Cis-regulatory elements controlling Gli3 expression during embryonic development

Thesis presented for the degree of PhD by Sarah Elizabeth Coy

The Centre for Developmental Genetics, Department of Biomedical Sciences, University of Sheffield

July 2008

Acknowledgements

Summary

Gli3 is an essential mediator of Sonic Hedgehog (Shh) response, acting both as a transcriptional activator and repressor. Mutations of *Gli3* are also responsible for a number of developmental defects characterised by mental retardation and skeletal abnormalities. In the early chick embryo, Gli3 is mainly expressed in the central nervous system (CNS), limb and paraxial mesoderm. Initially expressed in a widespread manner in these tissues, Gli3 mRNA become restricted by Shh from the notochord in the case of the neural tube and mesoderm, and from the zone of polarizing activity (ZPA) in the limb bud. Thus, a gradient of Gli3 expression is observed in relation to the Shh source. Wnt and BMP (bone morphogenic protein) pathways have also been implicated in Gli3 regulation. However, the mechanisms responsible for orchestrating Gli3 expression have not been investigated at the transcriptional level.

Here, I present a detailed analysis of the regulatory elements controlling Gli3 expression in the developing neural tube. In-silico analysis was used to determine the location of conserved non-coding elements, previously shown to contain enhancer modules. Eighteen non-coding elements were identified that are likely to regulate Gli3 expression. 5' RACE demonstrated that in mammals all putative enhancer elements are intragenic, owing to the identification of a novel untranslated exon upstream of the annotated transcript. The activity of each element was tested in vivo using chick neural tube electroporation. One element was of particular interest, since it drives reporter gene expression in a manner that mimics that of endogenous Gli3. Further analysis of this region established that it contains binding sites for, and is regulated by TALE family transcription factors. These findings implicate for the first time TALE family proteins in the regulation of Gli3 expression, and offer a mechanism for integrating Shh and BMP signaling in the regulation of this developmentally important gene.

Contents:

Figures:

Tables:

Chapter 1

Introduction

1.1 Neural tube development

The vertebrate neural tube is an epithelial structure running along the axial midline of vertebrates. It is divided into four portions along the Anterior-Posterior (AP) axis, which will develop into distinct regions of the CNS: the prosencephalon, the mesencephalon, the rhobencephalon and the spinal cord (Joyner, 2002). Here I focus on the region that will become the spinal cord

The neural tube forms by closure of the neural plate, which occurs in chicken embryos between Hamburger–Hamilton (HH) stage 9 – 12 (Hamburger and Hamilton, 1992). Two distinct signalling centres are established at opposite ends of the dorso-ventral (DV) axis. The floorplate is induced in the ventral neural tube by signals from an underlying mesodermal rod-like structure known as the notochord, whilst the roofplate is specified dorsally by signals from the overlying ectoderm (Jessell *et al.*, 1989; Placzek *et al.*, 1990; Price and Briscoe, 2004; Chizhikov and Millen, 2005). These structures secrete morphogens that determine the fate of initially unspecified populations of progenitors along the DV axes, whilst morphogens secreted from other tissues in close proximity to the neural tube provide positional information along the anterior-posterior (AP) axis (Fig. 1.1). By defining the transcriptional profile of progenitor cells, morphogen gradients specify cells to differentiate into discrete neuronal subtypes as the neural tube matures. Progenitor cells are maintained medially, in a region known as the ventricular zone. Lateral to this is the mantle zone, which harbours differentiating, and terminally differentiated neurons (Leber and Sanes, 1995). Along the neural tube, cells in the anterior are more mature, and are the first to differentiate.

1.1.1 Dorsoventral patterning of the neural tube

Progenitor domains can be distinguished by the expression of a characteristic combination of homeodomain (HD) and basic helix-loop-helix (bHLH) transcription factors (Briscoe *et al.*, 2000; Novitch *et al.*, 2001). Across the DV axis of the neural tube sub-populations of progenitor cells are formed in response to morphogen gradients emanating from the floorplate and roofplate (Fig. 1.2).

Sonic Hedgehog (Shh) is a secreted molecule expressed in the notochord and floorplate, and is required for specification of progenitor domains in the ventral neural tube (Chiang *et al.*, 1996; Ericson *et al.*, 1996). Graded activity of Shh in the neural tube has been demonstrated using various concentrations of the purified protein to elicit dose dependant response in intermediate neural plate explants (Yamada *et al.*, 1993; Ericson *et al.*, 1995; Marti *et al.*, 1995; Roelink *et al.*, 1995; Ericson *et al.*, 1997b; Briscoe *et al.*, 1999). These studies and others demonstrate that homeodomain proteins are induced at different concentrations of

Shh ligand, and specify neuronal identity (Briscoe *et al.*, 1999; Briscoe *et al.*, 2000). Indeed the Shh signal is sufficient for the specification of all ventral neuronal subtypes (Marti *et al.*, 1995; Roelink *et al.*, 1995; Ericson *et al.*, 1996; Ericson *et al.*, 1997b).

The roofplate secretes transforming growth factor β (TGFβ) proteins and Wnt proteins, both of which can act as morphogens. TGFβ family members, particularly BMP proteins, are important in the specification of dorsal neural type cell types, and appear to influence patterning in ventral regions by opposing the ventralising activity of Shh (Fig. 1.2; Liem *et al.*, 1995; Arkell and Beddington, 1997; Liem *et al.*, 1997). The TGFβ proteins BMP4, 5 and 7, GDF7, activin and dorsalin1 are all expressed in the roofplate (Lee *et al.*, 1998; Liem *et al.*, 2000). In vitro BMP proteins have been shown to alter the response of neural progenitor cells to a fixed concentration of Shh, causing a dorsal shift in the identity of neuronal progenitors, whilst downregulation of BMP signalling by exposure to Follistatin causes sensitisation to Shh signalling (Liem *et al.*, 2000). These data have been confirmed in vivo in the chick neural tube where BMP activity regulates the specification of dorsal neuronal subtypes, and in zebrafish embryos where depletion of BMP family members cause an expansion of ventral neurons (Barth *et al.*, 1999; Nguyen *et al.*, 2000; Timmer *et al.*, 2002). The TGFβ family member Nodal, expressed in the node and notochord, is also required for floor plate induction and its expression of Shh (Fig. 1.1; Sampath *et al.*, 1998; Muller *et al.*, 2000).

Early work implicated Wnt proteins in DV patterning of the neural tube (Wong *et al.*, 1994; Shimizu *et al.*, 1997; Robertson *et al.*, 2004). Consistent with this, several Wnt proteins are expressed during neurogenesis, and some Wnts are induced by BMP signalling (Chesnutt *et al.*, 2004). Wnt1 and Wnt3a are expressed in the roof plate of the vertebrate neural tube throughout neurogenesis (Parr *et al.*, 1993; Parr and McMahon, 1994; Hollyday *et al.*, 1995; Cauthen *et al.*, 2001; Robertson *et al.*, 2004), and neuronal Wnt3 expression is observed from an early stage, predominantly in the dorsal aspect of the neural tube (Roelink and Nusse, 1991; Salinas and Nusse, 1992; Bulfone *et al.*, 1993; Robertson *et al.*, 2004). Additionally Wnt1, Wnt4 and Wnt7a are expressed in the surface ectoderm (Parr and McMahon, 1995; Fan *et al.*, 1997a). Wnt1[-/-];Wnt3a[-/-] mice show a marked reduction in the numbers of neural crest cells, as well as defects in the specification of the dorsal most cells in the neural tube, accompanied by an increase in more ventrally located interneurons (Ikeya *et al.*, 1997; Ikeya and Takada, 1998; Muroyama *et al.*, 2002). Furthermore, Wnt signalling directly regulates the expression of Olig 3 and Nkx2.2, which are involved in determining the identity of neural progenitor cells (Lei *et al.*, 2006; Zechner *et al.*, 2007). However, the main function of Wnt signalling in the neural tube appears to be in regulating cell cycle

progression of neural progenitors (Dickinson *et al.*, 1994; Megason and McMahon, 2002; Zechner *et al.*, 2003; Chesnutt *et al.*, 2004).
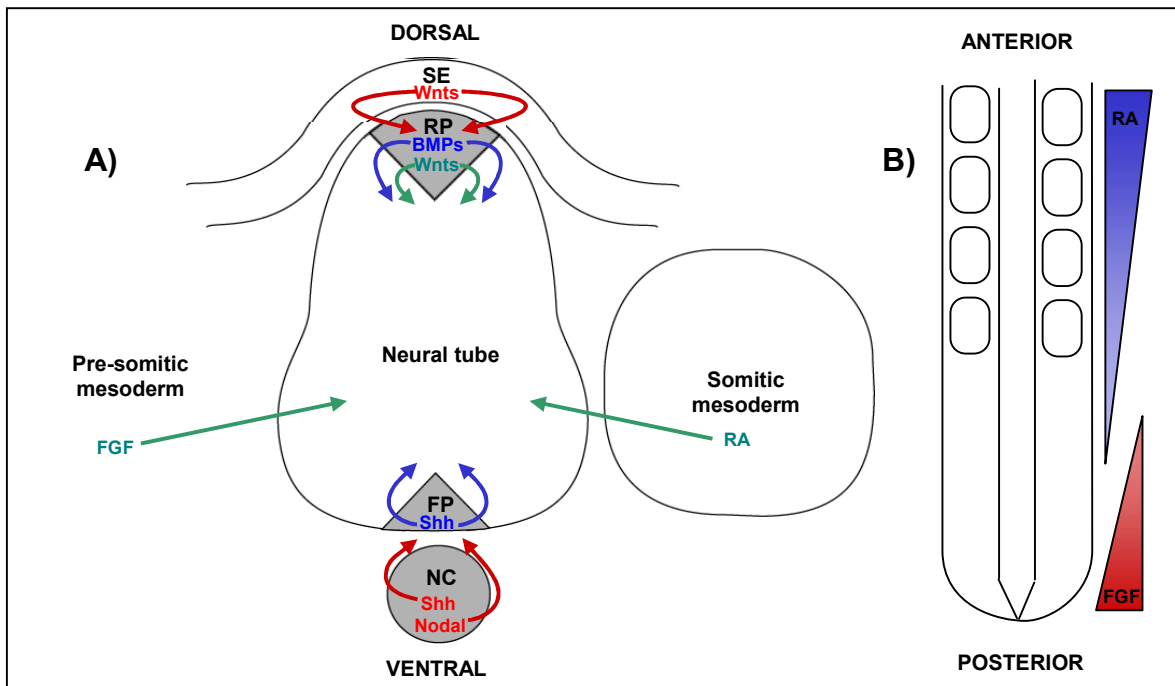
Figure 1.1: *Schematic diagram showing factors involved in neural tube patterning.* A) Cartoon showing the morphogens acting on the neural tube. Floorplate (FP) induction requires Nodal and Shh signalling from the underlying notochord (NC), whilst roofplate (RP) formation is induced by signals emanating from the surface ectoderm (SE), including Wnt signalling (red arrows). BMP proteins are involved in specifying neuronal subtypes in the dorsal neural tube, whilst a Shh emanating from the floorplate regulates neuronal identity in the ventral neural tube (blue arrows). In caudal regions of the neural tube FGF signalling represses neural differentiation and ventral neural genes. Retinoic acid (RA) is induced in the paraxial mesoderm around the time of somite formation, and promotes neuronal differentiation. Wnt signalling also regulates neuronal differentiation in the dorsal neural tube (green arrows). B) Cartoon showing the opposing gradients of FGF (red) and RA (blue) along the AP axis of the embryo. FGF is expressed in the caudal neural plate, paraxial mesoderm and regressing node. Retinoic acid expression is induced in the paraxial mesoderm around the time of somite formation. Along the AP axis, proliferation of neuronal subtypes is regulated by opposing gradients of FGF and Retinoic acid (RA) activity, the action of which is regulated in part by Wnt8a. In addition to their proliferative roles, RA and FGF gradients specify anterior-posterior identity by the regulation of Hox gene expression (not shown).

1.1.2 Anterior-posterior patterning of the neural tube

Patterning of the neural tube along the AP axis is regulated by opposing fibroblast growth factor (FGF) signals emanating from the regressing node, presomitic mesoderm and neural plate; and retinoids secreted from the somitic mesoderm and neural tube in response to retinoic acid (RA) signalling (Fig. 1.1; Diez del Corral *et al.*, 2003; Olivera-Martinez and Storey, 2007). Fgf8 is required to maintain a population of stem cells in the regressing primitive streak responsible for elongation of the body plan. In the neural tube, FGF signalling appears to repress the expression of Shh responsive proteins, thus maintaining a pluripotent state of progenitor cells (Novitch *et al.*, 2003). Conversely, RA signalling from the somitic mesoderm is required for neuronal differentiation to proceed, by inducing genes

that are required for the specification of the neural plate (reviewed in Diez del Corral *et al.*, 2003). The role of RA signalling in specifying AP identity is revealed upon mis-regulation. Excess RA causes a loss of anterior structures, and expansion of more caudal domains (reviewed in Maden 2002). In addition to their proliferative role, retinoids induce Class I proteins, which are repressed by Shh signalling (Pierani *et al.*, 1999; Novitch *et al.*, 2003; Wilson *et al.*, 2004). In the absence of RA signalling there is an expansion of ventral markers, at the expense of dorsal markers in the chick neural tube (Wilson *et al.*, 2003).

It has recently been shown that Wnt signalling mediates the transition from a progenitor to a proliferative state along the AP axis in response to FGF and Retinoic acid signalling (Olivera-Martinez and Storey, 2007). Wnt8c is expressed in the caudal neural tube from an early stage, and is regulated by opposing FGF and RA gradients in the paraxial mesoderm. In the presence of FGF signalling, Wnt signalling prevents premature differentiation. As FGF levels decline along the AP axis, Wnt is required to induce RA signalling, and thus promote differentiation (Olivera-Martinez and Storey, 2007). Thus, Wnt proteins appear to influence progenitor cell proliferation throughout the neural tube. They inhibit the generation of ventral neuronal subtypes by antagonising Shh signalling, and mediate the transition from a progenitor to proliferative state along the AP axis (Hollyday *et al.*, 1995; Megason and McMahon, 2002; Muroyama *et al.*, 2002; Chesnutt *et al.*, 2004; Robertson *et al.*, 2004; Lei *et al.*, 2006).

1.1.3 Specification of progenitor cell identity in the neural tube

Figure 1.2 shows a schematic diagram of neuronal cell specification across the DV axis of the neural tube. The ventral neural tube contains at least five progenitor domains, p0, p1, p2, pMN and p3, which differentiate to form V0, V1, V2, MN and V3 interneurons respectively (Ericson *et al.*, 1997a; Briscoe *et al.*, 1999). The transcription factors induced in ventral progenitors can be divided into two groups, according to their response to Shh. Class I proteins are repressed at distinct thresholds of Shh activity, such that Shh defines their ventral limits (Fig. 1.2; Briscoe *et al.*, 2000). Some, such as Dbx1 and Pax6 can be induced by retinoids (Gajovic *et al.*, 1997; Pierani *et al.*, 1999). Class II proteins are induced by Shh, and their dorsal expansion is defined by Shh levels. Cross-repressive interactions between complementary pairs of Class I and Class II proteins further helps specify their boundaries, and ensures that each progenitor domain expresses a unique combination of transcription factors (Ericson *et al.*, 1997b; Briscoe *et al.*, 2000; Sander *et al.*, 2000). Interestingly all Class 1/II transcription factors are HD proteins, with the exception of Olig2 which is a bHLH protein (Briscoe *et al.*, 1999; Novitch *et al.*, 2001). The Groucho family of co-repressors interact with Class I and Class II transcription factors, with the exception of Pax6,

to mediate their repressive activity (Muhr *et al.*, 2001). Pax6 instead represses Nkx2.2 indirectly and independently of Groucho (Muhr *et al.*, 2001; Briscoe and Novitch, 2008).

Neuronal specification in the dorsal neural tube is less well understood. Six progenitor domains (dp1-6) differentiate to produce six early born (dl1-6) and 2 lateborn (dlL^A and dlL^B) dorsal interneurons, which are categorised into Class A and Class B interneurons (Lee *et al.*, 2000; Chizhikov and Millen, 2005; Zhuang and Sockanathan, 2006). Class A interneurons (dl1-3) are roofplate-dependant, whereas Class B interneurons (dl4-6 and dlL$^{A/B)}$ are roofplate-independent (Lee *et al.*, 1998; Lee *et al.*, 2000; Liem *et al.*, 2000; Millonig *et al.*, 2000; Muroyama *et al.*, 2002; Zhuang and Sockanathan, 2006). TGF-β family proteins appear to be the major regulators of dorsal neural tube patterning, although Wnt signals emanating from the roofplate also play a role (Megason and McMahon, 2002; Muroyama *et al.*, 2002; Zhuang and Sockanathan, 2006; Zechner *et al.*, 2007). Contrary to ventral neural tube markers, Class A and B interneurons are specified predominantly by bHLH transcription factors. Distinct expression profiles can be used to identify individual progenitor populations in the dorsal neural tube. Class A progenitors express Olig3, which is proposed to specify Class A identity whilst repressing the Class B fate (Muller *et al.*, 2005; Zechner *et al.*, 2007). Post-mitotic Class B interneurons express Lbx1, which also represses Class A neuronal specification programmes. However, a universal determinant of Class B progenitor identity has not been found (Gross *et al.*, 2002; Muller *et al.*, 2002; Zhuang and Sockanathan, 2006). Math1, Ngn1/2 and Mash1 expressing progenitor cells give rise to dl1,dl2 and dl3-5 neurons respectively, although Mash1 and Ngn2, together with Olig3 are also expressed in some ventral cell types (Parras *et al.*, 2002; Takebayashi *et al.*, 2002). Additionally, dl3-5 express the HD protein Gsh2, and dl4/5 (but not dl3) express Gsh1 (Helms and Johnson, 1998; Gowan *et al.*, 2001). As in the ventral aspect, cross-repressive interactions are believed to refine progenitor domains. Although it has been shown that Ngn1 represses Mash1 and Gsh2 represses Ngn1, a transcription factor code for dl specification is not well defined (Gowan *et al.*, 2001; Chizhikov and Millen, 2005; Helms *et al.*, 2005; Kriks *et al.*, 2005). Some genes expressed in the dorsal neural tube, such as Ngn1, respond to specific thresholds of BMP signalling (Timmer *et al.*, 2002).
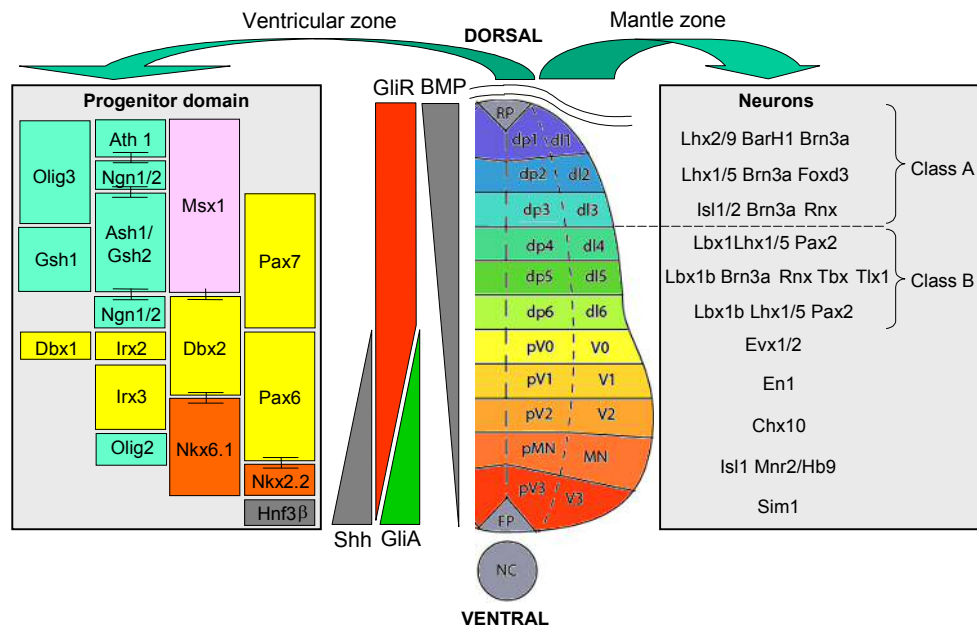
Figure 1.2: *Neuronal specification in the vertebrate neural tube.* Neuronal identity in the vertebrate neural tube is determined by cellular responses to secreted morphogens emanating from the ventral floorplate (FP) and notochord (NC), and from the dorsal roofplate (RP). Shh is expressed in the floorplate and notochord, and is involved in patterning the ventral neural tube. Shh signalling is transduced by activator or repressor forms of Gli proteins (GliA and GliR, respectively). BMP proteins expressed in the roofplate form a dorsal-high gradient that spans the entire neural tube. Progenitor cells are located in the medial ventricular zone, and terminally differentiated cells occupy the more lateral mantle zone. In ventral regions, progenitor domains pV0-pV3 and pMN give rise to V0-V3 interneurons and MN motor neurons respectively. High levels of Shh expression induce the expression of Class II homeodomain transcription factors in progenitor cells (orange boxes), whilst repressing Class I homeodomain transcription factors (yellow boxes). Cross-repressive interactions between complementary pairs of homeodomain proteins further define their expression boundaries. In dorsal regions, progenitor domains dp1-6 give rise to interneurons dl1-6 respectively. Dl1-3 represent Class A neurons that are dependant on roofplate signals, whilst dl4-6 are of the Class B family that are induced independently of the roofplate, and repressed by Olig3 (expressed in Class A progenitors). Dorsal progenitor subtypes are defined by basic Helix-loop helix family transcription factors (blue boxes), some of which display cross-repressive interactions. Progenitor domains are induced in response to different levels of BMP signalling, which is in part transduced by Msx homeodomain proteins (pink box, note that Msx2 is also expressed in the roofplate), and the bHLH proteins Ash1 and Ath1. Differentiated neurons can be distinguished by their expression of various combinations of transcription factors. (This schematic is adapted from figures presented in Wilson and Maden, 2005 and Briscoe and Novitch, 2008, see referenced therein. Additional information included in the schematic is mentioned in the text).

The bHLH proteins Cath1 (chicken atonal homologue 1) and Cash1 (chicken achaete scute homologue 1), as well as Msx HD proteins are also expressed in the dorsal neural tube (Fig. 1.2), and may define specific dorsal progenitor domains (Lee *et al.*, 1998; Ben-Arie *et al.*, 2000; Gowan *et al.*, 2001; Timmer *et al.*, 2002; Meyer and Roelink, 2003). BMP signalling defines the dorsal border of the Cash1 expression domain by repressing its expression at a threshold level of BMP activity (Lee *et al.*, 2000; Timmer *et al.*, 2002). Conversely, upregulation of BMP activity induces the expression of Cath1 throughout the dorsal neural tube, converting cells to a more dorsal fate. Msx2 expression, normally restricted to the roofplate, is induced throughout the neural tube upon electroporation of activated BMP

receptor, in agreement with vitro data (Liem *et al.*, 1995; Shimeld *et al.*, 1996; Timmer *et al.*, 2002). Similarly, the domain of Msx1 expression expands ventrally upon activation of BMP signalling, and is upregulated in endogenously expressing cells. Msx1 represses Dbx proteins, defining the dorsal limit of their expression (Timmer *et al.*, 2002). Repression of Shh signalling by electroporation of dominant negative Gli3 does not influence the expression of Msx1/2, Cath1 or Cash1, thus patterning of the dorsal neural tube appears to be independent of Shh signalling (Meyer and Roelink, 2003). BMP signalling appears to define neuronal identity in the dorsal neural tube by regulating protein intermediates that are not specific markers of neuronal identity.

1.2   Hedgehog signalling

1.2.1   Vertebrate hedgehog proteins

In vertebrates, Shh is one of three Hedgehog family members, each derived from a single gene present in Drosophila (Hedgehog, Hh) (Echelard *et al.*, 1993; Ingham and McMahon, 2001). In addition to neural tube patterning, Hh genes are involved in establishing left-right asymmetry along the midline, and anterior-posterior partitioning of the limb buds. They are critical regulators of proliferation, differentiation, tissue specification and morphogenesis, and mediate the development of organs such as the lungs, pancreas and eye.

Desert hedgehog (Dhh) shares greatest identity with Drosophila Hh, but has limited functions in vertebrates. Dhh is involved in germ-line development, formation of the peripheral nerve sheath and development of the eye. It is particularly important in spermatogenesis, and is expressed in the testis but not in the ovaries (Bitgood and McMahon, 1995; Bitgood *et al.*, 1996). Indian hedgehog (Ihh) is involved in the differentiation and proliferation of bone, cartilage, blood cells, mesenchyme, eye, gut and heart, and is involved in mediating left/right asymmetry (Bitgood and McMahon, 1995; Vortkamp *et al.*, 1996; Zhang *et al.*, 2001b). One of its main roles is in regulating the differentiation of chondrocytes during skeletal development (Vortkamp *et al.*, 1996).

The majority of functions performed by Hh in *Drosophila* have been retained by Shh, which is by far the most widely expressed of the three genes (Bitgood and McMahon, 1995). In the early vertebrate embryo, Shh is expressed in three key signalling centres: the zone of polarising activity (ZPA) in the limb bud, the floor plate and the notochord (Bumcrot *et al.*, 1995; Ingham and McMahon, 2001). Shh expressed in these centres acts as a morphogen, controlling cell proliferation and survival, cell differentiation and establishing axial patterning. Shh controls also left/right asymmetry (Levin *et al.*, 1995; Tsukui *et al.*, 1999; Zhang *et al.*, 2001b). It is also required for the correct formation of the somite derivatives, such as the axial skeleton and skeletal muscles (Fan and Tessier-Lavigne, 1994; Johnson *et*

al., 1994; Borycki *et al.*, 1998). In the limb bud, Shh plays a key role in regulating limb outgrowth and AP patterning (Riddle *et al.*, 1993; Marigo *et al.*, 1996; Ingham and McMahon, 2001). Its expression is required for organ morphogenesis of the heart, lung, prostate, gut and neural crest, and for the morphogenesis of hair follicles and feathers (Roberts *et al.*, 1995; Litingtung *et al.*, 1998; Chiang *et al.*, 1999; Chuong *et al.*, 2000; Berman *et al.*, 2004b; Washington Smoak *et al.*, 2005). Shh has multiple functions and can act either as a trophic factor controlling proliferation and survival, or as a morphogen controlling cell fate specification and differentiation (Ericson *et al.*, 1995; Borycki *et al.*, 1999; Briscoe *et al.*, 1999; Fan and Khavari, 1999; Rowitch *et al.*, 1999; Kenney and Rowitch, 2000; Charrier *et al.*, 2001; Kruger *et al.*, 2001; Dillon *et al.*, 2003). It is expressed in many adult tissues, notably the optic lamina, eye disc, gut, gonad, abdomen, and tracheal system (reviewed in Odent *et al.*, 1999; Ingham and McMahon, 2001, see references therein), and mis-regulation of Shh has been associated with the genesis or progression of multiple cancers (Fan *et al.*, 1997b; Berman *et al.*, 2002; Kenney *et al.*, 2003; Thayer *et al.*, 2003; Katoh and Katoh, 2005).

Loss of Shh activity in humans results in holoprosencephaly (HPE), a developmental defect resulting in severe facial abnormalities such as cyclopia and cleft lip and palate, and loss of ventral cell types in the central nervous system (Ingham and McMahon, 2001). HPE is a common cause of prenatal death, accounting for 1/250 induced abortions and 1/16000 live births (Cohen, 1989). Mouse embryos lacking Shh generally do not survive and die perinatally (Chiang *et al.*, 1996). Abnormalities include an indistinct midline, fused optic vesicles, a reduction in size of the brain and spinal cord, growth retardation, cranio-facial defects, severe skeletal defects including an absence of distal limb elements, and multiple organ abnormalities (Chiang *et al.*, 1996). In the neural tube most ventral cell types are lost, including motorneurons, interneurons and floorplate cells. Some V0 and V1 cells remain, suggesting that their development can be induced independently of Shh. This is thought to be due to redundancy with Ihh expressed in the underlying gut (Bitgood and McMahon, 1995). Smo null mice, which lack all Hh function, lack the ventral-most cell types in the neural tube (MN, V2 and V3 neurons) and die around E9.5 (Zhang *et al.*, 2001a; Wijgerde *et al.*, 2002).

1.2.2   The Hedgehog pathway

The Hedgehog pathway has been best-characterised in Drosophila, and appears to be highly conserved in vertebrates. One of the main differences is the duplication of the single transcriptional effector in Drosophila, Cubitus interruptis (Ci), into three vertebrate orthologues known as Gli1-3. Gli protein functions will be discussed in detail later.

However, it is worth noting that the cumulative function of Gli proteins, and the function of Ci, depends on post-translational modifications that produce activator or repressor forms of Gli family proteins. A schematic representation of the Hh pathway is shown in Figure 1.3.
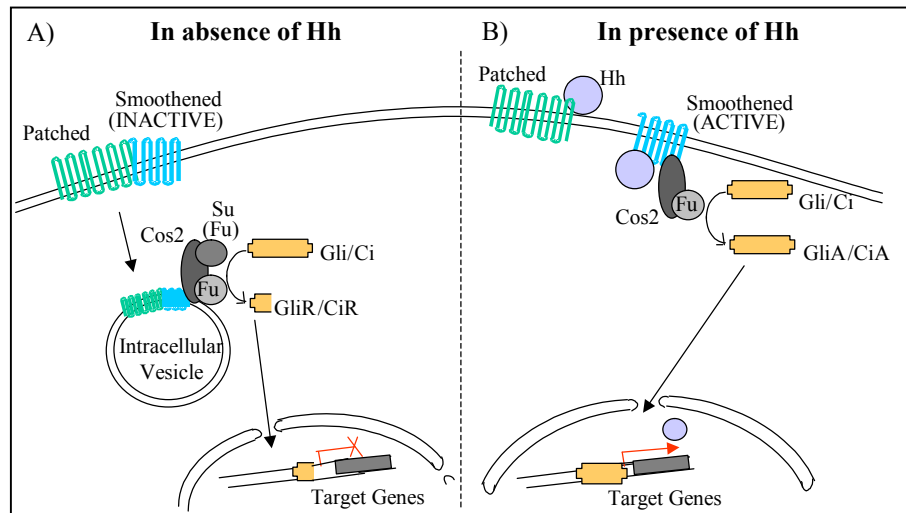
Figure 1.3: *A schematic representation of the Hh signalling pathway.* A) In the absence of Hh Patched inactivates Smoothened, and Gli/Ci transcription factors are processed into repressor forms, which repress the transcription of Hh target genes. B) In the presence of Hh the inhibition of Smoothened is relieved and Gli/Ci transcription factors are processed into activator forms, which activate transcription of Hh target genes.

In Drosophila, Hh itself is synthesised as a pro-protein, which upon cleavage produces an N-terminal fragment mainly bound to membranes (Lee *et al.*, 1994; Bumcrot *et al.*, 1995; Porter *et al.*, 1996). Hh release is thought to be aided by proteins such as Dispatched, and forms a gradient (Burke *et al.*, 1999; Caspary *et al.*, 2002; Kawakami *et al.*, 2002). Further modifications, including palmotyl and cholesterol binding alter the signal strength and range of action of Hh proteins. In the mouse limb, cholesterol modification is required for the long-range action of Shh signalling (Lewis *et al.*, 2001).

Hedgehog responsive cells express the protein Patched (Ptc) at their surface membrane (Chen and Struhl, 1996; Carpenter *et al.*, 1998). In the absence of Hh ligand, Ptc represses Smoothened (Smo), a 7 span transmembrane protein of the G-protein coupled receptor family, which is essential to the hedgehog response both in Drosophila and in vertebrates (Alcedo *et al.*, 1996; Chen and Struhl, 1996; Stone *et al.*, 1996; Alcedo *et al.*, 2000; Chen *et al.*, 2001). This inactivation affects the cellular processing of the Hh associated transcription factor Ci, which is responsible for the transcription of Hh target genes including Wg, Ptc and Hox genes as well as Hh itself. In vertebrates, over-expression of Smo results in the constitutive transcription of Hh target genes in vitro, but not when co-transfected with Ptc cDNA (Murone *et al.*, 1999; Hynes *et al.*, 2000). In Ptc null mice, Shh/Smo signaling is constitutively active (Goodrich *et al.*, 1997). Mice lacking the Smo protein have more severe

defects than those observed in Shh null mice, consistent with the Ptc-Smo pathway transducing all Hh signalling in vertebrates, including Dhh and Ihh response (Zhang *et al.*, 2001b).

Ci activator or repressor forms arise following post-translational modifications that depend on the presence of Hh. The complex that mediates Ci processing consists of Cos2, Fused (Fu) and Suppresser of fused (SuFu) (Monnier *et al.*, 2002; Jia *et al.*, 2003). Cos2 tethers the multiprotein complex to microtubules, and the complex directly associates with Ci (Robbins *et al.*, 1997; Sisson *et al.*, 1997; Wang and Holmgren, 2000; Lum and Beachy, 2004). This prevents translocation of Ci to the nucleus, whilst also acting as a docking point for Ci processing, which is initiated by phosphorylation by Protein kinase A (PKA), Glycogen synthsase kinase-β (GSK3β) and Casein Kinase I (CkI) (Zhang *et al.*, 2005). These kinases promote the processing of Ci into the 75kDa repressor form, which translocates to the nucleus where it inhibits transcription of Hh target genes (Fig. 1.3A; Alexandre *et al.*, 1996; Jiang and Struhl, 1998; Jiang, 2002). In the presence of Hh, the Ci/Cos2/Fu complex dissociates from SuFu, and moves to the membrane, where Cos2 associates with the cytoplasmic tail of Smo (Jia *et al.*, 2003; Ogden *et al.*, 2003). Ptc-mediated suppression of Smo is relieved, preventing the proteolysis of Ci. One mechanism by which the inhibition of Smo by Patched is relieved is by the targeting of Ptc to the proteosome upon Hh signalling (Nakano *et al.*, 2004). In the presence of Hh, full length (155kDa) Ci is converted into the activator form (Robbins *et al.*, 1997; Wang *et al.*, 1999; Methot and Basler, 2000), this results in the accumulation of Ci-activator in the nucleus, and transcription of Hh target genes (Fig. 1.3B; Wang and Holmgren, 2000).

Several differences have been identified between mammalian and Drosophila Hh signalling pathways (reviewed in Varjosalo *et al.*, 2006). One important difference is that mammalian Cos2 orthologues (Kif27 and Kif7, kinesin family) do not affect Shh signalling, the role of Cos2 appears to have been replaced by Su(Fu), that is essentially dispensable in Drosophila (Svard *et al.*, 2006; Varjosalo *et al.*, 2006). Mammals also have additional Hh binding proteins, such as Hip1 (Hedgehog interacting protein 1), which bind to secreted Hh and modify the shape of the gradient (Chuang and McMahon, 1999). Furthermore, mammalian hedgehog signalling has recently been shown to be reliant on cilia, and their intraflagellar transport proteins (IFTs; Scholey and Anderson, 2006). This is supported by the observation that Hh responsive cells are ciliated (Huangfu *et al.*, 2003). In the absence of IFTs, the functions of Gli2 and Gli3 are blocked, and cells are non-responsive to Shh. Furthermore, several members of the Hh pathway are enriched in cilia, including unprocessed Gli proteins and Su(Fu) (Corbit *et al.*, 2005; Haycraft *et al.*, 2005). Current models propose that in the presence of Shh, Smo is transported to the tip of the cilium, where it activates Gli2, and prevents processing of Gli3 to its repressor

form. In the absence of Shh, Ptc sequesters Smo, preventing it from entering the cilia (reviewed in Scholey and Anderson, 2006; Caspary *et al.*, 2007).

1.2.3 Gli proteins

1.2.3.1 Gli proteins are the vertebrate homologues of Ci

Gli family proteins derive their name from the initial identification of Gli1 in a <u>Gli</u>oblastoma (Kinzler *et al.*, 1987; Kinzler *et al.*, 1988). Gli2 and Gli3 were subsequently identified by cDNA hybridisation (Ruppert *et al.*, 1988). Homologues of the three Gli proteins have been identified in avians (Marigo *et al.*, 1996; Borycki *et al.*, 2000; Schweitzer *et al.*, 2000), in mammals (Kinzler *et al.*, 1988; Ruppert *et al.*, 1988; Hui *et al.*, 1994) and in Xenopus (Lee *et al.*, 1997). A high degree of homology with Drosophila Ci suggests that they are descendants of invertebrate Ci, having arisen by duplication, and the function of Ci appears to be distributed among them (Fig. 1.4). Indeed, expression of Gli1 and Gli3 transgenes in Drosophila can together rescue loss of Ci activity (von Mering and Basler, 1999). Each of the human Gli genes share 77-84% identity with Ci (Hui *et al.*, 1994).

The DNA binding domain of Gli proteins is located towards the amino terminus, and is composed of a cluster of five C2H2 zinc fingers (Fig. 1.4). Zinc fingers 3-5 are responsible for DNA binding and are highly conserved (Pavletich and Pabo, 1993). The DNA binding domain of human Gli3 is 95.6% identical to that of the chicken, 93% identical to that of Xenopus, and 63% of the residues share identity between Glis and Ci (Marigo *et al.*, 1996; Lee *et al.*, 1997). In mice, the DNA binding domain of Gli2 and Gli3 are 92% identical, each sharing 86-87% identity with Gli1 (Hui *et al.*, 1994).

While Gli1 and Gli3 recognise the same consensus site, 'GACCACCCA' with the same affinity in humans, Gli2 recognises a slightly different consensus, 'GAACACCCA' (Ruppert *et al.*, 1990; Tanimura *et al.*, 1998). The presence of these sequences in promoters of Shh target genes, suggests that Gli proteins are transcriptional effectors of Hh activity (Vortkamp *et al.*, 1995; Sasaki *et al.*, 1997; Dai *et al.*, 1999). Genes which have been shown to be directly regulated by Gli binding include Gli1, HNF3β, Ptch, Myf5, FoxE1, Hexokinase I (HKI), BCL2, and basonuclin (Bradac *et al.*, 1989; Sasaki *et al.*, 1997; Agren *et al.*, 2004; Cui *et al.*, 2004; Eichberger *et al.*, 2004; Ikram *et al.*, 2004; Regl *et al.*, 2004), any of these might act as intermediary's in regulating an indirect Shh response.
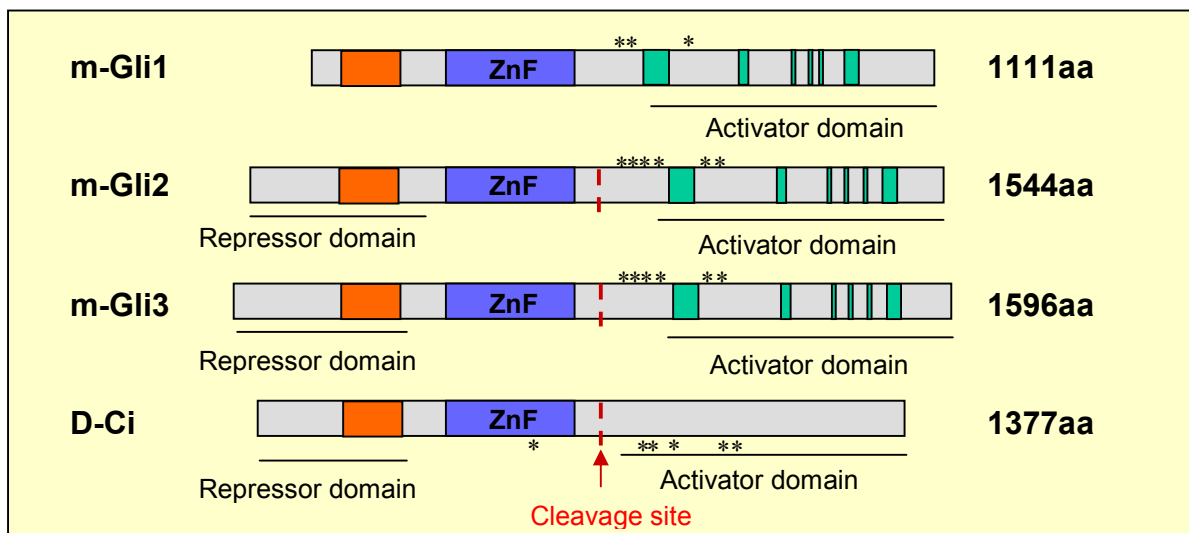
Figure 1.4: *Schematic diagram of the Gli protein family.* Gli2 and Gli3, along with Ci contain both activator and repressor domains, and can be proteolytically cleaved to yield a repressor protein that lacks the activator domain. Conversely, Gli1 cannot be cleaved, and does not contain a repressor domain. Sizes shown represent the full length proteins. Homology is greatest within the Zinc finger domain (ZnF, blue). Outside of this domain homology is limited, with the exception of a 55-57aa region in the N-terminus, in which 88% of residues are identical amongst mouse Glis, and 69% between Glis and Ci (orange). Mouse Glis contain a number of additional small conserved regions (8-44 aa) within the C-terminus (green). Asterisks show PKA phosphorylation sites. This figure is modified from (Matise and Joyner, 1999).

1.2.3.2 Evidence for activator and repressor function of Gli2/3

Investigation of repressor and activator functions demonstrates that whilst all three Gli proteins harbour activation domains in their carboxi termini, only Gli2 and Gli3 contain repressor domains (Fig. 1.4; Dai *et al.*, 1999; Ruiz i Altaba, 1999; Sasaki *et al.*, 1999). Gli2 and Gli3 transcriptional activity is governed by a similar mechanism to that of Ci. Indeed, ectopic expression of Gli proteins in the Drosophila embryo revealed that Gli2 and Gli3 can be processed to yield truncated forms in Drosophila (von Mering and Basler, 1999; Aza-Blanc *et al.*, 2000). Gli2 and Gli3 proteins contain six Protein Kinase A (PKA) phosphorylated sites, C-terminal to the zinc finger domain. Mutation of the PKA consensus sites or addition of PKA inhibitor prevents Gli3 processing in vivo, and stimulation of PKA activity triggers phosphorylation of Gli2 (Wang *et al.*, 2000). Like Ci, Gli3 interacts with vertebrate Su(fu) and the human homologue of fused, which mediate PKA phosphorylation of Gli3 in the absence of Shh (Pearse *et al.*, 1999). Gli2 cleavage requires further phosphorylation, and is independent of Hh signalling as demonstrated in *Drosophila* wing imaginal discs (Aza-Blanc *et al.*, 2000). Gli1 does not appear to be processed into a truncated form (Ruiz i Altaba, 1999), most likely because it lacks two of the six consensus PKA phosphorylation sites found in Gli2 and Gli3 (Fig. 1.4; Jiang and Struhl, 1998; Dai *et al.*, 1999; Sasaki *et al.*, 1999; Aza-Blanc *et al.*, 2000). Consistent with this, a non-cleavable form of Gli3, in which its C-terminal region is replaced by that of Gli1, has no repressor activity, confirming that cleavage of Gli3 is dependant on the C-terminal domain of Gli3 but

not Gli1 (Ruiz i Altaba, 1999). Supporting the processing of Gli2 and Gli3 into activator or repressor forms, the removal of the N-terminal repressor domain of Gli2 changes it into a more potent activator. Conversely, C-terminally truncated Gli2 and Gli3 peptides act as dominant-negative transcription factors (Ruiz i Altaba, 1999; Sasaki *et al.*, 1999).

1.2.4   Gli gene expression patterns

In general, Gli1 and Gli2 transcripts are observed close to the source of Hh expression, whereas Gli3 is expressed more distantly. Expression of all three Gli proteins is observed in endoderm, mesoderm and ectoderm derivatives. However, each Gli gene presents a distinct expression pattern. The differential expression of the Gli proteins is likely to play an important role in the diversification of developmental patterning.

1.2.4.1   Gli gene expression in the central nervous system

Gli genes have dynamic expression patterns along the AP axis of the neural tube as it matures. In E8.5 mice, all three Glis are expressed throughout the early neural plate (Lee *et al.*, 1997). In the anterior CNS, where neuronal tissue is more mature, Gli1 and Gli2 are excluded from the floorplate coinciding with the activation of Shh expression in this region. Gli1 expression is highest ventrally and gradually lost dorsally, whilst Gli2 is expressed uniformly throughout the neural tube/plate. In anterior regions, Gli3 expression is observed in an opposite gradient to that of Gli1 (Walterhouse *et al.*, 1993; Hui *et al.*, 1994; Lee *et al.*, 1997; Sasaki *et al.*, 1997). Expression of Gli1 and Gli3 is further restricted as development proceeds. By E10.5, Gli1 is only expressed in the ventral spinal cord near the floor plate. Gli2 is expressed throughout the spinal cord, except for the floorplate, its expression overlaps with both Gli1 and Gli3 expression domains. Gli3 expression forms a dorsal-high gradient spanning the DV axis (Bai and Joyner, 2001). Similar expression patterns have been reported in avian, Xenopus and zebrafish embryos (Lee *et al.*, 1997; Borycki *et al.*, 1998; Borycki *et al.*, 2000; Schweitzer *et al.*, 2000; Karlstrom *et al.*, 2003; Tyurina *et al.*, 2005).

