</td>
</tr>
<tr>
<td>DelmECR3-R</td>
</tr>
<tr>
<td>DelmECR2-R</td>
</tr>
<tr>
<td>RepmECR3-R</td>
</tr>
<tr>
<td>RepmECR3-F</td>
</tr>
<tr>
<td>RepmECR2-R</td>
</tr>
<tr>
<td>RepmECR1-R</td>
</tr>
</tbody>
</table>

either the 5′ or 3′ BAC. BAC recombination and removal of the tetracycline resistance gene by FLPe excision were performed as previously described (Chandler et al., 2007b). Finally, pulsed-field and fingerprint gel analysis and sequencing across the recombination sites were performed to verify the final BACs were intact and the GFP-IRESClgeo-SV40pa cassette was inserted correctly.

Creation of ECR-deletion BACs

The 5′ and 3′ Bmp4 GFPlacZ-BACs were modified to generate three deletion BACs using galk counterselection methods (Chandler et al., 2007b; Warming et al., 2005). Primers were designed to include homology arms to target each ECR and annealing ends to amplify the galk cassette from pGalk (Warming et al., 2005) as shown in Table 1. This allowed seamless deletion of each ECR by replacement of subsequent removal of galk. The galk targeting cassettes were amplified by PCR and 250 ng of each were transformed into SW102 cells containing either the 5′ or 3′ Bmp4 GFPlacZ-BAC. Recombinant colonies were selected on M63 minimal media galactose plates at 32 °C for 3–4 days and restreaked onto MacConkey agar indicator plates with 1% galactose to verify galk-positive clones. Correct ECR replacement was verified by restriction digest with MluI, which cuts the galk cassette. To delete galk cassettes, 100-mer “replacement” oligos were used that contain the two 50-bp homology arms flanking the ECR. These were transformed into recombination-compotent galk-positive BAC cells and the cells were plated onto M63 minimal media plates containing 0.2% 2-deoxy-galactose (2-DOG) at 32 °C for 3 days to select for loss of galk (Warming et al., 2005). Surviving clones were verified to have deleted galk as described above.

EGR-1globinlacZ and ECR2-Hsp68lacZ constructs

EGR-1globinlacZ and ECR2-Hsp68lacZ constructs

Table 2

<table>
<thead>
<tr>
<th>Identity and mouse genome coordinates of noncoding evolutionarily conserved regions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Length (bp)</td>
</tr>
<tr>
<td>------------</td>
</tr>
<tr>
<td>ECR1</td>
</tr>
<tr>
<td>ECR2</td>
</tr>
<tr>
<td>ECR3</td>
</tr>
</tbody>
</table>

Mouse genotyping

Bmp4 BAC transgenic mice were identified by PCR. A triplex PCR was performed on tail DNA samples using primers to detect Neo (galk cassette), CamR (BAC vector), and Gdf5 (as genomic DNA control, present in both transgenic and non-transgenic mice). Primer sequences are as follows: for Neo, 5′-TTTTCTATTGCGCCTGCGC-3′ and 5′-ACGGCTTGGCCTCCATCGCA-3′; for CamR, 5′-GAAATCTGTTGATTATTATCCTTCCTC-3′ and 5′-CCCAATGCTGCCCAAGA-3′; for Gdf5, 5′-TTGCGCACTCCGAGAAGTAC-3′ and 5′-TGGAGAGAAAGAGA-3′. PCR cycling conditions were: 94 °C 5 min +98 °C 5 s initial denaturation step; 94 °C 30 s/60 °C 1 min/72 °C 40 s (10 cycles); 94 °C 30 s/56 °C 1 min/72 °C 40 s (25 cycles); 72 °C 5 min final step. Bmp4 BAC transgene copy numbers and integrity in founder animals and progeny were analyzed as previously described (Chandler et al., 2007a). Bmp4lacZ mice were genotyped by visualizing lacZ expression in hair follicles of X-gal stained tail snips (see below).
Transgene expression, embryo processing and imaging

X-gal staining of Bmp4 BAC transgenics and Bmp4\textsuperscript{lacZneo} embryos was performed exactly as described previously (Chandler et al., 2007a). X-Gal stained embryos were cleared by staging through increasing glycerol (15%, 30%, 50%, 70%, 90% glycerol with 1X phosphate-buffered saline (PBS) pH 7.4, then twice with 100% glycerol). Each wash was performed at room temperature with agitation until embryos sank. Digital images for whole mount embryos and sections were recorded using an Olympus SZX-ILLD2-100 stereomicroscope and Olympus BX51 microscope.

**Histology**

X-Gal-stained embryos in 100% glycerol were staged through a series of glycerol/ethanol mixtures with increasing ethanol then repeated 30 min incubations in fresh 100% Citrisolv (Fisherbrand) until embryos become very clear. Embryos were then incubated at 60 °C for 1 h in 50% paraffin (Paraplast Plus, McCormick Scientific)/50% Citrisolv, then in 100% paraffin overnight, then transferred to fresh 100% paraffin prior to embedding. 10 μm paraffin sections were counterstained with either eoisin or nuclear fast red (Vector Laboratories).

**Results**

Multiple noncoding evolutionarily conserved regions (ECRs) are present in a gene desert encompassing Bmp4

We hypothesized that Bmp4 is controlled by numerous cis-regulatory elements, many of which are distant from the promoter. Comparative analysis can be an effective way to identify such elements, although inter-mammal sequence comparisons across hundreds of kilobases tend to identify large numbers of ECRs of unknown significance (Nobrega et al., 2003). In contrast, mammal/fish genome comparisons have proven highly sensitive at detecting intact reading frames or homology to any known proteins or mRNAs (data not shown).

By using the UCSC Genome Browser BLAT search tool (Kent, 2002), we confirmed that each ECR is located in the same order and orientation relative to Bmp4 in all three species (Fig. 1). The relative spacing was compressed in Fugu, consistent with its compact genome. In both human and mouse, Bmp4 lies in a “gene desert” such that the adjacent protein-coding genes (Cdkn3 and Ddhd) on 5′ and 3′ sides, respectively) are separated from Bmp4 by several hundred kilobases. In the Fugu genome assembly, however, the orthologous Ddhd gene is located 5′ and the 3′ gene is Lbh, suggesting that rearrangement in flanking genes has occurred during vertebrate evolution. The arrangement of ECRs within the Bmp4 locus is therefore an ancient vertebrate feature.

Expression of Bmp4-GFPlacZ-BAC transgenics suggests that multiple long-range enhancers are present within the BAC interval

We then employed a BAC-based strategy to test large segments of DNA containing Bmp4 for regulatory activity. Two overlapping BACs (referred to here as 5′ BAC and 3′ BAC for convenience) were selected such that both contained the Bmp4 transcription unit and together spanned a 398 kb segment of mouse chromosome 14 (Fig. 2a). Homologous recombination was used to insert a GFP-ires-beta-gal reporter cassette into the Bmp4 ATG start codon in each BAC. This cassette was designed to allow independent co-expression of the GFP and lacZ reporters, while precluding expression of the mature Bmp4 peptide. The dual reporter cassette is functional as demonstrated by coexpression of both GFP and lacZ (Supplemental Fig. 1).

Multiple founder mice were identified for both the 5′ and 3′ Bmp4 GFPlacZ-BAC transgenics, and used to establish breeding lines. When evaluating BAC transgenic mice for cis-regulatory analyses, it is critical to confirm structural integrity of the transgene in the lines analyzed. We have previously reported a detailed analysis of transgene integrity and copy number analyses in the Bmp4 GFPlacZ-BAC transgenic lines described here (Chandler et al., 2007a). This allowed us to identify multiple lines for each BAC that (1) expressed lacZ robustly, (2) were PCR-positive for a complete set of transgene-specific markers across the BAC, and (3) carried multiple copies of the BAC transgene, a characteristic that is highly

![Diagram](Image)
correlated with BAC integrity (Chandler et al., 2007a; Gong et al., 2003).

Two and five lines meeting these criteria were generated for 5′ Bmp4 GFP-lacZ-BAC and 3′ Bmp4 GFP-lacZ-BAC, respectively. From each line, we collected transgenic embryos at three stages that largely span the onset and the completion of organogenesis (9.5, 12.5, 15.5 dpc), and assayed them for lacZ activity. (For the 5′ Bmp4 GFP-lacZ-BAC, a third founder transmitted the transgene only to a single
12.5 dpc embryo, data from which is included in Fig. 2a). X-gal staining patterns were documented and compared to those of Bmp4lacZ embryos. While strength of lacZ expression varied among lines, we have shown this was largely related to transgene copy number (Chandler et al., 2007a).

Since each BAC shared a common overlapping region of approximately 56 kb (Fig. 2a), we expected to see some patterns of expression that were common to both BAC transgenes, indicative of enhancer elements in the overlapping domain. In addition, each BAC contained approximately 171 kb of unique genomic sequence (Fig. 2a). Therefore, we expected each BAC transgene would also direct an additional unique set of expression patterns, indicating long-range enhancers 5′ or 3′ to Bmp4. Indeed, each BAC drove a distinct subset of endogenous Bmp4 expression patterns (listed in Fig. 2a) each of which was observed in at least two lines for each BAC. Significantly, no patterns of ectopic BAC expression were observed that clearly did not overlap an obvious Bmp4 domain. This suggested there were little or no position effect(s) impacting expression. The transgene-driven patterns at 9.5, 12.5, and 15.5 were identical across multiple lines for each BAC. We therefore analyzed expression in one representative line for each BAC (5′ GFP-IRESlacZ BAC line L1a and 3′ GFP-IRESlacZ BAC line L15a) at each gestation day from 6.5–15.5 dpc to gain more detailed data.