1.2.4.2   Gli gene expression during somitogenesis

In the avian paraxial mesoderm, Gli genes are activated in co-ordination with somite formation. None of the Gli genes are activated in the presegmental mesoderm, and all become activated upon somite formation (Borycki *et al.*, 1998; Borycki *et al.*, 2000). Initially, Gli1 is expressed throughout the somite, but becomes restricted to a ventral-medial domain as somites mature. Upon differentiation Gli1 is expressed at high levels in the sclerotome, and at lower levels in the dermomyotome (Borycki *et al.*, 2000). Gli2 and Gli3 are initially expressed throughout newly formed somites, and then become restricted to the dermomyotome. Upon

differentiation, both genes are maintained in the myotome. However, whilst Gli2 is predominantly expressed in the medial myotome, Gli3 expression becomes more prominent in the lateral myotome (Borycki *et al.*, 2000; Schweitzer *et al.*, 2000).

Similarly, in mice Gli1 expression initiates at the time of somite formation. Gli2 and Gli3 are expressed slightly earlier, in the anterior pre-somitic mesoderm (psm). In newly-formed somites, Gli1 is strongly expressed in tissue adjacent to the notochord, and more diffusely in the dorsal sclerotome. As somites mature, expression expands both dorsally and laterally, but is excluded from the dermomyotome (McDermott *et al.*, 2005). Gli2 is expressed throughout newly formed somites, as the somite matures expression decreases in the sclerotome but is maintained at high levels in the myotome. At E10.5 expression is restricted to the dorsomedial lip of posterior, but not anterior somites. Gli3 is expressed laterally in newly formed somites, with weak expression in the sclerotome. As somites mature expression becomes restricted to the myotome and ventral dermomyotome, expression is also observed in the dorsomedial lip, but initiates slightly later than Gli2 expression. In E10.5 and E11.5 embryos expression maintained the dorsomedial lip and ventrolateral lip, but is weak in the medial myotome (Aruga *et al.*, 1999; McDermott *et al.*, 2005).

1.2.4.3 Gli gene expression during limb development

Limb bud expression of Gli genes has been investigated predominantly in mouse and chick embryos. In early limb buds (HH stage 20, mouse E 10.5), Gli1 is expressed in the posterior mesenchyme, whereas Gli2 and Gli3 are expressed broadly throughout the mesenchyme. All Glis are excluded from the ZPA (Zone of Polarising Activity). At later stages, Gli1 is expressed at low levels where Shh is most strongly expessed, and at high levels in the condensing mesenchyme (Marigo *et al.*, 1996; Masuya *et al.*, 1997; Mo *et al.*, 1997; Buscher and Ruther, 1998; Bai and Joyner, 2001). In the chick limb bud, a second proximal domain of expression is observed that is posteriorly restricted (Marigo *et al.*, 1996). Gli2 expression is upregulated between HH stages 22 and 24, but remains excluded from the ZPA and immediate surrounding area (Marigo *et al.*, 1996). In the limb buds of E12.5 mouse embryos, expression is confined to the region flanking the condensing mesenchyme (Buscher and Ruther, 1998; Bai and Joyner, 2001). Similarly, Gli3 expression in the limb bud is restricted as development proceeds. In chicken embryos, expression is progressively restricted from the posterior mesenchyme between HH stages 20-22, and a second domain of expression appears in the distal mesenchyme (Schweitzer *et al.*, 2000). By HH stage 24 only this distal domain remains (Schweitzer *et al.*, 2000). Likewise, by E12.5 murine Gli3 expression

becomes confined to a distal region surronding the condensing mesenchyme (Buscher and Ruther, 1998).

1.2.5 Mutant phenotypes reveal distinct functions of Gli proteins in the neural tube

Genetic analyses in the mouse have revealed the function of the Gli genes during development. These studies have also established that Gli proteins have redundant functions in the development of several organs. In particular, complimentary mechanisms have been found for Gli1 and Gli2 in the diencephalic region of the brain (Park *et al.*, 2000), Gli2 and Gli3 in somite derivatives (Mo *et al.*, 1997; Buttitta *et al.*, 2003; McDermott *et al.*, 2005), Gli1, Gli2 and Gli3 in lung development (Motoyama *et al.*, 1998; Park *et al.*, 2000), and Gli2 and Gli3 in tooth development (Hardcastle *et al.*, 1998).

In the neural tube, analyses of Gli mutant mice revealed specific defects in the patterning of various progenitor domains (summarised in Figure 1.5). In Shh$^{-/-}$ embryos, Class II progenitor proteins are not induced, and Class I proteins are not repressed, resulting in ventral expansion of their expression. As a result, embryos are cyclopic, fail to develop a floorplate, and lack ventral neuronal subtypes. Some pV0, pV1 and to a lesser extent pV2 domains remain, but all are reduced in number (Fig. 1.5 v; Chiang *et al.*, 1996; Litingtung and Chiang, 2000). Smo$^{-/-}$ embryos have a more severe phenotype, due to the concomitant loss of Dhh and Ihh signalling. This results in the loss of all ventral neuronal subtypes, demonstrating that Hh signalling is required for their induction. (Fig. 1.5 viii; Chiang *et al.*, 1996; Zhang *et al.*, 2001a; Wijgerde *et al.*, 2002)

Mice deficient in Gli1 do not display any phenotypic abnormality in neuronal specification, suggesting that other factors can compensate for its loss (Fig. 1.5 ii; Park *et al.*, 2000; Bai *et al.*, 2002; Bai *et al.*, 2004). Conversely, mice deficient in either Gli2 or Gli3 display severe developmental abnormalities. In the spinal cord of mice lacking Gli2 (specifically the DNA binding domain), the development of the most ventral neuronal cell types is disrupted (Fig. 1.5 iii). V3 neurons are reduced and the floor plate is absent, causing a ventral expansion of MNs (Ding *et al.*, 1998; Matise *et al.*, 1998). Gli2 null embryos also display reduced expression of the Hh target genes Ptc and Gli1, suggesting an activator role for Gli2 in regulating Shh target gene expression (Ding *et al.*, 1998; Bai *et al.*, 2002). Indeed, ventral neural tube defects observed in Gli2 null embryos can be rescued by replacing Gli2 with a Gli1 cDNA (Bai and Joyner, 2001; Bai *et al.*, 2002). Importantly, Gli2 function cannot be rescued by Gli1 in the absence of Shh, suggesting that in the absence of Gli2, the repressive function of Gli3 is dominant (Bai and Joyner, 2001).

Gli1 deficient mice are viable and do not display physiological or behavioural abnormalities (Park *et al.*, 2000). However, a requirement for this gene during normal development is revealed in Gli1 null embryos that additionally lack Gli2. Whilst Gli2 heterozygous embryos develop normally, severe defects are observed in Gli1$^{-/-}$;Gli2$^{+/-}$ embryos including small lungs, a distended gut and incomplete genitalia (Park *et al.*, 2000). This phenotype is similar to, but less severe than that of Gli2$^{-/-}$ single mutants, suggesting that Gli1 and Gli2 can compensate for each other in tissues where they are co-expressed (Mo *et al.*, 1997). However, some tissues display a dose-sensitive effect. For instance, different degrees of severity are observed in the lung phenotype as Gli1 and Gli2 alleles are deleted (Park *et al.*, 2000).

Gli3 acts predominantly as an inhibitor of Shh target genes in the neural tube. Gli3 null mice develop normally in the ventral neural tube, but display a reduction in the number of dorsal interneurons, specifically dl4 and dl5 neurons (Persson *et al.*, 2002). Concomitantly, V1, V0, and dI6 ventral interneurons expand dorsally into the domain normally occupied by dI5 neurons (Fig. 1.5iv). Thus, cells respond as if exposed to a higher concentration of Shh, such that progenitor cell patterning moves away from Hh secreting cells (Persson *et al.*, 2002). Expression of a Gli3 repressor construct from the endogenous locus rescues the neuronal phenotype, suggesting that the repressor form of Gli3 is sufficient for neuronal patterning (Persson *et al.*, 2002). However, mice homozygous for the Gli3 repressor allele die shortly after birth, with abnormalities that include an imperforate anus, abnormal gastrointestinal tracts and respiratory system, the absence of an adrenal gland, abnormal kidney development and skeletal abnormalities (Bose *et al.*, 2002). Therefore although Gli3 repressor can rescue the loss of Gli3 in the neural tube, in other tissues unregulated Gli3 repressor activity is not sufficient to substitute for the lack of full-length Gli3. This suggests that Gli3 has activator roles in some tissues, and further demonstrates that Gli activity is tissue dependent. In the neural tube, Gli3 repressor function is further demonstrated by overexpression of the Gli3 N-terminal repressor peptide. This results in a ventral to dorsal shift of progenitor cell identity and defects in motor neurons and V0-V2 interneurons (Fig. 1.5 ix). Class I progenitor domain proteins expand, whilst Class II proteins are repressed, including Nkx2.2, which leads to the loss of V3 interneurons (Meyer and Roelink, 2003). Dorsal most interneurons are not affected by upregulation of Gli3R, suggesting that these proteins are regulated by factors other than Shh signalling. Reduced expression of Ptc (a Hh target gene) in response to Gli3R is further evidence that Gli3 functions as a repressor of Shh target genes (Persson *et al.*, 2002).

Removal of Gli3 function in the absence of either Shh or Smo results in the recovery of MN, V2, V1 and V0 neurons, suggesting that Gli3 represses ventral cell fates in the absence of Hh signalling (Fig. 1.5 vi; Litingtung and Chiang, 2000; Wijgerde *et al.*, 2002). Importantly,

this indicates that one of the main functions of Shh is to prevent Gli3 processing. Moreover, V3 and floorplate cells are not recovered in the Gli3$^{-/-}$;Shh$^{-/-}$ double mutant mice, indicating that Shh is required to initiate their induction by Gli2 (Litingtung and Chiang, 2000). In Shh$^{-/-}$:Gli3$^{+/-}$ embryos there is less recovery than in the double homozygote, supporting a dose dependant mechanism of Gli3 function.

Compound Gli2;Gli3 homozygous mutant embryos fail to induce Gli1 expression and thus lack all Gli function (Lei *et al.*, 2004; McDermott *et al.*, 2005). Their phenotype reveals overlapping functions of Gli proteins in skeletal and lung development (Lei *et al.,* 2004). The absence of Gli1 expression in Gli2$^{-/-}$;Gli3$^{-/-}$ embryos, compared with its expression in mice homozygous for either single mutant, indicates that Gli3 activator function can compensate for the loss of Gli2 in the induction of Gli1 expression. However, comparison of Gli3$^{-/-}$;Gli2$^{-/-}$ and Gli2$^{-/-}$;Gli1$^{-/-}$ with single knockout embryos indicate that Gli2 activity is essential for floorplate formation (Park *et al.*, 2000). Similar to Gli3$^{-/-}$;Shh$^{-/-}$ or Gli3$^{-/-}$;Smo$^{-/-}$ embryos, those depleted of both Gli2 and Gli3 develop MN, V2, V1 and V0 neurons (Fig. 1.5 vii). However, in the absence of Gli function the spatial distribution of these cell types is disrupted. In Gli3$^{-/-}$;Shh$^{-/-}$ embryos, V2, V1 and MNs are intermingled, and progenitor domain boundaries are disrupted in Gli3$^{-/-}$ embryos (Litingtung and Chiang, 2000; Persson *et al.*, 2002). Despite this, individual cells retain discrete expression profiles of transcription factors associated with a particular neuronal cell type, suggesting that the role of Gli signalling is to determine progenitor cell identity at discrete positions along the DV axis, rather than inducing particular subsets of transcription factors (Briscoe and Novitch, 2008).

Together, the evidence presented above demonstrate that the role of Shh signalling in the neural tube is to establish a gradient of Gli activity by inhibiting Gli processing into repressors, and by inducing Gli activator function. This is summarised in Figure 1.6. Shh signalling induces the formation of Gli activator close to the Shh source, whilst inhibiting repressor function post-translationally, and perhaps the transcriptional level. In the absence of Shh expression, repressor function dominates and suppresses ventral cell fates. The cumulative activity of Gli proteins provides positional information to cells in a manner determined by the distance from the Shh source. Indeed, different levels of Gli activity recapitulate the response of cells in vitro to different concentrations of Shh (Stamataki *et al.*, 2005). In the vertebrate neural tube, Gli activator function is mainly provided by Gli2, whilst Gli3 functions as a transcriptional repressor of Shh target genes.
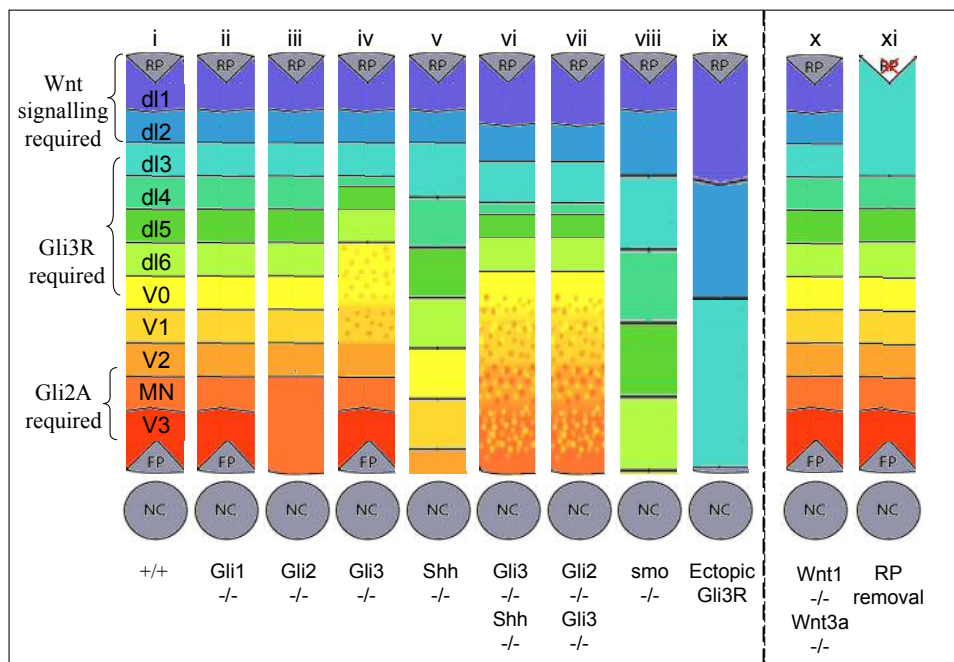
Figure 1.5: *Dorsoventral patterning in the neural tube is regulated by Shh signalling from the notochord and floorplate.* i, Neuronal domains are colour coded. In wildtype embryos (+/+) discrete domains are generated along the DV axis. ii-viii, Neuronal patterning along the DV axis of the neural tube of mice with different genetic mutations is shown. With the exception of Gli1, ablation of proteins involved in mediating Shh response in the neural tube results in the disruption of neural tube patterning. In some instances, specific neuronal subtypes are lost (as is the case for V3 neurons in Gli2$^{-/-}$ embryos). In Gli3$^{-/-}$ embryos discrete boundaries between domains are lost, resulting in an intermingling of several cell types. Neuronal progenitor cells in these embryos respond as if exposed to a higher level of Shh signalling. Notably, individual cells maintain their neuronal identity, as judged by the unique combination of progenitor proteins expressed. Intermingling of cell types is also observed in Gli2$^{-/-}$Gli3$^{-/-}$ and Gli3$^{-/-}$Shh$^{-/-}$ embryos. Loss of Gli3 partially recovers the phenotype of Shh$^{-/-}$, and Smo$^{-/-}$ embryos (not shown), for which single mutants display a dorsalisation of the neural tube. Ectopic expression of a repressor form of Gli3 has a similar effect (ix). x, mutations in Wnt1 and Wnt3a affect the specification of the dorsal most cell types. Removal of the roofplate has a similar effect (vi). This suggests that additional signals emanating from the roofplate control dorsal cell specification. (This figure is modified from schematics presented in Persson *et al.*, 2002; Jacob and Briscoe, 2003; Briscoe and Novitch, 2008, see text for references).
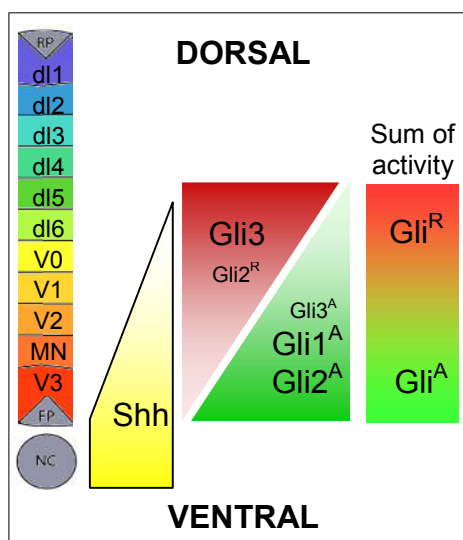
Figure 1.6: *Model of Gli activity in the neural tube.* Shh expressed in the notochord and floorplate foms a concentration gradient accross the ventral neural tube. At high concentrations of Shh Gli proteins are processed into activator forms. At low concentrations of Shh distal to the Shh source, Gli2 and Gli3 are processed into their repressor forms, resulting in an overall gradient of Gli activity. The combined transcriptional activity of Gli proteins results in a gradient of transcriptional activation in the ventral neural tube, and transcriptional repression in more dorsal regions. Adapted from (Jacob and Briscoe, 2003)

1.3   Transcriptional regulation of Gli3

Although significant progress has been made in our understanding of how Gli proteins function as transcriptional mediators of Shh signalling, little is known about how their expression is regulated. Gli binding sites in Gli1 regulatory sequences are required for its expression, and together with the absence of Gli1 expression in Gli2$^{-/-}$;Gli3$^{-/-}$ embryos, this suggests that Gli1 is a transcriptional target of Gli proteins (Lee *et al.*, 1997; Dai *et al.*, 1999; Ikram *et al.*, 2004; Lei *et al.*, 2004; McDermott *et al.*, 2005). In Shh$^{-/-}$ embryos low levels of Gli1 expression remain, suggesting that other Hh paralogues such as Ihh in the gut can induce Gli1 expression (Bai *et al.*, 2002). However, transcriptional control of Gli2 and Gli3, the main activator and repressor of Shh signalling have not been investigated. Gli3 is of particular interest because its misexpression is associated with a variety of human diseases, notably Pallister-Hall syndrome and Greig cephalopolysyndacyty syndrome (GCPS) (Kang *et al.*, 1997a; Kang *et al.*, 1997b; Kalff-Suske *et al.*, 1999). By understanding the mechanisms controlling Gli3 expression, we may gain an insight into the mechanisms underlying such diseases.

Several pathways have been implicated in the regulation of Gli3 transcription. Wnt signalling has been proposed to induce Gli3 expression, BMP signalling has been implicated in its maintenance, and Shh signalling has been suggested to restrict Gli3 expression. The evidence that led to these findings is outlined below. Importantly, none of these pathways have been shown to regulate Gli3 expression directly at the transcriptional level.

1.3.1   Wnt signalling is required for Gli3 initiation

The first evidence involving Wnt signalling in the activation of Gli3 expression came from embryological studies (Borycki *et al.*, 2000). Removal of surface ectoderm and axial tissues from unsegmented paraxial mesoderm causes a strong premature activation of Gli3 expression both in vitro and in vivo, demonstrating that Gli3 expression is a cell-autonomous property of segmental plate mesoderm in the absence of external stimuli. Gli3 repression in the presomitic mesoderm is restored by culturing segmental plate mesoderm explants lacking surface ectoderm in the presence of exogenous Wnt proteins. Further, this repression is also observed when explants lacking surface ectoderm are cultured in the presence of LiCl, which activates the canonical Wnt signalling (Klein and Melton, 1996; Hedgepeth *et al.*, 1997; Borycki *et al.*, 2000). Upon somite formation, Gli3 activation occurs in the presence of either ectoderm and neural tube tissue or in the presence of Wnt expressing cells, indicating that Gli3 is controlled by a β-catenin-dependent (canonical) Wnt signalling (Borycki *et al.*, 2000). More recently another study presented evidence that Wnt signalling may also control Gli3 expression in the developing neural tube (Alvarez-Medina *et al.*, 2008).

1.3.1.1 Canonical Wnt signalling

The canonical Wnt pathway is summarised in Figure 1.7. It is activated by proteins of the Wnt1 class, which comprises of Wnt1, Wnt3, Wnt3a, Wnt7a, Wnt7b, Wnt8a (Du *et al.*, 1995). Members of the Wnt5 class (Wnt4, Wnt5a, and Wnt11) are generally poor activators of the canonical pathway (Maye *et al.*, 2004).

Wnt proteins bind to the Frizzled (Fz) protein complex. Reception of the Wnt signal also requires the presence of the single span transmembrane protein Arrow in *Drosophila*, or LRP5 or LRP6 in vertebrates (reviewed in Logan and Nusse, 2004; Tamai *et al.*, 2004). Fz interacts with the ubiquitously expressed cytoplasmic protein dishevelled (Dsh), via a cytoplasmic motif (reviewed in Umbhauer *et al.*, 2000; Logan and Nusse, 2004).

In the absence of Wnt signalling, the cytoplasmic protein Axin associates with APC (Adenomatous Polyposis Coli) to form a β-catenin degradation complex that targets β−catenin, the main transcriptional activator of the Wnt signal, to proteosomal degradation (Latres *et al.*, 1999; Liu *et al.*, 1999b). In the nucleus, Lef1/Tcf transcription factors form a repressor complex that inhibits transcription of Wnt target genes (Fig. 1.7 A; Logan and Nusse, 2004).

Upon Wnt binding, Dsh becomes phosphorylated, and associates with Axin, which is also phosphorylated through its interaction with Fz/LRP (Mao *et al.*, 2001; Tolwinski *et al.*, 2003; Tamai *et al.*, 2004). This association prevents the formation of the β-catenin degradation complex, and levels of the transcriptional activator accumulate (Tamai *et al.*, 2004). Full length β-catenin localises preferentially to the nucleus where it activates Lef1/Tcf transcription factors that transcribe Wnt target genes (Fig. 1.7B; Behrens *et al.*, 1996; Molenaar *et al.*, 1996; van de Wetering *et al.*, 1997)

Lef1/Tcf protein family (Lymphoid enhancer-binding factor/T-cell factor) members are HMG box transcription factors (Clevers and Grosschedl, 1996). Alone they are poor transcriptional activators but upon binding of activator proteins they are converted into potent activators. Four homologues have been identified in vertebrates, Lef1, Tcf1, Tcf3 and Tcf4 (Eastman and Grosschedl, 1999). Lef1/Tcf proteins bind DNA via the HMG (high mobility group) domain, and in doing so bend the DNA helix (Giese *et al.*, 1992). Several different consensus DNA binding sequences have been published, such as the CTTTGWW sequence published by Eastman and Grosschedl (1999). Although Lef and Tcf family proteins share a high degree of conservation in the DNA binding HMG box domains, they have also been shown to bind sites that differ significantly from the consensus. DNA binding can be activated by proteins binding to sites adjacent to the Lef1/Tcf site, or by the binding of activator proteins such as β-catenin to the transcription factor itself (reviewed in Eastman and Grosschedl, 1999).

In the absence of Wnt activity, Tcf forms a complex with Groucho (Cavallo *et al.*, 1998), and represses Wnt target genes by histone modification, making DNA inaccessible to transcription factors (Chen *et al.*, 1999). Upon Wnt signalling, Groucho is displaced, and may be replaced by an activator such as the histone acetylase CBP/p300 (reviewed in Logan and Nusse, 2004). Lef1/Tcf activity is also controlled by a number of other factors such as Chibby (Takemaru *et al.*, 2003), ICAT (Tago *et al.*, 2000) and NLK/Nemo (Ishitani *et al.*, 1999).
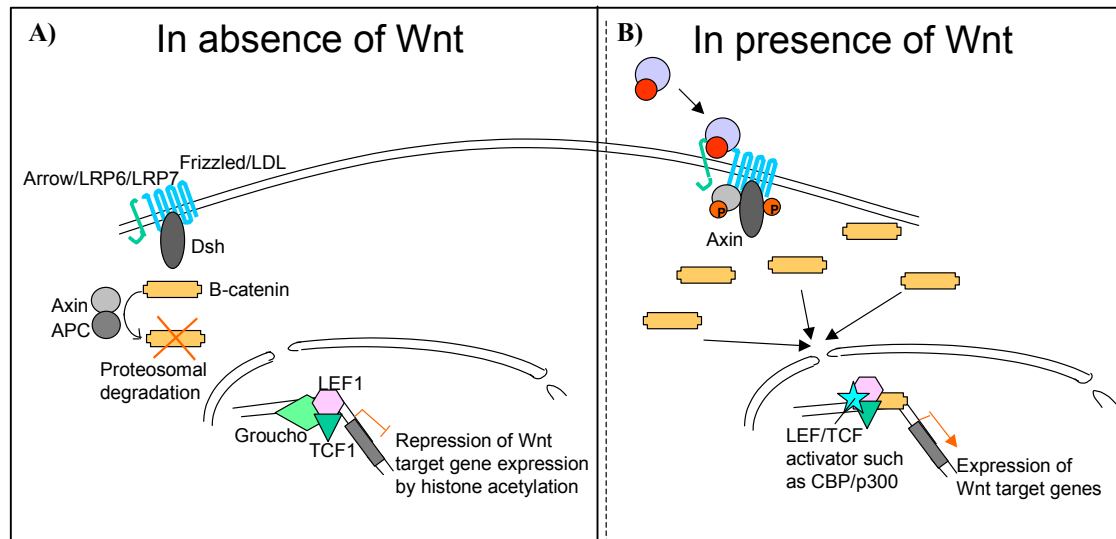
Figure 1.7: *Schematic representation of canonical Wnt signalling.* A) In the absence of Wnt signal unphosphorylated Axin associates with APC to form a β-catenin degradation complex. β-catenin cannot activate Lef1/Tcf, and instead repressor proteins such as Groucho inhibit Lef1/Tcf mediated transcription of Wnt target genes. B) Wnt binds Frizzled at the cell surface via the cysteine rich domain of Fz, in the presence of LRP/Arrow. Dishevelled (Dsh) bound to the cytoplasmic chain of Fz is phosphorylated, along with Axin, by its interaction with Fz. Phosphorylated Axin is no longer able to form the β-catenin degradation complex, and full length β-catenin accumulates. This is targeted to the nucleus where it activates Lef1/Tcf mediated transcription of target genes. Red circles represent post-translational modifications.

1.3.1.2 Further evidence for transcriptional regulation of Gli3 by Wnt signalling

Targeted mutations in mice show that Wnt1 is involved in midbrain patterning (McMahon and Bradley, 1990), Wnt3a in paraxial mesoderm formation (Takada *et al.*, 1994), and Wnt7a in limb formation (Parr and McMahon, 1995). Gli3 is also important in regulating development at each of these sites, such that phenotypes observed could be a consequence of Gli3 misexpression. Wnt1 and Wnt3a are expressed in adjacent cells at the dorsal midline of the developing neural tube (Ikeya *et al.*, 1997), consistent with the role of the neural tube in Gli3 expression in the paraxial mesoderm (Borycki *et al.*, 2000). Wnt1 is expressed dorsally in the early neural tube, adjacent and rostral to the last somite formed as neural tube closure continues caudally (Ikeya *et al.*, 1997; Capdevila *et al.*, 1998). Wnt3a is also expressed in the dorsal neural tube as it closes, at the level of somite formation (Capdevila *et al.*, 1998). Concomitant with somite formation there is a change in the expression profile of Lef/Tcf transcription factors both in mouse and chicken embryos (Galceran *et al.*, 1999; Galceran *et*

al., 2004; Schmidt *et al.*, 2004). Lef1 is expressed in the PSM and in the first somite, whereas Tcf3 is expressed in all somites but not in the PSM. Tcf3 is also expressed at earlier stages in the neural plate, offering a mechanism for Gli3 induction in neuronal tissue. Expression of Wnt8a in the caudal neural tube is regulated by opposing gradients of FGF and RA in the paraxial mesoderm, offering a potential mechanism for controlled initiation of Gli3 expression in neural tissue (Olivera-Martinez and Storey, 2007). Additionally, β-catenin, required for the activation of Lef/Tcf factors, is preferentially transcribed in the dorsal neural tube where Gli3 expression is greatest (Schmidt *et al.*, 2000). Robertson and colleagues have also shown that activation of canonical Wnt signalling results in an inhibition of the Shh response in neural tissue, which could be mediated by up-regulation of Gli3 repressor (Robertson *et al.*, 2004). Importantly, this is not due to a proliferative effect. In agreement with this finding, a reporter construct driven by canonical Wnt signalling is widely expressed in the dorsal neural tube of zebrafish embryos (Dorsky *et al.*, 2002). Finally, upregulation of the Shh antagonist Gas1 in response to Wnt signalling provides further evidence of a link between the two pathways, although stimulation of Wnt signalling in neural plate explants does not result in upregulation of Gas1 expression (Lee *et al.*, 2001; Robertson *et al.*, 2004).

1.3.1.3   Knockout studies

Wnt knockout mice display a wide range of defects from embryonic lethality to kidney and limb defects (Logan and Nusse, 2004). Wnt1[-/-] and Wnt3a[-/-] mice both have defects in neural crest derivatives, and Wnt3a[-/-] mice display paraxial mesoderm and tail bud defects. In the neural tube, Wnt1[-/-];Wnt3[-/-] mice present a reduction of dorsolateral neural precursor cells, and a reduction of neuronal crest cells (Ikeya *et al.*, 1997), similar to the phenotype observed for Lef1/Tcf1 knockout mice (Galceran *et al.*, 1999). More recent analyses showed that Wnt1[-/-];Wnt3a[-/-] embryos have a severe depletion of dI1 and dI2 interneurons, and a dorsal expansion of the dI3 class (Fig. 1.5 x; Muroyama *et al.*, 2002). Conversely, Wnt signalling induces dI2 interneurons and represses dI3 interneuron differentiation in vitro, without changing Bmp levels (Muroyama *et al.*, 2002). It would be interesting to test whether Gli3 expression is affected in these mutants, although neuronal patterning of the most dorsal neural tube does not appear to be affected in Gli3[-/-] embryos (Fig. 1.5 iv). Moreover, Wnt3a alone cannot induce or maintain Gli3 expression in intermediate neural tube explants (Meyer and Roelink, 2003), and removal of the roofplate results in a more severe phenotype than that observed in Wnt1[-/-];Wnt3[-/-] embryos (Fig. 1.5 xi), suggesting that other signals secreted

from the roofplate, such as BMPs, affect dorsal patterning in the neural tube (Lee *et al.*, 2000; Millonig *et al.*, 2000).

Single Lef1$^{-/-}$ and Tcf1$^{-/-}$ embryos do not have defects in mesoderm formation or somite patterning. Tcf1$^{-/-}$ impairs T cell differentiation and results in thymocyte differentiation defects and defects in limb development (Roose and Clevers, 1999). Lef1$^{-/-}$ mutants have defects in several organs, including hair follicles, mammary glands and teeth, and neural crest differentiation is disrupted (van Genderen *et al.*, 1994). The lack of mesodermal or paraxial mesoderm phenotypes in single mutant embryos indicates a redundant function between Lef1 and Tcf1. In agreement with this possibility, embryos expressing a dominant-negative Lef1 protein that can bind β-catenin but not DNA display a more severe phenotype than Lef1$^{-/-}$ mice, they present a severely malformed ribcage, suggesting that Lef1/Tcf transcription factors are involved in patterning of the paraxial mesoderm (Galceran *et al.*, 1999). Interestingly, expression of Pax1, a Shh target, is severely affected in both Lef1$^{-/-}$ and dominant negative mutants (Galceran *et al.*, 1999; Galceran *et al.*, 2004). Lef1$^{-/-}$Tcf1$^{-/-}$ embryos display a more severe phenotype than either single mutant. Somite formation is severely disrupted in these embryos, forming later, and with an abnormal morphology. Caudal regions are highly deformed, with multiple neural tube-like structures, reminiscent of the Wnt3a$^{-/-}$ phenotype (Galceran *et al.*, 1999). This is further suggestive that Wnt signalling is required to pattern the paraxial mesoderm. Furthermore, Tcf3$^{-/-}$ embryos have severe mesodermal structure defects, such as an expanded node and notochord, and a duplicated axis (Kim *et al.*, 2000; Merrill *et al.*, 2004). A similar duplicated axis phenotype results following overexpression of Lef1 in zebrafish embryos (Behrens *et al.*, 1996; Huber *et al.*, 1996), or early misexpression of Wnt8 in mouse embryos (Popperl *et al.*, 1997). Indeed, Tcf3 represses Wnt target genes in the absence of β-catenin (Brannon *et al.*, 1997; Roose *et al.*, 1998). This could be mediated by the association of Lef/Tcf complexes with known co-factors, which include Groucho and CBP (Cavallo *et al.*, 1998; Levanon *et al.*, 1998; Roose *et al.*, 1998; Waltzer and Bienz, 1998). Interestingly, the onset of Tcf3 expression in the somites co-incides with that of Gli3. It remains possible that in the context of Gli3 expression Tcf family members function indirectly. A possible mechanism for this is that the induction of Tcf3 expression causes repression of a factor that normally inhibits Gli3 expression, thus indirectly activating Gli3 transcription.

1.3.2   BMP signalling is required for the Maintenance of Gli3 expression

A gradient of BMP activity overlaps with the gradient of Gli3 expression in the dorsal neural tube (Liem *et al.*, 1995; Liem *et al.*, 1997; Barth *et al.*, 1999; Liem *et al.*, 2000). Likewise,

the expression of BMP family members is often associated with Gli3 expression in embryonic tissues, suggesting that BMP signalling could be a candidate for the regulation of Gli3 (Bitgood and McMahon, 1995; Faure *et al.*, 2002; Meyer and Roelink, 2003).

1.3.2.1 The BMP Pathway

BMP proteins are members of the TGFβ superfamily (Sekelsky *et al.*, 1995; Derynck *et al.*, 1996; Savage *et al.*, 1996; Miyazono, 1999). TGFβ proteins bind to Type I and Type II receptors at the cell membrane. Three Type-I receptors, Alk2, Alk3 (BRIa) and Alk6 (BRIb), and three Type-II receptors, BRII, ActRII and ActRIIB, are associated with BMP signalling in mammals (Nohe *et al.*, 2004). BMPs preferentially bind Type I receptors, associating with Type-II receptors with a lower affinity, furthermore individual BMPs display discrete binding affinities (ten Dijke *et al.*, 1994; Liu *et al.*, 1995; Rosenzweig *et al.*, 1995). Both Type I and Type II receptors contain intracellular serine/threonine kinase domains. Whereas Type II receptors are constitutively active, Type I receptors are activated by Type II receptors, which phosphorylate a conserved domain within BMP Type I receptors (Wrana *et al.*, 1994; Massague, 1998).

Intracellularly, Wnt signalling is transduced by members of the Smad (Sma and Mad) family of transcription factors. Eight Smad family proteins have been identified in mammals and are grouped into three classes according to their structure and function: receptor-regulated Smads (R-Smads), common-mediator Smads (co-Smads), and inhibitory Smads (anti-Smads). BMP signalling is distinguished from the signalling of other TGFβ family proteins by signal transduction via a specific set of R-Smads. Smad1, Smad5, and Smad8 are activated by BMP receptors, whereas Smad2 and Smad3 are activated by TGFβ and Activin receptors (Kretzschmar *et al.*, 1997; Nakao *et al.*, 1997b; Macias-Silva *et al.*, 1998; Tamaki *et al.*, 1998; Lebrun *et al.*, 1999; Miyazawa *et al.*, 2002). Anti-Smads (Smad6 and Smad7) are inhibitory, competing with R-Smads for binding to type I receptors. Smad7 inhibits both TGFβ activin and BMP signalling, whereas Smad6 preferentially inhibits BMP signalling (Hayashi *et al.*, 1997; Imamura *et al.*, 1997; Nakao *et al.*, 1997a; Souchelnytskyi *et al.*, 1998; Lebrun *et al.*, 1999; Hanyu *et al.*, 2001; Goto *et al.*, 2007) Smad4 is the only co-Smad identified in mammals (Hahn *et al.*, 1996).

Activation of BMP receptors results in the phosphorylation of R-Smads, which form hetero-oligomeric complexes with co-Smads that translocate to the nucleus and regulate the transcription of target genes (Kawabata *et al.*, 1998; Qin *et al.*, 2001; Wu *et al.*, 2001). Phosphorylated R-Smads can form oligomers and translocate into the nucleus in the absence

of co-Smad, but stabilisation of these complexes by co-Smads it thought to be necessary for efficient transcriptional activity (Lagna *et al.*, 1996; Liu *et al.*, 1997; Miyazono, 1999). Smad3 and Smad4 proteins bind DNA directly via their MH1 (Mad-homology 1) domains. Human Smad3 and Smad4 recognise an identical palindromic 8-mer sequence, GTCTAGAC (Zawel *et al.*, 1998), whereas Smad2 does not have a functional MH1 domain and is unable to bind DNA (Yagi *et al.*, 1999). The consensus sequence for the binding of Smads 1, 5, and 8 (the co-Smads activated by BMP signalling) remains to be determined (Miyazono, 1999). Smads can also interact with specific DNA sequences indirectly via other DNA-binding proteins, including both activators and repressors of transcription (Miyazawa *et al.*, 2002). For instance, FAST proteins play important roles in the transduction of certain Activin signals, as do AP1 family members (Chen *et al.*, 1997; Labbe *et al.*, 1998; Zhang *et al.*, 1998; Zhou *et al.*, 1998; Liberati *et al.*, 1999; Liu *et al.*, 1999a; Wong *et al.*, 1999; Inman and Hill, 2002). Upon ligand stimulation Smads 2 and 3 can also interact with transcriptional activators p300/CBP (Wotton *et al.*, 1999). Furthermore, Smad proteins have been shown to interact with β-catenin/ Tcf family members (Labbe *et al.*, 2000; Letamendia *et al.*, 2001; Hussein *et al.*, 2003), and an indirect association between Smad1 and STAT via p300/CBP suggests a possible interaction with the JAK/STAT signalling pathway (Nakashima *et al.*, 1999). Of particular interest, Smad2 has been shown to associate with the repressor form of Gli3. It remains unclear how such a complex might act, one possibility is that the Gli3-Smad complex may have novel binding or transcriptional specificity (Liu *et al.*, 1998). Known targets of BMPs identified in mammals include Smad6, Vent-2, Tlx-2 and Id (inhibitor of differentiation or inhibitor of DNA-binding) proteins 1-3 (Miyazono, 1999; Miyazono and Miyazawa, 2002). Induction of Id protein expression appears to be an important indirect mechanism through which BMPs are able to mediate transcription. By binding to the bHLH domain, Id proteins sequester ubiquitous bHLH transcription factors, and inhibit their transcriptional activity (Nakashima *et al.*, 2001; Goumans *et al.*, 2002). Recent evidence also suggests that BMP signalling activates the p38 MAPK pathway, by activation of Tak1/Tab1, and has also been shown to upregulate RAS and ERK (Nohe *et al.*, 2004).

1.3.2.2   Evidence for transcriptional regulation of Gli3 by BMP signalling

Interestingly, Gli3$^{+/-}$;Bmp4$^{+/-}$ embryos display a polydactyly phenotype more severe than that of single Gli3 or BMP heterozygote embryos (Dunn *et al.*, 1997; Aoto *et al.*, 2002). Additionally, in the absence of Shh, upregulation of Gli3 expression in the limb bud is accompanied by an increase in Bmp4 expression (Bastida *et al.*, 2004). This suggests that BMP4 might modulate Gli3 transcription.

In chick neural tube explants BMP4 is able to maintain Gli3 expression (Meyer and Roelink, 2003). Furthermore, upregulation of BMP expression in the chick neural tube results in a ventral expansion of Pax6 and Pax7, whose ventral limits are usually defined by Shh induced Nkx2.2. Thus, expansion of the Pax6 expression domain is consistent with an upregulation of Gli3R, which would antagonise Shh response in the ventral neural tube (Timmer *et al.*, 2002). Conversely, in the limb bud gain or loss of function of BMP4 expression does not significantly modify Gli3 expression or processing (Bastida *et al.*, 2004). Indeed, low levels of exogenous BMP4 caused a faint down-regulation of Gli3 expression in the distal limb mesoderm, suggesting a repressive role rather than maintenance. It is important to note that BMP4 expression domains in the limb bud vary significantly between chicken and mouse embryos. In mouse limb buds strong expression is observed in the posterior domain, whereas in the chick expression is higher in the anterior limb bud (Bastida *et al.*, 2004).

Redundancy between BMPs has impeded their investigation using mouse genetics (Dudley and Robertson, 1997; Katagiri *et al.*, 1998; Solloway *et al.*, 1998). Defects associated with single mutant mice include perturbed development of the kidney, eye and skeletal structures (Dudley *et al.*, 1995; Luo *et al.*, 1995; Winnier *et al.*, 1995; Miyazaki *et al.*, 2000). Disruption of Bmp6 or Bmp7, expressed in the roofplate, have not been reported to affect neurogenesis of the spinal cord (Solloway *et al.*, 1998; Muroyama *et al.*, 2002). Bmp4 embryos do not survive until a stage where neuronal development and protein function in the neural tube can be analysed (Winnier *et al.*, 1995; Muroyama *et al.*, 2002). However, BMPs are known to have an important role in patterning the neural tube (Liem *et al.*, 1997; Lee *et al.*, 1998; Chesnutt *et al.*, 2004; Ille *et al.*, 2007; Zechner *et al.*, 2007). Co-expression of BMP and Hh genes at many sites of the developing embryo, together with the ability of BMPs to maintain Gli3 expression in neural tube explants, suggests that BMP signalling may control Gli3 at the transcriptional level (Bitgood and McMahon, 1995; Meyer and Roelink, 2003).

1.3.3 Shh signalling represses Gli3 expression close to the Shh source

Mutually antagonistic interactions between Shh and Gli3 have been demonstrated in the limb bud (Masuya *et al.*, 1995; Buscher *et al.*, 1997; Schweitzer *et al.*, 2000), neural tube (Ruiz i Altaba, 1998) and somites (Borycki *et al.*, 2000). Shh represses Gli3 expression at the transcriptional level, and Gli3 represses the expression of Hh target genes, including Shh itself (Masuya *et al.*, 1995; Marigo *et al.*, 1996; Buscher *et al.*, 1997; Lee *et al.*, 1997; Ruiz i Altaba, 1998; Borycki *et al.*, 2000). This results in complimentary expression patterns, whereby Gli3 expression is usually repressed in regions where Shh targets are activated (Meyer and Roelink, 2003). The concentration gradient of Gli3 in the neural tube is the mirror image of the Shh

gradient, and in the limb bud Gli3 expression is greatest at a distance from the Shh source. This cross-repression is essential in the establishment of dorsoventral and antero-posterior patterning (Masuya *et al.*, 1995; Marigo *et al.*, 1996; Buscher *et al.*, 1997; Lee *et al.*, 1997; Ruiz i Altaba, 1998; Borycki *et al.*, 2000). In the absence of Shh signalling from the notochord, dorsalisation of Gli3 expression in the neural tube and somite does not occur (Borycki *et al.*, 1998; Borycki *et al.*, 2000; McDermott *et al.*, 2005), and in Shh[-/-] embryos Gli3 expression is maintained throughout the limb bud (te Welscher *et al.*, 2002). Misexpression of Shh in the chick limb has also been shown to down regulate Gli3 expression in nearby cells (Schweitzer *et al.*, 2000). Misexpression of two copies of Gli1 from the mouse Gli2 loci is embryonic lethal due to gain of function defects (Bai and Joyner, 2001). Reduced levels of Gli3 in these mice, but not in single Gli2 knockout embryos is further evidence that high levels of Shh activity antagonise Gli3 expression, but suggests an indirect mechanism (Bai and Joyner, 2001).

1.3.4 Model of Gli3 transcriptional control

Evidence to date suggests that in the neural ectoderm, Gli3 expression initiates prior to neural tube closure in response to Wnt signalling. It is repressed close to the Shh expressing notochord, and later the floorplate, resulting in a gradient of expression that opposes that of Shh activity. Repression is likely to be mediated by members of the Gli family. In the dorsal neural tube Gli3 expression is maintained by BMP signalling. Each of these pathways could act directly or indirectly to mediate Gli3 expression. The proposed mechanism for Gli3 regulation is summarised in Figure 1.8.
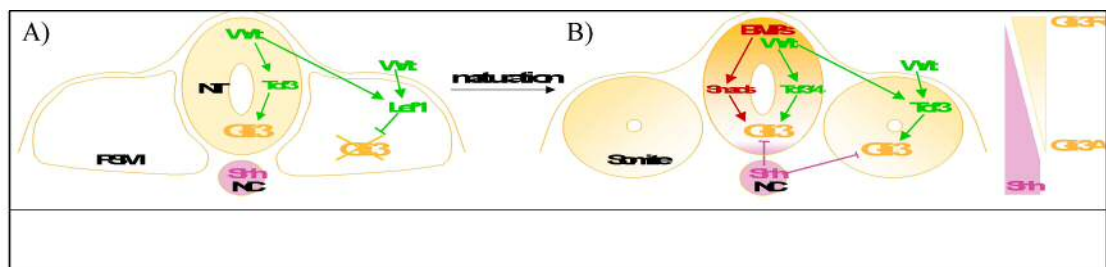
Figure 1.8: *Mechanisms for the regulation of Gli3 expression in the neural tube and paraxial mesoderm.* A) In the neural ectoderm Gli3 expression initiates prior to neural tube closure, whereas in the presegmental mesoderm Gli3 is not expressed. Wnt signalling emanating from the dorsal neural tube and surface ectoderm may regulate these expression patterns via the differentially expressed Lef1 and Tcf3 transcription factors. B) Upon somite formation Gli3 is initially expressed throughout the somite, initiated by Wnt signalling via Tcf3. As somites mature expression is repressed close to the Shh expressing notochord, resulting in graded expression with highest levels dorso-laterally. In the neural tube dorsal restriction also occurs, forming a gradient opposing that of Shh activity. Thus Shh is proposed to repress Gli3 expression. BMP signalling, emanating from the dorsal neural tube, is proposed to maintain Gli3 expression dorsally.

1.4   A detailed map of Gli3

Prior to initiating work on the transcriptional regulation of Gli3, I constructed a detailed map of the Gli3 locus and of the protein domains that regulate its activity (Fig. 1.9). I also reviewed publications that have reported genetic characterisation of Gli3 mutant phenotypes to establish whether such phenotypes are restricted to mutations within the known coding region.

1.4.1   Genomic structure of Gli3

Gli3 was originally identified by cDNA hybridisation to Gli (Ruppert *et al.*, 1988), and was shown to encode a 190kD protein from a 4797bp open reading frame on the human chromosome 7p13 (Ruppert *et al.*, 1990). The open reading frame identified encodes a transcriptional start site, which is preceded by a stop codon in the same reading frame 12 bases upstream. At the 3' end of the transcript a putative polyadenylation signal was identified that functions in vitro (ATTAAA, Fig. 1.9 B; Wilusz *et al.*, 1989; Ruppert *et al.*, 1990). Analysis of the human genomic Gli3 region shows that it contains at least 14 exons spread over a distance of 280kb (Fig. 1.9 A and B; Vortkamp *et al.*, 1994; Kang *et al.*, 1997b). Northern blot analysis suggested an approximately 3.5kb non-coding 5' region that has not been characterised, indicating that non-coding exons might exist upstream to those identified (Ruppert *et al.*, 1990; Vortkamp *et al.*, 1995). Indeed, Wild and colleagues have identified an additional 5' exon in the mouse transcript, which they have designated 'exon 0' in accordance with the nomenclature of the human transcript (Fig. 1.9B; Wild *et al.*, 1997). It is unclear whether exon 0 is unique to mouse, or present in other species. It might represent the true 5' end of the Gli3 transcript, or it might represent an alternative transcriptional start site that contributes to the regulation of Gli3 expression. In the absence of a characterised promoter, it is difficult to predict where the true 5' end of the transcript might lie. A 2kb CpG island has been identified upstream of the coding transcript which might mark the endogenous promoter, and five segments of DNA conserved between human and mouse might act as control elements (Fig. 1.9B; Vortkamp *et al.*, 1994; Vortkamp *et al.*, 1995; Wild *et al.*, 1997).
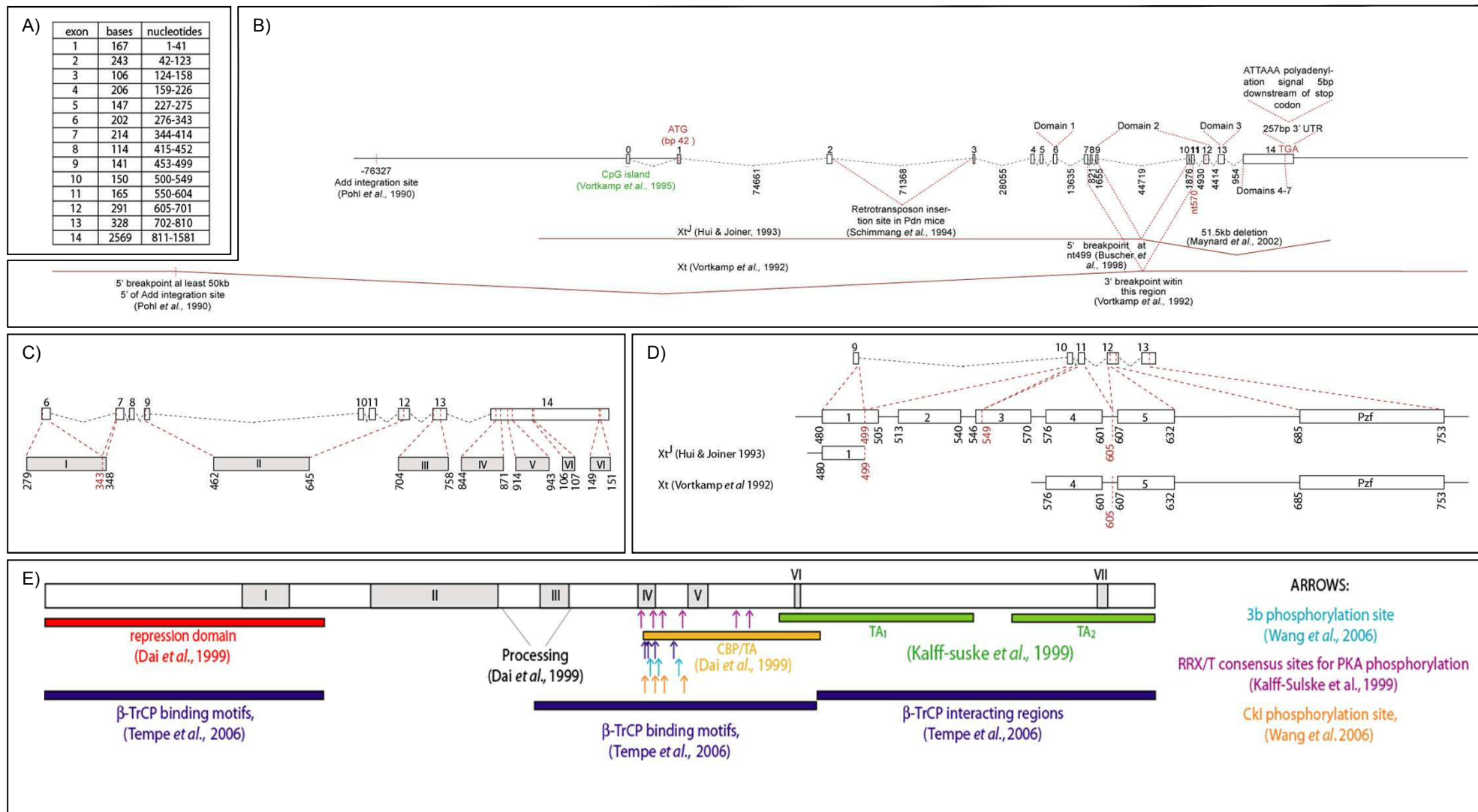
Figure 1.6, *Schematic representations of the Gli3 locus, proteins domains, and mutations identified in the mouse locus.* A) Exon sizes and corresponding residues in human Gli3. B) Genomic organisation of the human Gli3 locus. Intron and exon boundaries, along with other features identified by (Ruppert *et al.,* 1990) are shown, exon 0 represents a non-coding 5' exon that has been identified in mouse and may be present in the human transcript. Domains 1-7 represent regions of homology between Gli3 proteins described by (Hui and Joiner, 1990; Kang *et al.,* 1997). A CpG island has been identified at the 5' end of the transcript which overlaps with the putative exon 0. Breakpoints associated with mouse mutations within the locus are indicated, as is the integration site of the Add transgene. C) Exonic location of the seven conserved regions identified by Ruppert and colleagues(1990), the location of each region within Gli3 is also represented by residue number. D) Location of the 5 zinc finger motifs and the post zinc-finger motif, residue number and exonic organisation are shown. E) Functional domains within Gli3. All figures are drawn to an approximate scale. See text for references.

1.4.1.1   Alternative splicing in the Gli3 locus

The transcriptional start site within mouse exon 0 has not been determined. However, the exon is 62bp long, and is non-coding (Wild *et al.*, 1997). Supporting this finding the proximal end of mouse Gli3 cDNA was previously shown to extend 500bp upstream of the characterised human sequence (Ruppert *et al.*, 1990; Vortkamp *et al.*, 1992). Interestingly, initial sequencing of the human transcript identified 2 clones whose sequences diverge 42bp upstream of the predicted transcriptional start site, at a position corresponding to the 5' end of exon 1. One of these clones, the sequence of which has not been published, was shown to contain a sequence located in a similar region to mouse exon 0 (Wild *et al.*, 1997).

The divergence of the two human cDNA sequences identified by Ruppert *et al.* (1990) raises the possibility that alternative splicing at the 5' end, or the use of alternative transcriptional start sites, is an important mechanism in the regulation of Gli3. Further evidence for this was obtained through sequencing the quail gene. An alternative exon was identified between exons 2 and 3 that introduces a stop codon, and an in-frame methionine at position 201, located within a Kozak sequence, could produce a truncated protein lacking the N-terminal repressor domain (Borycki *et al.*, 2000).

1.4.2   Gli3 Protein Structure

The protein encoded by human Gli3 contains several functional domains (Dai *et al.*, 1999). Ruppert *et al.* (1990) identified 7 regions conserved amongst Gli proteins (I-VII in Figure 1.9C). At the C-terminus, a negatively charged helical structure (within region VII) has been implicated to have an activator role. Exons 9-12 encode five zinc-finger domains conserved across the Gli proteins (located within region II and shown in Figure 1.9D; Ruppert *et al.*, 1990; Kang *et al.*, 1997b). Further analysis of Gli3 revealed that it is able to act as an activator or as a repressor depending on its cellular context (Buscher *et al.*, 1997; Sasaki *et al.*, 1997; Dai *et al.*, 1999). Activator function requires the binding of CBP (CREB binding protein) to a region located at the C-terminus (Dai *et al.*, 1999). Two trans-activation (TA) domains have been identified towards the C-terminus, both of which are required for proper Gli3 function (Kalff-Suske *et al.*, 1999). TA_1 overlaps with the CBP binding domain, whereas TA_2 contains the helical structure proposed by Ruppert *et al.* to carry an activator role (Fig. 1.9 E; Ruppert *et al.*, 1990; Kalff-Suske *et al.*, 1999).

1.4.2.1   Gli3 processing

Cleavage of the Gli3 protein between amino acids 650 and 750 produces a 100kDa protein containing the DNA binding zinc finger domain and N-terminal repressor domain, and lacking the activator domain (Dai *et al.*, 1999). Similar to Ci, cleavage of Gli3 into its

repressor form requires a phosphorylation cascade mediated by Protein Kinase A (PKA), Casein Kinase 1 (CK1) and Glycogen Synthase Kinase 3 (GSK3β) (Fig. 1.9 E; Price and Kalderon, 2002). In the presence of Shh, cleavage does not occur (Zhang *et al.*, 2005). Phosphorylated Gli3 binds βTrCP (the vertebrate homologue of Slimb) of the ubiquitin ligase complex, targeting the protein to the proteosome (Tempe *et al.*, 2006; Wang and Li, 2006; Pan and Wang, 2007). It has recently been shown that three distinct regions of Gli3 are able to interact with βTrCP, all of which are required for efficient Gli3 processing (Tempe *et al.,* 2006; Fig. 1.9E). This implicates the proteosome in mediating proteolytic cleavage of Gli3, which is surprising since the proteosome usually degrades completely its protein substrates. The processing determinant domain (PDD), containing the first 197 amino acids at the C-terminus, overlaps with the presumed cleavage site, with approximately 50 residues lying upstream that are proposed to influence protein: proteosome interactions, and might prevent the degradation of the remaining protein (Pan and Wang, 2007).

1.4.3 Disease alleles

Gli3 has been associated with the Greig cephalopolysyndactyly syndrome (GCPS), an autosomal dominant disorder characterised by limb and craniofacial abnormalities in humans (Vortkamp *et al.*, 1991; Vortkamp *et al.*, 1992). GCPS can be caused by translocations in the Gli3 locus that interrupt the zinc finger DNA binding domain of the protein (Vortkamp *et al.*, 1991). It can also be caused by point mutations that disrupt the DNA binding or proteolytic processing of the protein, or disrupt splicing of the RNA (Wild *et al.*, 1997; Kalff-Suske *et al.*, 1999). In one case a translocation was mapped to approximately 10kb downstream of the Gli3 coding region, which could be due to disruption of a non-coding regulatory region (Vortkamp *et al.*, 1991). In addition to GCPS, several other human disorders are associated with mutations in Gli3. Pallister hall syndrome (PHS) is caused by single site frameshift mutations 3' of the zinc finger domains that in some instances introduce premature stop codons (Kang *et al.*, 1997a). PHS has a more severe phenotype than GCPS, probably because PHS mutations have a dominant-negative effect (Kang *et al.*, 1997a). Postaxial polydactyly type A results from frameshift mutations that cause premature termination of the protein after the zinc finger domain, but also after another highly conserved region of the protein that contains putative phosphorylation sites, the post zinc-finger domain (Pzf; Radhakrishna *et al.*, 1997; Wild *et al.*, 1997). This disease has a phenotype less severe than either GCPS or Pallister Hall syndrome, suggesting that the protein produced is more similar to the wildtype product. Acrocalossal syndrome is an autosomal recessive disorder and was in one case shown to be caused by a missense mutation in the final exon of Gli3 (Kalff-Suske *et al.*, 1999). Other candidate

disorders include frontonasal dysplasia (OMIM 136760), bifrid nose (OMIM 210 400 and 109 470), and simple hypertelomerism (OMIM 145 400) (Kang *et al.*, 1997a; Kang *et al.*, 1997b).

1.4.3.1 Mouse models of Gli3 disease alleles

GCPS is modelled by a naturally occurring mouse mutant, extra toes (Xt; Johnson, 1967; Winter and Huson, 1988; Schimmang *et al.*, 1992; Vortkamp *et al.*, 1992). This mouse mutant has a deletion of at least 80kbp, including the deletion of 415 - 570 nucleotides from the 5' end of the Gli3 coding sequence (Fig. 1.9B; Pohl *et al.*, 1990; Schimmang *et al.*, 1992; Vortkamp *et al.*, 1992). Xt homozygous mice do not express Gli3, suggesting that the Xt phenotype is caused by deletion of the Gli3 promoter and 5' coding region, or by the absence of conserved non-coding regulatory regions (Schimmang *et al.*, 1992).

Several other mouse lines have been developed that model GCPS, all of which affect the coding region of the gene and cause a truncation of the protein product. Brachyphalangy (Xt^{bph}, also referred to as Xt^{3H}) is a radiation-induced mutation, and Xt^{J} is a spontaneous Xt allele, both display a similar phenotype to Xt in heterozygotes and homozygotes (reviewed in Hui and Joyner, 1993). Xt^{J} is caused by a deletion from an intron within the first zinc finger domain (which does not determine DNA binding specificity in the human Gli protein) to a site after the 3'UTR of Gli3 (Fig. 1.9B). The Xt^{J} allele is believed to produce a non-functional protein (Fig. 1.9D; Hui and Joyner, 1993; Buscher *et al.*, 1998).

A recessive mutation in mice caused by the integration of a transgene, allelic digit pattern deformity syndrome (Add), has also been shown to be allelic to Xt (Pohl *et al.*, 1990). The integration site of Add, within 18kb upstream of Gli3, argues for a regulatory mechanism whereby the integration site exerts its phenotypic effect by interrupting a 5' control region of Gli3, namely the endogenous promoter or a cis-regulatory element (van der Hoeven *et al.*, 1993). Importantly, the transcript size is unaffected in mice carrying the Add insertion (van der Hoeven *et al.*, 1993). The transgene integration site of Add lies within the Xt mutation described previously (Fig. 1.9B; Vortkamp *et al.*, 1992). Xt mutant mice display a much more severe phenotype compared with Add mutants, which suggests that additional regulatory elements might be disrupted that could contribute to the observed phenotype. Add integration results in a phenotype characteristic of Gli3 misexpression restricted to the forelimbs. This suggests that the transgene insertion in Add mutants might interfere with a limb-specific enhancer of Gli3 (Pohl *et al.*, 1990).

Polydactyly Nagoya (Pdn) mutant mice are phenotypically similar to Xt mutant mice, causing anterior polydactyly (Thien and Ruther, 1999). Pdn has been shown to be allelic with Xt and Add mutations, and represents a dominant allele of Gli3 (Schimmang *et al.*, 1993). In Pdn

mutant embryos both the length and expression level of Gli3 appear unaffected. RFLP analysis did not reveal any insertions, deletions, rearrangements or point mutations (Schimmang *et al.*, 1993). The mutation is caused by the insertion of a retrotransposon of the Etn family into intron 2, between nucleotides 538 and 539 of the Gli3 coding sequence (Fig. 1.9B). This results in the generation of different splice variants. Pdn homozygous mice express multiple Gli3 transcripts and display premature termination of the coding sequence, or enlargement of the gene product, that are proposed to disturb protein function (Thien and Ruther, 1999).

The absence of Gli3 expression in Xt homozygous mice is likely to result from the deletion of its promoter and 5' coding region (Schimmang *et al.*, 1992). Furthermore, the limb-specific phenotype observed in Add mutants suggests that enhancer elements are contained within the 80kb upstream of mouse Gli3 that is disrupted in Xt mice(Pohl *et al.*, 1990).

1.5 Indentifying non-coding elements regulating Gli3 expression

This project aims to characterise cis-regulatory elements controlling Gli3 expression during embryogenesis. Eukaryotic gene transcription is initiated by recruitment of the basal transcription machinery to the core promoter of a gene. The basal machinery comprises RNA polymerase II, together with co-factors that are essential for transcription. Promoters themselves are regulated by enhancer elements. These are non-coding regions that bind transcription factors, and co-operate with promoter elements to regulate gene transcription. Through interaction with a number of enhancer elements, expression levels from a given promoter can be tightly regulated in a spatial and temporal manner. Thus, the expression pattern of a gene is the result of the transcriptional activity of a promoter, which is regulated by enhancer elements. Below I outline the current understanding of promoter and enhancer function. In particular I focus on methods that can be used to identify these elements.

1.5.1 Promoter elements

The core-promoter of a gene can be defined as the DNA region surrounding a transcriptional start site (TSS) that is sufficient for recruitment and assembly of the pre-initiation complex, and can thus drive basal transcription (Zhang, 2007). Spatial and temporal specificity are achieved by association of the core promoter with cis-regulatory elements, which will be discussed separately.

1.5.1.1 Characteristics of core promoters

1.5.1.1.1 Sequence motifs associated with promoter activity

Typically, mammalian core-promoters consist of 70-150bp and can be characterised according to sequence motifs contained within (Table 1.1; Sandelin *et al.*, 2007; Zhang, 2007). The

TATA box was the first motif to be associated with promoters (Breathnach and Chambon, 1981). Initial studies suggested that it is essential for transcription of protein-coding genes, but as more and more promoters are characterised the proportion shown to contain a TATA box diminishes. Current estimates show that around 22% of promoters contain a TATA box and 49% contain an Initiator (Inr) sequence. Around 25% of human promoters appear to lack all known core promoter elements (Gershenzon and Ioshikhes, 2005; Gross and Oelgeschlager, 2006).

Initiator sequences, positioned surrounding the TSS, help regulate its position in the absence of a TATA box (TATA-less) (Smale and Baltimore, 1989; Smale *et al.*, 1998). They are often found in conjunction with another element, the distal promoter element (DPE) (Kutach and Kadonaga, 2000). Like the TATA box, the DPE appears to direct precise transcriptional initiation, but it occurs downstream of the TSS and cannot function in the absence of an Inr sequence (Smale and Kadonaga, 2003).

Promoters can contain both TATA boxes and Inr sequences, or neither. Studies of Inr-TATA spacing have demonstrated that when separated by <30bp the two sites act together, the location of the TATA box defining the TSS. When separated by >30bp the two motifs function independently (O'Shea-Greenfield and Smale, 1992). A possible mechanism for this interaction was uncovered recently with the identification of the negative cofactor NC2, which antagonises expression of TATA mediated transcription by binding the Inr site (Malecova *et al.*, 2007). Table 1.1 summarises the sequence motifs recognised by core promoter elements, together with their positions relative to the transcription start site.

1.5.1.1.2 CpG Islands

Eukaryotic core promoters are often associated with CpG islands, defined as regions longer than 200bp, >50% G + C, with a CpG dinucleotide frequency that is at least 60% of that expected based on the nucleotide composition of the region (Gardiner-Garden and Frommer, 1987). Such regions usually lack consensus or near-consensus TATA boxes, Inr sequences or DPEs, and often contain multiple start sites spread over 100bp or more. They are thought to aid transcription by increasing the bendability and curvature of DNA (Fukue *et al.*, 2004). It is thought that the determinant of a TSS within a CpG island may be the binding of Sp1 40-80bp upstream of the initiation site (Smale and Kadonaga, 2003). Unlike other sequence motifs associated with enhancer activity, multiple Sp1 binding sites are a common feature of CpG islands. In 1993 it was estimated that 53% of human genes contain CpG islands, including all house-keeping genes and 40% of tissue restricted genes (Antequera and Bird, 1993).

1.5.1.1.3 Chromatin environment

The chromatin environment is also known to influence transcription from mammalian promoters. Densely packaged chromatin is non-conducive to transcription since it restricts access of the transcriptional machinery to DNA. Many yeast genes have been shown to contain an AT rich region upstream of the coding sequence that assembles poorly into nucleosomes (Meyer and Roelink, 2003; Yuan *et al.*, 2005; Williams and Tyler, 2007). Actively transcribed human promoters are also depleted of nucleosomes, and the length of DNA that wraps around a nucleosome (approximately 150bp) correlates with the maximal size observed of mammalian promoters (Luger *et al.*, 1997; Sandelin *et al.*, 2007). Histone density will also influence a promoters activity by affecting the transcription rate: in a genome-wide analysis in *Saccharomyces cerivisiae* it was demonstrated that core histone density within transcriptional units is inversely proportional to transcriptional rates (Schwabish and Struhl, 2004).

1.5.1.2 Methods used to predict core-promoter elements

1.5.1.2.1 In-silico promoter prediction

Various computational tools have been developed that use sequence information to predict the location of core-promoter elements. The tools differ in the statistical method used to score potential sites, including artificial neural networks (ANNs), relevance vector machines (RVMs), and quadratic discriminant analysis (QDA). They also differ in the promoter characteristics recognised. Bajic *et al.* (2004) compared the accuracy of various promoter prediction tools, analysing success by the proportion of elements identified that surround known TSSs present in the Database of Transcriptional Start Sites (DTBSS). The most accurate tool predicted two thirds of human core-promoters, but also retrieved one false prediction for every two hits. Predictions of core-promoters located within CpG islands were more accurate than those not associated with a CpG island (Bajic *et al.*, 2004).

1.5.1.2.2 Wet-lab approaches

Chromatin Immuno-Precipitation (ChIP) can be used to identify chromosome regions that are not bound by nucleosomes, and are in an active state for transcription. In-vivo protein:DNA interactions are captured by cross-linking and the DNA is then fragmented (either mechanically or enzymatically). DNA fragments bound to the protein of interest are retrieved by immuno-precipitation, and reversal of the cross-linking allows the bound DNA to be sequenced. ChIP technology can also be applied to high density microarrays (Chip-chip), allowing analysis of a larger genomic region (Ren *et al.*, 2000). This approach has been used on a genome-wide scale for the in-vivo mapping of pre-initiation complex formation at active core promoters (Kim *et al.*, 2005). Hybridisation of putative 5'-end regions of

labelled cDNA or cRNA to high-density genome tiling arrays is another approach, but becomes costly if only one gene is under consideration (Kapranov *et al.*, 2005).

For the identification of the 5' end of a particular transcript, 5' Rapid Amplification of cDNA Ends (5'RACE) is a successful tool (Frohman *et al.*, 1988). A gene-specific oligonucleotide that hybridises to a known sequence within a characterised coding region is used to prime reverse transcription. The cDNA produced is then amplified by PCR and subsequently sequenced. This approach was recently used in the comprehensive screening of promoters in 1% of the human genome, chosen for analysis in the first phase of the ENCODE (Encyclopaedia of DNA Elements) project (Birney *et al.*, 2007). A more recent approach for 5' characterisation of a cDNA is Cap Analysis of Gene Expression (CAGE; (Shiraki *et al.*, 2003). First strand cDNA is generated from total RNA primed with an olig-dT or random primer, and cap-trapping enriches cDNA/RNA hybrids through cap-structure-biotinylation. RnseI digestion is then used to remove any ssRNA linking the biotynilated cap and any double strand RNA/truncated cDNA, such that only RNA molecules hybridised to full-length cDNAs remain. Second strand synthesis is primed using an oligonucleotide that contains a MmeI restriction site, and subsequent digestion allows 5'cDNA fragments to be concatamerised, then sequenced. This approach has recently been used in the most extensive investigation into the 5' ends of mammalian cDNAs to date, carried out on a genome wide scale (Carninci *et al.*, 2006).

1.5.1.3 Transcriptional Start Sites

1.5.1.3.1 Databases of transcriptional start sites

Several databases of TSSs have been developed. The Eukaryotic promoter database (EPD) is a collection of promoter sequences derived from full-length 5'UTRs/cDNAs (Schmid *et al.*, 2004; 2006). Functional annotation of mice (Fantom) collates murine specific transcript information (Carninci *et al.*, 2006; Maeda *et al.*, 2006), and the Database of Transcriptional start sites (DBTSS) is constructed mainly of human TSSs (Suzuki *et al.*, 2002; Wakaguri *et al.*, 2008). Results from CAGE analysis have been used to generate two further publicly accessible database resources: CAGE Basic and CAGE Analysis (Kawaji *et al.*, 2006).

1.5.1.3.2 TSS specificity defines two classes of core promoters

In the most comprehensive analysis to date CAGE tags were often found overlapping on the same strand, forming a cluster of TSSs. Of 729,504 potential mouse TSSs, and 775,278 potential human TSSs, 593,290 and 629,716 sites respectively were shown to form clusters (Carninci *et al.*, 2006). 8,185 and 5,928 of these clusters respectively were supported by >100 CAGE tags. The clusters were grouped into various classes depending on the distribution of TSSs identified, which can be simplified into two categories, broad and sharp. Clusters

displaying sharp TSSs were associated with over-representation of TATA boxes, whereas broad TSS regions strongly associated with CpG islands. Around 90% of TSSs that were not associated with a TATA box were shown to occur within a CpG island. The study also demonstrated a preference for CA, CG and TG dinucleotides at the initiation site, which disagrees with the accepted paradigm that adenosine is the preferred starting base. This demonstrates that the Inr sequence is not an absolute determinant of transcription initiation.

Translation normally initiates from the first ATG in a mRNA. To enable efficient translation of the mRNA produced from broad promoters, premature ATG start codons must be excluded. A depletion of ATG start sites was recently observed in the broad promoter regions of MHC Class I genes. The authors estimate that ~82% of human genes are associated with ATG-deserts (Lee *et al.*, 2005).

1.5.1.3.3 Choice of transcriptional start site

Of particular interest in the study of Carninci *et al.* (2006), a relationship was demonstrated between the specificity of TSS and expression of the resulting transcript. Ubiquitously expressed transcripts were shown to be associated with broadly distributed TSSs and CpG islands, whereas sharper TSSs and TATA-box promoters were associated with more tightly regulated transcripts (Carninci *et al.*, 2006). This is in agreement with a study by Ponjavic and colleagues (2006), who showed that the position of the TATA box can determine the tissue specificity of the resulting transcript (Bailey *et al.*, 2006; Ponjavic *et al.*, 2006). An exception to this trend was identified for promoters associated with CNS-specific expression, which were found to be particularly CpG rich (Carninci *et al.*, 2006).

Taken together, the evidence obtained so far suggest that within broad promoters, that regulate expression of the majority of mammalian transcripts, the choice of TSS by RNA polymerase II is only loosely defined. In a nucleosome free region, the presence of a TATA box will direct transcription from a site located approximately 30bp upstream, which may or may not correspond to an Inr sequence. In the absence of a TATA box, transcription will preferentially initiate from a Py/Pu dinucleotide.

TATA-box containing promoters appear to evolve more slowly than broad promoters do. This is consistent with the observation that they are often associated with tightly regulated transcripts. Changes in the sequence surrounding the TSS will have regulatory consequences and are likely to disrupt gene expression. Within broad promoters the specific nucleotide sequence is less important, multiple redundant TSSs in these regions allows mutation of a particular site to be tolerated, facilitating adaptive evolution and allowing the fine tuning of promoter activity (Carninci *et al.*, 2006; Sandelin *et al.*, 2007).

Recent studies have shown that the selection of a particular TSS is subject to species-specific adaptation, with preferred characteristics differing between human and mouse genomes (Bajic *et al.*, 2006; Carninci *et al.*, 2006). Relationships have also been demonstrated between promoter type and gene ontology or tissue specificity of the resulting transcript. Broad promoters are often associated with ubiquitously expressed transcripts, whilst tightly regulated transcripts are associated with well-defined promoters, often containing a TATA box (Carninci *et al.*, 2006). Of particular interest to this study, CNS specific promoters were found to be an exception to this rule, and are often CpG rich. This supports the hypothesis that the CpG island identified upstream of Gli3 contains the endogenous promoter of Gli3 in the CNS.

Element	Optimal recognition sequence	Factors bound	Position relative to TSS	Method of action	Reference
TATA Box	TATAWAAR	TBP	~-30bp		(Bucher, 1990)
Inr (human)	YYA(+1)NWYY	TAF1/2, RNA polII, TFII-1, YY-1	-2 to +5		(Javahery *et al.*, 1994; Lo and Smale, 1996)
DPE	RGWYV, +G(+24)	TAF6/9	+28 to +32		(Kutach and Kadonaga, 2000)
TFIIB recognition element (BRE)	SSRCGCC	TFIIB (C-terminal motif unique to eukaryotes)	-37 to -32	Polar recognition of TATA box by TBP, may also repress basal transcription in vivo	(Lagrange *et al.*, 1998; Evans *et al.*, 2001)
Proximal sequence element (PSE)	TCACCNTNA-STNAAAAGK (human recognition sequence)	SNAPc complex (also known as PBP or PTF)	-45 to -60	Essential for transcription of small nuclear RNAs (snRNAs)	(Lobo *et al.*, 1991; Smale and Kadonaga, 2003)
Downstream core element (DCE)	3 sub elements: SI CTTC SII CTGT SII AGC	TAFI	+6 to +11 +16 to +21 +30 to +34		(Zhang, 2007)
MTE	CSARCSSAACGS	Unknown	+18 to +29		(Zhang, 2007)

Table 1.1: *Sequences involved in the regulation of core-promoter elements by TFIIB and TFIID,* adapted from Figure 1 of Zhang 2007, data is mainly compiled from Smale and Kadonga 2003, see references therein (Smale and Kadonaga, 2003; Zhang, 2007).

1.5.2 Enhancer elements regulate promoter activity

Typically enhancer elements are up to 300bp in length and are defined by their ability to drive reporter gene expression in a manner that recapitulates an aspect of the endogenous expression pattern of the gene they regulate (Arnone and Davidson, 1997). They are thought to direct precise spatial and temporal pattern of gene expression by association with regulatory proteins such as transcription factors, themselves often expressed in spatial and temporal manner. A single promoter can be regulated by one or more enhancer elements, thus the expression pattern of a gene is the result of the combinatorial effect of all enhancer elements acting upon it. Since a single functional module may act as an enhancer or silencer depending on its cellular context, and both act in a similar manner, I use the term enhancer to refer to regulatory

modules which exert positive and/or negative regulatory effects. Enhancers are thought to work by either modifying the chromatin structure (Non-contact; Martin *et al.*, 1996), or by direct association between factors binding the enhancer and those associated with the pre-initiation complex at a promoter (Contact; Majumder *et al.*, 1997). It has also been proposed that in some instances enhancers might act as entry points for factors that track along the DNA until they reach the promoter, where they are involved in the assembly of the initiation complex (the scanning model; Hatzis and Talianidis, 2002).

Insulator elements prevent enhancers from inappropriately engaging with promoter elements of nearby genes. They work either by regulating chromatin structure (preventing the spread of repressive chromatin or facilitating enhancer-promoter associations with the target gene), or by directly blocking the association of an enhancer with adjacent genes (Brasset and Vaury, 2005). Current data suggest that all mammalian insulators of the latter type work only when bound by the zinc finger protein 'CCCTC-binding factor' (CTCF; Ladomery and Dellaire, 2002).

1.5.2.1 Methods used to identify enhancers

1.5.2.1.1 Classical methods

Classical methods for the identification of mammalian regulatory elements rely on trial and error strategies. The most widely used approach has been deletion analysis of genomic regions capable of driving tissue specific expression of a reporter gene in transgenic animals. Mutations that do not alter the coding sequence of a gene, but are associated with a phenotype might be caused by deletion or interruption of an enhancer element, and have allowed the identification of several enhancers. An example is the long range enhancer element of the Shh gene, interrupted by a transgene insertion in the Sasquatch mouse mutant (Lettice *et al.*, 2002; Lettice *et al.*, 2003). Alternatively the entire region surrounding a gene of interest can be searched in a systematic manner for genomic fragments capable of producing tissue-specific reporter gene expression. This approach has been successfully used to identify enhancer elements of the Sox2 gene (Uchikawa *et al.*, 2003). Once a genomic fragment capable of directing reporter gene expression has been identified, the limits of the region can be investigated by progressive deletion. Although these approaches have proven useful in the identification of many enhancer elements that have critical roles in development, they are time and labour intensive, costly, and limited to a discrete genomic region.

1.5.2.1.2 In vitro approaches

Several in-vitro techniques have been developed to aid in the identification of cis-regulatory elements. DnaseI hypersensitivity assays detect genomic regions devoid of nucleosomes which are readily digested by the enzyme DnaseI, and are nearly always associated with cis-acting

DNA sequences (Wu, 1980; Gross and Garrard, 1988). ChIP can be used to identify enhancer elements associated with a particular transcription factor (Cohen-Kaminsky *et al.*, 1998; Kuo and Allis, 1999). However, success depends on the specificity of the antibody used and will only reveal binding within a discrete region, depending on the fragmentation conditions used, rather than indicating the exact position of protein:DNA interactions. Additionally, it is difficult to distinguish between sequences that are the result of specific protein:DNA interactions, and those which associate with the protein of interest via another protein intermediate. Kang *et al.* (2002) have devised a technique which avoids these caveats by combining ChIP with Dnase protection, whereby the binding site can be identified in addition to the interacting protein (Kang *et al.*, 2002). 'ChIP on ChIP' applies ChIP technology to microarray data, allowing binding site identification on a genome-wide scale (Ren *et al.*, 2000).

The function of an enhancer element relies upon the binding of transcription factors. However, the multitude of possible binding sequences, together with a high variability in their location relative to the gene they control, has made identification of functional binding sites troublesome. Several techniques have been developed that aim to locate the protein bound regions of DNA once a sequence suspected to contain an enhancer element has been isolated. Electrophoretic mobility shift assays (EMSAs) can be used to verify protein:DNA interactions. This technique relies on the resolving power of non-denaturing polyacrylamide gels to separate protein bound radio-labelled DNA molecules (around 30bp) from unbound molecules with a lower molecular weight (Garner and Revzin, 1981). Dnase protection assays combine the protein:DNA binding reaction of an EMSA with the cleavage properties of DnaseI, to establish the protein binding positions within a longer DNA probe (up to 500bp) (Galas and Schmitz, 1978). Alternatively, ligation-mediated PCR can be used to establish the precise location of protein bound sequences within an oligonucleotide (Mueller and Wold, 1989).

1.5.2.1.3 Phylogenetic footprinting

It is believed that enhancer elements have undergone 'purifying selection' to retain the ability to bind sequence specific regulatory proteins (Thomas *et al.*, 2003). Sequences composed of multiple protein binding regions, such as transcription-factor binding sites, will be under high levels of evolutionary constraint and will mutate at a slower rate than surrounding bases not under selection. The approach of identifying putative regulatory regions based on their sequence conservation is known as phylogenetic footprinting (Tagle *et al.*, 1988). Phylogenetic footprinting aims to distinguish functional regions under

purifying selection from those that have not yet diverged by evolutionary drift on a background of recently evolved sequence (Gumucio *et al.*, 1992; Stojanovic *et al.*, 1999).

A wet-lab approach to phylogenetic footprinting was used by Frazer and colleagues in a large scale screen for highly conserved DNA regions on human chromosome 21 (Frazer *et al.*, 2001). Genomic DNA fragments of one species were hybridised to high-density oligonucleotide arrays of another, allowing the identification of putative regulatory elements based on evolutionary conservation

In light of the publication of numerous genome sequences, various in-silico tools have been developed to facilitate the identification of enhancer elements by phylogenetic footprinting. This is now one of the preferred methods for the identification of enhancers. The ability to identify putative enhancers based on their high level of sequence conservation circumvents the need for tedious serial deletion studies. However, the successful identification of enhancer elements relies on the parameters used to identify candidate regions.

1.5.2.2   In-silico identification of enhancer elements

In silico enhancer prediction tools produce an alignment of the region of interest, and identify putative enhancer elements based on sequence conservation. Factors that need to be considered are the phylogenetic distance between organisms compared, the number of sequences included, the quality of an alignment, and sequence conservation parameters.

The degree of sequence similarity used to determine whether a region of DNA should be considered as a putative enhancer element is critical, and largely depends upon the phylogenetic distance that separates the organisms compared. Sequence similarity comprises of percentage identity between a given pair of sequences and the size of the conserved region, both of which are important in defining selection criteria.

Another important factor to consider is the genomic locus in which regulatory elements associated with a particular gene might lie. Studies have shown that cis-elements controlling gene expression can exist from within a few kb to around 1Mb upstream or downstream, or within intronic regions of the genes they control.

1.6   Outlook for this thesis

Since the publication of the first 'working draft' of the human genome in June 2000, uncovering the function of different sequence elements has been a growing challenge. Current estimates are that around 5% of the human genome is under evolutionary constraint, but only 32% of the constrained DNA sequence is protein coding (Waterston *et al.*, 2002; Birney *et al.*, 2007). 8% of the remaining DNA under selection can be accounted for by untranslated RNA coding regions (UTRs), leaving 60% of evolutionary constrained DNA in the non-coding fraction of the genome (Cooper *et al.*, 2004; Birney *et al.*, 2007). This conserved non-coding fraction of the genome has been shown to contain regulatory regions that control promoter activity, namely enhancers, silencers and insulators. Highly conserved non-coding elements containing such modules are specifically associated with transcription factors and genes involved in development (Ashburner *et al.*, 2000; Bejerano *et al.*, 2004; Ovcharenko *et al.*, 2004; Sandelin *et al.*, 2004b; Plessy *et al.*, 2005; Woolfe *et al.*, 2005; Vavouri *et al.*, 2006a). This makes Gli3 a good candidate for regulation by conserved cis-acting elements. Hence, I chose to adopt a bioinformatics approach for the identification of cis-regulatory elements that control Gli3 expression. In this thesis I describe the identification of numerous highly conserved non-coding regions in the Gli3 locus. The ability of these elements to regulate Gli3 expression in-vivo was analysed using a reporter gene system in chick neural tube. Coupled with the analysis of the transcription factor binding sites within the conserved regions, this has allowed a greater understanding of how Gli3 expression is controlled at the transcriptional level. Furthermore, detailed analysis of one of these regions has revealed a role for TALE family transcription factors in the regulation of Gli3 expression during neural tube development.

Chapter 2

Materials and Methods

2.1   Embryo Techniques

All chick embryos were harvested from embryonated White Leghorn chicken eggs, supplied by Winter Egg Farms. Embryos were staged according to Hamburger-Hamilton (HH; Hamburger and Hamilton, 1992).

2.1.1   Electroporation

Stage 10-11 embryos were electroporated in-ovo as described previously (Muramatsu *et al.*, 1997) under a MZ7.5 stereomicroscope (Leica). To maintain a humidified environment 2-3ml PBS solution containing 1x Antibiotic antimycotic solution (containing Penicillin, Streptomycin and Amphotericin B; Sigma) was applied over the embryo prior to electroporation. A small tear was made in the vitelline membrane at the site of the injection to facilitate access. DNA was injected using a pulled glass capillary needle (1mm O.D. x 0.5mm I.D. capillaries, Harvard apparatus, pulled using a Flaming Brown Micropipette puller, Model P-97, Sutter instruments). DNA used for electroporation was resuspended in PBS at 4μg/μl unless otherwise stated. For lacZ reporter assays 3μl of the test plasmid was mixed with 1.5μl pMES EGFP (3μg/μl) (Osumi and Inoue, 2001; Swartz *et al.*, 2001), giving a 3:1 ratio. To aid visualisation of injected DNA 0.5μl 4% fast-green (Sigma) was added to the solution (final concentration 0.4%). Where activator or repressor constructs were added to the electroporation mix, 5μl of the construct carrying the lacZ reporter was mixed with 3μl activator or repressor construct (or H_2O for control embryos), 1.5 μl pMES EGFP and 0.5μl 5% fastgreen.

Following injection into the neural tube, 5mm L-shaped gold plated electrodes (Genetronics inc.) were positioned either side of the embryo, separated by a 5mm gap. 5 pulses of 24volts, 30ms in length with a 500ms interval were delivered using a BTX ECM 830 square wave generator. Eggs were sealed with sellotape, and incubated for 12-24 hours at 39ºC. After incubation GFP levels were assessed under a Leica MZ160F fluorescence stereomicroscope (Leica). Embryos displaying good levels of GFP were harvested and photographed on a glass slide under the same microscope.

2.1.2   LacZ staining

For lacZ staining, embryos were fixed for 30-60 minutes at room temperature in 1% paraformaldehyde (PFA) in PBS. The fix was removed by rinsing three times, and washing three times for 15 minutes in lacZ rinse (5mM EGTA, 0.01% sodium deoxycholate, 0.02% NP40 (Igepal), 2mM $MgCl_2$ in PBS). Embryos were stained for 8 hours at 37ºC in lacZ rinse containing 10mM $K_3Fe(CN)_6$, 10mM $K_4Fe(CN)_6$, and 1mg/ml 5-Bromo-4-Chloro-3-Indolyl-β-D-Galactopyranoside (X-gal, Invitrogen). Following staining embryos were again rinsed three times and washed three times for 15 minutes in lacZ rinse buffer.

Table 2.1: *Electroporation constructs*

Plasmid name	Description	Supplier	Reference	Modifications/Notes
nP1230	Lacz reporter under control of human β-globin promoter	Peter Rigby	(Yee and Rigby, 1993)	Nuclear localisation signal of pCIG subcloned into NcoI site upstream of β-Galactosidase coding sequence
pCAB Smad6	Chicken Smad6 in pCAβ S6 vector (containing bicistronic IRES EGFP) under control of β-actin promoter	Claudio Stern	(Linker and Stern, 2004)	pMES GFP not added to electroporation mix
pCIG	Carries IRES-nuclear EGFP element, drives expression from a combination of chick β-actin and CMV enhancer	Andy McMahon	(Megason and McMahon, 2002)	
pMES	Carries IRES-EGFP under control of chicken ß-actin promoter/CMV-IE enchancer	Neva Meyer	(Y. Chen, 2004)	
pIRES2-EGFP-Meis1a-En1	Murine meis1a linked to *Drosophila* En1 repressor domain cloned into pIRES2-EGFP plasmid under control of the CMV promoter	Richard Maas	(Zhang *et al.*, 2002)	
pIRES2-EGFP-Meis1aVP16	Murine meis1a linked to the activation domain of VP16 (of herpes simplex virus) cloned into pIRES2-EGFP plasmid under control of the CMV promoter	Richard Maas	(Zhang *et al.*, 2002)	

2.1.3   In-situ hybridisation

2.1.3.1   Generating DIG-labelled RNA probes

Antisense digoxygenin-labelled RNA probes were made in a total volume of 20μl. 1μg of linearised DNA template was mixed with 1μl RNAsin RNAse Inhibitor (Promega), 4μl Trans5x RNA Polymerase Reaction Buffer (Promega), 2μl 10x Digoxygenin (DIG) RNA labelling mix (Roche), 0.2μl DTT (100mM), 1.5μl RNA polymerase (T7 or SP6 depending on the vector, and the inserts orientation) and made up to 20μl with distilled H_2O. Reactions were incubated at 37ºC for 90 minutes. Template DNA was degraded by incubation with 3μl RQ1 Dnase (Promega) at 37ºC for 15 minutes. Riboprobes were precipitated by the addition of 2μl of 200mM EDTA, 2.5μl 4M LiCl and 70μl 100% RNAse-free ethanol, followed by incubation at -70ºC for 1 hour. To recover the riboprobe the precipitated solution was centrifuged at 13000rpm for 30minutes at 4ºC. Pellets were washed with 100μl 70% RNAse free ethanol, followed by centrifugation at 13000rpm for a further 10 minutes at 4ºC. Pellets were dried using a speed vacuum, and resuspended in 50μl DEPC H_2O containing 1.5μl RNAsin and 0.5μl DTT. To verify the success of RNA synthesis 1μl of the probe was run on a 1% TBE/Agarose gel.