Each Bmp4 BAC clearly directed some shared expression patterns. For example, each BAC directed lacZ expression in the genital tubercle, digit tips, dorsal root ganglia and whisker hair shafts (Figs. 2b, c). In contrast, sometimes each BAC directed expression in the same organ, but in different patterns; for example, both BACs directed expression in the kidney at 15.5 dpc. However, the 5′ BAC directed expression in both kidney mesenchyme and epithelial cells while the 3′ BAC directed expression solely in epithelium (Fig. 2c). This suggests that separate cis-regulatory elements exist that control cell-type-specific kidney expression.

The 5′ Bmp4 BAC directed numerous sites of expression never seen in any of the 3′ Bmp4 BAC lines, but were recapitulated by the Bmp4 lacZ knock-in. These include expression in posterior lateral plate mesoderm, foregut, and outflow tract of the developing heart at 9.5 dpc (Figs. 2b, c). By 10.5 dpc in 5′ BAC line 1a, lacZ expression was detected in the forebrain and in limb bud, both in the apical ectodermal ridge (AER) and zone of polarizing activity (ZPA) (Fig. 2b). However, 3′ BAC age-matched embryos were devoid of these patterns. Later in development, the 5′ Bmp4 BAC drove expression in the distal epithelium of the branching lung from 12.5 (data not shown) to 15.5 dpc (Fig. 2c), as well as in the pelage hair follicles in a dramatic spotted pattern (Fig. 2b). In addition, the 5′ BAC alone directed expression in tooth, bladder, ventral pawpads, forebrain, bone, kidney mesenchyme, thymus, stomach and gut at 15.5 dpc (Fig. 2c and Supplemental material). Taken together, these results suggest multiple cis-regulatory elements controlling developmental patterns at 9.5, 12.5 and 15.5 dpc for comparison to full-length BAC reporter data. Analysis of multiple lines for ECR1 and ECR3 deletion BACs failed to reveal any alterations in staining patterns as compared to uncleaved BACs, nor was ectopic expression observed (data not shown), suggesting these are not obviously required for activating or repressing Bmp4 expression at these stages. To test these for enhancer potential, we also generated and analyzed transgenic lines carrying either ECR1 or ECR3 in the context of a β-globin promoter-lacZ minigene cassette (N=6 and N=4 lines, respectively). Neither of these constructs showed any reproducible LacZ expression in mouse embryos from 9.5–15.5 dpc. For the ECR2 deletion BAC (Deletion 2), two lines (L7 and L8a) were established. Both lines L7 and L8a had transgene insertions of more than 20 copies as estimated by quantitative PCR (not shown), and each also were positive for a set of BAC-specific markers as previously described (Chandler et al., 2007a). Expression in heart and foregut remained similar in 9.5 dpc 5′ GFPPlaccZ-BAC and Deletion 2 BAC L7 embryos (Fig. 3a). However, Deletion 2 BAC L7 embryos showed no lacZ activity in 9.5 dpc posterior lateral plate mesoderm, whereas age-matched 5′ GFPPlaccZ-BAC and Bmp4lacZ embryos clearly had reporter expression in this region (Fig. 3a). The selective loss of lateral plate mesoderm expression in Deletion 2 embryos was more clear at 10.5 dpc, when they retained expression in developing limbs and forebrain (Fig. 3a, arrowheads). Therefore, ECR2 seems critical for Bmp4 expression in lateral plate mesoderm at 9.5–10.5 dpc.