2.1.3.2   Wholemount in-situ hybridisation

Wholemount in situ hybridisations were carried out according to a protocol supplied by Domingos Henrique and David Ish-Horowicz (ICRF Dev.Biol.Unit, Oxford), modified from protocols of Ron

Conlon (Mt Sinai, Toronto), Phil Ingham (Department of Biomedical Science, Sheffield) and David Wilkinson (NIMR, London). Embryos were dissected out in QPBS (1.25mM MgSO₄, 0.14mM CaCl₂, 137mM NaCl₂, 5.37mM KCl, 1.1mM KH₂PO₄, 1.1mM Na₂HPO₄ in DEPC-treated H₂O) and fixed overnight at 4ºC in 1ml of 4% formaldehyde, 2mM EGTA in QPBS, then washed three times for 5 minutes in PTW (QPBS + 0.1% Tween 20). Embryos were then treated in 1ml Proteinase buffer (100mM Tris HCL pH8, 50mM EDTA pH8 in DEPC H₂O) containing 20μg/ml Proteinase K (Roche) for 3-8 minutes (according to HH stage) at 37ºC, followed by three rinses in PTW. Embryos were post-fixed for 20 minutes in 4% formaldehyde, 0.1% gluteraldehyde in PTW. The fixing agent was removed by rinsing three times in PTW, followed by three 5 minute washes in PTW. Embryos were then rinsed in 1:1 PTW:hybridisation buffer, and allowed to sink, before transferring to hybridisation buffer. Hybridisation buffer consisted of 50% formamide, 1.3x SSC pH 7.5, 5mM EDTA pH 8.0, 50μg/ml tRNA, 0.2% TWEEN 20, 0.5% CHAPS, and 100μg/ml Heparin in DEPC treated H₂O.

For prehybidisation, embryos were incubated in hybridisation buffer for 2 hours at 70ºC in a rotating hybridisation oven. Hybridisation buffer was then replaced with 1ml fresh hybridisation buffer containing 3μl DIG labelled antisense RNA probe, hybridisation was performed overnight at 70ºC in a rotating oven.

Table 2.2: *RNA probes used for in-situ hybridisation*

Plasmid name	Insert size & plasmid	Restriction site used to linearise	RNA polymerase	Provider (reference)
qGli3-T3	900bp insert in pCRII-TOPO	HindIII	T3	Made by N. van-Hateren contact A.G. Borycki
qGli3-900	900bp insert in pCRII-TOPO	HindIII	T7	Made by N. van-Hateren contact A.G. Borycki
cMeis1	Full coding sequence of cMeis1a (1.5kb) inserted to pGEMT easy vector	ClaI	SP6	Miguel Torres (Mercader *et al.*, 1999)
cMeis2	500bp from N terminus to meis domain of cMeis2 inserted to pGEMTeasy vector	NcoI	SP6	Miguel Torres (Mercader *et al.*, 1999)
cPbx1a	901bp cPbx1a in pCRII-TOPO	EcoRV/NotI	SP6	This thesis
cPbx1b	788bp cPbx1b in pCRII-TOPO	EcoRV/NotI	SP6	This thesis
cPbx3	1073bp cPbx3 in pCRII-TOPO	EcoRV/NotI	SP6	This thesis
cPbx4	1064bp cPbx4 in pCRII-TOPO	EcoRV/NotI	SP6	This thesis
cPrep1	901bp cPrep1 in pCRII-TOPO	EcoRV/NotI	SP6	This thesis
cPrep2	1059bp cPrep2 in pCRII-TOPO	BamHI	T7	This thesis

To remove the probe, embryos were rinsed once and washed twice (30 minute washes) with 1ml hybridisation buffer at 70ºC. Embryos were then equilibrated to the next solution by washing for 20 minutes at 70ºC in 1ml 1:1 hybridisation buffer:TBST (137mM NaCl, 2.68mM Kcl, 25mM TrisHCl pH 7.5, 0.1% TWEEN 20 in DEPC treated H₂O).

Embryos were washed twice for 30 minutes at room temperature in 1ml TBST, then rinsed twice with 1ml MABT (100mM Tris HCl pH 7.5, 0.15M NaCl, 0.1% TWEEN 20 in DEPC treated H₂O). To block non-specific antibody detection, embryos were incubated for 2 hours at room temperature in 1ml of 2% blocking solution (2% Dig DNA labelling and detection kit blocking reagent [Roche], 10mM Tris HCl pH 7.5, 0.15M NaCl, 10% Horse Serum, 0.1% TWEEN 20). This solution was then

replaced with 1ml of 2% blocking solution containing 0.5μl Anti-Digoxygenin Alkaline Phosphatase conjugated antibody (Roche). To allow binding, embryos were incubated overnight at 4ºC.

Non-bound antibody was removed by rinsing three times with MABT, and washing three times for 1 hour at room temperature with 2ml MABT. Embryos were then washed twice for 10 min in 2ml NTMT (10mM Tris HCl pH 9.5, 5mM $MgCl_2$, 0.15mMNaCl, 1% TWEEN 20 in DEPC treated H_20). Embryos were stained in the dark at room temperature in 1ml NTMT + 3.5μl 5-bromo-4-chloro-3-indolyl phosphate (X-phosphate, Roche) + 4.5μl 4-Nitro blue tetrazolium chloride (NBT, Roche) until strong staining was achieved. Where high background was observed, the staining was paused intermittently, and embryos were washed overnight with NTMT before staining was continued. The colour reaction was stopped by washing three times for 5 minutes in PTW.

2.1.4   Embedding and sectioning

2.1.4.1   Vibratome sectioning

Following in-situ hybridisation or lacZ staining, embryos were embedded in 2% Agarose in PBS, and incubated at 4ºC (2 hours-2 days). Vibratome sections were taken at 80μM using a Vibratome 1500 sectioning system (Vibratome ®), and mounted under glycergel (DakoCytomation). Slides were left to dry overnight at room temperature before photographing.

2.1.4.2   Cryostat sectioning

Embryos were fixed in solution I (154mMNa$_2$HPO$_4$, 46mM NaH$_2$PO$_4$.H$_2$O, 0.12mM CaCl$_2$, 4% sucrose) with 0.2% PFA overnight, rinsed three times and incubated overnight at 4ºC in solution I (without PFA). Embryos were equilibrated to the embedding solution by washing three times and incubating overnight at 4ºC in phosphate buffer (120mM phosphate in distilled H_2O) containing 15% sucrose. Embryos were incubated in phosphate buffer with 15% sucrose and 7.5% gelatin for 1hour at 37ºC, and embedded in the same solution by freezing in a dry-ice chilled isopentane bath. Embryos were stored at -80ºC until sectioning (Bajanca *et al.*, 2004).

Embryo blocks were mounted onto chucks using OCT compound (BDH), and sectioned on a Bright cryostat at 15μm (Bright Instruments). Sections were mounted on super frost slides (Menzel-Glaser), and stored in the dark at -20ºC.

2.1.5   Immunohistochemistry

2.1.5.1   Processing of sections for immunohistochemistry

Cryostat sections were re-hydrated for 1 hour at room temperature with 500μl PBT (PBS containing 0.05% Triton X and 10% heat inactivated goat serum [HINGS]). Primary antibody was applied, diluted to the desired concentration in 500ml PBT, and incubated overnight in a humidified environment at 4ºC. Primary antibody was removed by rinsing the slides three times with 500μl PBT, and soaking the slides three times for 5 minutes in PBT. Secondary antibodies were applied at the desired concentration, diluted in 500μl PBT, and incubated for 1 hour at room temperature. Secondary antibody was removed in the same way as the primary antibody. Slides were dried by

dabbing onto tissue paper. Slides were mounted by applying 2-3 drops of MOWIOL mounting media (13.3% Mowiol 4-88 [Sigma], 33.3% glycerol, 2.5% DABCO [Sigma], in 130mM Tris-HCl pH 8.5), and coverslips were held in place using nail polish.

Table 2.3: *Antibodies used for Immunohistochemistry*

Antibody	Supplier	Species derivation	Dilution	Secondary antibody used
Anti β-Galactosidase	Promega	E-Coli monoclonal IgG raised in mouse	1:1000	Anti-mouse Alexa Fluor 594 *
MNR2	Hybridoma bank	Pan MNR2 monoclonal IgG raised in mouse	1:50	Anti-mouse Alexa Fluor 594
En1	Hybridoma bank	Chicken monoclonal IgG raised in mouse	1:50	Anti-mouse Alexa Fluor 594
Nkx6.1	Hybridoma bank	Rat monoclonal IgG raised in mouse	1:200	Anti-mouse Alexa Fluor 594
Isl1	Hybridoma bank	Mouse monoclonal IgG raised in mouse	1:50	Anti-mouse Alexa Fluor 594
Pax6	Hybridoma bank	Chicken monoclonal IgG raised in mouse	1:1000	Anti-mouse Alexa Fluor 594
Anti-rabbit Alexa Fluor 594	Invitrogen	Rabbit IgG raised in Goat	1:1000	
Anti-mouse Alexa Fluor 594	Invitrogen	Mouse IgG raised in Goat	1:1000*	

*Anti-mouse Alexa Fluor 594 was used at a dilution f 1:2000 when used with Anti β-Galactosidase

2.1.6   Image capture and manipulation

2.1.6.1   Wholemount images of LacZ stained/ in-situ embryos

Images of wholemount embryos were captured under a MZ12.5 stereomicroscope (Leica) using a SPOT® INSIGHT Colour camera, with SPOT Advanced digital image capture software (Diagnostic Instruments).

2.1.6.2   Wholemount fluorescence of electroporated embryos

Fluorescent images of GFP positive wholemount embryos were captured using a DFC300FX camera (Leica) with Leica FireCam Mac V3.1.0 (Leica), under a Leica MZ160F fluorescence stereomicroscope (Leica).

2.1.6.3   Imaging of Vibratome sections

Vibratome sections were visualised on a Leica DMR microscope (Leica) mounted with a DMR DC300FX digital camera (Leica). Images were captured using Leica IM50 Image capture software v1.20 (Leica)

2.1.6.4   Confocal imaging of immunohistochemistry sections

Images were captured on a Leica SP1 confocal microscope using Leica confocal software v2.61 Build 1537 (Leica Microsystems). Images were stacked and processed using ImageJ 1.38x software (NIH, USA).

2.1.6.5 Image manipulation

All images were manipulated in Adobe Photoshop CS version 8.

2.2 Molecular Biology Techniques

2.2.1 PCR

2.2.1.1 Primer design

All primers were designed with the aid of primer premier 5, version 5 (www.PremierBiosoft.com). Specificity was tested by searching the nucleotide collection nr/nt using the nucleotide Blast tool at the NCBI website (http://www.ncbi.nlm.nih.gov/BLAST).

2.2.1.1.1 Standard primer design

Parameters used for the generation of standard primers were that they should be 18-25 BP in length, have roughly an equal distribution of A/T vs C/G, be flanked at the 5' and 3' end by at least 1 C/G, and have a Tm of 55-60ºC.

2.2.1.1.2 Primers including restriction sites

Primers including restriction sites for amplification of individual regions of interest were designed using the following parameters:

➢ Sequence homology: 17-23bp containing a roughly even distribution of A/T:G/G, with at least 1 G/C at the 5' and 3' end.

Incorporation of the KpnI site (GGTAC^C) at the 5' end, proceeded by 3xC/G (as suggested by Stratagene technical tools; Cleavage Activity Near DNA Termini)

➢ Lowest possible score for the generation of primer dimers, and stable hairpins, as calculated by primer premier 5.
➢ Tm <80ºC as calculated by primer premier 5.

2.2.1.1.3 Primers for 5'RACE

First strand cDNA synthesis was carried out using oligo dT and SMART II™ A Oligonucleotide primers described in the SMART™ RACE cDNA Amplification Kit (Clontech) handbook. Rapid Amplification of cDNA Ends (RACE) was carried out using the universal primer mix (also described in the handbook) along with a gene specific primer for Gli3. Gene specific primers were designed as specified in the SMART™ RACE cDNA Amplification Kit handbook:

➢ GC content 50-70%
➢ Tm >70ºC
➢ Avoidance of primer dimers, and stable hairpins.

2.2.1.1.4 Primers for site directed mutagenesis

Primers used for site directed mutagenesis were designed using Stratagenes QuikChange® Primer Design Program (http://www.stratagene.com/sdmdesigner), with parameters set for the

QuikChange® protocol. The sequence of the DNA to be used as a template was inputted and desired base mutations were selected (see 2.2.1.4). The program designed primers of 25 - 45bp in length with Tm >78ºC, GC content >40%, carrying the desired mutation around the middle of the primer and terminating with one or more G/C at each end as recommended in the Stratagenes QuikChange® handbook. Tm's were verified using the Stratagene Quikchange Primer T_m Calculator (http://www.stratagene.com/QPCR/tmCalc.aspx).

2.2.1.2 Standard PCR protocol

For amplification of products for cloning, PCR conditions were optimised using BioMix™ Red (Bioline) in a 25μl reaction:

Sequencing mix:

Forward primer (20μM)	1μl
Reverse primer (20μM)	1μl
Bioline RM Taq	12.5μl
Template DNA	1μl
H_2O	up to 20μl*

* In some cases 1μl DMSO or 1μl 50mM $MgCl_2$ were added to the reaction mix and the volume of water adjusted accordingly (see Table 2.4)

Once optimised PCRs were repeated using the same PCR conditions, with Expand High Fidelity PCR system (Roche) in a 50μl volume.

Sequencing mix:

Forward primer (20μM)	1μl
Reverse primer (20μM)	1μl
Expand High Fidelity TAQ	25μl
dNTPs (10mM each)	1μl
Template DNA	1μl
H20	up to 50μl*

* In some cases 1μl DMSO or 1μl 50mM $MgCl_2$ were added to the reaction mix and the volume of water adjusted accordingly (see Table 2.4)

All PCRs were performed using an Eppendorf® mastercycler gradient PCR machine (Eppendorf). For standard reactions the program used was:

1.	Denaturation	94ºC	2min
2.	Denaturation	94ºC	30 secs
3.	Annealing	see Table 2.4	30 secs
4.	Elongation	72ºC	30secs
5.		Repeat steps 2-4 34x	
6.	Final Elongation	72ºC	10min
7.	HOLD		4ºC

Conditions were optimised by adjusting the annealing temperature using a gradient program. 10μl PCR product was run on a 1% Agarose/TAE gel alongside 3μl GeneRuler™ 1Kb Ladder

(Fermentas). PCR products of successful high fidelity PCRs were cloned into the pCR$^{®}$II-TOPO$^{®}$ plasmid (section 2.2.2).

2.2.1.3 PCR based screening of positive recombinants

To identify positive recombinants of ligation reactions, colonies were picked and used to inoculate 40μl ddH$_2$0, followed by inoculation of 1ml LB including 50μM Ampicillin. Inoculated H$_2$0 was heated at 99ºC for 10 minutes, and subsequently centrifuged for 10 minutes at 13,000 RPM. This was used as a template in PCR reactions.

Sequencing mix:

Region specific primer R (100μM)	0.1μl
Primer T3 60ºC (100μM)	0.1μl
Bioline RM Taq	6.25μl
Template (Inoculated H$_2$O)	4μl
H20	2.05μl

Touchdown PCR:

1.	Denaturation	94ºC	5 min
2.	Annealing	72ºC	1 min 15 sec
3.	Denaturation	94ºC	30 secs
4.	Annealing	70ºC	30 secs
5.	Elongation	72ºC	45 secs
6.	Repeat steps 5-7 1x		
7.	Denaturation	94ºC	30 secs
8.	Annealing	68ºC	30 secs
9.	Elongation	72ºC	45 secs
10.	Repeat steps 9-11 1x		
11.	Denaturation	94ºC	30 secs
12.	Annealing	66ºC	30 secs
13.	Elongation	72ºC	45 secs
14.	Repeat steps 13-15 1x		
15.	Denaturation	94ºC	30 secs
16.	Annealing	64ºC	30 secs
17.	Elongation	72ºC	45 secs.
18.	Repeat steps 16-19 1x		
19.	Denaturation	94ºC	30 secs
20.	Annealing	62ºC	30 secs
21.	Elongation	72ºC	45 secs.
22.	Repeat steps 21-23 1x		
23.	Denaturation	94ºC	30 secs
24.	Annealing	60ºC	30 secs
25.	Elongation	72ºC	45 secs.
26.	Repeat steps 21-23 1x		
27.	Denaturation	94ºC	30 secs
28.	Annealing	58ºC	30 secs

29.	Elongation	72ºC	45 secs.
30	Repeat steps 21-23 1x		
31.	Denaturation	94ºC	30 secs
32.	Annealing	56ºC	30 secs
33.	Elongation	72ºC	30 secs.
34	Repeat steps 21-23 1x		
35.	Denaturation	94ºC	30 secs
36.	Annealing	54ºC	30 secs
37.	Elongation	72ºC	30 secs.
38.	Repeat steps 25-27 17x		
39.	Final Elongation	72ºC	10 min
40.	HOLD	4ºC	

10μl of each reaction was run on a 1% Agarose/TAE gel for analysis alongside 3μl GeneRuler^{TM} 1Kb Ladder (Fermentas). Colonies displaying an amplified band of the expected size were grown up to make a miniculture. Minicultures were analysed by KpnI digest to confirm they carried the desired insert, and subsequently sequenced.

2.2.1.4 Site directed mutagenesis

Site-directed mutagenesis was carried out following a modified version of the protocol described in the Quikchange Site-Directed mutagenesis kit (Stratagene), using Phusion™ Hot Start High-Fidelity DNA Polymerase (finnzymes). Primers were designed as described (section 2.2.1.1.4) to incorporate the desired mutation flanked by neighbouring sequences homologous to the template. The following PCR mix was used:

Primer F (10μM)	1μl (approx. 125ng)
Primer R (10μM)	1μl (approx. 125ng)
10x Buffer	5μl
Template DNA	5 or 50ng
dNTPs (10mM each)	1μl
Phusion Polymerase	0.5μl
[MgCl$_2$ (50μM)	1μl]*
[DMSO	1μl]*
H$_2$O	Up to 50μl

* DMSO and MgCl$_2$ were added to optimise reactions, and were not present in all PCRs (see Table 2.4).

PCR conditions:

1.	Denaturation	95ºC	2 min
2.	Denaturation	95ºC	30 sec
3.	Annealing	55ºC	1 min
4.	Elongation	68ºC	8 min
5.	Repeat steps 2-4 18x		
6.	Final Elongation	68ºC	15 min
7.	HOLD	4ºC	

At this point 10μl was run on a 1% Agarose/TAE gel alongside the same amount of template used and 3μl GeneRuler™ 1Kb Ladder (Fermentas) to analyse the success of amplification. The remaining reaction, as well as the equivalent amount of unamplified template DNA was incubated for 1 hour at 37ºC with 1μl DpnI enzyme (New England Biolabs) to digest parental DNA that did not contain the inserted mutation. To deactivate the DpnI enzyme, reactions were incubated at 65ºC for 60 minutes then placed on ice. Again 10μl of each reaction was run on a 1% Agarose/TAE gel alongside 3μl GeneRuler™ 1Kb Ladder (Fermentas) and the digested template DNA to assess the success of parental plasmid degradation. Samples that appeared to have undergone successful amplification were transformed using the standard procedure (2.2.5). 4-5 colonies were analysed by KpnI digestion and those displaying bands of the expected insert and vector sizes were sequenced to confirm that they carried the desired mutation.

2.2.2 TA cloning of PCR products

PCR products were cloned into pCR®II-TOPO® plasmid using the TOPO TA Cloning® kit (Invitrogen). 2μl PCR Product was mixed with 1μl salt solution, 1μl sterile H_2O, and 1μl TOPO® vector. The reaction was incubated at room temperature for 5-30 minutes, and subsequently chilled on ice for 5 minutes. 0.5-2μl of the ligation reaction was transformed into TOP10F' competent cells (Invitrogen) as described in section 2.2.4. To allow blue/white screening of bacterial colonies 40μl of 40mg/ml X-gal (Promega), and 40μl 100μg/μl IPTG (Sigma) were applied to plates, and left to dry before plating. 4-5 white colonies were selected for restriction analysis of miniprep DNA.

2.2.3 Ligation

2.2.3.1 Preparation of insert DNA

Insert DNA was recovered (usually from the TOPO vector) by digestion with the relevant enzyme (usually KpnI):

Miniprep DNA	30-50μl
10x BSA	6.5μl
10x Enzyme buffer	6.5μl
Restriction Enzyme	1.5μl
H20	up to 65μl

The reaction was incubated at 37ºC for 1 hour, and incubated at 65ºC for 15 minutes to inactivate the enzyme. The whole reaction was run at 40V for 3-4 hours on a 1% Agarose/TAE gel containing 1.5μl Ethidium Bromide (10mg/ml). Released insert DNA was extracted from the gel using Qiaquick spin columns (Quiagen), and following the Gel Extraction Spin protocol. DNA was eluted in 50μl distilled H_2O and 6μl was set aside to estimate the amount of DNA in the eluate, the remainder (~40μl) was precipitated for 1 hour at -20ºC (see 2.2.9). DNA was resuspended in distilled H_2O at a concentration of 10ng/μl.

2.2.3.2   Preparation of vector DNA

Vector DNA was digested with the relevant enzyme (usually KpnI):

Midiprep DNA	100ng
10x BSA	8.5μl
10x Enzyme buffer	8.5μl
Restriction Enzyme	1.5μl
H20	up to 85μl

The reaction was incubated at 37ºC for 90 minutes, then at 65ºC for 15 minutes to inactivate the restriction enzyme. Shrimp Alkaline Phosphatase (SAP, Invitrogen) was used to dephosphorylate template DNA ends and prevent self-ligation. The following were added to the inactivated digest reaction:

SAP 10x Buffer	10μl
SAP	1.5μl
H_2O	up to 100μl

The reaction was incubated for 1 hour at 37ºC, and the enzyme was deactivated by incubation at 65ºC for 15 minutes. DNA concentration (expected to be 10ng/μl) was verified by running 5μl and 1μl of the final reaction mix on a 1% Agarose/TAE gel alongside 3μl GeneRuler™ 1Kb Ladder (Fermentas).

2.2.3.3   Ligation reaction

Ligations were carried out using T4 DNA ligase (Invitrogen, High concentration T4 ligase). Unless otherwise stated 20ng (2μl) insert DNA and 50ng (5μl) dephosphorylated vector DNA were used in the following reaction:

Insert DNA	2μl
Vector DNA	5μl
5x ligation buffer	4μl
1x T4 ligase	1μl
H_2O	up to 20μl

The reaction was incubated at room temperature for 6 hours, and cooled on ice for 5 minutes prior to transformation into ultracompetent XL-Gold cells (Stratagene).

Table 2.4: *Primer sequences.* Sequences give run from 5' to 3', for primers designed to contain restriction enzyme cleavage sites the sequence recognised by the restriction enzyme is underlined and shown in italics. For primers designed for site directed mutagenesis, nucleotides that are mutated relative to the wildtype sequence are underlined and shown in bold. Where PCR parameters differed from standard reaction conditions the annealing temperature and any additives used are indicated.

Primer Name	Sequence	Supplier	Length (bp)	Tm (ºC)	%GC	Tm of homologous region (ºC)	Annealing Temp/ Additives used
5'RACE primers							
gli3exon25'RACE	CACAGTCCCACGGTAAGGGAGAGAGGG	MWG	27	71	63		
ChickGli35'RACE	GGCCCTTTCCTCGCTAGATGTTGAAGGC	MWG	28	75.1	57.1	50.4	
DarkBlueRACE	CTCAGTATCTCTTGGCCTTTCTTGCCTC	MWG	28	68.2	50		
OrangeRACE	GGATCCCAGGGCGCTGCGCTGAGGG	MWG	25	79.3	76		
Primers containing restriction sites for cloning of regions							
region1F	CGC*GGTACC*GGAGCTCAACTTTTGCAAC	MWG	28	78.8	57.1	53.8	60ºC +DMSO
region1R	CCC*GGTACC*GCTTTGGCTGTCAGGTCC	MWG	27	80.5	66.7	56.4	60ºC +DMSO
region2F	GGC*GGTACC*CTTGGTTGTAGATTCTGGC	MWG	28	76.3	57.1	49.8	55ºC
region2R	CGC*GGTACC*TTCAGCAGAGTGCACGTAC	MWG	27	77.3	60.7	50.6	55ºC
region3F	GCC*GGTACC*GAACTGTGCTCTGCCATCTGA	MWG	30	78.6	60	54.2	54.4ºC
region3R	GGC*GGTACC*CCTTAGGCGGAGAGTTTC	MWG	27	79.6	63	55.1	54.4ºC
Region4 F	GCC*GGTACC*CATGCTGCAAGAGTTACAGG	MWG	29	77.8	58.6	52.4	52ºC
region4 R	CGC*GGTACC*GTAGAAAACTGACAATGTGGG	MWG	30	77	53.3	52.2	52ºC
Region5Forward	CGG*GGTACC*ATTTCAGGTTGTCTTCTCATTAG	MWG	32	68.2	46.9	50.8	57ºC
Region5Reverse	CGG*GGTACC*AAACAATCCAGACAACAAAGT	MWG	30	66.8	46.7	50.9	57ºC
region6f	GGC*GGTACC*CCATTGAAGGCTATGTTGTCTGA	MWG	32	77.2	53.1	55	53ºC
region6R	GGC*GGTACC*CACAGAGGCAGCTAAGGAATA	MWG	30	78.8	56.7	55.9	53ºC
Region7Forward	CGG*GGTACC*TATTCCTTAGCTGCCTCTGTG	MWG	30	70.9	56.7	55.2	55ºC
Region7reverse	GCC*GGTACC*CCTCCATAGCAAAGCAGCC	MWG	28	72.4	64.3	57.7	55ºC
Region7bF	GCC*GGTACC*GCAATTAAGCCTTGTTACTG	MWG	29	77.4	51.7	53.2	55ºC
Region7bR	CGG*GGTACC*TTCAGATGCAGGGTCTACTT	MWG	29	75.4	55.2	51.8	55ºC
Region8F	CGG*GGTACC*TGCTCTGAGTGCTGCTCTG	MWG	28	77.1	64.3	52.1	55ºC
Region8R	CGG*GGTACC*TTGGCAGTAACTCTGATGGTG	MWG	30	76.7	56.7	54	55ºC
Region9F	CGG*GGTACC*AGTTTCAGGGGAGGTTGAC	MWG	28	77.3	60.7	55.2	53ºC
Region9R	GGC*GGTACC*AGGGCTTCAGTTCCACAGG	MWG	28	79.6	64.3	57.6	53ºC
Region1consAR	CCG*GGTACC*GCATCAGTTTCAGGGATCCTGT	MWG	31	78.2	58.1	60.4	Touchdown
Region1plusconsBrev	CGG*GGTACC*GCACACAAGCCGCTCCGAC	MWG	28	85.4	71.4	65.5	
Primers for site-directed mutagenesis							
region1pbxmutationF	CCACCCGCGGTTGAACC**GAA**AG**AGAACTGTCACTCAGGG	Thermo Scientific	39	86.7	59		DMSO
region1pbxmutationR	CCCTGAGTGACAGTTCT**CT**TT**C**GGTTCAACCGCGGGTGG	Thermo Scientific	39	86.7	59		DMSO
region1meispbxmutationF	CGCGGTTGAACCAAATCAGAA**T**TGT**G**ACT**G**AGGGCCGGTG	Thermo Scientific	40	84.4	55		DMSO
region1meispbxmutationR	CACCGGCCCT**C**AGT**C**ACA**A**TTCTGATTTGGTTCAACCGCG	Thermo Scientific	40	84.4	55		DMSO

pbxmutregion1meisconsAF	CCACCCGCGGTTGAACC**GAAAG**AGAATTGTGACTGAGGG	MWG	39	84.8	56.4		MgCl$_2$
pbxmutregion1meisconsAR	CCCTCAGTCACAATTCT**CT**TT**C**GGTTCAACCGCGGGTGG	MWG	39	84.8	56.4		MgCl$_2$

Primers for the generation of RNA probes

cPbx1F	CGGACCCTCAGCTGATGCGG	MWG	20	64.2	70		61.9 + DMSO
cPbx1R	GCACTGATGCCCTGCGGACT	MWG	20	64	65		61.9 + DMSO
CPbx3F	GACGAGGCGCAAGCAAATAAG	MWG	21	62.5	52.4		61.9 + DMSO
CPbx3R	GTCACAGAAGATGGAGTAGTTGCG	MWG	24	60.8	50		61.9 + DMSO
CPbx4F	CAGAGCTTGGATGAGGCCCAG	MWG	21	60.1	61.9		Touchdown
CPbx4R	GTGATGGATGAAGGGGTGGTCG	MWG	22	60.4	59.1		Touchdown
Cprep1F	CTACACAAGGATCAGAAGGCACAAC	MWG	25	62	48		Touchdown
Cprep1R	CTGAACGGCCAATGTAGCACC	MWG	21	62.7	57.1		Touchdown
Cprep2F	CGCTGCTCTTTGAGAAGTGTGAG	MWG	23	60.8	52.2		Touchdown
Cprep2R	CGTCCTCCTCAGTGCCATCTAG	MWG	22	60.3	59.1		Touchdown

Primers used for EMSA

EMSA_F	GTTGAACCAAATCAGAACTGTCACTCAGGG	MWG	30				
EMSA_R	ACCGGCCCTGAGTGACAGTTCTGATTTGG	MWG	29				
EMSAPbxmut_F	GTTGAACCGAAAGAGAACTGTCACTCAGGG	MWG	30				
EMSAPbxmut_R	ACCGGCCCTGAGTGACAGTTCTCTTTCGG	MWG	29				
EMSAMeismut_F	GTTGAACCAAATCAGAATTGTGACTGAGGG	MWG	30				
EMSAMeismut_R	ACCGGCCCTCAGTCACAATTCTGATTTGG	MWG	29				
EMSAMPmut_F	GTTGAACCGAAAGAGAATTGTGACTGAGGG	MWG	30				
EMSAMPmut_R	ACCGGCCCTCAGTCACAATTCTCTTTCGG	MWG	29				
Pbx2mutf	GTTGAACCAAATCAGAACTGTCATAGAGGG	MWG	30				
Pbx2mutR	ACCGGCCCTCTATGACAGTTCTGATTTGG	MWG	29				
Mppbx6mutf	GTTGAACCGAAAGAACTTGTTGATAGAGGG	MWG	30				
Mppbx6mutr	ACCGGCCCTCTATCAACAAGTTCTTTCGG	MWG	29				

Other

T360deg	CGCAATTAACCCTCACTAAAGGG	MWG	23	63.9	47.8		Touchdown
NotchF	CCAACCGAGACATCACGGACCAC	MWG	23	65.7	60.9		68ºC
Notch R	GCCGAGGAGTAACAGCTGTGCTG	MWG	23	68.5	60.9		68ºC
mGli3 4000	GCCTACTGCATAATCGCAAGG	MWG	21	59.9	52.4		57ºC
mGli3 4929	GGGATTCAAGGGTCCAAGC	MWG	19	57	57.9		57ºC
SP6	ATTTAGGTGACACTATAGAA	MWG	20	44.2	30		
P1230R	CTCGACCTGCAGGCTAG	MWG	17	64.7	64.7		
T3	CAATTAACCCTCACTAAAGGG	MWG	21	55.6	42.9		
T3upstream	GACCATGATTACGCCAAGCGCG	MWG	22	66.7	59.1		
M13F	GTAAAACGACGGCCAG	Invitrogen	16	52.6	56.3		
M13R	CAGGAAACAGCTATGAC	Invitrogen	17	41.2	47.1		

2.2.4 Recovery of plasmid DNA

Filter paper carrying plasmid DNA was cut up and soaked overnight in 1x TE pH7 at 4ºC. 2µl of TE containing resuspended plasmid was used in transformation reactions.

2.2.5 Transformation

50µl chemically competent *E.coli* cells (ultracompetent XL-Gold cells, Stratagene, unless otherwise stated) were defrosted on ice. 0.5-2µl plasmid DNA or ligation reaction was added to the cells, followed by 20 minutes incubation on ice. Cells were heat shocked for 45 seconds at 42˚C, and left to recover on ice for 5 minutes. 200µl pre-warmed SOC medium (Sambrook *et al.*, 1989) was added, followed by incubation at 37˚C in a shaking incubator (200rpm) for 45-60 minutes. Cells were plated on pre-warmed LB-Agar plates (Sambrook *et al.*, 1989) containing the appropriate antibiotic (50µg/ml Ampicillin or Kanamycin). Plates were incubated overnight at 37˚C.

2.2.6 Sequencing

Sequencing reactions were carried out using BigDye Terminator v3.1 Cycle Sequencing Kit (Applied Biosystems). The reaction mix was made up as follows:

0.5-1µg Plasmid DNA	~5µl of a miniprep
sequencing primer (3.2µM)	1µl
5x ABI sequencing buffer	3µl
BigDye® Terminator v3.1	1µl
ddH$_2$O	up to 10µl

Sequencing reactions were carried out according to the following cycling conditions:

1.	Denaturation	96ºC	5min
2.	Denaturation	96ºC	30 secs
3.	Annealing	50ºC	15 secs
4.	Elongation	60ºC	4 min
5.	Repeat steps 2-4 44x		
6.	HOLD		4ºC

3µl 3M Sodium Acetate pH5.3, 62.5µl 100% ethanol and 24.5µl ddH$_2$O were then added to the reaction, which was vortexed and incubated at room temperature for 20 minutes. The reactions were then centrifuged for 20 minutes at 13,000 rpm. The supernatant was removed and replaced with 100µl 70% ethanol. Samples were spun for a further 20 minutes, after which the supernatant was removed and pellets left to air-dry. Sequencing gels were run by the University of Sheffield Core Genomics sequencing service in an ABI 3730 capillary sequencer (Applied Biosystems).

2.2.7 DNA preparation

2.2.7.1 Bacterial Artificial Chromosome (BAC) DNA

Bacterial Artificial Chromosome (BAC) DNA was prepared using standard Qiagen buffers, following a protocol supplied by Jaime Carvajal (ICR, London). Cultures were grown overnight in a 32ºC shaking incubator (200RPM) in 10ml LB medium containing 12.5µg/ml chloramphenicol. 1.7ml of the culture

was pelletted by centrifugation at 13,000RPM for 1min. The supernatant was removed, replaced with 100µl buffer P1 and the cells were resuspended by vortexing. 250µl buffer P2 was added, and the lysis reaction was incubated for 4-5 minutes at room temperature. The reaction was neutralised by the addition of 150µl buffer P3, inverted 7/8 times, then incubated on ice for 5 minutes. Reactions were centrifuged for 5 minutes at 13,000RPM, and the supernatant was transferred to a fresh tube. The DNA was precipitated by the addition of 1ml ethanol, and samples were incubated in a dry-ice/ethanol bath for 15 minutes, followed by centrifugation at 13,000RPM for 30 minutes at 4ºC. DNA pellets were washed with 200µl 70% ethanol, followed by centrifuging for a further 10 minutes at 13,000RPM. Pellets were air-dried, and resuspended in 50µl H_2O.

2.2.7.2   Plasmid DNA

Plasmid preparations were made using Qiaprep columns (Qiagen), following Qiaprep spin protocols for mini (<20µg DNA), midi (<100µg DNA) and maxi (<500µg DNA) sized preparations. Cultures were grown overnight in a 37ºC shaking incubator (200RPM) in 3ml, 50ml and 100ml LB respectively containing the appropriate antibiotic. This culture was subjected to centrifugation at 6000RPM for 5 minutes (miniprep) or 15 minutes (midi or maxiprep). The supernatant was removed, and plasmid DNA was isolated from the cells using spin protocols listed in the QIAGEN® Plasmid Purification Handbook. Pellets were dried, and resuspended in 200µl (midiprep), or 500µl (maxiprep) H_2O. DNA to be used for electroporation was subject to phenol/chloroform extraction and DNA Precipitation.

2.2.8   Phenol/Chloroform extraction

For phenol/chloroform extraction the initial volume of DNA in solution was taken to be 1 volume. First, 1 volume of room temperature saturated phenol was added to the DNA and the solution was vortexed for 30 minutes, followed by centrifugation at 13,000RPM for 20 minutes. The aqueous phase was recovered into a fresh tube and 0.5 volumes phenol + 0.5 volumes chloroform/isoamylic alcohol (24:1) were added, again followed by 30 minutes vortexing, and 20 minutes centrifugation. The aqueous phase was recovered into a fresh tube, to which 1 volume chloroform/isoamylic alcohol (24:1) was added. The solution was vortexed for 30 minutes and then centrifuged for 20 minutes. Following recovery of the aqueous phase, 1 volume chloroform/isoamylic alcohol (24:1) was added and the sample was subjected to a final 10 minute vortex followed by 10 minutes centrifugation. The aqueous phase was recovered and 6µl was set aside for estimation of DNA concentration. The remaining DNA was precipitated and resuspended to the desired concentration.

2.2.9   DNA precipitation

Unless otherwise specified DNA was precipitated by the addition of 1/10 volumes of 3M Sodium Acetate (pH 5.3), 2 volumes 100% ethanol and 1µl glycogen, followed by incubation at -20ºC overnight. DNA was pelletted by centrifugation at 13,000RPM for 30 minutes at 4ºC, washed with 200µl 70% ethanol, and centrifuged for a further 10 minutes (13,000RPM, 4ºC). The supernatant was removed and pellets were left to air-dry at room temperature. DNA was resuspended to the desired

concentration in H$_2$O, with the exception of DNA to be used for electroporation which was resuspended in PBS.

2.2.10 Restriction digestion

Restriction digests were carried out using restriction enzymes supplied by New England Biolabs, Promega or Roche, with the corresponding buffers. For DNA analysis 4μl of a miniprep, or 100μg of a midi/maxiprep was digested with 0.5μl restriction enzyme (~5U), 2μl 10x buffer, 2μl 10x BSA in a 20μl reaction, at 37ºC for 1 hour. 10μl of the digest reaction was run on a 1% Agarose/TAE gel and compared to GeneRuler^{TM} 1Kb Ladder (Fermentas).

2.2.11 Estimation of DNA concentration

To estimate DNA concentration, 1μl, 0.1μl and 0.01μl were run on a 1% Agarose/TAE gel with 1.5μl 10mM Ethidium Bromide, flanked by 3μl and 6μl of GeneRuler^{TM} 1Kb Ladder (Fermentas). DNA concentration was estimated by comparing the intensity of sample DNA bands to that of ladder bands of known concentration.

2.2.12 RNA isolation

Total RNA was isolated from HH stage 12-13 chick embryos or E9.5 mouse embryos using TRIzol® reagent (Gibco BRL®) following the protocol supplied for RNA isolation from tissues, or from a T-75 flask of confluent cells following the protocol supplied for cells grown in a monolayer. RNA was resuspended in 5μl DEPC treated H$_2$O and stored at -80ºC. RNA concentration was measured by spectrometry on a Beckman DV® 520 General purpose UV/Vis spectrophotometer (Beckman), measuring the optical density at 260nm. For 5'RACE on chick cDNA the neural tube and notochord, or mesoderm were isolated from the embryos and used for RNA isolation.

2.2.13 cDNA synthesis

cDNA was prepared using Superscript^{TM} First-Strand Synthesis System (Invitrogen) for RT-PCR, following the protocol provided for cDNA synthesis from total RNA.

2.2.14 Visualisation and image capture of agarose gels

All gels made contained 1.5μl Ethidium Bromide (10mg/ml) per 50ml. Samples were viewed under UV light and images captured using Uvidoc system (Uvitec).

2.2.15 5'RACE

2.2.15.1 5'RACE PCR standard conditions

1.	Denaturation	94ºC	2 min
2.	Denaturation	94ºC	30 secs
3.	Annealing	72ºC	3 min
4.	Repeat steps 2-3 4x		

5.	Denaturation	94°C	30 secs
6.	Annealing	70°C	30 secs
7.	Elongation	72°C	3 min
8.	Repeat steps 6-7 4x		
9.	Denaturation	94°C	30 secs
10.	Annealing	68°C	30 secs
11.	Elongation	72°C	3 min
12.	Repeat steps 10-11 26x		
13.	Final elongation	72°C	10 min
14.	HOLD 4°C		

2.2.15.2 cDNA synthesis

2.2.15.2.1 Gli3 specific cDNA synthesis and control reactions

1μg total RNA was used to prepare 5'RACE ready cDNA using Clontech SMART^{TM} RACE cDNA amplification kit and following the protocol supplied (Barnes, 1994; Cheng *et al.*, 1994). To confirm the presence of Gli3 transcript in mouse cDNA generated, a control PCR was carried out using mGli3 4000 (Forward) and mGli3 4929 (Reverse) primers on 1μl cDNA using standard PCR conditions. To check the quality of chick RNA, a control PCR was carried out using standard PCR conditions with primers against the Notch 2 transcript (Table 2.4). PCR conditions used for the 5'RACE PCR were tested by carrying out the positive control RACE PCR experiment using RACE-Ready cDNAs generated from Control Human Placental Total RNA as described in the protocols handbook.

2.2.15.2.2 5'RACE from mouse Gli3 exon 2

2.2.15.2.2.1 Initial amplification of mouse Gli3 5'transcript from RACE ready cDNA

Amplification of Gli3 5' transcripts by RACE was carried out as described using Gli3exon25'RACE primer with standard 5'RACE PCR conditions and yielded several faint bands (primary products). To amplify the signal of transcripts the PCR was repeated under the same conditions using 1μl of the primary product as a template. Amplification of the primary product using standard PCR conditions (10 cycles annealing at 68°C) produced the same sized DNA fragments. The amplification procedure (consisting of the initial amplification of RACE ready cDNA and reamplification of the primary product) was repeated using slightly modified conditions (modifications to standard RACE PCR protocol: step 4, repeat 6x; step 10, 69°C; step12, repeat steps 10-11 32x). The second set of amplifications yielded the same products as before, with the addition of some slightly larger products.

2.2.15.2.2.2 Amplification from exonic sequences identified in the initial reaction

From the results of the initial 5' RACE from mouse exon 2, a further 2 gene specific primers were designed in order to verify that the transcripts were not too long for the true 5' end to be identified by the original procedure. 5' RACE PCR was carried out on the primary product using these internal primers (Orange RACE and Dark Blue RACE) using the standard RACE PCR protocol with the following modifications:

10.	Annealing	69ºC	30 secs
11.	Elongation	72ºC	3 min
12.	Repeat steps 10-11 4x		
13.	Denaturation	94ºC	30 secs
14.	Annealing	68ºC	30 secs
15.	Elongation	72ºC	3 min
16.	Repeat steps 10-11 29x		
17.	Final elongation	72ºC	10 min
18.	HOLD 4º		

2.2.15.2.3 5'RACE from chick mesoderm or neural tube and notochord cDNA

Amplification of Gli3 5' transcripts by RACE was carried out as described using ChickGli35'RACE primer with standard 5'RACE PCR conditions.

2.2.15.3 Cloning of products

1μl of each of the reamplification PCRs carried out on the primary 5' RACE PCR product were TA cloned as described (2.2.2). Additionally, in some instances specific band sizes were targeted for cloning by gel extraction. Following transformation colonies were selected and analysed by EcoRI digestion of miniprep DNA. Several representative colonies of each fragment size observed were sequenced as described (2.2.6) using M13 F and M13R primers.

2.2.16 Preparation of nuclear extract

For the preparation of nuclear extracts all reagents and plastic ware were prepared in advance and cooled to 4ºC prior to use. For buffers A and C, DTT and protease inhibitors were added immediately before use. For preparation of Nuclear extract from chicken embryos 100-150 chicken embryos were harvested at HH stage 10-12 in ice-cold PBS, and transferred to a 2ml tube. Tissues were pelletted by centrifuging for 2 minutes at 3000rpm at 4ºC. Pellets were rinsed in ice-cold PBS, and pelletted as before. Pellets were resuspended in 0.4ml Buffer A (10mM Hepes pH7.9, 10mM KCl, 0.1mM EDTA, 0.1mM EGTA, 1mM DTT, 2μg Aprotin, 2μg Leupeptin, 750μM PMSF), and tissues were homogenised.

For preparation of nuclear extract from cell lines, cells were harvested from 8-12 confluent T75 flasks. Cells were rinsed 2x in 50ml ice-cold PBS and pelletted by centrifuging for 5 minutes at 3000rpm. Cells were transferred into a 2ml tube, and resuspended in 0.4ml Buffer A by pippetting. All samples were treated in the same manner from this point

Samples were incubated on ice for 15 minutes, followed by the addition of 16.7μl 10% NP40. Tubes were briefly vortexed, and centrifuged at 4000RPM at 4ºC for 5 minutes. At this stage the supernatant containing cytoplasmic extracts was removed, aliquotted and immediately frozen in liquid nitrogen. Pellets were resuspended in 50μl buffer C (20mM Hepes pH7.9, 0.1mM EDTA, 0.1mM EGTA, 0.4M NaCl, 1mM DTT, 5μg Aprotin, 2μg Leupeptin, 2mM PMSF) by gentle

vortexing at 4ºC for 15 minutes. Samples were pelleted by centrifugation at 4000RPM for 15 minutes at 4ºC. Supernatants containing nuclear extract were recovered, aliquotted and immediately frozen in liquid nitrogen. Nuclear and cytoplasmic extracts were stored at –80ºC until use. In each case 5μl was set-aside prior to freezing in order to calculate protein concentration. Protein concentrations were calculated using Bradford reagent (Biorad) as described by the manufacturer, following the standard procedure supplied and using BSA protein standards prepared from a 10mg/ml stock (Promega).

2.2.17 Electrophoretic Mobility Shift Assay (EMSA)

2.2.17.1 Labelling of oligonucleotides

10pmol forward and reverse primers were annealed in 20μl 1x restriction enzyme buffer (diluted in H_2O), by incubation at 37ºC for 20 minutes, followed by incubation at room temperature for 10 minutes. For the fill-in reaction the following reagents were added to the sample:

dATP, dTTP, dGTP (5μM)	1.5μl	
Klenow buffer 4ml	4μl	
α^{32}PdCTP (3000Ci/mmol)	5μl	
Klenow enzyme	2μl	
H_2O	7.5μl	

The reaction was incubated for 40 minutes at 30ºC, followed by the addition of 3ml dCTP (0.5mM) and incubation for 10 minutes at room temperature. TE was added to a final volume of 50μl. Oligos were purified by filtering through a Sepadex G-25 column by centrifugation at 1000RPM for 2-3 minutes. Radioactivity was measured using a scintillation counter.

2.2.17.2 Preparation of cold competitor oligonucleotides

1500pmol forward and reverse primers were annealed in a 30μl 1x restriction enzyme buffer (10X React 2, Invitrogen, diluted in H_2O) by incubation at 37ºC for 20 minutes, followed by incubation at room temperature for 10 minutes. Various amounts of cold competitor were used in the binding reaction to give the desired excess.

2.2.17.3 Binding reaction

For the binding reaction 1.5mg nuclear extract was combined with 2μl Poly dI:dC (1μg/μl), 2μl H_2O, and made up to a final volume of 22μl using BC100 (20mM Hepes pH7.9, 100mM KCl, 5mM MgCl2, 1mM DTT, 17% glycerol). Where relevant, competitors were also added at this stage. For cold-oligo competitors 2μl oligo was added a concentration calculated to give the desired excess. For antibody competition experiments 3μl of the relevant antibody was used. Samples were incubated for 15 minutes at room temperature, followed by the addition of 2μl labelled oligo ($4x10^5$ cpm). Samples were incubated for a further 15 minutes at room temperature to allow binding, and subsequently loaded onto pre-cast gels.

2.2.17.4 Antibodies

Meis1/2 (N-17), Pbx1/2/3 (c-20) and Meis1/2 (H-80) antibodies (Santa Cruz Biotechnology inc.) were used as competitors in EMSA experiments.

2.2.17.5 Gel preparation and running

Samples were run on 5.5% Polyacrylamide:bis gels (9.2ml 30:polyacrylamide:bis, 1.25ml 10x TBE, 260µl 25%APS, 20µl TEMED, 39.3ml H20). Gels were cast 20 minutes prior to loading and were run in 0.25X TBE. Gels were run at room temperature at a voltage of 110V for 3.5 hours.

2.2.17.6 Detection

Gels were transferred to membranes, labelled and overlaid with cling film. Gels were dried under a vacuum on a heated bed for 1 hour at 80ºC. Once dried the cling film was removed and gels were exposed in the dark to Biomax XAR film (Kodak) at –80ºC for 4-6 hours or overnight. Images were scanned using a Epson Perfection 4490 Photo scanner (Epson) using ABBYY FineReader 6.0 Sprint Plus software (ABBYY Software Ltd.), and manipulated using Adobe Photoshop CS version 8.

2.2.18 Western blotting

Western blotting was performed essentially as described by Sambrook *et al.* 1989, using the Mini-PROTEAN Tetra Electrophoresis System (Bio-Rad). Unless otherwise stated reagents were supplied by Bio-Rad. For each sample 7.5µg nuclear extract was thawed on ice and reconstituted to 15µl with loading buffer and water. Samples were denatured at 100ºC for 10 minutes and collected by centrifugation. Samples were loaded into pre-cast SDS-12% Polyacrylamide gels, overlaid with a 5% stacking gel (Sambrook *et al.*, 1989). Electrophoresis was performed at 50V until the sample entered the resolving gel, and gradually increased to 120V in 5V intervals, the final voltage was maintained for approximately 3 hours. Samples were transferred to nitro-cellulose membrane using a current of 90mA at 4ºC overnight. Membranes were incubated in 5% blocking solution (5% dried milk in TBST) for 2-3 hours whilst gently agitating, followed by overnight incubation at 4ºC in fresh blocking solution containing the appropriate primary antibody (Meis1/2 [N-17], 1:1000; Pbx1/2/3 [c-20], 1:500, Santa Cruz Biotechnology inc.) Non-bound antibody was removed by washing three times for 30 minutes in TBST while agitating. For detection, membranes were incubated for 2 hours at room temperature in 5% blocking solution containing horseradish peroxidase-conjugated secondary antibody (Peroxidase labelled AntiMouse IgG, 1:20000; Peroxidase labelled AntiGoat IgG 1:5000, VECTOR laboratories). Non-bound secondary antibody was removed by washing three times for 30 minutes while agitating. The antibody was detected using Supersignal® WestPico (Pierce), and gels were exposed in the dark to Biomax XAR film (Kodak) for 10 seconds - 20 minutes. Images were captured as for EMSA experiments.

2.3 Cell Culture Techniques

All cells were grown at 37ºC, 5% CO_2 in T-75 flasks containing 10ml media. PC12 cells were grown in RPMI 1640 media (Sigma) supplemented with 10% FBS (Invitrogen), 5% FHS (Invitrogen), 1% L-Glutamine (Invitrogen) and 1% PenStrep (Sigma). Cells were split 1:4 every 3-4 days. Media was discarded and cells were washed with 3ml PBS by gentle agitation. The PBS was discarded, and replaced with another 10ml PBS. The flask was incubated at 37ºC incubator for 5-15 minutes and intermittently tapped to detach the cells. Cells were pelletted in a 50ml falcon tube by centrifugation at 500rpm for 5 minutes. Cells were split into clean flasks containing fresh media.

DAOY cells were grown in Dulbeccos Modified Eagle Medium (DMEM, Gibco) supplemented with 10% FBS (Invitrogen), 1% L-Glutamine (Invitrogen) and 1% PenStrep (Sigma). Cells were split 1:4 every 2-3 days. Media was discarded and cells were washed with 10 ml PBS by gentle agitation. The PBS was discarded, and replaced 3ml Trypsin. The flask was incubated at 37ºC for 5-10 minutes and intermittently tapped to detach the cells. 10ml media was added to inactivate the Trypsin, and cells were pelletted in a 50ml falcon tube by centrifugation at 500rpm for 5 minutes. Cells were split into clean flasks containing fresh media.

2.4 In-silico analysis

2.4.1 Genomic sequence retrieval

The genomic location of Gli3 in various species were identified using the Ensembl project; www.ensembl.org (Hubbard *et al.*, 2007). Their Ensembl gene and transcript identifiers, and genomic locations are (respectively):

Chick ENSGALG00000012329, ENSGALP00000020121, Chr. 2: 50,832,042-51,026,884

Human ENSG00000106571, ENST00000265526, Chr. 7: 41,970,205-42,229,420.

Rat ENSRNOG00000014395, ENSRNOT00000019396, Chr. 17: 57,594,126-57,853,747.

Mouse ENSMUSG00000021318, ENSMUST00000021754, Chr. 13: 15,254,867-15,517,860.

Xenopus ENSXETG00000001856, ENSXETT00000003920, scaffold_57: 312,820-442,672.

Zebrafish ENSDARG00000052131, ENSDART00000058992, Chr. 24: 7,858,744-8,146,034.

Fugu SINFRUG00000153715, SINFRUT00000163565, scaffold_210: 236,104-293,133.

Chimp ENSPTRG00000019117, ENSPTRT00000035323, Chr. 7: 41,987,133-42,268,279.

Dog ENSCAFG00000003535, ENSCAFT00000005687, Chr 18: 10,978,911-11,053,613.

Cow ENSBTAG00000010671, ENSBTAT00000014119, Chr 4: 74,032,783-74,345,332.

The Enseml browser is regularly updated and these identifiers were subject to change during the progression of this thesis. Identifiers and locations supplied are accurate as of Ensembl release 46, August 2007. For alignment purposes the genomic sequence of the gene, along with 100kb upstream of the annotated transcriptional start site was downloaded and used in subsequent analyses.

2.4.2   Alignment tools

2.4.2.1   Genomic alignment tools

Genomic sequences were aligned using the Vista tool for comparative genomics (http://genome.lbl.gov/vista; Mayor *et al.*, 2000), using parameters specified in the text. AVID alignments were produced independently using the MAVID multiple alignment server (http://baboon.math.berkeley.edu/mavid/; Bray and Pachter, 2004). Regions of interest identified using the Vista browser were annotated on the multiple alignment file produced. To produce a more accurate alignment of regions of interest, shorter sequences were aligned. Alignments produced by MAVID using the new sequences were compared to those produced by ClustalW (http://www.ebi.ac.uk/Tools/clustalw/; Thompson *et al.*, 1994), using default parameters. In some instances flanking sequences were shortened to optimise the quality of the alignment.

2.4.2.2   General purpose alignments

DNA sequences were aligned using the ClustalW tool (http://www.ebi.ac.uk/Tools/clustalw/; Thompson *et al.*, 1994), with standard parameters.

2.4.3   Binding site identification

Transcription factors previously proposed to be involved in transcriptional regulation of Gli3 were identified in the initial alignment of Gli3 genomic sequences using rVista (Loots *et al.*, 2002). rVISTA scans the Vista alignment for transcription factor consensus binding sites selected from the TRANSFAC database. LEF1_Q2, LEF1_Q6, TCF11, TCFP_Q6, TCF4_Q5, SMAD3, SMAD4 and GLI consensus matrices were selected, and used to search human:rat, human:mouse and human:chick alignments. Binding sites identified were annotated on alignments in Microsoft Word 97 sr-1 (Microsoft) to identify those conserved across all species.

To identify binding sites of other transcription factors, individual regions of interest for each species were inputted to the MatInspector tool of the Genomatix server (http://www.genomatix.de/; Quandt *et al.*, 1995) using the following parameters:

<div align="center">

Library version: Matrix library 6.3

Matrix group: vertebrates

</div>

Sites identified were recorded on the alignment file in Microsoft Word 97 sr-1 (Microsoft). Transcription factor matrices used by MatInspector were originally derived from the Transfac database (http://transfac.gbf-braunschweig.de), which is the main transcription factor knowledge-base. An alternative transcription factor database, Jaspar (http://jaspar.cgb.ki.se/; Sandelin *et al.*, 2004a) was also used, and sequences annotated as above. Binding sites of interest were determined as those conserved amongst human, mouse, rat and chick sequences. Sites which occurred in each species' sequence within the conserved region, but whose position varied, were also noted.

Literature searches were carried out to gain further insight into binding sites of particular interest, the results of which are cited elsewhere. These literature searches also identified mutations used previously

to abolish binding to individual sites, or in some cases to modify the binding potential. Previously described mutations were used to design primers for site directed mutagenesis.

2.4.4 Annotation and analyses of sequence files

Sequence files were assembled and manipulated in ApE (A plasmid editor v1.10.4, http://www.biology.utah.edu/jorgensen/wayned/ape/). This tool was also used for restriction enzyme identification, plasmid annotation and viewing chromatogram files. Sequence alignments were annotated in Microsoft word 97 sr-1 (Microsoft).

2.4.5 Promoter search

Web-based promoter prediction tools used in this study are listed in Table 4.1. Ensembl and UCSC browsers were used for CpG island identification (Gardiner-Garden and Frommer, 1987; Karolchik *et al.*, 2003; Hubbard *et al.*, 2007). Promoter elements identified were annotated on the sequence in Microsoft Word 97 sr-1 (Microsoft).

2.4.6 Blast tools

Sequence identity of plasmid and insert DNA was confirmed using the nucleotide blast tool against the nucleotide collection nr/nt (http://www.ncbi.nlm.nih.gov/BLAST; Altschul *et al.*, 1990) or using a 2-way blast (http://www.ncbi.nlm.nih.gov/BLAST/bl2seq/wblast2.cgi; Tatusova and Madden, 1999) against an expected sequence. Where the species was known (e.g. in RACE reactions) the genomic location of sequenced DNA was identified using the Ensembl blast tool (http://www.ensembl.org/Multi/blastview).

Chapter 3

Use of bioinformatics to identify and analyse Gli3 regulatory regions

3.1 Introduction

Whilst genomic tools for the annotation of protein coding DNA regions are well established, parameters for identification of non-coding regulatory elements are not. Here I shall outline factors that should be considered when attempting to pinpoint regulatory elements of genes by phylogenetic footprinting. Although previous studies agree on the concept that regulatory elements are conserved at a greater level than surrounding DNA not under selection, they disagree in the choice of organisms that should be compared and the threshold level of conservation that must be achieved to predict regulatory potential.

3.1.1 Phylogenetic distance between organisms compared

In a whole genome comparison of human and fugu sequences Woolfe *et al.* (2005) identified a set of 1373 conserved non-coding elements (CNEs). When the search was extended to include invertebrates no significant homology could be found between vertebrate CNEs and invertebrate sequences, although the invertebrate homologues of vertebrate genes associated with CNEs were often identified (Woolfe *et al.*, 2005). Conserved non-coding elements identified in invertebrate genomes have been found to be smaller and less frequent than vertebrate CNEs (Vavouri *et al.*, 2006a). Thus, novel set of CNEs may have evolved in vertebrates to orchestrate specific spatial and temporal patterns of gene expression in the vertebrate body plan. Taking this into consideration, together with the observation that Gli3 is a gene unique to vertebrates, I chose to restrict my sequence comparisons to vertebrate genomes. Within the vertebrate lineage, neutral substitution rates and phylogenetic distance are important parameters that influence the potential to identify regulatory regions. With time, neutral substitutions accumulate in a genome, whilst purifying selection protects regions of functional significance. Although comparisons of distantly related species can be useful in the identification of critically conserved regions, they often fail to identify true enhancer elements (Aparicio *et al.*, 1995). This is because increasing evolutionary distance also raises the likelihood that gene expression pattern, and thus the regulatory elements governing it, will have diverged. In a genome-wide screen Dickmeis *et al.* identified 45 genes expressed at the midline of zebrafish embryos. Of these, only 10 genes were associated with non-coding regions conserved between zebrafish and fugu (Dickmeis *et al.*, 2004). Only one of the conserved regions identified in this study acted as an enhancer element in the zebrafish embryo. Similarly, Thomas *et al.* (2003) found that human:fish alignments yield mostly coding sequence (Thomas *et al.*, 2003). Together these studies and others (such as Dickmeis *et al.*, 2004; Plessy *et al.*, 2005) indicate that the evolutionary distance between human and fish is too great to permit the identification of many non-coding regulatory elements. It has been suggested that elements

conserved through such long evolutionary distances may have a function other than transcriptional regulation. They may be involved in chromosome architecture or chromatin function, acting as modulators of chromatin structure, matrix attachment regions, insulator regions, regions involved in chromosome pairing or origins of replication. Alternatively, they could represent unidentified transcripts that work at the RNA level, or control gene expression by acting as promoters or silencers rather than enhancers (Cremer and Cremer, 2001; Eddy, 2002; Cooper and Sidow, 2003; Pennacchio, 2003; Cooper *et al.*, 2004; Woolfe *et al.*, 2005; Paparidis *et al.*, 2007). However, the compact nature of the fugu genome (around an eighth of the size of the human genome), results in non coding conserved elements residing on average 90% closer to their gene counterpart in fugu than in mammals (Pennacchio, 2003). This makes the fugu genome a useful tool for the identification of distant regulatory elements. Human:fugu comparisons were recently used to identify five conserved non-coding elements in 18kb upstream of fugu Sox9, homologous to regions located up to 290kb upstream of the human orthologue (Bagheri-Fam *et al.*, 2001). Whilst fish genomes can be useful to identify regulatory regions conserved over extreme evolutionary distances, or far from the genomic location of the gene they control, absence of a particular region of conservation in fish genomes should not negate interest in its potential to regulate gene expression amongst amniotes.

Several studies have focussed on finding an optimal evolutionary distance to allow regulatory regions to be identified. In a study of 1.8Mb surrounding the CFTR locus on human chromosome 7q31.3, Thomas *et al.* (2003) investigated the benefits of including various vertebrate species, ranging from fish to primates, in alignments with human. Human:chicken alignments were identified as being of particular interest, since they produced a detection rate higher than any other pairwise species comparison. Human:chicken alignments detected almost all (>98%) coding sequences, and accurately predicted regions conserved across multiple species (Thomas *et al.*, 2003). In a study of the Sox2 locus, Uchikawa *et al.* (2003) used mammal:chicken genome comparisons to identify 25 conserved non-coding elements, 13 of which were shown to have regulatory effects on transgene expression in vivo. Human:mouse alignments did not permit the identification of any element, as there were long stretches of conservation. In contrast, only 2 of the conserved elements were conserved between human and fish alignments (Uchikawa *et al.*, 2003). This demonstrates that the chicken genome serves as a useful intermediate to bridge the gap in sequence conservation between mammalian genomes and other vertebrates.

3.1.2   Inclusion of multiple species in an alignment

The studies described above rely on pairwise alignments. Such alignments preclude the identification of elements lost during evolution that result in expression pattern differing between species (Fougerousse *et al.*, 2000). Thus, no species can be completely informative of the regulatory elements controlling expression of an orthologous gene in another species. Introducing multiple species to an alignment has proven beneficial for a number of reasons. First, it reduces the chance of mis-alignment that can occur in pairwise studies. Comparisons of three different human:mouse alignments revealed that less than 80% of conserved regions overlapped between alignments (Waterston *et al.*, 2002; Couronne *et al.*, 2003). Second, independent lineage deviation of each individual genome since divergence from the last common ancestor serves as an essentially additive effect on the total evolutionary distance between species. This allows substantial knowledge of conserved regions to be obtained from even closely related species. (Stojanovic *et al.*, 1999; Boffelli *et al.*, 2003). By allowing one to compare closely related species to those separated by a longer evolutionary distance, the inclusion of multiple species in an alignment also permits the identification of lineage specific variations (Gottgens *et al.*, 2000; Gillemans *et al.*, 2003; Thomas *et al.*, 2003).

3.1.3   Bioinformatic tools used to identify control regions

Three main classes of pairwise alignment tools have been developed: global, local and glocal. In global alignments one string is transformed into another, whereas in local alignments all regions of similarity between the two sequences are returned. Both methods have advantages and disadvantages. Global alignments are less prone to demonstrating false homology, whilst local alignments are better adapted to cope with rearrangements between non-syntenic orthologous sequences. Glocal alignments combine the two methods, transforming one sequence into the other while allowing for rearrangement events (Brudno *et al.*, 2003b).

Most multiple alignment tools publicly accessible over the Internet limit the size of input sequences, so the investigation of conserved regulatory elements over large regions of DNA requires specialised programmes. MAVID software (Bray *et al.*, 2003) uses the AVID2.2 algorithm to produce a multiple local alignment of input sequences without size limitations.

Literature searches revealed two programmes regularly used for the identification of regulatory non-coding regions of DNA (these tools are reviewed and compared in Nardone *et al.*, 2004). Pipmaker software produces BlastZ alignments that are pairwise local alignments around short sites of exact homology (Brudno *et al.*, 2003a). Results are displayed as a percentage identity plot (pip), a dot plot in pdf format, or as a conventional textual alignment. VISTA software produces pairwise alignments of the species of interest, all relative to a user-defined base

sequence, and allows the user to choose between local (using AVID; Bray *et al.*, 2003), global (using MLAGAN; Brudno *et al.*, 2003a), and glocal (using shuffle-LAGAN; Brudno *et al.*, 2003b) alignments. Results of the alignments are presented in the form of a graph, where the height of the peak at any point represents the percentage conservation of nucleotides between species. Graphs of sequence conservation levels between various species and the base organism are stacked to ease comparison. User defined threshold cut-off values are used to identify conserved genomic elements. The VISTA software used in this thesis offers several advantages over Pipmaker:

➢ Results are displayed in a clearer manner.

➢ Various alignment tools are available, allowing local, global or glocal alignments.

➢ An interactive browser has been developed to aid user manipulation of graphical outputs, and access to alignment details following the initial alignment.

➢ Genome VISTA allows direct addition of sequences from multiple species.

Regulatory VISTA (rVISTA; Loots *et al.*, 2002) can also be used to search alignments for the presence of transcription factor binding sites taken from the Transfac database (Wingender *et al.*, 1996).

3.1.4 Conserved enhancer elements contain transcription factor binding sites

For conserved regions to act as enhancers they must contain binding sites for regulatory proteins. Occupation of these sites by the relevant proteins serves as an input to the regulatory element, which directs a response by mediating the expression pattern of an associated gene (Arnone and Davidson, 1997). To identify which regulatory proteins are capable of binding to a DNA sequence, the sequence can be searched for the presence of characterised binding sites. Binding site information is available in two main forms: consensus binding sites and position/frequency weight matrices.

Consensus binding sites record the preferred base at each position of a binding site. Position frequency matrices record the frequency that each base (A, T, C or G) occurs at a particular position in the binding site, thus accounting for the fact that some sites are more conserved than others (Stormo, 2000; Wasserman and Sandelin, 2004). Query sequences are scored by the frequency at which each position is occupied by the same base in the matrix. Position weight matrices are similar to position frequency matrices, but frequency values are converted to a logarithmic scale. Binding site information both in the form of consensus sites and position weight/frequency matrices is available through a number of publicly available databases, the most widely used being TRANSFAC (Wingender *et al.*, 1996) and JASPAR (Sandelin *et al.*, 2004a). In conjunction with web-based search tools such as MatInspector (Cartharius *et al.*,

2005), ConSite (Sandelin *et al.*, 2004c), MATCH (Kel *et al.*, 2003) and PMSearch (Su *et al.*, 2006) this information can be used to locate the presence of transcription factor binding sites in a DNA sequence (reviewed in Wasserman and Sandelin, 2004; Dickmeis and Muller, 2005). Previous data from our lab and others implicated the Wnt signalling pathway in the initiation of Gli3 expression (Borycki *et al.*, 2000; Mullor *et al.*, 2001), the BMP pathway in its maintenance (Kuschel *et al.*, 2003; Meyer and Roelink, 2003), and the Sonic Hedgehog pathway in mediating Gli3 repression (Marigo *et al.*, 1996; Borycki *et al.*, 1998; Borycki *et al.*, 2000). At the time of this study it was unknown how these pathways regulate Gli3 at the molecular level. They might mediate Gli3 expression directly via their associated transcription factors. If this is the case, one would expect transcription factor binding sites to be conserved within regulatory regions. Alternatively, they could act indirectly by regulating uncharacterised intermediates that in turn regulate Gli3 expression.

3.2   Results

3.2.1   Identification of Gli3 orthologues

In order to generate an alignment, I started by investigating the sequence data available surrounding the Gli3 locus in various species. The Ensembl genome server (v27, Dec 2004) was used to search for orthologues of Gli3 in various genomes. The search result for "Gli3" identified the Ensembl gene entry ENSG00000106571 in *Homo Sapiens*, the sequence of which was shown to correspond to the protein sequence deposited in the NCBI protein database (accession P10071). The orthologue prediction tool was used to identify the corresponding gene in *Pan troglodytes* (ENSPTRG00000019117), *Canis familiaris* (ENSCAFG00000003535), *Mus musculus* (ENSMUSG00000021318), *Rattus norvegicus* (ENSRNOG00000014395), *Gallus gallus* (ENSGALG00000012329), *Xenopus tropicalis* (ENSXETG00000001856), *Danio rerio* (ENSDARG00000004073) and *Tagifugu rubipes* (SINFRUG00000153715). The search term 'Gli3' identified 2 copies annotated as Gli3 in *Gallus gallus* (ENSGALG00000012329 along with ENSGALG00000011630). ENSGALG00000012329 was identified as the true orthologue by alignment of the coding region with the human sequence (Thompson *et al.*, 1994).

3.2.2   Selecting the radia for phylogenetic footprinting analysis

In order to identify putative regulatory elements surrounding Gli3 by phylogenetic footprinting it was necessary to produce a good quality sequence alignment. To permit this it was important to establish that the annotated Gli3 locus for each genome contained the entire coding region. The exon structure in each species was compared to that of human Gli3, since this was the best annotated sequence at the time of the study. Orthologues of human exon 1 were identified in

Canis familiaris (dog), *Rattus norvegicus* (rat), *Xenopus tropicalis* (frog), *Gallus gallus* (chicken) and *Pan troglodytes* (chimp). In *Mus musculus* (mouse) the orthologue of human exon 1 was preceded by a 5' UTR, annotated as an additional exon. This exon (designated exon 0) may constitute the true transcriptional start site and is addressed elsewhere (see Chapter 4). At the time of the study a corresponding exon 0 was not annotated in any of the other species investigated. A sequence orthologous to human exon 1 could not be identified in the *Tagifugu rubipes* (fugu) and *Danio rerio* (zebrafish) genomes.

Protein sequence comparison showed that the zebrafish sequence was missing the first 2 exons. Various sequence search and alignment programmes were used in an attempt to identify the true first exon and thus identify the transcription start site, but a region corresponding to the first two exons of human Gli3 could not be identified. Similarly, the most 5' exon annotated at the Gli3 locus in *Tagifugu rubipes* corresponded to exon 4 of the human gene, the scaffold (1460) carrying this exon terminates 600bp upstream, precluding isolation of 5' sequences. Therefore these organisms were not included in my study.

Human Gli3 spans over 258 kb, and cis-regulatory elements have been shown to reside from within a few kb to more than 1Mb up or downstream, or within intronic regions of the genes they control (Lettice *et al.*, 2002; Bien-Willner *et al.*, 2007). However, regulatory elements are more likely to occur within a few kilobases 5' of the transcription start site, becoming more sparse with increasing distance from the gene of interest (Wasserman *et al.*, 2000; Uchikawa *et al.*, 2003; Uchikawa *et al.*, 2004). Furthermore, analysis of Add and Xt mouse mutants suggest that non-coding regulatory elements are likely to lie within the 80kb region upstream of Gli3 (Pohl *et al.*, 1990; Schimmang *et al.*, 1992). Although the investigation of a discrete area surrounding the Gli3 locus might not identify all enhancer regions, studies of small genomic loci have on numerous occasions proven successful in the identification of such elements (Epstein *et al.*, 1999; Uchikawa *et al.*, 2003). I therefore decided to initially investigate 100kb upstream of human exon 1. Non-coding conserved regions are equally likely to occur within intronic regions (intragenic; Whiting *et al.*, 1991). In order to limit the size of the region under investigation I chose to limit my intragenic search to the first intron of Gli3 (74661bp in human). Therefore, this study investigated approximately 175Kb of DNA surrounding human exon 1.

3.2.3   Generating an alignment
3.2.3.1   Organisms included

Figure 3.1 shows the evolutionary relationship of organisms for which the sequence of interest was successfully extracted from Ensembl. They include examples of warm and cold-blooded vertebrates, consisting of primates, carnivores, rodents, birds and amphibians (Fig. 3.1). Together they reflect evolutionary distances ranging from around 6 million years (Goodman *et al.*, 1998; Chen and Li, 2001) to around 360 million years (Hedges, 2002; Fig. 3.1).
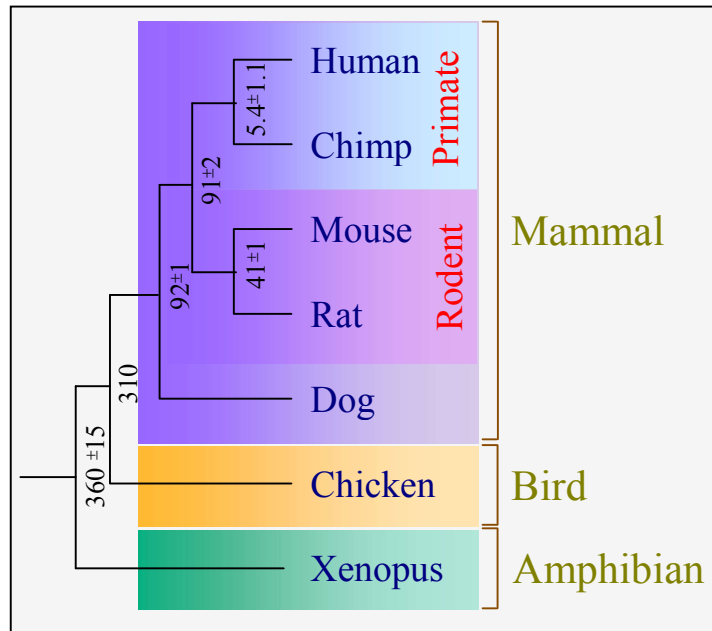
Figure 3.1: *Relationships and divergence times of organisms used in this study*. The figure is adapted from Figure 6 in Hedges *et al.*, 2002, and is based on several multi-gene and multi-protein studies. Numbers shown represent the divergence time between each species (millions of years ago (Mya) ± one standard error). The molecular clock was calibrated using the fossil divergence time of birds and mammals (310 Mya). Branch lengths are not to scale.

I chose to carry out a multiple alignment of all species for which the sequence of interest was available, using human as a base since this was the best annotated genome. Figure 3.2 shows a VISTA alignment produced using the AVID2.2 algorithm (which allows draft sequences) and visualised using standard conservation parameters of 75% homology over 100bp (Bray *et al.*, 2003).

Alignments of human and chimp sequences show a high degree of homology throughout the inputted sequence both upstream of Gli3 and within the first intron. Thus, comparison of these two primates is not informative to identify elements under selection. Interestingly though, some regions do not align to the human genome, such as a region around 12kb upstream of the human transcriptional start-site (indicated by a star in Fig. 3.2A). It is worth noting that a short region of high homology is preserved in the middle of the gap identified at 12kb in the human:chimp alignment. This region is equally conserved in other mammals studied but not in birds or amphibians (Fig. 3.2A).
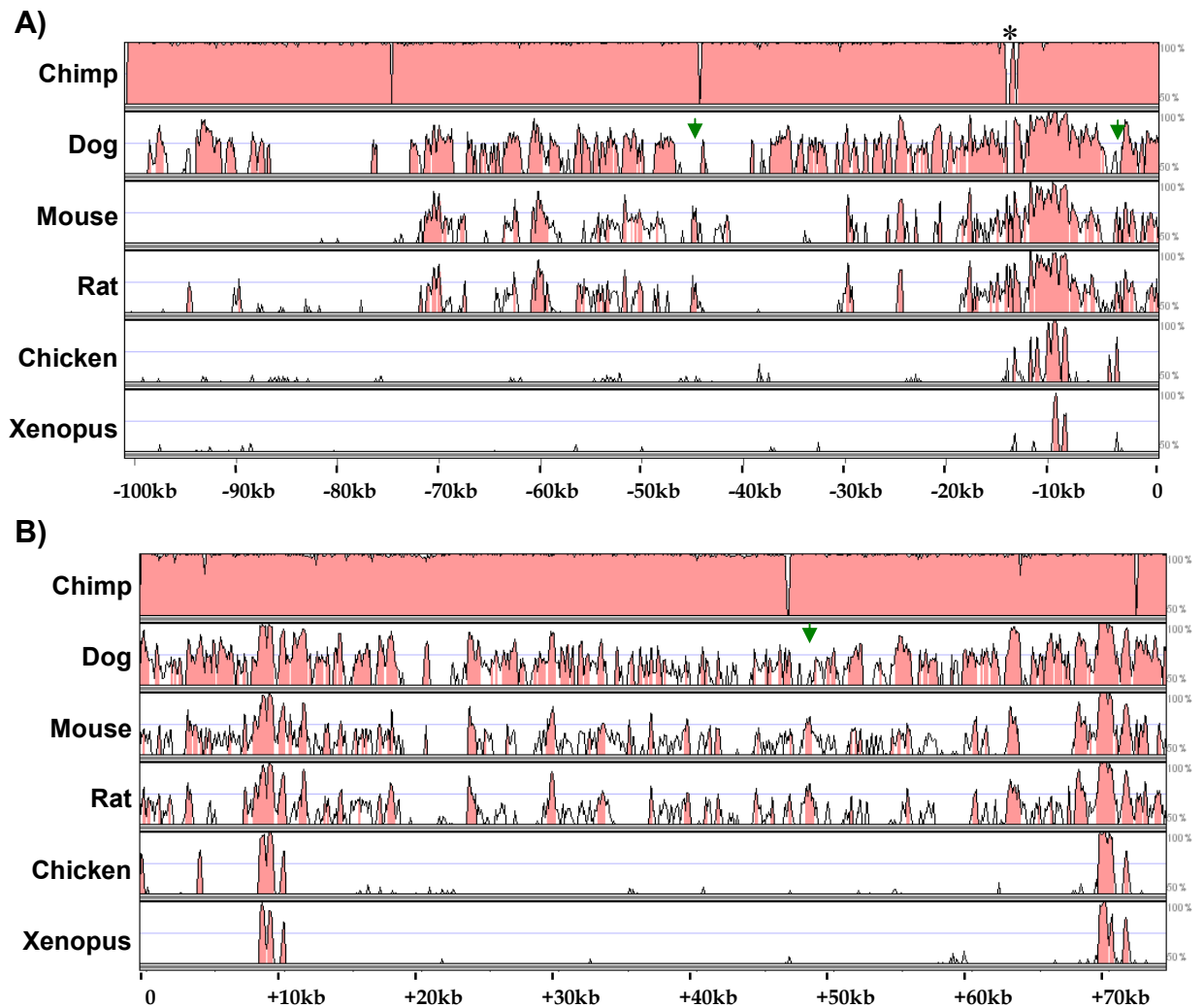
Figure 3.2: *VISTA genome alignment of the genomic region surrounding the first exon of Gli3 in various species.* A) 100kb directly upstream of the human first exon. B) 1st intron of Gli3 as annotated on the human sequence. Each row represents the alignment between the corresponding organism (labelled on the left) and the human base genome. The height of the peak shows the level of homology to the base genome (human). Where the level of homology is greater than 70% identity over 100bp the peak is shaded pink. The position of a peak on the x-axis represents the position of the corresponding alignment on the human genome relative to the beginning (A), or end (B) of exon 1. Green arrows are used to regions that appear conserved amongst mammals but are absent in the human:dog alignment. '*' indicates a conserved element within a gap region in the chimp genome.

Humans last shared a common ancestor with dogs 92 million years ago, and with rodents 91 million years ago (Fig. 3.1; Hedges, 2002). Although these evolutionary distances are similar, the human:dog alignment displays many more regions of homology than either human:mouse or human:rat alignments (Fig. 3.2). This reflects that rodent species have undergone a higher rate of neutral substitution since the divergence, partly due to a shorter generation time (Li *et al.*, 1996; Pennacchio, 2003; Thomas *et al.*, 2003). Indeed, most of the elements associated with peaks of conservation in both human:mouse, and human:rat alignments also appear in the human:dog alignment, with the exception of those marked by green arrows in Figure 3.2. Regulatory elements residing within these regions may account for aspects of Gli3 expression that are specifically altered in the canine lineage. A major

physiological difference between humans and rodents compared to dogs is observed in the digits, which are elongated in both rodents and humans with respect to canines. Since digit specification is affected in Gli3 mutants, it is interesting to hypothesise that regulatory elements absent in the canine lineage might influence Gli3 expression in the developing foot/limb bud. This could be readily tested by transgenic analysis.

In the human:bird alignment, the number of conserved regions drops significantly from those already discussed, dropping further still in the human:Xenopus alignment. This is expected from the greater evolutionary distance that separates the species compared, and illustrates the importance of including species distantly related in phylogenetic footprinting studies. Elements that are highly conserved during evolution may represent enhancers that control Gli3 expression in a highly conserved pattern, directing the basic outline of expression common to all vertebrates. Elements absent birds and amphibians may represent enhancers that control mammalian-specific domains of expression. Alternatively, rather than accounting for differences in expression, it is also possible that an element absent in a specific lineage has been replaced by another element located elsewhere.

3.2.3.2 Refining the search region

Because of variation in the genome size of organisms studied, it is unlikely that the arbitrary sequence length of 100kb upstream of the transcriptional start site will include similar proportions of regulatory sequence for each genome. All upstream elements conserved beyond mammals (i.e. in chicken and Xenopus) are located within around 15kb of human exon 1 (Fig. 3.2A). This may be due to poor alignment upstream of this anchor point that obscures the identification of conserved regions. In an attempt to improve the quality of the alignments I decided to look at a smaller locus. A region of 31kb was chosen, since this includes some peaks observed in mammalian alignments, and terminates before a region of low conservation between human and rodents (Fig. 3.2A). To identify the region of each genomic sequence that corresponded to -31kb in the human sequence, multiple alignments were carried out in MAVID (Bray and Pachter, 2004), successively removing upstream portions of the various sequences to optimise the resulting alignment. Pairwise alignments of human:chicken, and human:Xenopus were also carried out to identify the corresponding locus. The human sequence spanning 31kb upstream of exon 0 was found to correspond to approximately 29kb in rat, 33kb in mouse, 20kb in chicken and 16.7kb in Xenopus.

3.2.3.3 Adjustment of conservation parameters

A recent study of the Sox2 locus demonstrated that a conservation threshold of >60% identity over 100bp between chicken and mammals is a reliable indicator of DNA blocks

containing important regulatory regions (Uchikawa *et al.*, 2003). To determine whether this was a reasonable threshold to use in my study, the number of blocks displaying sequence identity over 100bp was assessed at various identity thresholds, both in the 31kb upstream of human Gli3, and in the first intron (Fig. 3.3). Below 60% identity the number of conserved regions detected increases significantly both in the 31kb upstream of Gli3, and in the first intron. Within the intronic region, all sequence blocks conserved at the 60% level were also observed at the 80% identity threshold, and only above this did the number of conserved elements begin to decline. A similar trend was observed in the 31Kb upstream of Gli3, with all elements identified at the 60% threshold containing blocks of >70% identity. However more elements were identified using the 65% identity threshold than by the 60% threshold, since increasing the specificity parameters split one of the regions into 2 discrete blocks separated by a sequence of lower sequence identity. The 60% sequence identity threshold between human and chicken identifies a total of 16 elements within the region of interest (Fig. 3.3, grey dashed lines; these regions correspond to those indicated by green blocks in Fig. 3.4). This is a manageable number of regions to investigate, and eliminates a large number of elements that are less well conserved. I chose to only investigate regions that are also conserved >65% amongst mammals, since genome-wide this threshold estimates the percentage of the genome shown to be under purifying selection (2002; Margulies *et al.*, 2003; Cooper *et al.*, 2004).
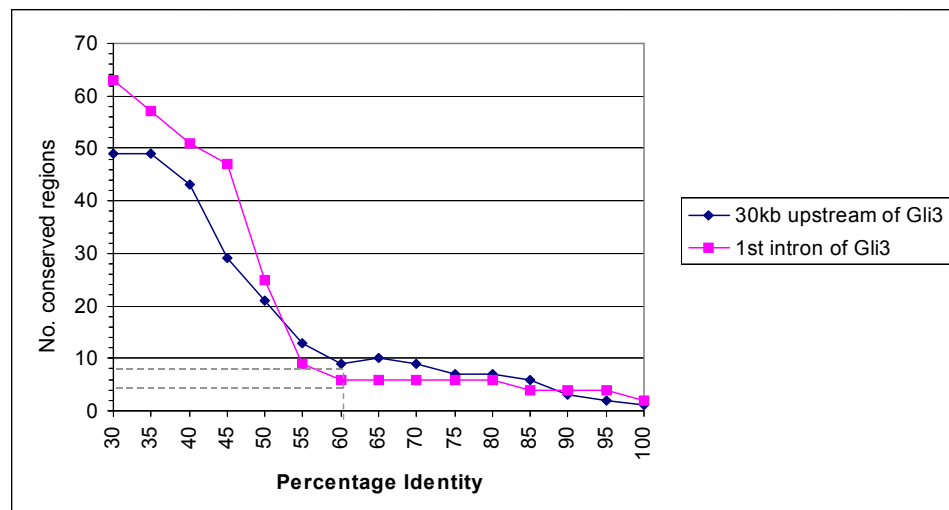
Figure 3.3: *Adjustment of Conservation Parameters.* The number of blocks conserved above various identity thresholds over 100bp are shown. Conservation parameters were adjusted between 30% and 100% identity over 100bp in AVID alignments of the 31kb upstream of Gli3, and of the 1st intron. The number of sequences meeting the threshold in each criteria set were counted in the VISTA browser.

3.2.4 Selection of regions of interest

The AVID algorithm relies on identifying short regions of exact sequence identity and uses these as anchors to align intervening sequences (Brudno *et al.*, 2003a). Although this method is well adapted to comparing highly similar sequences, more distantly related sequences

contain few regions of exact similarity that can be used as anchor points. The LAGAN algorithm detects local alignments using multiple short inexact words instead of longer exact words, and is better suited to the alignment of more divergent sequences (Brudno *et al.*, 2003a). LAGAN also introduces a maximum permitted distance between consecutive pairs of anchors, improving the alignment of intervening sequences. I decided to select my regions of interest based on MLAGAN alignments, which are the only truly multiple alignments available through VISTA (http://genome.lbl.gov/vista/mvista/instructions.html; Brudno *et al.,* 2003a). Comparison of MLAGAN and AVID alignments of the region targetted for investigation confirmed that the MLAGAN algorithm pruduces higher quality alignments. A VISTA alignment produced using the MLAGAN algorithm is shown in Figure 3.4, conserved regions chosen for investigation are illustrated alongside the alignment (filled boxes). All conserved regions identified by VISTA using the AVID algorithm were also identified using MLAGAN (Fig. 3.4, green boxes). However, using MLAGAN alignments VISTA identified an extra region of homology upstream of Gli3, and two within the first intron that were not identified even as weakly conserved using AVID alignments (Fig. 3.4, brown blocks). Additionally, the MLAGAN algorithm allowed successful alignment of the orthologous rodent sequences of Region 8, which were not apparent in AVID alignments.