Although heart outflow tract expression was observed at 9.5 dpc in Deletion 2 BAC line L7, no expression in the heart outflow tract was observed at 9.5 dpc (Fig. 2a). X-gal staining from the representative BAC lines 5′ BAC L1a and 3′ BAC L15a, as well as age-matched embryos from the Bmp4 lacZ knock-in line. Insets in 5′ BAC 9.5 and 15.5 dpc embryos show close ups of heart outflow tract and forelimb bud. (c) Localization of selected lacZ expression sites in 5′ and 3′ BAC lines. Panels g, j, o and q are whole mount images while the rest are sections. All are 15.5 dpc except g’ and h’, which are 9.5 dpc. Arrowheads indicate localized X-gal stain. (a′–h′): X-gal-stained embryos from 5′ BAC line L1a reveal expression in (a′) lung epithelia, (b′) kidney epithelia and mesenchyme, (c′) whisker hair shaft (hs) and dermal papilla (dp), (d′) gut mesenchyme, (e′) upper tooth dermal papilla (left), lower tooth dermal papilla (right), (f′) rib bones (rb), (g′, h′) lateral plate mesoderm. (i–q′): X-gal-stained embryos at 15.5 dpc from the 3′ BAC line L15a reveal expression in (i′) pulmonary artery in lung, (j′) kidney epithelium but not mesenchyme, (k′) whisker hair shaft but not dermal papilla, (l′) dura mater (dm), (m′) craniofacial mesenchyme (cm) and whisker hair shaft (wh), (n′) roof palate mesenchyme (rp mes), (o′) vertebral column, where it is expressed in a segmented pattern along the vertebrae (vc), (p′) ventral ribs (vr), and the (q′) umbilical artery (ua). br = brain; sc = spinal cord.
was observed at 10.5 dpc in lines L7 or L8a (Fig. 3b) despite maintenance of this staining in 5′GFPlacZ-BAC embryos. Therefore, ECR2 may also be required for maintaining Bmp4 expression in heart beyond 9.5 dpc.

At earlier stages, Bmp4 is dynamically expressed in chorionic, amnionic and allantoic bud portions of the extraembryonic mesoderm and these were recapitulated by 5′GFPlacZ-BAC and Bmp4lacZneo embryos (not shown). At 7.5 dpc, Deletion 2 BAC embryos from both lines were lacking lacZ expression in chorionic and amnionic mesoderm, while expression in the allantoic bud was maintained (not shown).

ECR2-containing sequences are sufficient to direct mesoderm expression in mouse

We then tested ECR2-containing sequences for ability to direct mesoderm expression. Three ECR2-containing sequences of varying length (220, 467, and 668 bp; see Fig. 4a) were cloned into a lacZ minigene vector containing a minimal β-globin promoter (Maconochie et al., 1997), and these constructs were used to generate transgenic mouse embryos or lines (Fig. 4b). Five and three transiently-generated transgenic 8.5 dpc embryos were obtained with the 668 bp and 467 bp ECR2-β-globinlacZ constructs, respectively. Of these, 5/5 and 3/3 had lacZ expression in lateral plate mesoderm (Fig. 4b). This closely recapitulated expression directed by the 5′ GFPlacZ-BAC transgene (Fig. 4b). A separate construct containing the 668 bp sequence linked to an Hsp68 promoter-lacZ minigene also drove expression in lateral plate mesoderm, showing that the ECR2 enhancer can function in the context of a different promoter (N=2/2 transgenic embryos, although one embryo was highly mosaic; Supplemental Fig. 2 and data not shown). To examine an earlier stage when Bmp4 is expressed in extraembryonic mesoderm, we collected transgenic embryos at 7.5 dpc. Similar to the 5′ GFPlacZ-BAC, the 467 bp ECR2-β-globinlacZ transgene was sufficient to direct reporter expression in extraembryonic mesoderm in 5/5 transgenic embryos at 7.5 dpc (Fig. 4c).

However, 9/9 breeding transgenic lines generated from a 220 bp ECR2-β-globinlacZ transgene failed to drive reporter expression in either extraembryonic mesoderm or lateral plate mesoderm at 7.5–8.5 dpc (data not shown). Therefore, although deletion of the 220 bp sequence containing ECR2 from the BAC context resulted in partial ablation of extraembryonic mesoderm and complete loss of lateral plate mesoderm expression, this sequence alone is not sufficient to direct mesoderm expression. Other sequences contained in the 467 bp segment are therefore critical for full mesodermal enhancer activity.

Putative binding motifs for mesodermal factors in ECR2

Enhancer elements often contain multiple binding sites that allow a combination of transcription factors to bind the DNA and elicit or repress transcription of the target gene (Carey and Smale, 2000). To search in the ECR2 region for putative factor binding motifs in TRANSFAC, the weight matrix-based MATCH tool was utilized (Kel et al., 2003; Matys et al., 2003). This analysis revealed numerous potential binding motifs for vertebrate factors within the mouse ECR2-containing sequences tested by deletion or minigene assays. To filter these for factors of highest relevance to mesodermal regulation, a gene expression data query was performed using the Mouse Genome Informatics database (http://www.informatics.jax.org) (Eppig et al., 2005; Hill et al., 2004) for transcription factor genes expressed in embryonic or extraembryonic mesoderm during the developmental window when these structures first appear (6.25–8.0 dpc). The resulting list of genes (n=215) was compared to the potential binding motifs identified as described above. Five

Fig. 3. ECR2 is required for BAC-directed expression of lacZ in mesoderm. (a) X-gal stained 9.5 dpc 5′ GFPlacZ-BAC and Bmp4lacZneo embryos show expression in lateral plate mesoderm (arrowheads). In Deletion 2 BAC line L7 embryos the lateral plate mesoderm is ablated (middle panel). Both 5′ GFPlacZ-BAC and Deletion 2 BAC embryos show expression in outflow tract similar in pattern to that of Bmp4lacZneo embryos (arrows). (b) X-gal stained 10.5 dpc embryos from Deletion 2 BAC lines L7 and L8a show a loss of lacZ expression in posterior mesoderm (arrows) and in outflow tract as compared to 5′ GFPlacZ-BAC embryos. Expression in forebrain, limb buds and inner ear is similar in Deletion 2 BAC and 5′ GFPlacZ-BAC embryos.
mesodermal factors (Nfe2l1, Hand1, Zic3, Gata4, and Cdx1) had predicted binding motifs in the ECR2 region. Several of these factors are known to be critical for mesodermal development (see Discussion). Interestingly, this analysis predicted several multiple binding motifs for mesodermally-expressed factors in the core conserved 220 bp element we originally identified by mouse/fish comparisons.
as well as motifs in the extended regions shared by both larger fragments (467 bp, 668 bp). To visualize the extent of conservation in mesoderm-specific binding motifs, binding sites were annotated on the MULAN-generated alignment of the mouse 668 bp sequence to other vertebrates (Fig. 5). A total of 18, 11 and 6 predicted binding motifs for the mesodermal factors were found in the 668 bp, 467 bp, and 220 bp sequences respectively. The core 220 bp region of mammal/fish homology contains a cluster of binding motifs for Nfe2l1, Zic3, Gat4a, and Cdx1, several of which span highly conserved bases (Fig. 5). Interestingly, several predicted Hand1/47 heterodimer motifs were found in the region outside the 220 bp core but within the 467 bp sequence, suggesting these may be critical for enhancer activity.

**Discussion**

Here we have for the first time begun dissecting the greater cis-regulatory landscape surrounding Bmp4. Bmp4 has been implicated in many developmental processes that hinge on its patterned expression. Our findings will provide a valuable framework for further efforts to identify individual enhancers controlling a wide variety of Bmp4 expression domains, in teeth, lung, bone, and other organs and tissues. Since global deletion of Bmp4 causes early lethality, targeted deletion of regulatory elements at the endogenous Bmp4 locus may be an elegant alternate approach to Cre-based strategies for selectively removing Bmp4 expression in certain tissues/sites. Furthermore, we tested the hypothesis that sequence comparisons between mouse and fish would fine-tune the detection of cis-regulatory elements that might be critical for developmentally patterned Bmp4 expression. Using this approach, we were able to use a focused set of transgenes to pinpoint a key regulatory element that is likely directly related to its fundamental roles in mesodermal development. Specifically, Bmp4 expression within early mesoderm is critical for supporting primordial germ cells and establishment of the embryo/placental vascular connection through the allantois (Fujisawa et al., 2001; Lawson et al., 1999). Bmp4 expression in lateral plate mesoderm is also important for supporting the left/right asymmetry cascade by regulating Nodal (Fujisawa et al., 2002; Mine et al., 2008). ECR2 seems to coordinate Bmp4 expression in mesoderm during these events.

**The Bmp4 cis-regulatory landscape**

The diverse roles of Bmp4 are consistent with a model that it is embedded in a large ‘regulatory landscape’ of noncoding control elements spread around the flanking gene desert. Thus, prior studies that have focused on the proximal promoter region have only examined a very small portion of the regulatory elements needed for its diverse roles. Minimal Bmp4 promoter fragments have been previously tested in mouse, but failed to direct many known patterns of Bmp4 expression (Feng et al., 2002; Zhang et al., 2002) suggesting cis-regulatory elements reside beyond the minimal promoter. Our data strongly confirm that distant 5' and 3' intervals flanking Bmp4 harbor many unique cis-regulatory elements most of which are distant from the promoter. Remarkably, a territory of 56 kb around the promoter (shared by both BACs in this study) is insufficient to drive Bmp4 transcription in most of its normal expression sites during embryogenesis. Our results are in keeping with similar findings of large cis-regulatory domains for other BMP genes. Such domains are often associated with developmental regulator genes like transcription factors (Nobrega et al., 2003; Ovcharenko et al., 2005b). Several BMPs are now clearly in the category of genes embedded within large cis-regulatory landscapes. Like other regulatory genes, BMPs are “molecular toolkit” players that have been co-opted during evolution to have many context-dependent functions.