As indicated in Figure 3.4, ten elements located upstream of Gli3, and eight regions located in the first intron (which I refer to with the prefix 'I') were chosen for investigation. Alignment characteristics of these regions, produced by VISTA, are shown in Table 3.1. With the exception of Region I3, all meet the selection criteria of >60% identity over 100bp between human and chicken sequences and >65% identity between human and rodent sequences in MLAGAN alignments (Table 3.1). Although the length of the aligned sequence between human and chicken Region I3 shown in Table 3.1 is only 98bp, this region displays an identity of 78.6%, and conservation extends further at a slightly lower level. Thus I chose to include Region I3 in my analysis. Eleven of the regions selected for analysis are conserved >60% between human and Xenopus (Fig. 3.4).
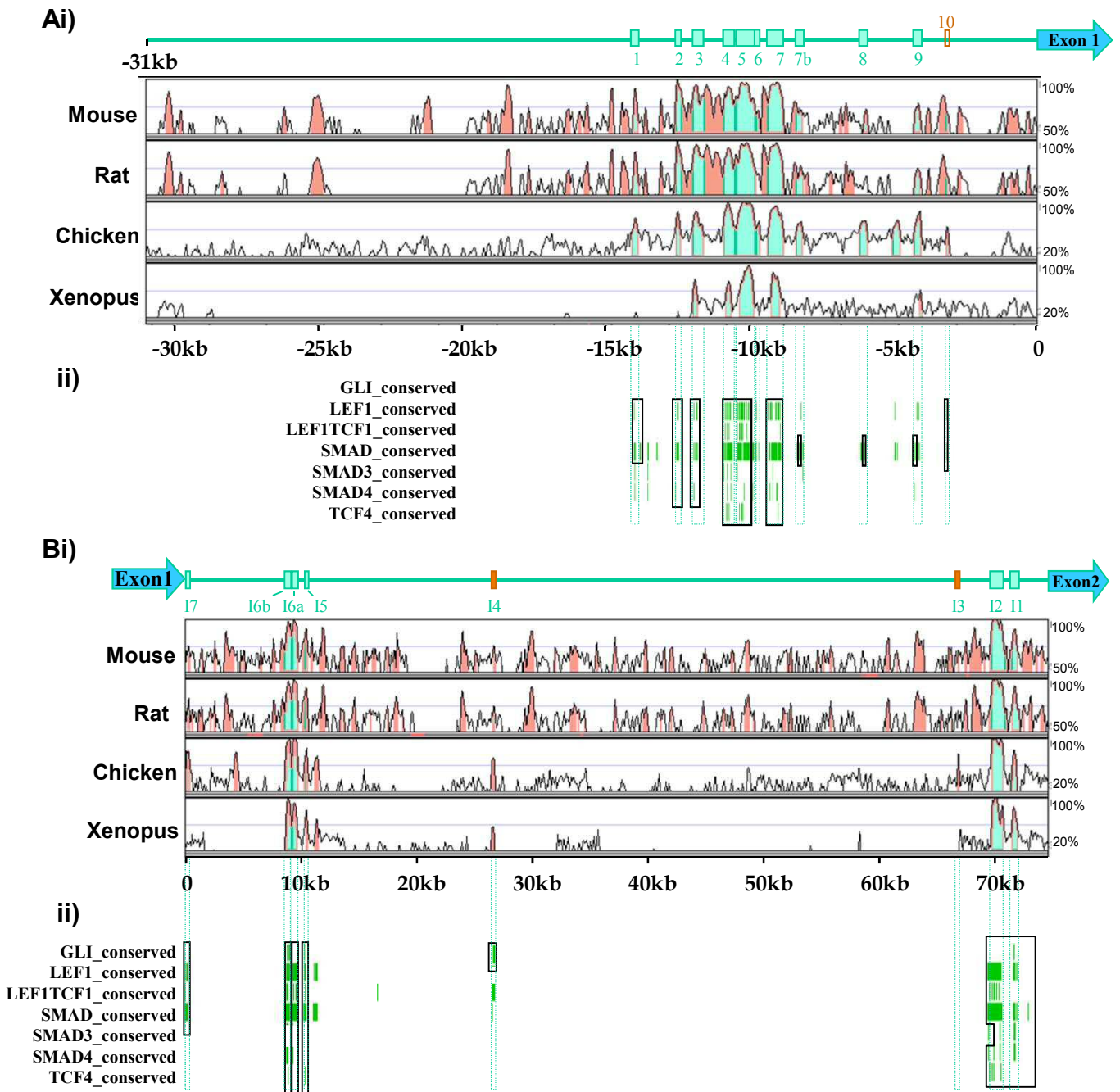
Figure 3.4. *Conserved elements surrounding exon 1 of human Gli3.* Chick, mouse and rat sequences orthologous to the 31kb immediately upstream of human Gli3 exon 1 (A), or the first intron (B) were extracted from Ensembl and aligned to the human sequence using the MLAGAN alignment algorithm on the VISTA server. i) VISTA representations of conserved regions. Peaks indicate the percentage sequence identity shared with the human genome at a given position. Peaks are shaded pink where conservation levels are greater than 65% identity over 100bp in mouse and rat alignments, or greater than 60% identity over 100bp in chicken and Xenopus alignments. Regions shaded in green meet the conservation criteria in alignments with mouse, rat and chicken sequences. Above the conservation profiles is a schematic representation of the Gli3 locus, indicating the position of conserved regions identified. Green text below the blocks indicates regions chosen for investigation. Brown blocks correspond to conserved regions that were not identified in alignments produced using the AVID algorithm. An open box is shown for Region 10, which could not be identified in the corresponding textual alignment. ii) Conserved binding sites for transcription factor effectors of signalling pathways known to affect Gli3 expression. Overlapping 20kb sequence blocks were scanned at 5kb intervals using rVISTA. Lef/Tcf, Smad and Gli consensus binding sites included in the TRANSFAC professional v9.4 database are shown. Vertical green lines indicate binding sites conserved between human and chicken genomes. Boxed binding sites are also conserved in rat and mouse genomes. Note that several well-conserved binding sites are found in the putative enhancer regions identified.

	Human:mouse conservation			Human:rat conservation		Human:chicken conservation	
Region	Position relative to human exon 1 (5'end)	Length (bp)	Percentage identity	Length (bp)	Percentage identity	Length (bp)	Percentage identity
1	-13678	83/232	75.9/76.7	315	74.6	138/93	68.8/62.4
2	-12296	370	83.2	313	87.5	132	83.3
3	-11514	535	82.4	535	84.7	406	68.5
4	-10494	562	81.5	561	80.6	455	79.8
5	-11913	813	68.5	819	86.1	720	87.4
6	-9627	237	68.4	284	65.8	244	61.5
7	-8816	709	89.6	713	88.6	542	79.9
7b	-8068	362	69	353	70.3	214	64.5
8	-5929	122	71.3	101	65.3	226	61.9
9	-4080	214	73.4	184	69	203	80.3
I1	+71395	718	76	721	74.8	433	71.1
I2	+69586	1222	92	1318	88.8	1090	93.1
I3	+66862	193	65.3	107	67.3	98	78.6
I4	+26674	278	66.5	286	67.8	271	67.5
I5	+11248	448	76.6	482	75.1	229	76.9
I6a	+8784	1139	86.5	1267	84.6	487	90.1
I6b	+8773					483	89.6
I7	+325	184	69	181	74	355	69

Table 3.1: *Characteristics of conserved regions selected for investigation.* Values were calculated by VISTA using MLAGAN (Brudno *et al.*, 2003a; Frazer *et al.*, 2004). For Region 1 '/' is used to distinguish between two conserved regions separated by a sequence that does not meet the identity threshold in human:mouse or human:chicken alignments.

Several discrepencies are observed between the visual VISTA output shown in Figure 3.4 and the conservation statistics shown in Table 3.1. First, Figure 3.4 indicates that Region 8 and Region I3 do not meet the selection criteria of >65% identity over 100bp in the human:rat alignment, however, VISTA calculations show that they have 65.3% and 67.3% homology respectively between human and rat sequences (Table 3.1). Second, Region 10 is detected in visual outputs, but is not recognised as conserved in the corresponding textual alignment, consequently conservation statistics were not produced. A well conserved sequence could not be observed by eye in the expected region in any of the alignments produced by VISTA, nor in a MAVID alignment of human, mouse, rat and chicken. These discrepancies demonstrate that the visual output generated by the VISTA tool does not always correlate precisely with textual alignments produced. The lack of sequence information precluded the investigation of Region 10.

Text alignments generated by MLAGAN through VISTA, together with the conservation statistics shown in Table 3.1 showed that Regions 4, 5 and 6 are separated into 3 distinct regions by lower regions of conservation, as are intronic Region I6a and Region I6b. Region 1 is separated into two regions in human:mouse and human:chick sequences, but not in human:rat sequences. I decided to study this region as a single conserved element.

3.2.5 Identification of regions of interest in a multi-species alignment

The boundaries of homologous regions in each species were annotated on a multiple pairwise alignment. The MAVID alignment tool was chosen because at the time of the study it was the only multiple alignment tool identified that allowed input sequences >20kb. With the exception of Region I4, all of the regions chosen for investigation align well across species. Region I4 has already been identified as one of the most poorly conserved regions (Table 3.1), and did not appear on initial alignments generated using the AVID algorithm. Since MAVID software uses this algorithm it is not surprising that the chicken sequence for Region I4 does not align.

3.2.6 Conservation of transcription factor binding sites

3.2.6.1 Conserved transcription factor binding sites cluster in conserved regions

Regulatory VISTA (rVISTA) allows one to search pre-computed VISTA alignments for a selection of transcription factor binding sites available in TRANSFAC Professional 9.2, but limits the size of each inputted sequence to 20kb. All conserved intergenic regions identified in my alignments occur within 20kb of the transcriptional start site of human Gli3. This region was searched using rVISTA for the presence of vertebrate specific Tcf, Smad and Gli binding sites conserved between human:mouse, human:rat and human:chicken, using standard core similarity values of 0.75 and matrix similarity values of 0.70. The intronic region was investigated using overlapping 20kb sequence blocks. Binding sites displayed in Figure 3.4B are those conserved at >80% over a 24 bp window between human and chicken. A single green line represents one binding site. Boxed regions contain binding sites that were also conserved in human:mouse, and human:rat alignments. Transcription factors identified within each region are recorded in Table 3.2. Searches using different consensus sequences selected from the Transfac v7.4 database showed similar results.

No conserved Gli binding sites were identified upstream of Gli3 using rVISTA, but a number were identified in the first intron. (Fig. 3.4). Multiple Smad and Lef/Tcf sites were identified in both aligned regions. These occur individually as well as in clusters, and are often conserved between human, mouse, rat and chicken sequences (Table 3.2). They are often found overlapping, and run in both directions on the DNA.

The conservation of binding sites is much more likely in functional elements than in non-functional conserved elements, and clustering of transcription factor binding sites within conserved non-coding elements significantly enhances the likelihood that such an element will act as an enhancer (Berman *et al.*, 2004a). Therefore, these clustered binding sites can be seen as an indication that the regions in which they occur are likely to contain elements

with enhancer activity. Using this logic, it seems unlikely that Region I3, which contains no conserved binding sites for the transcription factors investigated, has enhancer activity. Very few conserved transcription factor binding sites were identified outside of the putative enhancers, suggesting that the parameters used successfully identify regions with regulatory potential. Alternatively, the presence of conserved transcription factor binding sites may be a consequence of sequence conservation.

With the exception of Region I3, all of the regions chosen for investigation contain one or more conserved Smad binding site (Table 3.2). This supports the possibility that BMP signalling may be acting directly via the Smad proteins to regulate Gli3 expression. However, since the Transfac database used by VISTA (v9.4) is not accessible to the public, it is not clear whether the site labelled Smad_consensus is that of factors associated with BMP signalling. Binding sites for Smad4, which transduces BMP signalling are only conserved across all species investigated in Region 2, 3, I1 and I2. These are the prime candidates in my study for direct regulation by BMP signalling.

Sixteen of the eighteen regions chosen for investigation contain one or more Lef/Tcf binding site conserved between humans and birds. The absence of any conserved Lef/Tcf binding sites in Region 8 or I3, and the lack of conserved Lef/Tcf binding sites in Region 7b, 9 and I4, suggests that if Wnt signalling acts via these elements to regulate gene expression, it does so indirectly. Region I4 is of particular interest because the only transcription factor binding site conserved between all four species at this locus is that of a Gli protein. This implicates Region I4 as a putative enhancer element regulated directly by Shh signalling. Mutation of this Gli binding site in a reporter construct used in transgenic analysis could be used to establish if the Gli binding site in Region 4 is functional. An in vitro approach would be to compare binding to wild-type and mutated binding sites by EMSA analysis.

Name	Lef/tcf binding sites			Smad binding sites			Gli binding sites		
	Present in human:chicken alignments	Conserved amongst all species investigated	Clustered	Present in human:chicken alignments	Conserved amongst all species investigated	Clustered	Present in human:chicken alignments)	Conserved amongst all species investigated	Clustered
Region 1	✓	✓	x	✓	✓	✓	x	x	x
Region 2	✓	✓	✓	✓	✓	✓	x	x	x
Region 3	✓	✓	✓	✓	✓	✓	x	x	x
Region 4	✓	✓	✓	✓	✓	✓	x	x	x
Region 5	✓	✓	✓	✓	✓	✓	x	x	x
Region 6	✓	✓	✓	✓	✓	✓	x	x	x
Region 7	✓	✓	✓	✓	✓	✓	x	x	x
Region 7b	✓	x	x	✓	✓	✓*	x	x	x
Region 8	x	x	x	✓	✓	✓*	x	x	x
Region 9	✓	x	x	✓	✓	✓*	x	x	x
Region I1	✓	✓	✓	✓	✓	✓	✓	✓	x
Region I2	✓	✓	✓	✓	✓	✓	x	x	x
Region I3	x	x	x	x	x	x	x	x	x
Region I4	✓	x	x	✓	x	x	✓	✓	x
Region I5	✓	✓	✓	✓	✓	✓	✓	✓	x
Region I6a	✓	✓	✓	✓	✓	✓	x	x	x
Region I6b	✓	✓	✓	✓	✓	✓	✓	✓	✓
Region I7	✓	✓	✓	✓	✓	✓	x	x	x

Table 3.2: *Lef/Tcf, Smad and Gli binding sites in highly conserved regions of the Gli3 locus.* Binding sites indicated are those displayed in Figure 3.4.
'*', only 1 of the clustered binding sites is conserved amongst mammals.

3.2.6.2 Regulatory pathways associated with Gli3 expression might act indirectly

Combinatorial control involving 4-8 diverse inputs is a common mechanism used by developmental cis-regulatory modules (Davidson *et al.*, 2003). The identification of transcription factor binding sites associated with each of the pathways implicated in Gli3 regulation suggests that direct control of Gli3 expression is possible. However, we cannot rule out an indirect mechanism of action of Wnt, BMP and Shh on Gli3 expression. Pathways known to act as master regulators during development provide good candidates for Gli3 regulation. In addition to Wnt, Shh and BMP signaling, these include the Retinoic acid pathway, the Notch pathway, FGF pathway and Hox proteins. Additionally, BMP signaling has been shown to act indirectly via Id family proteins, by sequestering bHLH containing proteins (Nakashima *et al.*, 2001; Goumans *et al.*, 2002). bHLH domain containing proteins are numerous, including transcription factors involved in myogenesis (MyoD, myogenin and their partner E47) neurogenesis (NeuroD, neurogenin) and haemopoiesis (SCL) (Kewley *et al.*, 2004). Investigation of transcription factor binding sites associated with indirect pathways at this stage would be laborious. I decided not to pursue transcription factor binding site analysis at this point, and await for functional studies before carrying out such analyses.

3.3 Discussion

Cis-acting elements that regulate gene expression are conserved in the non-coding fraction of the genome, and have been shown to cluster around transcription factors and genes involved in early development. They occur up to 1Mb upstream or downstream of the genes they control, as well as in intronic regions. Using phylogenetic footprinting I have searched 175kb surrounding the 1st exon of human Gli3, and uncovered multiple blocks of non-coding sequence that are conserved between mammals and birds. Several of these elements are also conserved in amphibians. Clustering of transcription factors previously associated with Gli3 regulation within the conserved regions supports the hypothesis that they are enhancer elements.

3.3.1 Location of regions of interest

Conserved non-coding elements have been shown to cluster close to the gene they regulate and become more sparse with increasing distance from the genomic locus (Lemos *et al.*, 2004; Uchikawa *et al.*, 2004; Woolfe *et al.*, 2005). I observed a similar trend upstream of Gli3, where conserved elements cluster within 15kb of the transcriptional start site, and are not identified further upstream. Similarly, within the intronic region studied conserved elements were found to cluster close to coding regions, and become more sparse towards the

centre of the intron (Fig. 3.4, Table 3.1). The location of conserved regions proximal to the transcriptional start site of human Gli3 correlates with a recent study by Sandelin *et al.* (2004). The study found that C2H2 zinc finger domain proteins (such as Gli3) display an over-representation of conserved non-coding elements extending up to 150kb away, relative to other transcription factors investigated, suggesting a conserved mechanism for transcriptional regulation of these proteins (Sandelin *et al.*, 2004b). The clusters of conserved non-coding elements observed for Gli3 and other zinc-finger proteins may function by creating a favourable environment for transcription, for example by modification of the chromatin structure. Alternatively they might function by recruiting factors required for transcription.

3.3.1.1   Intragenic conserved regions cluster close to exon boundaries

It is interesting to note that the genomic structure of Gli3 has been conserved, including a long first intron. In human and rat, the first intron is ~ 75kb long, ~70kb in mouse, ~50kb in chicken and ~34kb in Xenopus. However, all conserved regions in intron 1 are located within 11Kb of an exon boundary in the human sequence, with the exception of Region I4. The intervening region that lacks homology may be required as a spacer region. In contact models of enhancer function, looping out of intervening DNA has been proposed as a mechanism of bringing together an enhancer element and its associated promoter (Hatzis and Talianidis, 2002). Alternatively, two enhancer elements might be brought together in a similar manner to form a complex able to influence transcription. RNA tagging and recovery of associated proteins (RNA TRAP) is a technique that has been developed to test whether enhancer elements are brought into contact with their promoters by a looping mechanism (Carter *et al.*, 2002). Biotin tyramide is used to label molecules within the vicinity of a pre-mRNA of interest. The amount of labelling is proportional to the physical proximity of an element to the transcript, such that enhancer elements that have been in contact with a proximal promoter will contain higher levels of labelling than surrounding sequences. This technique could be applied to the Gli3 locus to identify enhancer elements that act by a looping mechanism.

The presence of conserved elements in Gli3 intronic regions might also be unrelated to transcriptional control. For instance, clustering of conserved elements around exons may be associated with the regulation of alternative splicing. A possible mechanism for this is the formation of secondary structures that modulate the transcriptional machinery (Sorek and Ast, 2003). Alternative splicing has been proposed as a mechanism for post-transcriptional regulation of Gli2, by the inclusion of a novel non-coding exon (Speek *et al.*, 2006). Indeed, analysis of avian Gli3 identified an alternative exon expressed in stage 12 embryos that introduces stop codons to the N-terminal domain of the protein, and an in-frame methionine

was identified that that could generate a truncated protein without the N-terminal repressor domain (Borycki *et al.*, 2000). To determine whether alternative splicing is a mechanism used to modulate Gli3, I have used 5'RACE to investigate Gli3 isoforms (see Chapter 4).

3.3.1.2 Identification of more distal regulatory elements

Inclusion of a more distant species in my alignment was hampered by a lack of sequence information, which also made long range searches for regulatory elements difficult. A development to the VISTA browser since I initiated my study is the addition of a number of whole genome alignments (Genome VISTA; Couronne *et al.*, 2003). To investigate whether any conserved regions upstream of Gli3 have been missed in my study, I used Genome VISTA to examine up to 1Mb upstream of the human transcriptional start site. In this region, fifteen new elements were identified with >60% identity over 100bp between human and chicken alignments, that were also present in human:mouse comparisons (Fig. 3.5). Beyond 100kb upstream of Gli3 these conserved regions do not cluster and are distributed evenly up to 700kb away from the gene (located at -118, -168, -230, -232, -255, -281, -293, -362, -403, -522, -542, -571, -603 and -695 kb relative to the transcriptional start site in humans; Fig. 3.5). An additional element located at -91kb was also identified by genome VISTA that was not detected in my initial alignment. Presumably this element was missed because of the alignment method used. Genome VISTA produces SLAGAN alignments of the whole genome, whereas my initial alignments of the 100kb upstream of Gli3 used AVID alignments. This demonstrates the importance of comparing multiple alignment algorithms.

Since enhancer elements have been identified that regulate the expression of a gene located >1Mb away it is difficult to assign a limit as to where non-coding regulatory elements might lie. A recent study estimates that around half of conserved non-coding elements occur more than 250kb away from the genes they control, indicating that the distal elements identified by Genome VISTA could be involved in the transcriptional control of Gli3 (Vavouri *et al.*, 2006b). However the same study estimates that around 30% of intergenic conserved elements occur within 100kb of the genes they control, suggesting that the regions I have chosen to investigate should contain a significant proportion of conserved regulatory elements associated with Gli3 (Vavouri *et al.*, 2006b). In humans the Gli3 locus is flanked by PSMA2 700kb upstream, and INHBA 270kb downstream of the coding sequence. It is conceivable that some of the more distal conserved regions identified upstream of Gli3 might be associated with the regulation of PSMA2.
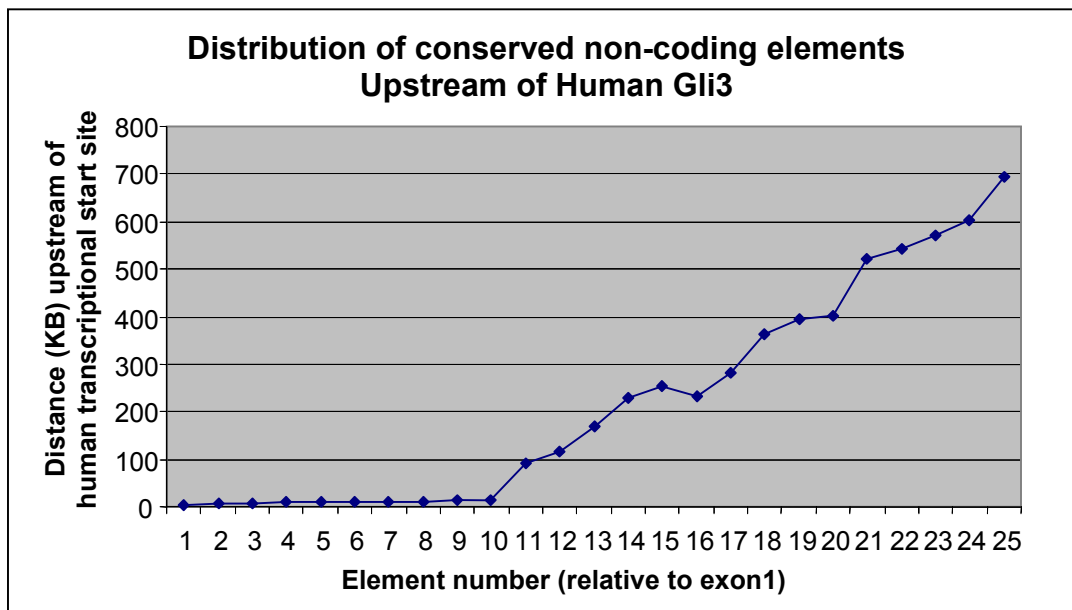
Figure 3.5: *Conserved elements upstream of human Gli3 cluster close to the 1ˢᵗ exon.* Elements displaying >60% over 100bp between human and chicken sequences are numbered according to their position upstream of Gli3 (1 being closest to exon 1). The distance upstream of the 5' end of human exon 1 is plotted.

To determine whether these new regions identified are sufficient to drive Gli3 expression a BAC reporter construct incorporating the genomic locus covered by the conserved regions identified (between -15kb and exon 2) could be generated and tested in transient transgenesis. Reproducible reporter gene expression in a similar spatial and temporal pattern to that observed for endogenous Gli3 would be indicative that the BAC contains all necessary enhancer elements required to regulate Gli3 expression. In the event that aspects of Gli3 expression pattern are missing, similar constructs spanning different genomic regions, containing conserved regions identified further upstream of the Gli3 locus, could be used to establish where additional elements might lie.

3.3.2 Possible functions of the conserved elements identified

The non-coding conserved regions identified in this study have been selected for their potential to regulate Gli3 expression by acting as an enhancer element. However, conserved non-coding DNA might also have a role beyond gene regulation, such as in the regulation of chromosome pairing or condensation, replication, or higher order chromatin structure (Cremer and Cremer, 2001; Pennacchio, 2003).

Matrix attachment regions (MARs), which mediate the attachment of chromatin loops to the nuclear matrix or scaffold, reside within the non-coding fraction of the genome and have been attributed insulator function (Laemmli *et al.*, 1992). Up to 11% of non-coding conserved elements have been proposed to contain MARs, making it highly likely that they are present in some of the regions chosen for investigation (Glazko *et al.*, 2003). However, the limited

sequence conservation of MARs, and a lack of understanding of their evolution makes identification difficult with current tools (Glazko *et al.*, 2003).

Some of the non-coding regions identified might represent previously uncharacterised non-coding RNAs such as antisense regulatory RNAs, small nucleolar RNAs (SnoRNAs) and microRNA sequences (Eddy, 2002; Margulies *et al.*, 2003; Cooper *et al.*, 2004; Woolfe *et al.*, 2005). SnoRNAs are often found within the introns of protein coding genes, and the snoRNA U1 has been identified as a core splicing component (O'Gorman *et al.*, 2006). Additionally, the introns of ion channel genes have been reported to contain RNA-editing targets, and are enriched for conserved sequences in the chicken genome (Chicken Genome Sequencing Consortium, 2004). This suggests that non-coding RNAs might be particularly associated with intronic regions of conservation. Some estimates suggest that non-coding RNAs may be as common as protein coding genes (Lim *et al.*, 2003; Cawley *et al.*, 2004; Chicken Genome Sequencing Consortium, 2004). Since the Gli3 core-promoter has not been identified, it remains possible that one of the conserved non-coding elements identified in my study represents the endogenous promoter of Gli3.

3.3.3 Transcription factor binding site conservation

The conserved regions identified contain numerous potential transcription factor binding sites (Fig. 3.4B, Table 3.2). These appear preferentially located within regions of high conservation, which is to be expected since the conservation of a particular binding site requires conservation of the underlying alignment. However the clustering of binding sites within conserved regions might indicate that the sequence has been maintained by purifying selection. Wasserman *et al.* (2000) estimated that sequence-specific regulatory sites are more than 320 times more likely to occur within conserved regions.

The number of sites identified appears to represent the variation tolerated in consensus sequence. The Gli consensus (GACCACCCA) allows less variation than Smad or Lef/Tcf consensus sequences. Consistent with this, conserved Gli binding sites were only identified in four of the conserved regions identified, and only cluster in two of the conserved regions, as oppose to Smad and Lef/Tcf binding sites which were each clustered in most of the conserved regions identified (Table 3.2, Fig. 3.4). Alternatively the low number of Gli binding sites identified in my search may indicate that Shh signalling acts indirectly on Gli3 expression. Consistent with this possibility, Gli3 was not identified in a genome-wide screen for genes whose enhancer elements contain Gli binding sites (Hallikas *et al.*, 2006), nor in a screen for direct target genes of Shh expression in zebrafish embryos (Bergeron *et al.*, 2008). To my knowledge, no such screens have been performed to identify targets of BMP or Wnt signalling.

Initial searches for transcription factor binding sites suggest that Wnt and BMP pathways may act directly on a large proportion of the regions identified. Indeed, binding sites associated with these two pathways cluster together in many of the conserved regions identified (Table 3.2). The significance of the co-clustering of Tcf and Smad binding sites is unclear. However, adjacent Tcf and Smad consensus binding sequences are required for transcription from the Myc and Msx2 promoters, and have also been identifed in an enhancer element of the Emx2 gene (Theil *et al.*, 2002; Yagi *et al.*, 2002; Hu and Rosenblum, 2005). These data suggest that BMP and Wnt pathways interact directly at the transcriptional level. Indeed, Smad3/Lef1 complexes have been reported (Hussein *et al.,* 2003, Labbe *et al.*, 2000). The clustering of Lef/Tcf binding sites together with smad binding sites in a number of the conserved regions identified in this study suggests that Wnt and BMP pathways might act synergistically in the regulation of Gli3 expression.

3.3.4   Other studies of Gli3 transcriptional control

Since I initiated my study, a number of articles have been published that investigated regulatory elements surrounding the Gli3 locus using phylogenetic footprinting (Abbasi *et al.*, 2007; Paparidis *et al.*, 2007; Alvarez-Medina *et al.*, 2008). Comparisons of the bioinformatic tools and criteria used in these studies, together with the genomes chosen, provide interesting lessons and have confirmed most of my predictions.

Abbasi *et al.* (2007) report the investigation of 1Mb DNA surrounding Gli3 in Human, Chimpanzee, Mouse, Rat and Fugu, using a percentage identity cut-off of 50% over 60bp between human and fugu alignments. These parameters identified 11 conserved non-coding elements, all located within intragenic regions of mouse Gli3 (taking mouse exon 0 as the transcriptional start site and thus including the elements reported as upstream in my study, Fig. 3.5). Of these 11 conserved elements, only three enhancers (CNE 12, 1 and 2) are comprised in the genomic region I studied. The remaining eight conserved elements are located downstream of exon 2. The regions identified by Abbasi *et al.* correlate with some of the most highly conserved regions identified in my investigation (compare Fig. 3.6 with Table 3.1 and Table 3.3): CNE 12 corresponds to Region 5, CNE 11 to Regions I6a andI6b, and CNE2 to Region I2. However, other regions that I found to be well conserved were not identified, such as Region 2. Interestingly, none of the conserved regions identified by Abbasi *et al.* have been reported to drive reporter gene expression in the somites or neural tube of zebrafish embryos transfected with reporter constructs, which are both well characterised sites of Gli3 expression both in tetrapods and in fish (Hui *et al.*, 1994; Lee *et al.*, 1997; Borycki *et al.*, 1998; Borycki *et al.*, 2000; Schweitzer *et al.*, 2000; Tyurina *et al.*, 2005). Elements encoded within the regions

identified in my study that have not been investigated elsewhere are good candidates for mediating Gli3 regulation at these sites. The fact that these regions were missed in other investigations suggests that the criteria used may have been too stringent to identify enhancer elements associated with Gli3 expression patterns that are perhaps amniote specific. For instance, it is likely that enhancers controlling Gli3 expression in the paraxial mesoderm may have diverged between amniotes and amphibians/fish, as Gli3 is expressed in co-ordination with somite formation in the former (Borycki *et al.*, 1998), whereas it is expressed in the unsegmented mesoderm in the latter (Marine *et al.*, 1997). Consistent with this observation, a recent study by Thomas *et al.* concluded that comparisons between mammal and fish alignments are not sufficient to identify most non-coding enhancers conserved amongst amniotes (Thomas *et al.*, 2003; Plessy *et al.*, 2005).

The most highly conserved of the enhancer elements identified by Abbasi *et al.* (CNE2) corresponds to Region I2 identified in my study, and shows a high degree of sequence identity between human and Xenopus (Abbasi *et al.*, 2007; Fig. 3.4, 3.6). Its level of conservation, together with the clustering of multiple Smad and Tcf binding sites (Fig. 3.4) suggest that this region is likely to function as a Gli3 enhancer. Indeed, a more recent study shows that this element drives reporter gene expression in the midbrain, hindbrain and forebrain of transiently transfected zebrafish embryos, as well as in the developing forebrain and first branchial arch of transgenic mice (Paparidis *et al.*, 2007). Thus it seems that Abbasi and colleagues have identified some functionally conserved enhancer elements. However, they did not identify enhancer elements responsible for driving reporter gene expression in several well characterised domains of Gli3 expression. Indeed, the elements investigated drove reporter gene expression in ectopic sites not normally associated with Gli3 expression, such as the notochord. One possibility is that by using such high levels of conservation Abbasi *et al.* preferentially identified elements with other functions than enhancers. The variable expression reported for a number of the conserved regions might indicate that they are acting as anti-silencers without directly affecting levels of transcription (Cooper and Sidow, 2003). Alternatively, they may display an activity that would normally be co-ordinated by the presence of other regulatory elements, such that regulation is disrupted when the element is taken in isolation. For example, the activity of elements that drive reporter gene expression in sites not normally associated with Gli3 expression might normally be suppressed by repressive elements.

Alvarez-Medina *et al.*, (2008) investigated four of the regions similar to those identified by Abbasi *et al.*. Three of these regions (HCNR1-3) correspond to my Regions 4-7; I6a, I6b and I5; and I1-2 respectively, while HCNR4 corresponds to CNE3, not identified in my study

because it lies downstream of exon 2 (Fig. 3.6). All contain Tcf binding sites, suggesting that they may be responsive to Wnt mediated induction of Gli3 expression. Consistent with this possibility, dominant negative Tcf and constitutively active β-catenin constructs co-electroporated with reporter constructs affected reporter gene expression in the chick neural tube. The responsiveness of HCNR1, HCNR2 and HCNR3 suggests that Tcf binding sites present in each region may directly mediate the control of Gli3 by Wnt signalling. Mutational analysis of these binding sites is needed to establish whether Wnt signalling controls enhancer activity directly, and to identify the Lef/Tcf binding sites involved in this control. In addition, my own analysis has shown that the regions investigated by Abbasi *et al.* (2007) and Alverez-Medina *et al.* (2008) contain multiple Smad binding sites, and HCNR2 (CNE 1) and HCNR3 (CNE 2) contain conserved Gli binding sites, these may also influence reporter gene expression patterns.

The conservation parameters used in my study have allowed me to identify several novel conserved regions of non-coding DNA, including nine elements not investigated previously. These include Regions 1-3, 7b, 8, 9, I3, I4, and I7. These elements contain clusters of transcription factor binding sites suggesting they have enhancer function. Transgenic analysis using in-ovo electroporation will establish which regions are capable of driving reporter gene expression, and may identify novel enhancer elements critical for the regulation of Gli3 gene expression amongst amniotes.
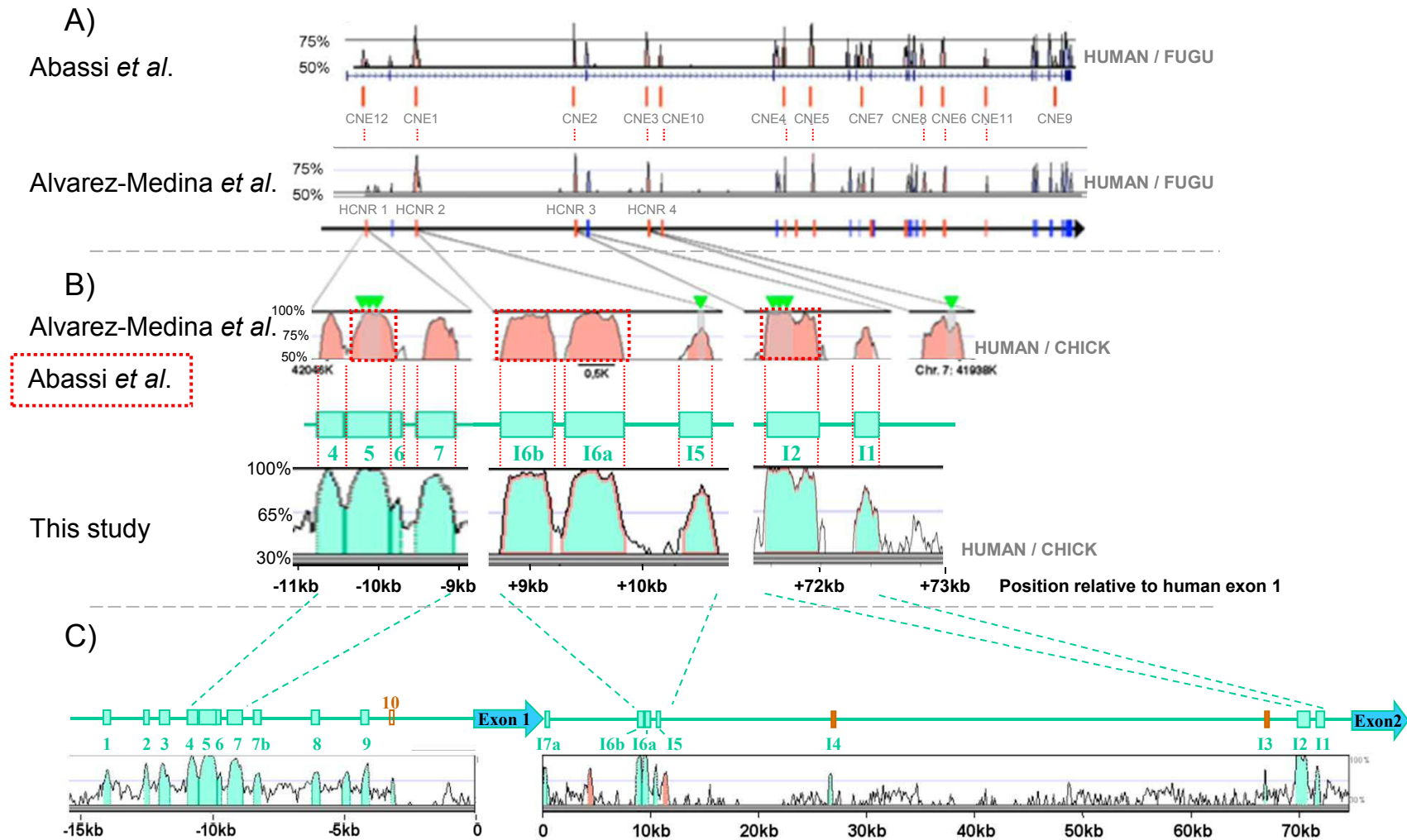
Figure 3.6, Comparison of the conserved regions identified in this study to those identified in other studies of Gli3 regulation. A) Human/fugu conservation plots produced by (Abassi *et al.,* 2007) and by (Alverez-Medina *et al.,* 2008) identified similar conserved regions. CNE 12, CNE1, CNE2, and CNE3 of Abbasi *et al.* correspond to regions HCNR 1- 4 identified by Alvarez-Medina *et al.*, respectively. B) Magnified views of human/chick conservation plots of HCNR 1- 4 identified by Alvarez-Medina *et al.* are aligned to the corresponding regions identified in my study. HCNR1 correlates with Regions 4-7, HCNR2 corresponds to Regions I6a, I6b and I5, and HCNR3 corresponds to Regions I1 and I2. Red dotted lines surround areas of HCNRs that correspond to the elements identified by Abassi *et al.*. Generally, shorter regions of homology were studied in the latter. C) Regions aligning with those identified in other studies are shown in the context of the locus investigated in this study. Human/chick conservation plots are shown using the same parameters as those used in Figure 3.4.

Chapter 4

Promoter search and 5'RACE

4.1 Introduction

At the time of this study, the promoter region of Gli3 has not been characterised. Identification of the promoter region would help identify the true 5' end of the transcript, and thus would indicate where enhancer elements might lie. It is also preferable to use the endogenous promoter when studying a gene by reporter gene expression, since it will give a more accurate read out of the expression profile. In this chapter, I attempt to identify the endogenous promoter and 5'limit of the Gli3 transcript. In-silico tools are used in an attempt to isolate the endogenous promoter, followed by characterisation the 5' end of the Gli3 transcript by EST analysis and 5'RACE.

4.1.1   In-silico identification of promoter elements

Various computational tools have been developed that use sequence information to predict the location of core-promoter elements. However the variation in core promoter composition has hindered their identification, and a combinatorial regulatory code for promoter activity remains elusive. The various in-silico programmes incorporate different characteristics associated with promoter activity, including the identification of core-motifs, their spacing, the composition of underlying DNA and the presence of transcription factor binding sites. The success of such programmes is measured by their accuracy in predicting promoter elements surrounding characterised TSSs, although experimental validation is necessary.

At the time of this study it was estimated that independently none of the existing programmes were able to accurately predict the location of functional promoter elements, without also producing a large number false predictions (Bajic *et al.*, 2004). However, by combining results generated using different tools, it was demonstrated that current tools can be used to accurately predict around three quarters of known core-promoters (Bajic *et al.*, 2004). I decided to use a similar approach in searching for promoter elements associated with Gli3 activity. The programmes identified for prediction of promoter elements are shown in Table 4.1. They differ in the statistical methods used and in promoter characteristics recognised.

Program	Statistical method and characteristics analysed	Restrictions	Reference
CpGProD (CpG island Promoter Detection	CpG island, RVM		(Ponger and Mouchiroud, 2002)
McPromoter	ANN, mechanical properties, sites scored by an optimal set of pwm's	Maximum input sequence 20kb	(Ohler *et al.*, 2000)
NNPP2.2	ANN, TATA-box motif, core-promoter elements		(Reese, 2001)
Promoter2.0	ANN, TATA-box motif	Copy available upon request from author	(Knudsen, 1999)
TSSG/TSSW	ANN, TATA-box motif, C+G content, transcription factor binding sites		(Solovyev and Salamov, 1997)
Promoter scan	ANN, TATA-box motif, C+G content		(Prestridge, 1995)
Fprom	LDF, uses functional motifs and oligonucleotide composition		(Solovyev and Salamov, 1997)
Core promoter	TATA-box motif	Maximum sequence length 2kb	(Zhang, 1998)
Promoter inspector	http://www.genomatix.de/products/ElDorado/	Subscription to the ElDorado package of genomatix required, trial version available with limited output	(Scherf *et al.*, 2000)
Cister	TATA-box motifs, other core-promoter elements, transcription factor motifs present in the transfac database		(Frith *et al.*, 2001)
ARTS	ANN, mechanical properties CpG islands not distinguished	Only became available 2006	(Sonnenburg *et al.*, 2006)
Coreboost	TATA, Inr, and CCAAT- box motifs, GC content, and mechanical properties considered, RVM	Only became available 2007	(Zhao *et al.*, 2007)
DragonGSF	ANN, G+C content, CpG island	Not publicly available, uses binding site information from the TRANSFAC database	(Bajic and Seah, 2003)
DragonPF	ANN, G+C content	Not publicly available, uses binding site information from the TRANSFAC database	(Bajic *et al.*, 2002)
Eponine	G+C content, TATA-box motif, RVM	Maximum input sequence 1,024,000nt	(Down and Hubbard, 2002)
FirstEF	G+C content, CpG island, QDA		(Davuluri *et al.*, 2001)

Table 4.1: *Programmes available for the prediction of core-promoter elements*. Adapted from (Bajic *et al.*, 2004), incorporating a number of programmes developed since this publication. Programmes listed in black in the top part of the table have been used in this thesis; those shown in red have not.

4.2 Results

'Promoter inspector', available through Genomatix (Table 4.1), was used to search for promoter elements surrounding Gli3 in human, mouse, rat and chicken genomes. A promoter region was identified in each species that appeared to correspond to the position of mouse exon 0. However, analysis of the sequence surrounding the promoter predicted for each species did not show any identifiable similarities. Interestingly, all predicted promoter elements were located within the previously reported CpG island (Vortkamp *et al.*, 1994). This indicates that the CpG island is likely to contain the core promoter of Gli3.

4.2.1 CpG Island Characterisation

The limits of the CpG islands present in each species were investigated. The UCSC browser predicts CpG islands using the CpG programme developed by Gos Micklem (Gardiner-Garden and Frommer, 1987). Statistics of the CpG islands present in each species are shown in Table 4.2. The position of the CpG island appears to be conserved between species, and a high level of conservation is observed. The chicken sequence is not fully characterised in this region. Mouse exon 0 is located within the CpG island, approximately 400bp upstream of the 3' end.

Species	Human	Mouse	Rat	Chick
UCSSC genome release	May 2004	March 2005	Nov 2004	May 2006
Genomic size (bp)	1847	1665	1456	1036
CpG count	154	161	151	133
C count plus G count	1012	1052	954	708
Percentage CpG	16.7	19.3	20.7	25.7
Percentage C or G	54.8	63.2	65.5	68.3
Ratio of observed to expected CpG	1.14	0.98	0.97	1.1

Table 4.2: Sequence composition statistics for the CpG upstream of Gli3. Statistics listed are for the CpG islands identified by the UCSC browser (http://genome.ucsc.edu). The CpG count is the number of CG dinucleotides in the island. The Percentage CpG is the ratio of CpG nucleotide bases (twice the CpG count) to the length. The ratio of observed to expected CpG is calculated according to the formula cited in (Gardiner-Garden and Frommer, 1987).

CpGProd was used to predict the likelihood of promoter elements occurring within each CpG island. 100kb upstream of exon 1 in each species were inputted to the CpGProd server, results are shown in Table 4.3. This predicts that there is an approximately 70% chance of the transcriptional start site occurring within the CpG island.

Sequence name	Number	Begin	End	Length (bp)	G+C frequency	CpG o/e ratio	Start-p	AT skew	GC skew	Strand (strand-p*)
Chick	1/1	86462	88207	1596	0.6378	0.9613	0.7103	-0.1349	0.0177	+ (0.8523)
Human	1/2	84652	87469	2818	0.6498	0.8612	0.7660	0.0740	0.1109	+ (0.6691
Human	2/2	94658	95720	1063	0.5673	0.7485	0.3247	0.0217	0.0116	- (0.5137)
Mouse	1/1	84554	86887	2334	0.6482	0.8761	0.7241	0.0037	0.0919	+ (0.7727)
Rat	1/1	85382	87382	2256	0.6352	0.8480	0.6748	0.0158	0.1012	+ (0.7712)

Table 4.3: *CpG island prediction in the 100kb upstream of exon 1.* 100kb of DNA upstream of Gli3 exon 1 in chick, human, mouse and rat genomes was analysed by CpGProd (Ponger and Mouchiroud, 2002). 'Sequence name' gives the species under investigation; 'Number' shows the number of CpG islands identified in each sequence; 'begin', 'end', 'length', 'G+C frequency', 'CpG o/e ratio', 'start-p', 'AT skew', 'GC skew' and 'strand' each refer to the CpG island described.