Comparative analyses were instrumental in our identification of the ECR synteny amongst vertebrates. This finding substantiates the hypothesis that Bmp4 resides in a “stable gene desert” (Ovcharenko et al., 2005a). The rearrangement of adjacent genes in pufferfish further supports the conservation of the Bmp4 linkage to the ECRs described here and lends weight to their suggested functional significance. Examination of Bmp4 flanking genes in the chick genome assembly also suggests that Dldhd is the 3' neighbor as in mammals, while the 5' genes are clearly different but still separated from Bmp4 by a large intergenic region (not shown).

Our transgenic evidence suggests that numerous cis-regulatory elements are located within either BAC clone. However, some sites of Bmp4 expression were not recapitulated by either the 5' or 3' BAC (see Results). This could in theory be due to the separation of cooperative elements that must work together to induce Bmp4 transcription in that particular cell type. Alternatively, since the BAC interval we tested only spans a portion of the gene desert surrounding Bmp4, elements required for Bmp4 expression in the eye or other structures may simply be located beyond the interval tested. In support of this, significant noncoding conservation is present in the desert outside these BACs (data not shown). These regions would be interesting to test in future studies.

**Potential impact of modular Bmp4 cis-regulation on discrete target tissues**

While global Bmp4 deletion causes embryonic lethality, tissue-specific variation in Bmp4 expression might have biological effects on any of the organs or tissues where Bmp4 is functional by altering BMP signaling output. Many cell types indicate exquisite sensitivity to BMP signaling levels, often with context-specific effects. Bmp4 haploinsufficiency causes a spectrum of sub-lethal developmental abnormalities in mice and/or humans including microphthalmia, digital defects and brain anomalies (Bakrania et al., 2008; Dunn et al., 1997). If genetic variants within Bmp4 cis-regulatory elements affected their function(s), the ensuing signaling effects would likely be restricted to certain organs or cell types that could predispose to specific birth defects or pathologies. Given the function of mesodermal Bmp4 in regulating left–right asymmetry, we speculate that ECR2 mutations may be involved in some human cases of situs inversus.

**Role of ECR2 in coordinating Bmp4 in mesoderm**

In early post-implantation mouse embryogenesis, Bmp4 is expressed in extraembryonic ectoderm followed closely by expression in extraembryonic and embryonic mesoderm (Lawson et al., 1999; Winnier et al., 1995). Interestingly, Bmp4 transcription in extraembryonic mesoderm may be induced by Bmp4 itself, produced by the adjacent extraembryonic ectoderm. Our results show that transcriptional activation of Bmp4 in extraembryonic mesoderm and ectoderm is controlled by distinct cis-regulatory sequences: ECR2, and a separate ectodermal element probably located outside the BACs we tested. Comparative analysis of genomic sequence surrounding Bmp4 in fish and mouse found three ancient noncoding sequences present in the BAC clones we tested (Fig. 1). Therefore, we hypothesized these functioned as tissue-specific enhancers. To test this, we deleted each ECR from its respective GPFlacZ-BAC and tested the Deletion BACs in vivo. Deletion of ECR2 resulted in partial ablation of expression in extraembryonic mesoderm, and complete failure to express and/or maintain expression in lateral plate mesoderm. Further minigene

---

**Fig. 5.** Multiple sequence alignment of ECR2-containing sequences as tested in Fig. 4. (220, 467, 668 bp), annotated with predicted binding motifs of mesodermally-expressed transcription factors according to expression data from the Mouse Genome Informatics database. Predicted motifs identified by MATCH analysis are underlined. Light gray and dashed bars respectively indicate the 467 bp and 220 bp mouse sequences.
analysis of ECR2 showed its enhancer function in these sites. Thus, ECR2 seems at least partly necessary, and sufficient, for expression in both extraembryonic and embryonic mesoderm. That the ECR2 BAC deletion only resulted in partial ablation of extraembryonic mesoderm expression might be due to the deletion design (see below). Alternatively, extraembryonic mesoderm is subdivided into distinctly specified tissues (amnionic, chorionic, and allantoic mesoderm) (Hogan et al., 1994) that may allow ECR2 to direct expression in the chorionic and amnionic portion of extraembryonic mesoderm, but not the allantoic portion. This suggests differential regulation of Bmp4 in distinct extraembryonic mesoderm compartments. In the future it may be useful to test larger segments spanning ECR2 by deletion analysis to rule out a requirement for the mammal-specific conserved sequences in regulating extraembryonic mesoderm expression. Modular elements in and around ECR2 may be required for controlling separate extraembryonic mesoderm domains (e.g. chorionic, amnionic, allantoic). Some elements may be partly redundant, as is the case for redundant cis-elements described in genes such as Shh (Jeong et al., 2006).