The prediction of TSSs within a CpG island is troublesome, since they are often not associated with well characterised sequence motifs such as the TATA box and Inr sequence (Carninci *et al.*, 2006). I decided to focus my search on a 20kb region upstream of human exon 1, along with the corresponding region of mouse, rat and chicken genomes. In each species this region incorporates the entire CpG island, the characterised TSSs of Gli3, and all of the conserved elements identified upstream of exon 1 in Chapter 3.

McPromoter incorporates data on DNA structure and binding site information generating a graph that displays the likelihood of a promoter occurring at each position (Fig. 4.1). In each species a region within the CpG island (enclosed by grey lines in Fig. 4.1) was identified as the most likely location of a promoter element. However, at the threshold used a promoter element was only identified in mouse, predicted to lie within the CpG island with a TSS at nucleotide 5391. This TSS has a predictive score of 0.00999, on a range of -0.5 to 1 in which higher values indicate a greater likelihood of promoter activity. This demonstrates that although the CpG islands are the most likely regions to contain promoter elements, putative promoters only score weakly in comparison to documented promoter elements. Interestingly, in the human sequence a second region, located around 5kb upstream of exon 1, also has a high likelihood of containing a promoter element. This appears to correlate with the small second CpG island predicted by CpGProd in the human sequence (Table 4.3).

4.2.2   Transcription Start Site Prediction

The same region containing 20kb upstream of exon 1 was analysed using various programmes listed in Table 4.1. Here I describe the results of the promoter prediction tools that generated the most insightful results. Some tools were found to give an output that was too vague to be critically included in my study, others were not accessible at the time. Figure 4.2 summarises the TSS's predicted by these programmes, showing regions of particular interest that are referred to in the text.
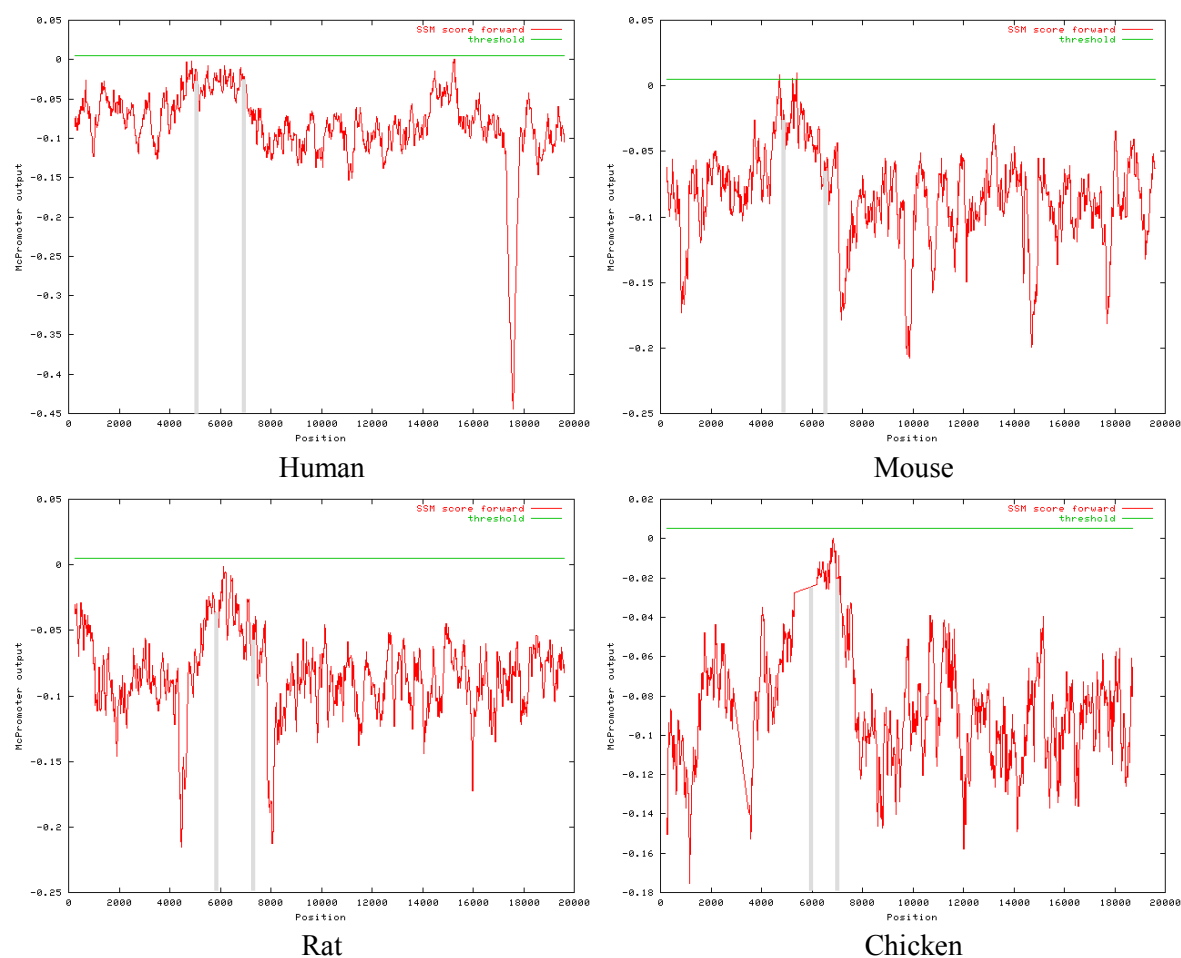
Figure 4.1: *McPromoter analysis of the 20kb region upstream of exon 1 in each species.* Grey lines indicate the boundaries of CpG islands identified in each species using the UCSC browser. The height of the graph at any point indicates the likelihood of a promoter occurring at that site, on a scale of -0.5 to 1 with 1 being optimal. The threshold used is for an intermediate sensitivity of 50%, for which the threshold is +0.005. At this threshold a promoter is only predicted in the mouse sequence. However, in each species the CpG island is identified as the most likely to contain a promoter element. Lowering the threshold will increase the sensitivity and the number of promoters predicted, but will also increase the chance of false-positives.

4.2.2.1 Comparison of programmes that use a linear discriminate function

Fprom, TSSG and TSSW each identify functional motifs in a sequence and compare them with the nucleotide composition of the region. Predictions are discriminated as to whether or not they contain a TATA box, and are given a linear discriminant function (LDF) score, the higher the LDF the greater the predictive value (Solovyev and Salamov, 1997). TSSG and TSSW also search for selected transcription factor binding sites, and incorporate the density of functional sites into their predictions. They differ in the selection of transcription factor binding sites used. Promoters predicted by these programmes, along with LDF scores are shown in Table 4.4A. Within each species, no TSS was predicted by all three programmes, but several sites are supported by more than one prediction within a 30bp region (boxed predictions in Table 4.4A).

For each species except chicken, the highest LDF produced by each programme (shown in bold in Table 4.4A) occurs approximately 14kb upstream of exon1, within the CpG island (grey highlighting in Table 4.4A). The highest scores assigned to chicken promoters by TSSW and TSSG fall slightly outside of this region but are still within the CpG island limits reported by CpGProd, and would be included in CpGProd predictions. It is also possible that the chicken promoter lies within the unsequenced region in the CpG island of the chicken sequence. Attempts to clone and sequence this region were hindered by the high GC content. For each species multiple TSSs were predicted within the CpG island, but none align at similar position between all species (boxed regions in Figure 4.4A represent binding sites that align within 30bp of one another on a MAVID alignment of the region). This suggests that the position of the Gli3 transcriptional start site might not be conserved between species. This might also explain why exon 0 has only been characterised in mouse. TSSW predicts a TSS 2bp upstream of mouse exon 0, with a high predictive score of 10.42 this is the most likely candidate for a promoter element responsible for transcription initiation from exon 0. TSSs were not identified at this position in any other species, but human and rat TSSs were predicted within exon 0 (Fig. 4.2C). The nearest TSS in chick was predicted 500bp downstream of exon 0.

4.2.2.2 The CpG island might constitute a broad promoter

The prediction of several TSSs within the CpG island fits with the hypothesis that the CpG island constitutes a broad promoter with multiple transcriptional start points. Within such a promoter, the presence of a TATA box will preferentially define the transcription start site. With the exception of chicken, TATA box containing promoters (shown in red in Table 4.4A) were predicted within the CpG island for each species. However these do not align, and are each only identified by a single tool. The TATA-containing promoters predicted in the mouse CpG island do not align with exon 0, and are not supported by other predictions.

The increased evolutionary rate of broad promoters relative to sharp promoters is thought to facilitate species specific variations. The presence of multiple transcriptional start sites in such regions would allow the loss of one TSS to be accommodated by the use of another. Otherwise, a novel TSS may evolve in a specific lineage and become the preferred start site. Interestingly, a TSS located at nucleotide 5260 in the mouse sequence is supported by TSS predictions in human and rat alignments (Fig. 4.2B). Furthermore, transcription start sites in this region are predicted both TSSW and TSSG in mouse and rat sequences.

4.2.2.3 Comparison with other promoter prediction tools

To establish whether Fprom, TSSW and TSSG give accurate predictions of promoter elements, other programmes were compared. NNPP2.2 is based on a time-delay neural network that predicts promoters by recognising TATA boxes and Inr sequences, and produces a prediction score between 0 and 1 for each unit identified (Reese, 2001; Table 4.4B). In several instances NNPP2.2 predicts promoters at similar positions to the other programmes. TSSs predicted by NNPP2.2 are shown in bold and are underlined in Figure 4.2. At nucleotide 8121 in the chicken sequence NNPP2.2 predicts a TSS that aligns with a site predicted by Fprom in both mouse and rat sequences (Fig. 4.2D). Similarly, a site predicted by TSSG at position 11500 in the mouse sequence is supported by two predictions made by NNPP2.2 (Fig. 4.2E). A prediction made by NNPP2.2 at position 12387 in the mouse sequence is flanked by predictions made by Fprom (nt13101) and TSSW and TSSG (nt13099) in the rat sequence (Fig. 4.2F). The support of predictions from all programmes investigated at this site suggests that it could represent a true promoter element, but the prediction is not supported by any characterised transcript. Sites that show a strong agreement between the different tools analysed each occur outside of the CpG island.

Multiple novel promoter elements are also predicted by NNPP2.2, including two sites at which TSSs predicted in each species align exactly (shown in bold in Table 4.4B). Interestingly, in each species NNPP2.2 predicted a transcriptional start site that aligns within a few (≤12) bp of the 5' end of mouse exon 0 (Fig. 4.2C). This would allow for exon 0 to constitute the true 5' end of the Gli3 transcript in all species, but the lack of support from other programmes indicates that these sites might be false predictions. None of the promoter elements predicted by NNPP2.2, Fprom, TSSW or TSSG correlate with the TSS predicted by McPromoter at nucleotide 5391 in the mouse sequence (Fig. 4.2B).

4.2.2.4 Characterisation of conserved sequence motifs surrounding predicted TSSs

Cister and Promoter Scan were next used to establish whether any of the TSSs identified correlate with clusters of sequence motifs associated with core promoters. Cister produces a graphical output in a similar format to that produced by McPromoter, but also indicates the presence of characterised elements associated with promoter activity. Results are shown in Figure 4.3, and in Table 4.4.C.

Human Fprom		Human TSSW		Human TSSG		Mouse Fprom		Mouse TSSW		Mouse TSSG		Rat Fprom		Rat TSSW		Rat TSSG		Chicken Fprom		Chicken TSSW		Chicken TSSG	
												1054	1.762	109	0.6					135	0.46		
3873	-0.304	3871	0.57			3043	0.741					3743	-0.809										
4779	2.887											4862	0.074	4862	0.95								
				5001	5.02	5049	11.227																
		5417	10.8	5311	5.34			5260	6.13	5267	5.55			6063	6.11	6073	4.92						
						5463	0.075																
				5643	6.88			5572	10.72											5338	4.1		
										5644	9.41												
		5837	6.03					5977	4.77							6432	6.83			6575	4.87		
6203	11.519											6793	9.851										
		6252	11.45	6244	12.46																		
		6469	0.84			6215	9.711							7008	5.57			6906	11.193				
		6856	6.77					6461	10.45														
				7131	9.27	7072	-0.643					7846	0.762							7077	8.92	7065	10.93
7936	0.118					8180	0.51					8966	1.068										
						9477	-0.846					10277	0.64	10275	0.48			10109	-0.937				
11114	2.427									11500	4.03	12741	0.191										
				13240	5.37							13101	0.294	13099	0.55	13099	4.22						
						13306	2.924			13269	9.24	14021	2.218										
		15007	0.48															15671	-0.42				
		15523	0.6	15520	11.24																		
		15536	3.73																				
						16359	1.554																
						16946	-0.197																
						18729	1.142											18660	-0.521				

Table 4.4A: *Location of TSSs predicted by Fprom, TSSW and TSSG.* For each programme the first column represents the position of the TSS in the input sequence (20kb immediately upstream of exon 1), and the second column shows the LDF score. TSSs shown in green represent TATA-less promoters, those shown in red predict the presence of a TATA box. Boxes are used to link promoters that occur within 30bp of one another in a MAVID alignment of the region. The TSS producing the highest LDF score for each programme is shown in bold. Grey highlighting shows the location of the CpG island as annotated in the UCSC browser.

Table 4.4B: *Location of TSSs predicted by NNPP2.2.* The location and probability score for TSSs predicted by NNPP2.2 is shown. Sites shown in bold align exactly between species, those shown in blue occur within 30bp of a TSS predicted in the same species by Fprom, TSSW or TSSG. Green boxes represent CpG boundaries in each species.

Table 4.4C: *Location of TSSs predicted by Cister.* Green boxes represent CpG boundaries in each species. Colour coding is used show different binding sites in Figure 4.3 Str. = strand, P = position.

Table 4.4B: NNPP2.2 results

human		mouse		rat		chick	
NNPP2.2		NNPP2.2		NNPP2.2		NNPP2.2	
737	0.94	310	0.92	199	0.84	108	0.86
1233	0.91	1327	0.88	330	0.97	3723	0.81
1328	0.81	1903	0.91	996	0.85	4440	0.8
1369	0.84	2348	0.94	2037	0.81	6332	0.88
2082	0.84	2789	0.84	2984	0.92	6426	0.8
2190	0.98	3002	0.86	3480	0.88	6627	0.83
2378	0.97	3030	0.95	3704	0.93	7371	0.84
2654	0.85	3268	0.96	4014	0.84	7494	0.93
3829	0.85	4284	0.99	6673	0.81	7963	0.81
4567	0.97	4409	0.93	6685	0.88	8081	0.93
4740	1	4687	0.82	6716	0.81	8462	0.87
5370	0.8	5440	0.82	8747	0.86	9548	0.99
5387	0.81	5470	0.92	9945	0.81	9903	0.94
5677	1	5907	0.82	10236	0.99	10598	0.96
6124	0.84	5937	0.8	10889	0.89	10879	0.86
6146	0.81	7174	0.84	10966	0.92	10965	0.97
6224	0.81	7186	0.84	11297	0.99	14063	0.92
6438	1	7964	0.97	11787	0.91	14720	0.84
6915	0.81	9448	0.99	11835	1	15186	0.95
7002	0.85	9795	0.91	13520	0.84	15256	0.88
7093	0.9	10242	0.83	13981	0.94	15302	0.83
7228	0.93	10573	0.97	14180	0.89	15770	0.97
7608	0.97	11458	0.84	14844	0.94	15919	0.98
7895	0.9	11468	0.99	15470	1	18163	0.87
8232	1	11515	0.83	15518	0.99		
9754	0.99	11632	1	16337	0.95		
10108	0.9	12344	0.99	16458	0.92		
10314	0.82	13158	0.82	16763	0.89		
10413	0.95	13264	0.91	16896	0.89		
10820	0.99	13473	0.96	17703	0.98		
11075	0.95	13840	0.82	17742	0.89		
11146	0.97	14885	0.87				
12443	0.8	15506	0.82				
13403	0.86	15523	0.84				
14686	1	15813	0.92				
14978	0.88	16116	0.87				
15180	0.82	16348	0.98				
15206	0.84	16420	0.97				
15760	0.91	16876	0.9				
16271	0.98	17490	0.89				
16721	0.97	18681	0.98				
		19641	0.93				

Table 4.4C: Cister resuts

Cister results for human 100kb (1 to 99999 out of 99999

Type	Position	Str.	Sequence	P
Ets	01724 - 01734	-	tccttcctggc	0.21
NF_1	01728 - 01745	+	tcctggcagctggccagt	0.22
NF_1	01729 - 01746	-	cctggcagctggccagtc	0.1
CCAAT	01736 - 01751	+	gctggccagtcagcgt	0.12
Ets	01786 - 01796	-	tgtttcctttc	0.22
SRF	01855 - 01867	-	tccttacatgcag	0.21
ERE	01899 - 01912	+	aaatcaccaggacc	0.47
ERE	01947 - 01960	+	aggtgagcctggcc	0.62
Myc	01976 - 01985	-	tccacttggg	0.1
Myf	02034 - 02045	-	atgcagctgcct	0.21
Tef	02051 - 02062	+	caccttcctcgg	0.28
CCAAT	05886 - 05901	-	aagtcgattggatgcc	0.2
SRF	05960 - 05972	+	aaccaaatcagaa	0.19
AP_1	05988 - 05998	-	tgtgagtcaga	0.2
LSF	10800 - 10814	-	ctgcccggcccagc	0.4
AP_1	10815 - 10825	-	cattagtcatc	0.28
ERE	10847 - 10860	-	ggccatgctgccca	0.4
LSF	10921 - 10935	+	tctgatttccgcttg	0.14
Sp1	10952 - 10964	-	gtcccctccctac	0.23
LSF	11119 - 11133	-	ctggcctgaaacagc	0.28
Sp1	11161 - 11173	+	gaggggaggggc	0.64
Sp1	11181 - 11193	+	tggggggagggtgg	0.4
Ets	12521 - 12531	-	tgcttcctgca	0.17
Myc	12544 - 12553	-	ggcacgtggg	0.18
Sp1	13193 - 13205	-	cccccgcccaac	0.16
Ets	13239 - 13249	+	gagaggaaatg	0.35
Ets	18383 - 18393	-	tacttcctccc	0.31
E2F	18388 - 18399	-	cctcccgcccaa	0.12
ERE	18410 - 18423	+	agatctcactgccc	0.3
ERE	18445 - 18458	+	agatcactgaggct	0.12
Myf	18459 - 18470	+	cagcaacaggtt	0.2
LSF	18490 - 18504	-	ccagctagccccagg	0.2
ERE	18516 - 18529	+	aggtagccctgccc	0.36

Cister

Type	Position	Str.	Sequence	P
Myc	14770 -	+	gccatgtgcc	0.19
NF_1	14821 - 14838	-	ctgggaagtgggccaagt	0.18
LSF	14966 - 14980	+	gctggttcagactac	0.16
NF_1	16589 - 16606	+	ccttggcaaggtgccacg	0.11

Cister results for mouse 100kb (1 to 99981 out of 99981

Type	Position	Str.	Sequence	P
LSF	10536 - 10550	-	ctgccctagcccagc	0.27
AP_1	10551 - 10561	-	cattagtcatc	0.23
ERE	10583 - 10596	-	ggccatgctgcccg	0.39
LSF	10672 - 10686	+	gctggtttgaggtaa	0.13
Sp1	10687 - 10699	-	gtccctccctcc	0.29
NF_1	10774 - 10791	+	tgttggcatcctaccaca	0.2
Ets	16275 - 16285	-	tacttcctcc	0.15
Ets	16565 - 16575	+	tacaggaagca	0.11
NF_1	16584 - 16601	+	ccttggcaagttgcatt	0.14
Sp1	17854 - 17866	-	ggcgccacccca	0.14
CCAAT	18817 - 18832	+	ataagccaatcagcag	0.13

Cister results for chick 100kb (1 - 99399 out of 99399

type	position	Str.	sequence	P
CRE	02652 - 02663	+	tgtgacgatagc	0.11
Myf	02699 - 02710	+	cagcaacagcag	0.34
Myf	02718 - 02729	-	gaccagctgccc	0.14
NF_1	02772 - 02789	-	tgaggtggggagccaaaa	0.21
Ets	02852 - 02862	+	gagaggaagtg	0.4
Sp1	04355 - 04367	-	gagccctcccctc	0.14
Myc	04539 - 04548	-	ggcacatgcc	0.16
Myc	04561 - 04570	+	gccctgtgcc	0.13
CCAAT	04626 - 04641	-	ttgcccattggtatgc	0.14
NF_1	04650 - 04667	+	cgctggctcccagactgc	0.17
LSF	04676 - 04690	-	ctgctgcaagccagc	0.16
LSF	04697 - 04711	-	ctggatctccccagt	0.11
AP_1	06614 - 06624	+	tctgattcatc	0.16
AP_1	06614 - 06624	-	tctgattcatc	0.39
Myc	06647 - 06656	-	gacacgtgag	0.13
Myc	06647 - 06656	+	gacacgtgag	0.17
TATA	06666 - 06680	-	acggtgtttattt	0.4
Ets	06676 - 06686	-	tatttccttct	0.24
GATA	06783 - 06795	-	ccctttttctttt	0.1
SRF	06822 - 06834	+	aaccaaatcagaa	0.45
AP_1	06836 - 06846	-	tgtcagtcagt	0.33
AP_1	06850 - 06860	+	agtgagtcaga	0.11
AP_1	06850 - 06860	-	agtgagtcaga	0.53
GATA	06865 - 06877	+	atcagattacaag	0.32
TATA	16727 - 16741	-	cacgcacttttatag	0.14
Ets	16865 - 16875	+	ggcaggaactg	0.14
NF_1	16894 - 16911	+	cgctggcaggagccctgt	0.17

Figure 4.2*: Map of promoter regions located within the CpG island upstream of Gli3.*

A) The location of each conserved element identified in Chapter 3 upstream of Exon 0 is shown (green and brown boxes; see Fig. 3.4). Region 1 lies within a CpG island (grey box).

B-E) A MAVID alignment was produced of the 20kb immediately upstream of Gli3 in human (h), mouse (m), rat (r) and chick (c) genomes (Bray and Pachter, 2004). The regions shown are those predicted by various computational prediction tools to contain promoter activity, their position relative to the conserved elements identified in Chapter 3 is indicated in (A). Regions of the alignment shaded in grey form part of a CpG island (Gardiner-Garden and Frommer, 1987). Transcriptional start sites predicted by Fprom, TSSW, TSSG and NNPP2.2 tools are shown; In B and C promoter elements identified by Promoter Search are sindicated (see key below). In addition, transcriptional start sites identified by 5' RACE and EST analysis are indicated. The sequence shown in C contains exon 0, which has previously been reported in the mouse genome (shown in italics and underlined). The sequence shown in B contains an exon identified by 5' RACE on mouse E9.5 cDNA. The sequence of this exon identified in my longest RACE product is shown in italics and is underlined. Note that the 5' end of this transcript lies within a Sp1 site conserved between mouse and rat genomes. All experimentally verified transcriptional start sites within the sequence shown in B produce transcripts that splice from the 3' end of exon 0b to exon 1 (not shown). Two regions (Element I and Element II) were predicted by multiple prediction tools to contain transcriptional start sites, and are likely to contain promoter elements regulating transcriptional start sites located in exon 0b and exon 0, respectively. A 300bp region containing exon 0b, and that corresponds to Element I, has previously been shown to contain promoter activity, and is shown in italics (Paparidis, 2005). Furthermore, this region contains transcriptional start sites associated with transcripts expressed in E9.5 mouse embryos, suggesting that it contains the endogenous promoter of Gli3 at this stage of development.

<u>**KEY**</u>

```
N TATA-less promoter prediction TSSW/TSSG/fprom
N TATA promoter prediction TSSW/TSSG/fprom
N NNPP2.2 predicted start site

Promoter Search results:
N = AP2 binding site  N = TTV inverted repeat
N = GCF binding site  N = PUF binding site
N = T-Ag binding site N = Krox binding site
N = SP1 binding site  N = UCE.1 binding site
N = Early-seq1          N = c-fos.5 binding site
N = JVC_repeated sequence
```

NNNNNNNNN = 300bp promoter sequence previously tested for enhancer activity in vivo (Paparidis, 2005).

N = 5' end of 5'RACE products generated from E9.5 mouse cDNA

N = exon identified by 5'RACE on human placental cDNA (Paparidis, 2005).

ENSMUSTTxxxxx or BYxxxxx: 5' end of EST transcript

Element I/II: Promoter element
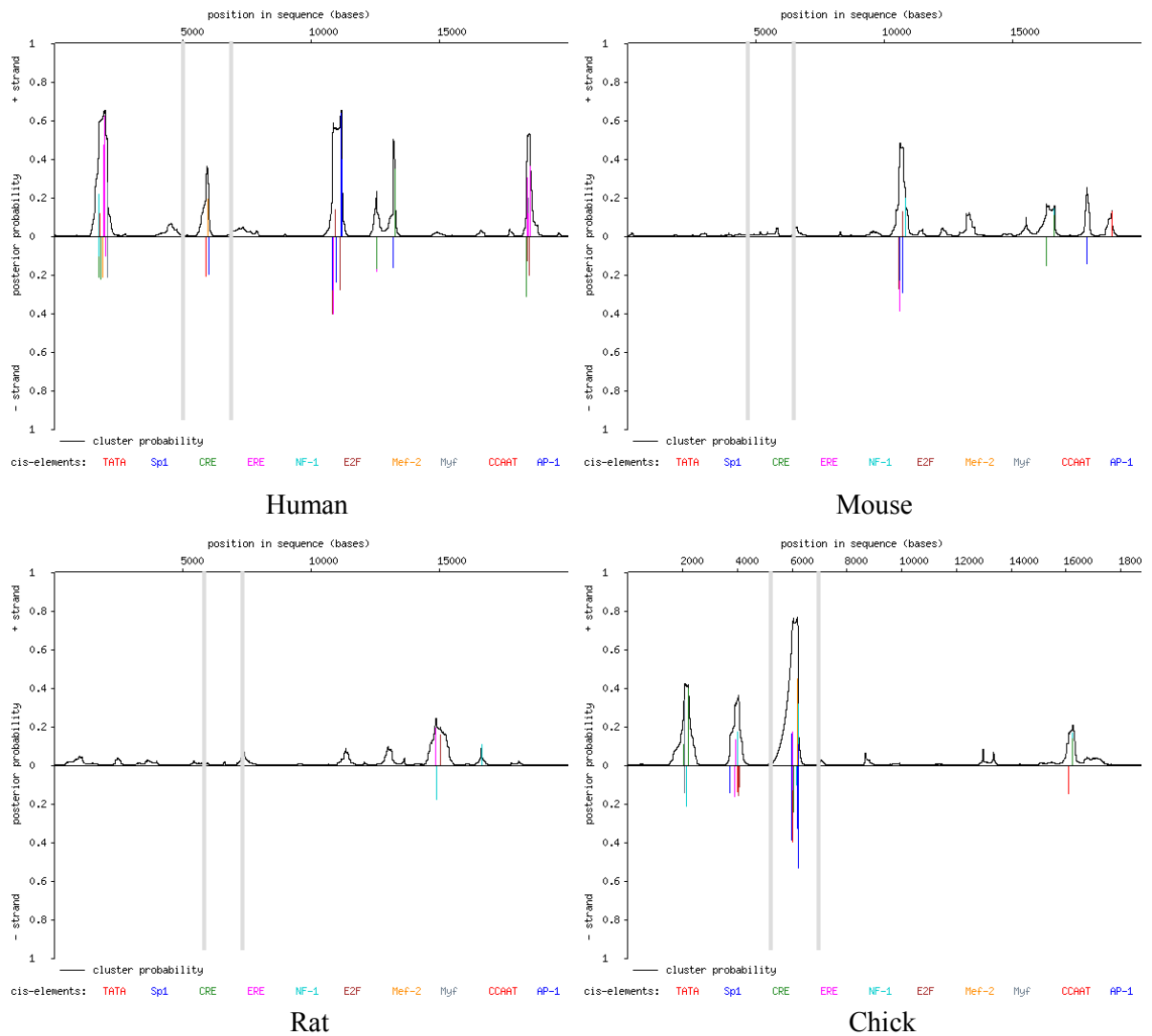
Figure 4.3: *Cister predictions of promoter location upstream of Gli3.* The likelihood of a core promoter occurring at various positions within the 20kb upstream of Gli3 in various organisms is shown. Grey lines indicate the boundary of the CpG island in each species, as annotated on the UCSC browser. Coloured lines represent the location of promoter elements associated with core promoters, further details are shown in Table 4.4C.

Based upon clustering of core-promoter motifs, the CpG island does not contain the region most likely to convey promoter activity. Within the CpG islands only human and chicken sequences are predicted to contain core-promoter elements, none of the core promoter elements predicted elsewhere are conserved across all species. However, Promoter Scan identifies a larger number of sequence motifs associated with promoter elements than any of the programmes previously discussed. The sequence analysed by Promoter Scan was chosen to be the CpG island +1kb flanking either side. In each species Promoter Scan predicted one or more promoters within the CpG island, and comparison of the functional motifs identified indicated synteny, this is most readily observed in the boxed regions in Figure 4.4.

The motifs present on the forward strand identified by Promoter Scan were marked on the MAVID alignment (Fig. 4.2), they indicate that amongst mammals clusters of motifs

associated with promoter activity are conserved. Interestingly, promoters were predicted by Promoter Scan in mouse, rat and human sequences that fall within region B in Figure 4.2. The chick sequence aligns poorly in the corresponding region, but some of the same elements are observed (Fig. 4.4). Within this region, several sites are observed that were predicted by other promoter prediction tools. These include sites that were identified by both TSSW and TSSG, and that occur within a similar position between species, NNPP2.2 also predicts several TSSs in the corresponding region of the human sequence (Fig. 4.2B).

Interestingly, a second promoter predicted by Promoter Scan in human and chicken sequences surrounds mouse exon 0 (Fig. 4.2C), and correlates with promoters predicted by Cister in human and chicken sequences but not in rodents. Of particular interest a cluster of UCE.1, AP2 and SP1 sites in the human sequence also incorporates a TSS predicted by NNPP2.2 at a position aligning with the reported 5'end of mouse exon 0 (Fig. 4.2C). The absence of promoter motifs in rodent sequences of this region is puzzling, especially since it correlates to a transcriptional start site that has been identified in mouse. The motifs present in human and chicken sequences might act to enhance transcription from this start site in some cellular contexts, whilst transcription might initiate from an alternative site in other circumstances. The rodent lineage may have lost the alternative TSS during evolution, such that even in the absence of promoter motifs, transcription will preferentially occur from the characterised TSS.

The in-silico promoter prediction presented here suggests Gli3 transcription likely to initiate within the CpG island upstream of the protein coding sequence. This is supported by the previous identification of a non-coding exon in the mouse genome (exon 0). Within the CpG island two promoter regions are supported by multiple prediction methods. These can broadly be classified as Element I and Element II, indicated in Figure 4.2. These two elements may represent alternative promoter elements that are differentially used to promote Gli3 expression in different cellular contexts.

```
>mouse
Promoter Score: 58.11
Promoter Cut-off = 53.000000)
Significant Signals:
 Name          Strand Lcn    Weight
AP-2             +    1301   1.863000
GCF              +    1303   2.361000
PuF              +    1323   1.082000
JCV_rep_seq      +    1327   1.427000
AP-2             -    1329   1.091000
T-Ag             +    1331   1.086000
APRT-CHO_US      -    1335   1.628000
TTR_inv_repeat   +    1342   2.151000
GCF              -    1344   2.284000
KROX24           +    1346   2.151000
EGR-1            -    1354   2.294000
JCV_rep_seq      +    1412   1.427000
T-Ag             +    1416   1.086000
Sp1              +    1417   3.013000
Sp1              +    1417   7.086000
AP-2             -    1419   1.091000
Sp1              -    1422   3.061000
EARLY-SEQ1       -    1424   5.795000
UCE.2            -    1472   1.216000
GCF              +    1477   2.361000
Sp1              +    1491   3.292000
Sp1              -    1496   3.361000
AP-2             -    1515   1.091000
AP-2             +    1537   1.108000
Sp1              +    1538   2.755000
Sp1              -    1543   2.772000


>rat
Promoter Score: 54.01
(Promoter Cut-off = 53.000000)

Significant Signals:
 Name          Strand Lcn    Weight
AP-2             +    1084   1.863000
GCF              +    1086   2.361000
PuF              +    1106   1.082000
JCV_rep_seq      +    1110   1.427000
AP-2             -    1112   1.091000
T-Ag             +    1114   1.086000
APRT-CHO_US      -    1118   1.628000
TTR_inv._rep     +    1125   2.151000
```

```
GCF              -    1127   2.284000
KROX24           +    1129   2.151000
EGR-1            -    1137   2.294000
JCV_rep_seq      +    1195   1.427000
T-Ag             +    1199   1.086000
Sp1              +    1200   7.086000
Sp1              +    1200   3.013000
AP-2             -    1202   1.091000
Sp1              -    1205   3.061000
EARLY-SEQ1       -    1207   5.795000
Sp1              +    1236   2.755000
AP-2             +    1237   1.355000
Sp1              -    1241   2.772000
UCE.2            -    1255   1.216000
GCF              +    1260   2.361000
Sp1              +    1274   3.292000


>chicken
1) Promoter Score: 74.38
(Promoter Cut-off = 53.000000)
TATA found at 524, Est.TSS = 554

Significant Signals:
 Name          Strand Lcn    Weight
T-Ag             +    323    1.086000
UCE.2            +    325    1.278000
AP-2             -    326    1.091000
UCE.2            -    332    1.216000
ETF              -    337    3.933000
AP-2             -    338    1.672000
Sp1              +    340    2.755000
GCF              -    344    2.284000
Sp1              -    345    2.772000
Sp1              +    362    3.013000
Sp1              +    363    3.191000
Sp1              -    367    3.061000
Sp1              -    368    3.119000
AP-2             +    389    1.355000


2) Promoter Score: 53.87
(Promoter Cut-off = 53.000000)

Significant Signals:
 Name          Strand Lcn    Weight
AP-2             +    1382   1.355000
UCE.2            -    1394   1.216000
```

```
UCE.2            +    1428   1.278000
AP-2             -    1453   1.091000
GCF              -    1467   2.284000
AP-2             +    1469   1.108000
GCF              +    1481   2.361000
T-Ag             +    1483   1.086000
AP-2             -    1494   1.672000
Sp1              +    1511   3.191000
Sp1              -    1516   3.119000
APRT-mouse_US    -    1518   7.604000
EARLY-SEQ1       +    1546   6.322000
Sp1              +    1548   3.292000
Sp1              -    1553   3.361000
Sp1              -    1553   7.086000
JCV_rep_seq      +    1553   1.658000
T-Ag             +    1605   1.086000
ETF              -    1607   3.933000
AP-2             -    1607   1.091000


>human
1) Promoter Score: 62.62
(Promoter Cut-off = 53.000000)

Significant Signals:
 Name          Strand Lcn    Weight
AP-2             +    1097   1.108000
PuF              +    1106   1.082000
JCV_rep_seq      +    1110   1.427000
AP-2             -    1112   1.091000
T-Ag             +    1114   1.086000
APRT-CHO_US      -    1118   1.628000
GCF              +    1123   2.361000
GCF              -    1128   2.284000
JCV_rep_seq      +    1196   1.427000
T-Ag             +    1200   1.086000
Sp1              +    1201   3.013000
Sp1              +    1201   7.086000
AP-2             -    1203   1.091000
Sp1              -    1206   3.061000
EARLY-SEQ1       -    1208   5.795000
Sp1              +    1277   3.013000
JCV_rep_seq      +    1277   1.427000
Sp1              +    1278   3.191000
AP-2             -    1282   1.672000
Sp1              -    1282   3.061000
Sp1              -    1283   3.119000
```

```
EARLY-SEQ1       -    1284   5.795000
(Sp1)            -    1284   6.819000
AP-2             -    1284   1.091000


2) Promoter Score: 75.03 (Promoter
Cut-off = 53.000000)
Significant Signals:
 Name          Strand Lcn    Weight
JCV_rep_seq      +    2003   1.427000
T-Ag             +    2018   1.086000
AP-2             -    2021   1.091000
c-fos.5          +    2024   1.912000
GCF              -    2038   2.284000
Sp1              +    2062   2.755000
Sp1              -    2067   2.772000
GCF              +    2082   2.361000
AP-2             +    2100   1.108000
GCF              -    2114   2.284000
UCE.2            +    2115   1.278000
AP-2             +    2117   1.355000
Sp1              +    2117   3.292000
UCE.2            -    2118   1.216000
EGR-1            +    2118   5.736000
Sp1              +    2122   3.361000
KROX24           -    2126   5.378000
Sp1              +    2141   3.191000
AP-2             -    2142   1.064000
TTR_inv._rep     -    2144   3.442000
Sp1              -    2146   3.119000
AP-2             +    2187   7.211000
UCE.2            +    2195   1.278000
UCE.2            -    2198   1.216000
GCF              -    2202   2.284000
GCF              +    2224   2.361000
```

Figure 4.4: Functional elements present in promoters predicted by Promoter Scan. Colours represent the colour of the corresponding site as annotated in Figure 4.2. Lcn=location in the inputted sequences.

4.2.3 EST search

To further establish where the 5' end of the Gli3 transcript might lie, data was collated from various EST databases, including EBI-EMBL, Ensemble and NCBI. ESTs identified within the 20kb upstream of murine Gli3 are annotated on the MAVID alignment (Fig. 4.2), they all lie within the CpG island. ENSMUSTT00003857423 and ENSMUSTT00003857424 are all products of ENSMUSESTG00003740645 (Ensemble release 41, October 2006; ESTMUSESTG00000028830 in Ensemble release 49, March 2008). Another transcript, ENSMUSTT00003857422 was also identified but has since been removed from the database and is not shown in Figure 4.2. BY734506 and BY753055 are EMB-EBI entries. ENSMUSTT00003857423 (also listed as BY734506) and ENSMUSTT00003857424 initiate from an exon located Element I, and splice from the same 3' end to the beginning of exon 1. I shall refer to the exon shared by these ESTs within Element I as exon 0b. ENSMUSTT00003857423 contains an additional 30bp upstream of the ENSMUSTT00003857424 start site suggesting that exon 0b may contain multiple TSSs. This is consistent with its location in a broad promoter. BY753055 initiates upstream of Element II, and contains exon 0. It begins upstream of the characterised exon 0 but contains the same 3' end and splices to exon 1. Thus, the two promoter elements predicted within the CpG island are both supported by EST data. The data suggest that two alternative non-coding exons exist at the 5' end of the mouse Gli3 transcript. The use of an alternative first exon may add to the regulatory control of Gli3 expression, and may result in distinct tissue specific transcripts. Indeed, both ESTs containing exon 0b were isolated from a cDNA library derived from lung tissue, whereas transcript BY753055 that contains exon 0 was derived from RNA extracted from the adult mouse inner ear.

4.2.4 5'RACE

To establish the true 5' end of mouse and chick Gli3 at the developmental stage under investigation, 5' RACE was carried out on RNA extracted from either E9.5 mouse embryos or HH Stage 10 chick embryos. To establish whether transcripts are tissue specific, separate reactions were carried out on RNA extracted from neural tissue (notochord and neural tube) and mesoderm (paraxial mesoderm) of chicken embryos. To increase my likelihood of isolating the 5' end of Gli3 transcripts, primers were designed within exon 2. The initial RACE reaction generated very weak bands when visualised on an ethidium bromide stained agarose gel. However, where these samples were used in a second round of 5' RACE a series of discrete

bands of sizes varying between 200bp and 2.25kb were obtained (Fig. 4.5). To demonstrate that these RACE products were present in the original mRNA pool, the remaining cDNA product was amplified for a further 10 cycles at a high annealing temperature of 68°C. RACE products observed were similar to those generated by the re-PCR reaction shown in Figure 4.5.
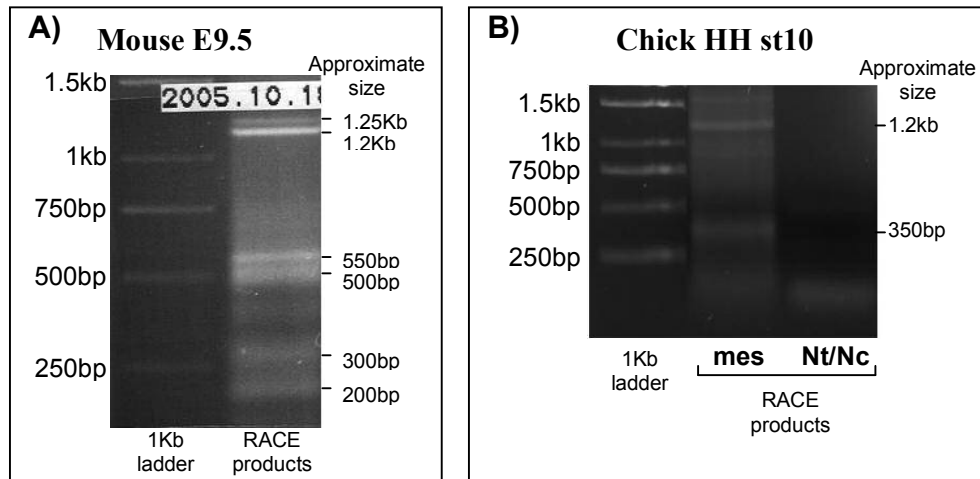
Figure 4.5: *Products produced by 5'RACE from exon 2 of mouse and chick Gli3. 5'RACE was carried out on RNA extracted from E9.5 whole mouse embryos (A), and from mesoderm or neural tube and notochord of HH stage 10 chick embryos (B). mes, mesoderm; Nt/Nc, Neural tube/notochord. . Approximate sizes of the main products are shown.*

E9.5 mouse cDNAs contain several RACE products of varying size, indicating that different transcripts are generated in the embryo. The major product is approximately 1.2kb long, although a slightly larger transcript of ~1.25kb is also observed (Fig.4.5A). In addition, four other RACE products of 200bp, 300bp, 500bp and 550bp were obtained. HH stage 10 chick cDNAs showed unequal success. In RNA extracted from mesoderm, the maximal product size is similar to that observed in mouse (~1.2kb, Fig. 4.5B). However, various transcript sizes amplified from mouse RNA were not identified in chicken embryos. Indeed, in chick embryos only 2 discrete transcript sizes were observed (Fig. 4.5B). This suggests that fewer alternative products are generated in chick embryos. Interestingly, amplification of Gli3 5' cDNA from RNA extracted from the notochord and neural tube of HH stage 10 chick embryos was unsuccessful. The most likely cause for this is that not enough RNA was extracted to produce the quality of cDNA required for this technique.

To characterise the RACE products generated, each product was sub-cloned into the pCR®II-TOPO® vector (Invitrogen), and inserts were analysed by EcoR1 restriction digest. Inserts that correlated in size with the PCR products generated by 5'RACE were sequenced and results are shown in Figure 4.6. Splice products identified by 5'RACE from mouse and chicken cDNAs

are summarised in Figure 4.7. The location of exons relative to the conserved regions identified in Chapter 3 is indicated.

Two clones were isolated from 5'RACE products amplified from mouse E9.5 cDNA that are products of splicing from exon 0b to exon 1. The splice junction correlates to that observed in ENSMUSTT00003857423, BY734506 and ENSMUSTT00003857424. No transcript containing exon 0 was identified. Thus, murine Gli3 transcripts containing exon 0b seem to be more abundant than those containing exon 0 at E9.5. On two occasions an alternative transcript was observed that splices from exon 2 to a novel exon located 3635bp upstream of exon 2, within the first intron of Gli3 (Fig. 4.6A, exon 1b). This exon does not overlap with the location of any of the conserved elements identified in Chapter 3. However, it is located 92bp downstream of Region I2, suggesting that this highly conserved non-coding element might be involved in the regulation of alternative splicing. In summary, these results show that the murine Gli3 locus harbours at least three alternative 5' non-coding exons, two of which have not been published.

The size of each cloned transcript is indicated in Figure 4.6. It should be noted that primers used in the 5'RACE reaction result in the addition 52bp to each amplified sequence, such that the bands observed in Figure 4.7 are larger in size than the corresponding cDNA transcript. Thus, the sequences shown in Figure 4.6A can account for most of the bands observed in Figure 4.5. However, clones correlating to the two largest products were not identified. Inclusion of all the exons identified upstream of exon 2 would result in a transcript of 665bp. A product of this size is not observed in Figure 4.5. The identity of the largest transcripts observed in Figure 4.5 has not been determined at this stage.