The 467 bp ECR2-β-globinlacZ transgene was able to direct extraembryonic mesoderm expression at ~7.5 dpc as well as lateral plate mesoderm expression at ~8.5 dpc. In contrast, a shorter 220 bp ECR2-β-globinlacZ transgene failed to direct any mesoderm expression. However, the same 220 bp deletion from the 5′ GFPlacZ-BAC transgene (Deletion BAC 2) abolished lateral plate mesoderm expression (Fig. 3). Taken together, this suggests there are critical binding sites needed for enhancer activity that reside in the additional sequence provided by the 467 bp fragment. We cannot rule out the possibility that the 467 bp fragment may actually represent two distinct enhancer modules for extraembryonic mesoderm and lateral plate mesoderm regulation. Although the 220 bp deletion resulted in complete loss of lateral plate mesoderm expression, partial expression in extraembryonic mesoderm (allantoic bud) was retained. This may be due to remaining functional enhancer-like sequences flanking the 220 bp element. This element was originally defined based on conservation between mouse and pufferfish, and did not include flanking sequences that contain inter-mammal conservation (Figs. 3 and 4). Therefore, some critical sequences necessary for ECR2 enhancer function in mouse reside beyond the confines of the pufferfish/mouse conservation. Although mammal/fish conservation has proven to be a beacon for identifying many noncoding enhancer elements, our results strongly suggest it can be advantageous to test large fragments containing the ECR in enhancer assays.

Motif analysis predicted Nfe2l1 (also known as Lcrf1/Tcf11) binding sites in the sequence shared by the ECR2 fragments tested in vivo. Interestingly, both embryonic and extraembryonic mesodermal formation is ablated in Lcrf1-null embryos, suggesting Nfe2l1/Lcrf1 is absolutely essential for mesoderm development (Farmer et al., 1997). In addition to Nfe2l1/Lcrf1, all three ECR2-containing sequences contained binding motifs for Cdx1, Zic3, Gata4 and Hand1. Cdx1 is expressed in mesoderm (Meyer and Gruss, 1993) although it is not required for early mesoderm development in mouse (Subramanian et al., 1995). Zic3 is expressed in embryonic mesoderm and primitive streak, but not in extraembryonic mesoderm (Elms et al., 2004). Zic3-null embryos exhibit variable phenotypes, including gastrulation defects, failure to develop mesoderm and defects in primitive streak patterning (Ware et al., 2006). Gata4 is expressed in mesoderm at 7.5 dpc (Saga et al., 1999) and approximately 33% of Gata4-null embryos fail to gastrulate (Molkentin et al., 1997). Finally, Hand1 is expressed in extraembryonic mesoderm and, later in development, in lateral plate mesoderm (Cserjesi et al., 1995). Hand1-null embryos also exhibit defects in extraembryonic mesoderm (Furulli et al., 1998). Given these results, we hypothesize a combination of sites present in the ECR2 element work cooperatively to elicit Bmp4 transcription in mesoderm. Future studies testing the functional significance of putative binding sites will allow researchers to understand what combination of factors bind ECR2 to enable Bmp4 transcription. We hypothesize that ECR2 is essential for Bmp4 expression in early mesoderm, and thus is probably critical for normal mouse development. Targeted mutagenesis and/or deletion of the endogenous ECR2 element will be required to conclusively test whether it functions non-redundantly.

**Transgenic approaches for defining Bmp4 cis-elements**

Interestingly, ECR2 can drive mesodermal expression in the context of either the minimal β-globin or Hsp68 promoter fragments (although only two transgenic embryos were generated with the latter). While both promoters are frequently used in enhancer assays, they are rarely tested in parallel and are rather different structurally. The β-globin promoter fragment is merely 51 base pairs spanning the TATA-like box and transcript start site (~40/+11 relative to the major start site). In contrast, the Hsp68 (official name: Hspa1a) promoter fragment is 878 bp (~652/+226 relative to its RefSeq annotated transcript; DPM, unpublished). Our results suggest that ECR2 contains sufficient regulatory motifs to activate either promoter with tissue and temporal specificity.

Deletion of ECR1 and 3 failed to reveal requirements for Bmp4 expression at 9.5–15.5 dpc. Minigene constructs also failed to detect enhancer activity for either ECR1 or ECR3 at 9.5–15.5 dpc, although earlier time points were not analyzed. Although each ECR is highly conserved, it is possible that they only function postnatally, or at prenatal time points not analyzed in this study, or they only function in context with other unknown Bmp4 enhancers (or each other) not tested in our assay, or they are not Bmp4 cis-regulatory elements. Because we tested only the minimally-defined ECR1 and ECR3 sequences (Table 2) in minigene transgensics, it is also possible that in each case additionally flanking sequence might be required for enhancer function, as seen for ECR2. While our deletion and minigene analyses could not test all potential hypotheses for ECR function, they successfully identified ECR2 as a critical Bmp4 cis-element active in development.

**Evolution of the Bmp2/4 gene pair and their mesodermal functions**

Bmp2 and Bmp4 are close paralogs that arose by duplication from an ancestral BMP gene early in vertebrate evolution. Bmp2 is also expressed in early extraembryonic mesoderm where it is required for closure of the proamniotic canal (Zhang and Bradley, 1996). Do Bmp2 and Bmp4 share an ancestral mesodermal enhancer? For a few other vertebrate gene pairs or clusters (e.g. Fox), paralogous enhancers have been found that duplicated along with the ancestral gene, probably early during vertebrate evolution (Lehoczyk et al., 2004). It is possible that ECR2 indicates a mesoderm enhancer that evolved before duplication of the ancestral Bmp2/4 gene. We could not detect noncoding conservation between vertebrate Bmp2 and Bmp4 loci using BLASTZ-based alignments (not shown). In vertebrates, Bmp2 may have evolved (or re-evolved) mesoderm regulation independently. Alternatively, a paralogous ECR2 enhancer may remain in the Bmp2 locus but has diverged too much at the sequence level to be alignable to Bmp4 ECR2. If so, this enhancer might be located in a similar position in the 5′ region of the Bmp2 locus as compared to Bmp4. Further insight into these questions could be gained by identification of Bmp2 mesodermal regulatory elements and determination of the signals directly controlling Bmp4 and Bmp2 in mesoderm.

**Conclusions**

Taken together, our results indicate that Bmp4 cis-regulatory elements are spread over a large genomic region. Among these elements a 467 bp noncoding DNA sequence is sufficient to function in a context-independent manner as a mesodermal enhancer and is
likely a critical Bmp4 cis-regulatory element. To our knowledge, this is the first tissue-specific Bmp4 enhancer identified far from the transcription start site (46 kb 5′ to the promoter). The significance of this ancient, long-range Bmp4 mesoderm enhancer is increased by the knowledge that Bmp4-null mice fail to develop mesoderm and, as a result, fail to complete embryogenesis (Winnier et al., 1995). ECRI2 is probably a critical genomic circuit that, through Bmp4, coordinates BMP signaling at several stages during mesoderm development. In addition to roles in mesoderm, Bmp4 has clearly been re-deployed in different ways to drive evolution of various organs, structures and morphology. Indeed, the modification of craniofacial Bmp4 expression probably shaped evolution of beak morphology in a subset of Darwin’s finches (Abzhanov et al., 2004). Our studies provide a first framework for teasing apart the cis-regulatory landscape of this critical gene.

Acknowledgments

We thank Yue Hou for technical assistance and members of the Mortlock Lab for discussions regarding this work. We kindly acknowledge Bridg Hogen for permission to use the Bmp4 knock-in mice and Mark de Caestecker for providing a breeding pair. We thank Trish Labosky and Ray Dunn for advice regarding early embryo dissections. We thank Maureen Cannon, Bridg Hogen, Trish Labosky, Kristina Roberts and Michelle Southard-Smith for helpful comments on the manuscript. Kelly J. Chandler was supported by NIH Genetics Training Grant 1T32GM07588 and by NIH Grant 1R01HD47880. Ronald L. Mortlock was supported by NIH Developmental Biology Training Grant 5T32HD07502. Douglas P. Mortlock was supported by NIH Grant 1R01HD47880. Transgenic mice were generated by the Vanderbilt University Transgenic and ES Cell Shared Resource, which is supported by the Vanderbilt Cancer, Diabetes, Kennedy and Vision Centers. We acknowledge the use of the VUMC CHGR DNA Resources Core Facility.

Appendix A. Supplementary data


References

Chandler was supported by NIH Developmental Biology Training Grant 1T32GM07588 and by NIH Grant 1R01HD47880. Ronald L. Mortlock was supported by NIH Developmental Biology Training Grant 5T32HD07502. Douglas P. Mortlock was supported by NIH Grant 1R01HD47880. Transgenic mice were generated by the Vanderbilt University Transgenic and ES Cell Shared Resource, which is supported by the Vanderbilt Cancer, Diabetes, Kennedy and Vision Centers. We acknowledge the use of the VUMC CHGR DNA Resources Core Facility.

Appendix A. Supplementary data


References