Transcripts were found to terminate at different positions within each exon. To rule out that the alternative TSSs identified are due to aborted 5'RACE, new primers were designed close to the 5' end sequence identified for each RACE product. 5'RACE was repeated using these primers. However, the new RACE products did not differ in their start point from those previously identified.

A) Mouse:

```
Exon 0b (84bp)
GGCCCCTCAGCGCAGCGCCCTGGGATCCGCGCGCGCGGAGCGGGACCCCGCCGGGCGCGGCCTGGAAAGGAGC
GGGAAAGCAAA
```

1x (449bp) 1x (408bp)

```
Exon 0 (132bp)
GCCGGCCCACGCCACACCATGCCCGCCGGGGCCGCACGCTGGGATCAGTCAGGCCATCCACGGTAGCCCGAGT
CCGCCTGCCGCTCCGCCGGCGCAGGTCTGTGGATTGGGACCTGACAGCCAAAGCCGCAG

Exon 1 (165bp)
GGATTCCTTTGAGAAACAAGCTGAAGTAATGAGAAGACATTATGGAGGCCCAGGCCCACAGCTCTACGGCGAC
TGAGAGGAAGAAAGCTGAAAATTCCATTGGGAAATGTCCCACGAGAACAGATGTCAGCGAGAAGGCCGTGGCC
TCTAGTACCACTTCCAATG
```

1x (244bp)

1x (316bp)
```
Exon 1b (116bp)
GGAGAGAGAGAGAGGCAAGAAAGGCCAAGAGATACTGAGAATGTTCCTAGAGGGCCGGCCCAGCTAGAGAAAC
TAAGTTTCTAGAAAGTAGATAGCCAGAGTCCAGGTACTGAGTT
```

1x (270bp)

```
Exon 2 (From 5'end to end of primer = 200bp)
AGGATGAAAGTCCTGGACAGATCTATCACCGAGAGAGAAGAAACGCAATCACTATGCAGCCTCAGAGTGTGCA
GGGTCTCAACAAAATCAGTGAGGAGCCCTCGACGTCTAGTGATGAGAGGGCCTCGCTGATCAAGAAAGAGATC
CATGGCTCTCTACCACATCTGGCGGAGCCCCCTCTCCCTTACCGTGGGACTGTG
```

2x (137bp)
1x (59bp)
1x (192bp) 2x (45bp)

B) Chicken:

```
Exon X
GCGGGACGGACAAGTTTTCCGACAC
```

1x (429bp)

```
Exon 1 (163bp, extends 39bp beyond published start site)
GTATTAGTTCAAGGAGAGCTGAAGTAATGAGAAGTCACCATGGAGGCGCAGTCCCATAGCTCTACCACGACAG
AGAAGAAAAAAGTGGAGAATTCCATTGTGAAATGCTCCAATCGAACAGATGTCAGTGAGAAAGCTGTTGCCTC
CAGCACAACTTCCAATG
```

1x (263bp)   1x (365bp)   1x (172bp)
1x (141bp)

```
Exon c1i (50bp)
TGAGCCAGCCCTACCCGTACCCTACTCTAAAGCAAATGATCTGCAAGTGC

Exon c1ii (66bp)
CTGCATACATAGGAATACTTACTGCATGTGTGTGCCAGTAAGTGGAAGAAGCGTGCTTAGGGCATG

Exon 2 (From 5' end to end of primer = 125bp)
AGGATGAAAGTCCTGGACAGACCTATCACAGAGAGAGAAGGAACGCAATCACAATGCAGCCACAAGGTGGGCA
AGGCCTCAGCAAGATCAGTGAGGAGCCTTCAACATCTAGCGAGGAAAGGGCC
```

1x (68bp)

Figure 4.6: *Sequences of products identified by 5' RACE on mouse and chick cDNAs.* 5'RACE products isolated from E 9.5 mouse mRNA (A), or HH stage 10 chick mesoderm mRNA (B) are shown. Coloured lines represent the different splice forms. Arrows corresponding to the colours of the lines represent the 5' ends of each transcript, the number of times each transcriptional start site was observed is also shown. Exon sizes, and the length of each product are shown in brackets. Primer sequences are underlined and shown in italics. The sequence, and relative position of mouse exon 0 is shown, but was not identified in RACE products.
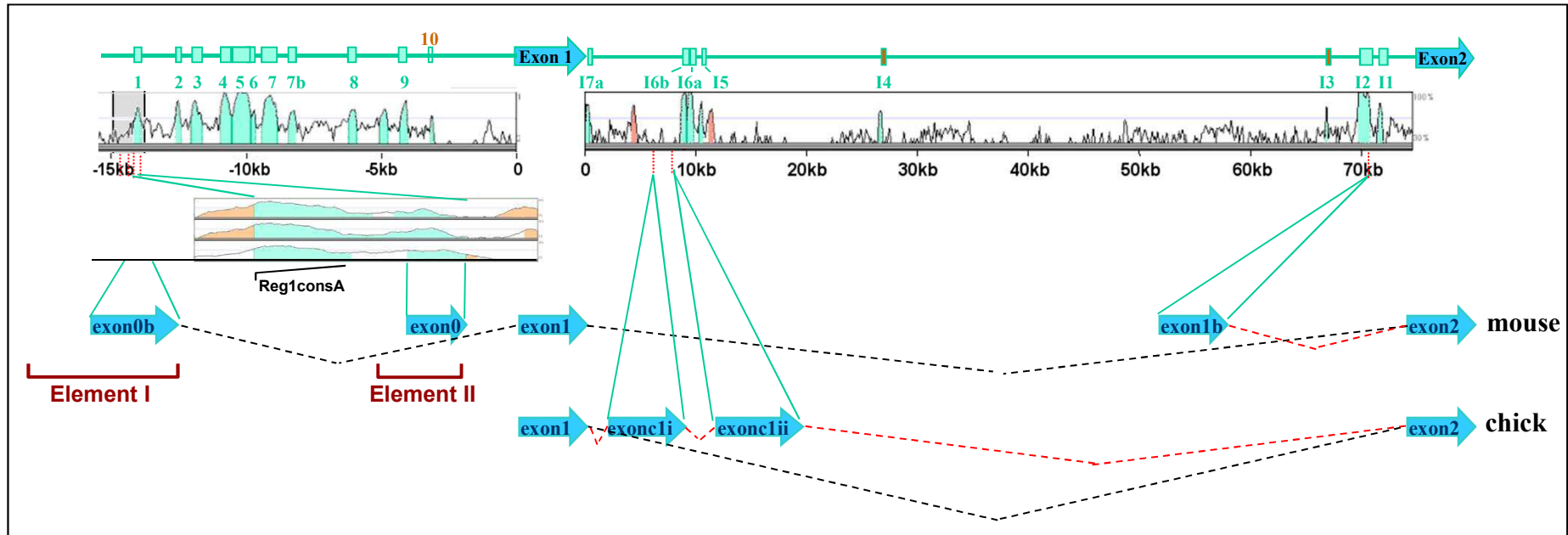
Figure 4.7: *Map of splice products identified by 5'RACE on mouse and chick cDNAs.*

The positions of various exons upstream of exon 2 relative to the conserved regions identified in Chapter 3 are shown. The region shaded grey surrounding Region 1 represents the CpG island. Exon 0b and exon 0 are both located in this region. A magnified view of Region 1 reveals two distinct regions of homology (also detected in Table 3.1). The 3' region of homology overlaps with exon 1 and Element II, and may contain promoter elements. The 3' region of homology lies upstream of Element II, I refer to this 3' region as Region 1consA. None of the other exons are associated with the conserved regions identified, although exon 1b lies in close proximity to Region I2. Exons contained in different splice variants are linked by dashed lines. Mouse cDNAs initiate from two alternative first exons (exon 0b and exon 1b). A transcript containing exon 0 was not identified in this study. Two alternatively spliced products were obtained from 5'RACE reactions performed on chick mesoderm cDNA, which differentially incorporate exons c1i and c1ii.

Sequencing 5' RACE products isolated from HH stage 10 chick mesoderm revealed a splice variant containing two novel exons located between exons 1 and 2 (labelled exon c1i and exon c1ii in Fig. 4.6B and Fig. 4.7). On one occasion a transcript containing these two exons was shown to extend beyond exon 1, and splice to an unknown sequence (labelled exon X in Fig. 4.6). A sequence corresponding to exon X could not be identified in the chicken genome. This raises the possibility that exon X is located within the unsequenced region of the chicken genome in the CpG island upstream of Gli3, although comparison of the exon X sequence with genomic sequences of other organisms in this region did not show any apparent similarity. I have already postulated that this uncharacterised region may contain a promoter of chick Gli3.

On three occasions, the previously reported splice form containing exons 1 and 2 was identified in RACE products generated from chick mesoderm cDNA. In one case the corresponding transcript had the same TSS as that previously reported (Genbank accession no. XM_418866, Borycki *et al.,* 2000), whilst in another case the 5' end of chick exon 1 was shown to extend a further 14bp. None of the 5' RACE products generated from neural tissue produced sequences that could be associated with Gli3, precluding the analysis of whether different transcripts represent tissue specific isoforms. The results demonstrate that within the paraxial mesoderm of chicken embryos, splice products differ considerably from those observed in mammals. The most abundant transcript contains two exons that have not been described previously, and were not identified in mouse cDNA. Conversely, the exons corresponding to exon 0, and 0b were not identified in chick cDNA. Transcription of chick Gli3 appears to initiate at exon 1, which extends 39bp upstream of the published sequence and contains multiple TSSs. As for the mouse cDNA, none of the RACE products generated from chicken cDNA corresponded to the largest RACE products identified in Figure 4.5. However, on multiple occasions chick genomic DNA was cloned, suggesting that the large fragments may be an artefact of the RNA extraction technique.

I next investigated whether any of the alternative transcripts identified for chicken and mouse Gli3 will affect the protein sequence. None of the transcripts identified upstream of exon 1, either in mouse or in chicken cDNA contain a start codon, suggesting that they do not influence the protein product. However, alternative exons were also identified within intron 1 (mouse exon 1b and chick exon c1i and exon c1ii). 5'RACE from within mouse exon 1b did not identify any sequence 5' to that shown, suggesting that it might represent an alternative first exon. It is located slightly downstream of Region I2 identified in Chapter 3 (Fig. 4.7). A three-way translation of exon 1b is shown in Figure 4.8. Stop codons are encoded in all frames, which will cause termination of any transcripts initiating upstream (asterisks in Fig. 4.8). Only one

Methionine is encoded within exon 1b, and is followed by a stop codon eight residues downstream. Since translation normally initiates at the first start codon encountered in a transcript, it seems that the function of the novel exon might be to disrupt the production of a functional protein. The next in-frame Methionine is located sixty residues into the protein sequence. If the inclusion of exon 1b in the cDNA causes translation to initiate at this residue, part of the repression domain described by Dai *et al*. (2002) would be excluded from the protein product, and would result in a protein devoid of repressor activity.
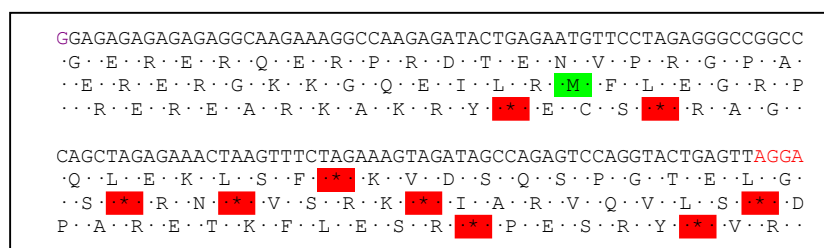
Figure 4.8: *Three-frame translation of mouse exon 1b.* Start and stop codons are highlighted. The sequence shown in red at the 3' end shows the first 4bp of exon 2.

Exons c1i and c1ii are located within intron 1 of chick Gli3 (Fig. 4.7). Since translation initiates in exon1, the inclusion of these exons will influence the coding sequence. A three-way translation is shown in Figure 4.9. Incorporation of exons c1i and c1ii into the transcript generates a novel open reading frame of 67 residues. The production of this protein may have a similar affect to that proposed for mouse exon 1b. Alternatively it may produce a functional protein. Interestingly the 5'end of exon 2 is flanked by two start codons in the open reading frame of Gli3. These may produce a truncated form of the protein from mRNAs that incorporate exons c1i and ii. Neither of the novel exons of chick Gli3 identified here occur within any of the conserved regions identified in Chapter 3 (Fig. 4.7).

```
GTATTAGTTCAAGGAGAGCTGAAGTAATGAGAAGTCACCATGGAGGCGCAGTCCCATAGC
·V··L··V··Q··G··E··L··K··*··*··E··V··T··M··E··A··Q··S··H··S·
·Y··*··F··K··E··S··*··S··N··E··K··S··P··W··R··R··S··P··I··A
··I··S··S··R··R··A··E··V··M··R··S··H··H··G··G··A··V··P··*·

TCTACCACGACAGAGAAGAAAAAAGTGGAGAATTCCATTGTGAAATGCTCCAATCGAACA
·S··T··T··T··E··K··K··K··V··E··N··S··I··V··K··C··S··N··R··T·
··L··P··R··Q··R··R··K··K··W··R··I··P··L··*··N··A··P··I··E··Q
L··Y··H··D··R··E··E··K··S··G··E··F··H··C··E··M··L··Q··S··N··

GATGTCAGTGAGAAAGCTGTTGCCTCCAGCACAACTTCCAATGTGAGCCAGCCCTACCCG
·D··V··S··E··K··A··V··A··S··S··T··T··S··N··V··S··Q··P··Y··P·
··M··S··V··R··K··L··L··P··P··A··Q··L··P··M··*··A··S··P··T··R
R··C··Q··*··E··S··C··C··L··Q··H··N··F··Q··C··E··P··A··L··P··

TACCCTACTCTAAAGCAAATGATCTGCAAGTGCGATATAAAACCTCAGCTTCTGCATACA
·Y··P··T··L··K··Q··M··I··C··K··C··D··I··K··P··Q··L··L··H··T·
··T··L··L··*··S··K··*··S··A··S··A··I··*··N··L··S··F··C··I··H
V··P··Y··S··K··A··N··D··L··Q··V··R··Y··K··T··S··A··S··A··Y··

TAGGAATACTTACTGCATGTGTGTGCCAGTAAGTGGAAGAAGCGTGCTTAGGGCATGAGG
·*··E··Y··L··L··H··V··C··A··S··K··W··K··K··R··A··*··G··M··R·
··R··N··T··Y··C··M··C··V··P··V··S··G··R··S··V··L··R··A··*··G
I··G··I··L··T··A··C··V··C··Q··*··V··E··E··A··C··L··G··H··E··

ATGAAAGTCCTGGACAGACCTATCACAGAGAGAGAAGGAACGCAATCACAATGCAGCCAC
·M··K··V··L··D··R··P··I··T··E··R··E··G··T··Q··S··Q··C··S··H·
··*··K··S··W··T··D··L··S··Q··R··E··K··E··R··N··H··N··A··A··T
D··E··S··P··G··Q··T··Y··H··R··E··R··R··N··A··I··T··M··Q··P··

AAGGTGGGCAAGGCCTCAGCAAGATCAGTGAGGAGCCTTCAACATCTAGCGAGGAAAGGGCC
·K··V··G··K··A··S··A··R··S··V··R··S··L··Q··H··L··A··R··K··G·
··R··W··A··R··P··Q··Q··D··Q··*··G··A··F··N··I··*··R··G··K··G·
Q··G··G··Q··G··L··S··K··I··S··E··E··P··S··T··S··S··E··R··A·
```

Figure 4.9: Three-frame translation of chicken transcript incorporating exons c1i and c1ii. Coloured text indicates exon number: Blue, exon 1; green, exon c1i; red, exon c1ii; black, exon 2. Start and stop codons are highlighted.

4.2.4.1 5'RACE on Human Gli3 cDNAs

Since conducting this study it has come to my attention that a similar study was performed to identify the 5' end of the human Gli3 transcript in placental tissue (Paparidis, 2005). Echoing my findings, the author identified a novel exon that aligns with murine exon 0b, and found that transcripts started at multiple positions within this exon. The longest transcript identified was cloned on two separate occasions and is indicated in Figure 4.2B (purple box). A primer extension assay was also used to establish the true 5' end of the human transcript, which in agreement with 5'RACE predicted a TSS at the 5' end of human exon 0b (Fig. 4.2B). A 300bp region upstream of the presumed TSS was used as a minimal promoter, and correlates with Element I, the strongest promoter prediction produced by Promoter Scan in my investigation (Fig. 4.2B; Fig. 4.4). A further transcription start site was identified upstream of exon 2 within a conserved region (CNE 2 in Fig. 3.6, corresponding to region I2), but promoter activity of this region was not investigated. Although not identified in my study, this TSS is located 568bp

upstream of exon 1b, and may represent the true 5' end of the associated transcript. Region I2 may therefore contain elements required to initiate transcription from an alternative exon.

4.3 Discussion

Investigation of the 5' end of the Gli3 locus has identified numerous potential transcriptional start sites. Two putative promoter regions were identified that are supported by multiple promoter prediction tools, each containing several potential transcription start sites. Consistent with this, recent findings revealed that transcription is rarely associated with a discrete initiation site. Transcription of a given gene is more likely to initiate from a number of sites within a broad promoter region (Bajic *et al.*, 2004). Indeed, most mouse and human protein-coding genes (>58%) are associated with more than one promoter (Carninci *et al.*, 2006; Sandelin *et al.*, 2007). This appears to be the case for Gli3, as data presented in this study demonstrate that initiation of Gli3 transcription from within the CpG island occurs from multiple transcriptional start sites, and can also initiate from a start site identified within the first intron. Interestingly, within the CpG island two clusters of promoter predictions were identified, encompassing exons 0 and 0b (Fig. 4.2, Elements I and II). These may be used differentially to produce tissue specific transcripts, or used at different stages of development. Together my 5'RACE data and EST analyses support that exon 0b defines the preferred 5' end of murine Gli3 at E9.5, although a second transcript initiating within intron 1 is also produced. Furthermore, the most 5' end I identified by 5'RACE on murine cDNA is located at the level of a binding site for Sp1, which is commonly associated with promoter regions of genes expressed during early development (Marin *et al.*, 1997).

4.3.1   A broad promoter is located within the CpG island

The majority of Gli3 transcripts appear to be under the control of a broad promoter located within the CpG island, which I refer to as Element I. The data presented by Paparidis and colleagues suggests that Element I can act as a functional promoter in vitro. However, mutation of transcription factor binding sites within the promoter did not influence reporter gene expression, perhaps due to the use of alternative transcriptional start sites within the 300bp construct (Paparidis, 2005). Recent evidence has demonstrated that broad promoters regulate the majority of mammalian transcripts, and are particularly associated with ubiquitously expressed genes, and those involved in early development, such as Gli3 (Bajic *et al.*, 2006; Carninci *et al.*, 2006). Within these promoters, TSSs recognised by RNA polymerase II are

generally loosely defined (Sandelin *et al.*, 2007). The existence of multiple redundant TSSs in broad promoter containing genes allows mutation of a particular site to be tolerated. It also facilitates adaptive evolution and allows the fine tuning of promoter activity (Carninci *et al.*, 2006; Sandelin *et al.*, 2007). In support of Element I functioning as a broad promoter, numerous transcriptional start sites were identified in the region, both by 5' RACE and EST analysis. Different TSSs within this exon might represent the preferred TSSs in different spatial and/or temporal contexts. However, the identification of a second promoter element (Element II) within the CpG island, that is associated with an exon, together with the identification of a further transcriptional start site within intron 1 suggests that the Gli3 locus contains multiple alternative promoters. These might also regulate the production of distinct transcripts in different cellular contexts.

It has recently been shown that use of a particular TSS is subject to species-specific adaptation, with TSS properties relative to promoter elements differing between human and mouse genomes (Bajic *et al.*, 2006; Carninci *et al.*, 2006). This might account for the differences in transcriptional start site identified by 5' RACE performed on chick and mouse cDNA.

4.3.2 Region 1 lies within the CpG island and may contain promoter activity

Figure 4.7 shows that Region 1, identified in Chapter 3 is located within the CpG island, and contains exon 0. This raises the question of whether Region 1 contains a promoter of Gli3. Since Region 1 lies downstream of the transcriptional start site identified by 5'RACE in E9.5 mouse embryos, it is unlikely to contain the promoter of Gli3 transcripts expressed at that stage. However, EST analysis identified a TSS located within Region 1. This EST was cloned from a cDNA library derived from RNA extracted from the mouse inner ear. Together with its location within the CpG island, this suggests that Region 1 may form part of a promoter associated with Gli3 expression at other embryonic stages of development or in adult tissues. It is likely to contain regulatory elements that govern the use of a TSS associated with exon 0.

Chapter 5

In-ovo screen of conserved regions

5.1 Introduction

To establish whether the conserved non-coding elements identified in Chapter 3 are involved in Gli3 regulation, their ability to drive reporter gene expression was assessed by in-ovo electroporation of the chick neural tube. In this Chapter I describe the initial screening of each region in this manner. I begin by outlying the advantages of this technique, particularly in the high-throughput analysis of putative enhancers, and describe the design of the reporter construct used.

5.1.1   Choice of model organism

Factors that determine the use of a model organism include generation times, proven experimental techniques, and maintenance costs. These factors have made Drosophila a commonly used tool for the investigation of enhancer activity. However, it is not useful for the investigation of vertebrate-specific genes such as Gli3. In vertebrates, the mouse has been the most widely used model organism, but studies are hindered by the long generation time, cost of upkeep, and difficulty of introducing exogenous DNA. The chicken embryo has recently become a widely used model. Its advantages as a model organism are the ease of access and manipulation of the embryo, low maintenance costs, and limited space/ equipment requirements. The chicken embryo shares developmental and structural similarities with the mouse embryo, and is thus preferable to other widely used model organisms such as zebrafish.

Transgenic studies have shown that the high degree of evolutionary conservation of enhancer elements allows the enhancer elements of one organism to be regulated by the transcriptional machinery of another. For example, Fugu Hox-4 and Wnt-1 enhancers drive reporter gene expression transgenic mice that recapitulate the expression patterns produced by their murine homologues (Aparicio *et al.*, 1995; Rowitch *et al.*, 1998 respectively). Such trans-species analyses reveal that the factors required to regulate spatially restricted gene expression have been conserved through vertebrate evolution (Westerfield *et al.*, 1992; Aparicio *et al.*, 1995). However, differences in transgene expression between homologous enhancers derived from different species are often observed in mouse transgenesis, which must be a consequence of sequence differences (Aparicio *et al.*, 1995; Rowitch *et al.*, 1998). In these cases the expression pattern observed mimics that observed in the species from which the enhancer is derived.

Timmer and colleagues recently used the high-throughput technology of in-ovo electroporation to study a series of characterised murine enhancers that generate restricted patterns of gene expression along the dorsal-ventral axis of the neural tube (Timmer *et al.*,

2001). They found that each of these enhancers drove developmentally appropriate gene expression in specific regions of the chicken neural tube. The advantage of studying mouse enhancers rather than chick is a twofold one: first the mouse genome is better characterised and more tools are available for its study, second it is more closely related to human, permitting a greater degree of the understanding of gene regulation in humans (Hedges, 2002).

In-ovo electroporation is a powerful technique, permitting efficient delivery of exogenous DNA to target tissues with limited consequences on embryo viability (non-toxic), an additional advantage is that non-electroporated tissue provides an internal control of expression levels (Muramatsu *et al.*, 1997; Itasaki *et al.*, 1999). For the analysis of gene regulation it offers a quick, high throughput, transient transgenic-like approach (Muramatsu *et al.*, 1997). Region-specific enhancers have been shown to generate expected patterns of gene expression following electroporation into chick embryos (Itasaki *et al.*, 1999; Timmer *et al.*, 2001), and co-electroporation of activator or repression constructs can be used to investigate transcriptional regulation. Aside from enhancer studies, electroporation can be used to investigate interactions between regulatory pathways, for fate mapping, and for transient knockdown studies by the use of RNAi vectors (Das *et al.*, 2006). This study aimed to investigate the regulation of Gli3 expression in the neural tube, which is particularly well suited to electroporation studies. The tubular structure allows DNA to propagate easily, and once injected it is held in the lumen until the current is delivered.

5.1.2 Endogenous expression patterns of Gli3

The expression pattern of Gli3 is conserved amongst vertebrates. During early embryogenesis, Gli3 is expressed in three main sites: the paraxial mesoderm, neural ectoderm and limb bud (Hui and Joyner, 1993; Marigo *et al.*, 1996; Marine *et al.*, 1997; Sasaki *et al.*, 1997; Borycki *et al.*, 2000; Meyer and Roelink, 2003). At later stages, expression is confined to the CNS, mesoderm derived structures and organs, such as the developing eye, gut, kidney, gonads and hindbrain (Schimmang *et al.*, 1992; Hui and Joyner, 1993; Hui *et al.*, 1994).

Subtle differences in Gli3 expression in the neural ectoderm are observed amongst species. In *Xenopus,* it is first expressed in the animal cap before Shh signalling, upon receipt of the Shh signal it is expressed throughout the neural plate, but is excluded from the midline. Highest levels of expression are observed at the lateral edges and in anterior regions. As development proceeds, Gli3 transcripts remains absent from the floorplate and become restricted to the dorsal neural tube, forming a gradient complementary to that of Shh

concentration (Lee *et al.*, 1997; Ruiz i Altaba, 1998). Expression remains low in the roofplate. In the mouse neural tube, Gli3 expression has not been observed prior to Shh signalling. Otherwise expression is similar to that of Xenopus: in posterior regions of E8.5-E8.75 embryos it is expressed weakly throughout the neural plate and is absent from the ventral midline, which lies above the Shh expressing notochord. In anterior regions, expression levels are higher and form a gradient complementing that of Gli1 (Lee *et al.*, 1997; Sasaki *et al.*, 1997). In the anterior neural tube of E9.5 embryos, expression is restricted to dorsal ventricular zone and adjacent cells. As in Xenopus, expression is low in the roofplate (Lee *et al.*, 1997). Similarly, in the avian embryo, Gli3 is first expressed throughout the primitive streak and neural plate. As the neural tube begins to close, expression becomes laterally restricted in the neural plate, and is thus restricted to the dorsal aspect of the neural tube. Unlike mouse and Xenopus, Gli3 expression in avians does not appear to be excluded from the roofplate (Borycki *et al.*, 2000; Meyer and Roelink, 2003). As development proceeds expression is refined to the dorsal ventricular zone (Borycki *et al.*, 2000).

During development, progenitor domains of the neural tube are known to be patterned in response to Shh signalling from the notochord, and BMP/Wnt signalling from the roofplate (Liem *et al.*, 1995; Chiang *et al.*, 1996; Ericson *et al.*, 1996; Arkell and Beddington, 1997; Ericson *et al.*, 1997a; Liem *et al.*, 1997). Interpretation of the Shh gradient relies on signal transduction by the Gli factors. As the main repressor of the Shh response, Gli3 expression is critical in participating in the steepness of the gradient. An understanding of the regulation of Gli3 expression in the neural tube during development will aid with our understanding of how the Shh gradient is interpreted, and of how important gradients interact to determine neural subtype identity.

5.2 Results

5.2.1 Gli3 expression profile during chicken development

I chose to assess the activity of the conserved elements identified in Chapter 3 by electroporation of the developing chick neural tube. If the elements possess enhancer activity, they are expected to direct a promoter to drive reporter gene expression in a manner that recapitulates aspects endogenous Gli3 expression. To serve as a basis for future comparisons, I began by carrying out a detailed analysis of Gli3 expression at the developmental stages that will be investigated in this study. To determine the transcriptional profile of Gli3 between HH stage 10-16, RNA expression was analysed by in-situ hybridisation analysis (Fig. 5.1).
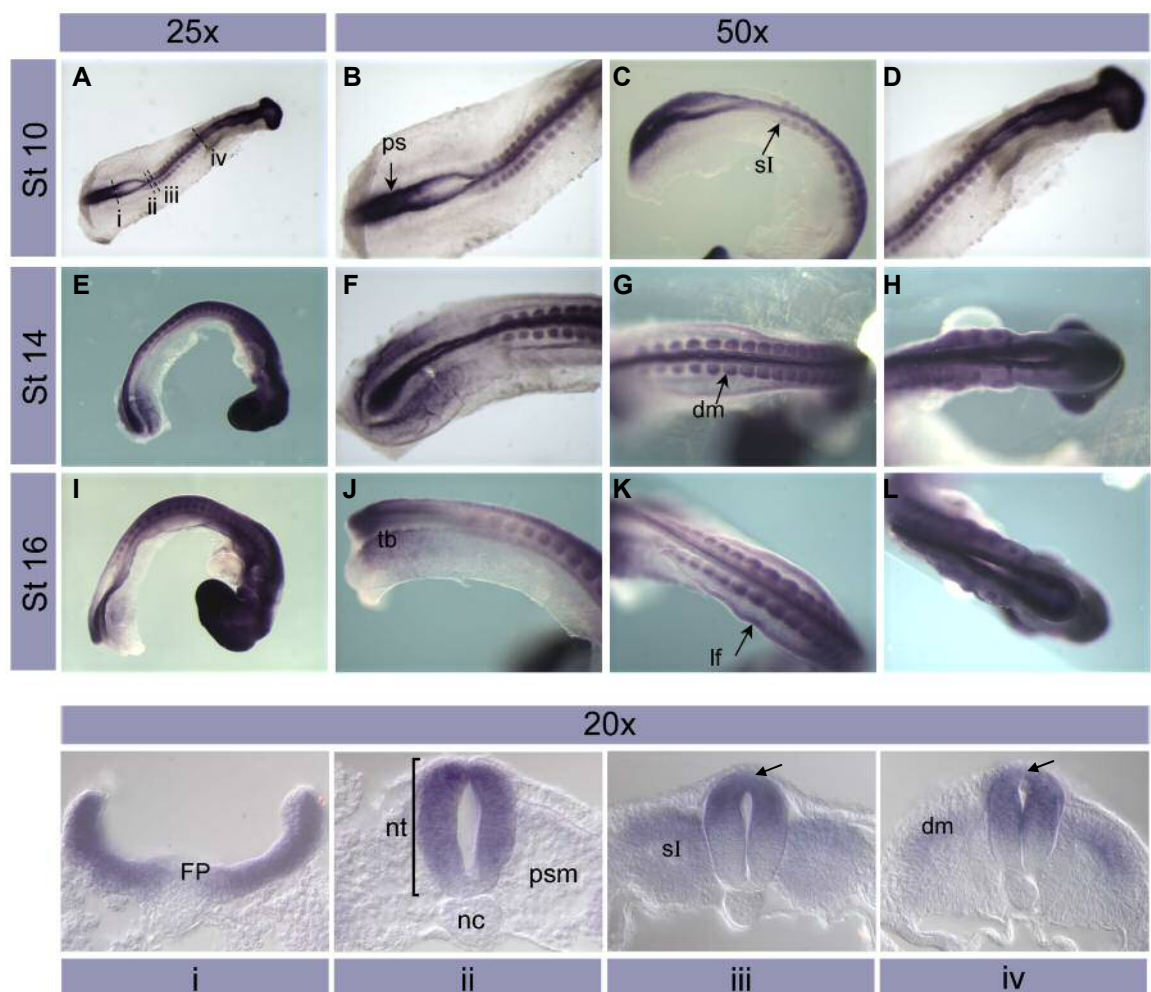
Figure 5.1: *Analysis by wholemount in-situ hybridisation of Gli3 expression in the chick embryo between HH stages 10 - 16.* A-L) wholemount images, anterior is to the right. i-iv) 80μm vibratome transverse sections of the HH stage 10 embryo shown. The position of transverse sections is indicated by dotted lines in panel A. Regions of interest referred to in the text are labelled, they are: primitive streak (ps), somite I (sI), dermomyotome (dm), tail bud (tb), limb field (lf), floor plate (FP), neural tube (nt), presomitic mesoderm (psm), notochord (nc).

Within the neural tube of the developing chicken embryo, Gli3 expression pattern displays some key characteristics along the anterior-posterior (AP) axis and the dorso-ventral (DV)

axis. Expression is initially observed in the primitive streak and throughout the neural plate, with the exception of the floor plate (Fig. 5.1 A, B, i). Upon neural tube closure expression is observed throughout the neural tube (Fig. 5.1 ii) but is later excluded from the ventral neural tube, forming a dorsal-high gradient of expression (Fig. 5.1 iii, iv). Expression becomes more dorsally restricted in older embryos as the neural tube matures (not shown). As the embryo develops, Gli3 expression is maintained in the tail-bud (Fig. 5.1 J).

In the paraxial mesoderm, Gli3 is not expressed in the presomitic mesoderm, and becomes activated in the first somite (sI, Fig. 5.1 C, F, ii, iii). As somites mature, Gli3 expression becomes restricted to the dermomyotome in the dorsal somite (Fig. 5.1 iv), and then to the dorsal dermomyotome. This can be observed on wholemount in-situ hybridisation at later stages (Fig. 5.1 G, K), and in anterior somites at HH stage 10 (Fig. 5.1 D, iv). From HH stage 16 onward, Gli3 expression can be observed in the limb field in the lateral mesoderm (Fig. 5.1 K). This expression pattern resembles that reported previously in avian embryos (Borycki *et al.*, 2000; Schweitzer *et al.*, 2000; Meyer and Roelink, 2003). Exclusion from the roof plate as reported in Xenopus and mouse studies is not immediately apparent, although expression in the most dorsal neural tube is reduced (Fig. 5.1 iii and iv; Lee *et al.*, 1997).

5.2.2   Designing a reporter construct

In order to investigate the function of the conserved regions identified in Chapter 3 in the regulation of Gli3 expression during development of the chick neural tube, a suitable reporter system was required. Factors that need to be considered in choosing a reporter construct are the sensitivity of the reporter gene, and ability to discriminate between basal levels of transcription and those influenced by enhancer activity. The success of a reporter system will be determined by the choice of reporter gene, and the promoter element used.

5.2.2.1   Choice of reporter gene

Together with GFP, the E.coli lacZ gene encoding β-Galactosidase is the most widely used reporter gene used in mouse transgenesis. Whereas GFP offers the advantage that expression levels can be directly visualised, lacZ detection requires a histo-chemical reaction in which β-Galactosidase catalyses the cleavage of a glycosidic linkage in a substrate protein. The most commonly used substrate for detection is X-gal (5-bromo-4-chloro-3-indolyl-β-D-galactosidase), which upon cleavage produces an insoluble, stable blue compound that clearly marks expressing tissues (Cepko *et al.*). Thus, once stained embryos expressing lacZ can be kept indefinitely, whereas fluorescence fades. Timmer *et al.* showed that under the control of a β-Globin promoter, the enhanced GFP (EGFP) reporter is more sensitive than lacZ for detecting expression in the chick neural tube, low levels of EGFP could be detected

where no lacZ staining was found (Timmer *et al.*, 2001). However, this problem can easily be overcome by using an anti β-Galactosidase antibody to enhance the sensitivity of detection. The use of lacZ as a reporter gene has the additional advantage that GFP can be used to assess the success of electroporation. This system has been adopted in my study.

5.2.2.2 Characterisation of basal promoters

Eukaryotic gene transcription requires the assembly of a pre-initiation complex on the core promoter of a gene. Furthermore, one method by which enhancer elements are thought to work is by moderating the activity promoter elements (Majumder *et al.*, 1997). Therefore, to assess the function of enhancer elements in a reporter system, a suitable promoter is required. Ideally, the endogenous promoter of a gene should be used, since it will give a more accurate readout of enhancer activity. However, in Chapter 4 I showed that the Gli3 locus contains multiple alternative promoters, and an individual promoter responsible for neural expression could not be determined. Therefore I resolved to the use of a constitutive promoter. Constitutive promoters may drive different levels of expression depending in the system they are used in. For instance, previous studies have shown that the HSP-68 promoter, often used in mouse transgenics, drives high levels of basal transcription such that when enhancers are linked to the promoter they have only marginal effects (Kamachi *et al.*, 2001; Timmer *et al.*, 2001; Muta *et al.*, 2002; Uchikawa *et al.*, 2004). I compared two promoters for the basal transcription levels they drive in in-ovo electroporation. The pSKTnlacZPa plasmid contains the constitutive promoter for thymidine kinase (Tk) driving β-Galactosidase expression (Hadchouel *et al.*, 2003), and P1230 carries the same reporter gene, driven by the human β-Globin promoter (Yee and Rigby, 1993). To determine the efficiency of electroporation, each plasmid was co-electroporated with the pMES plasmid, which carries a stabilised form of GFP (EGFP) under control of chicken β-actin promoter/CMV-IE (Y. Chen, 2004). This drives EGFP expression in the electroporated side of the neural tube, whilst no expression is observed on the contralateral side.

Electroporations were carried out at HH stages 10-11, and expression levels were analysed 24 hours later. pSKTnlacZPa produced high levels of lacZ expression in the electroporated side of the neural tube (Fig. 5.2 B, b'), indicating that in this assay the promoter produces a high basal activity. Therefore, the thymidine kinase promoter is inappropriate for expression studies in the chicken neural tube. However, its basal activity makes this promoter suitable for the analysis of conserved elements with suspected repressor function. Indeed, inhibition of basal levels of transcription driven by the Tk promoter in transfected cell-lines was used to demonstrate the repressor role of Gli3 (Marine *et al.*, 1997). The β-Globin promoter drives

lower levels of lacZ expression (Fig. 5.2 A, a') making it a suitable promoter for expression analyses in the chick neural tube, since up/down regulation will indicate the activity of both enhancer/repressor elements respectively. In addition, the β-Globin promoter functions well in transgenic mice (Yee and Rigby, 1993; Maconochie *et al.*, 1997).

5.2.2.3 Introduction of a NLS

To further enhance the specificity of the reporter assay, I subcloned the 3x NLS of the pCIG vector upstream and in frame of the β-Galactosidase gene in P1230 (using NcoI restriction sites) to create nP1230. Expression analysis demonstrated that nP1230 produces lower levels of reporter gene expression than the original plasmid P1230 (Fig. 5.2 C, c'), suggesting that in the absence of the NLS, staining leaches to surrounding cells. A map of the nP1230 reporter plasmid, showing the cloning site of enhancer elements is shown in Figure 5.3.
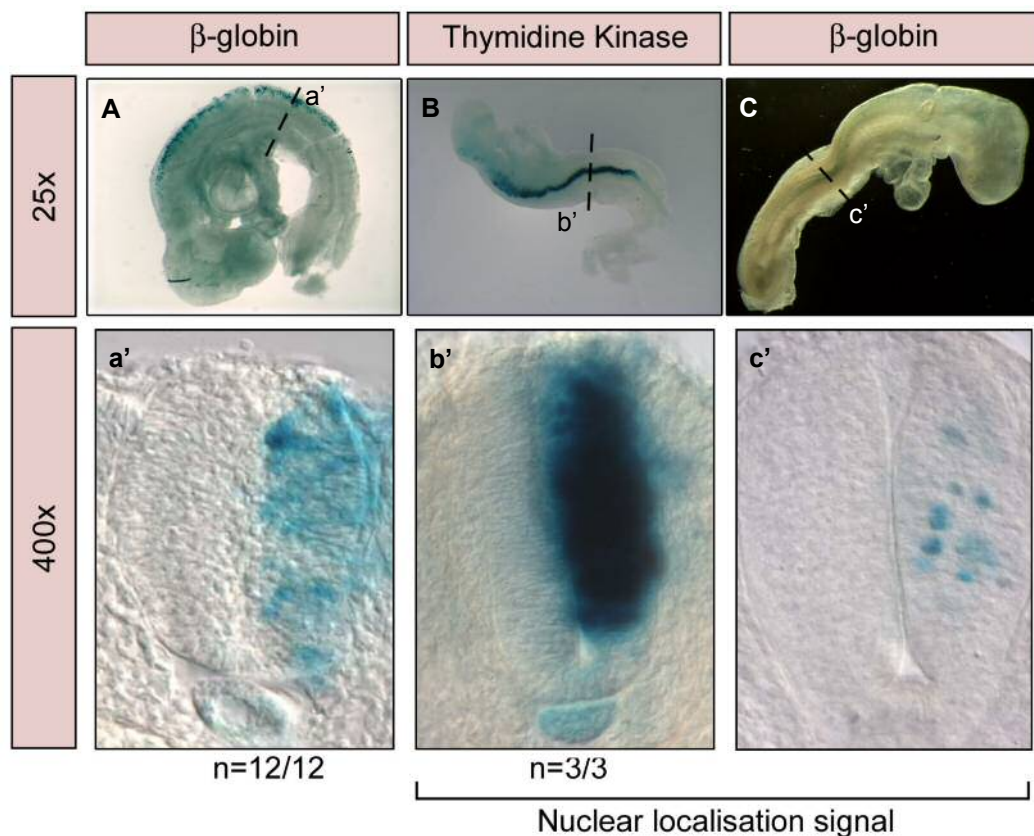
Figure 5.2: *Comparison of β-Galactosidase expression levels driven by Thymidine Kinase and β-Globin constitutive minimal promoters.* Embryos shown displayed similar levels of EGFP expression. A-C, wholemount embryo images following lacZ staining. a'-c', transverse sections taken at the positions indicated. The effect of addition of a nuclear localisation signal to the reporter protein is also shown.
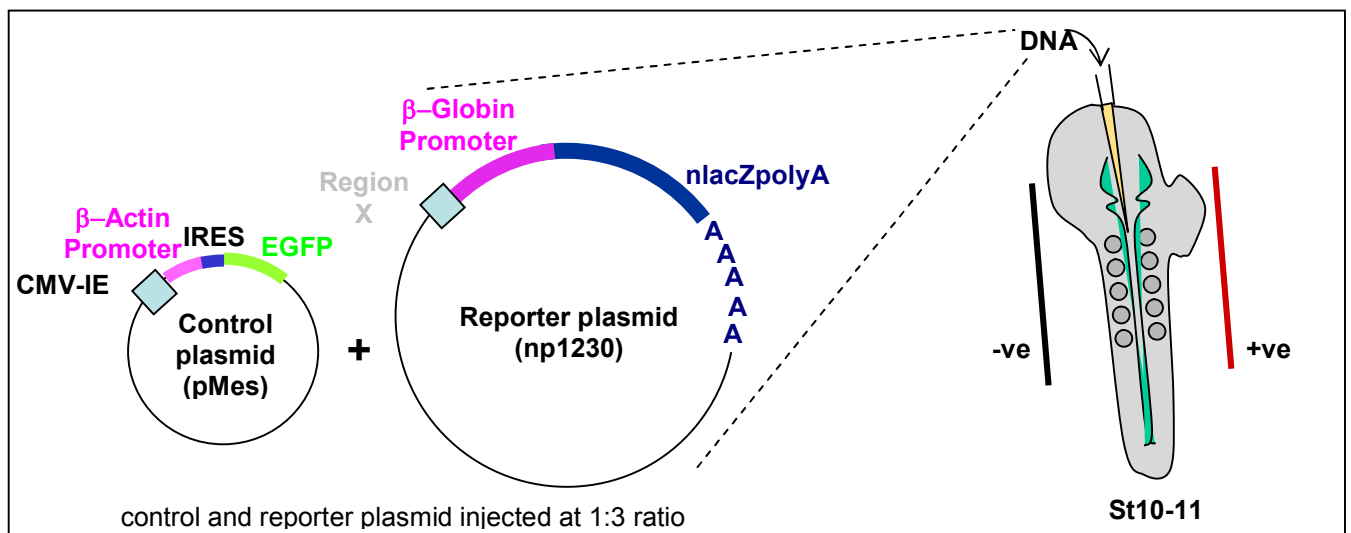
Figure 5.3: *Schematic diagram of the system used to investigate the function of putative cis-regulatory elements in-ovo.* Murine DNA of the conserved regions identified in Chapter 3 were ligated upstream of the β-Globin promoter in nP1230, at the position indicated by 'Region X'. The resulting plasmid was introduced to cells on one side of the neural tube (+ve) by electroporation of HH stage 10-11 chicken embryos. The contra-lateral side (-ve) was used as a control. A control plasmid (pMes) was co-electroporated at 1/3 the concentration of the reporter plasmid.

5.2.3 High throughput screen for enhancer activity

The murine sequence of each of the 10 conserved regions identified upstream of exon 1 in Chapter 3 (Fig. 3.4) were cloned upstream of the β-Globin promoter in nP1230 (Fig. 5.3). Region 1 consists of two regions of high conservation. The 3' portion was shown in Chapter 4 to contain mouse exon 0, and has not been studied in this chapter. Region1consA analysed here consists of the 5' region of homology shown in Table 3.1 and in Figure 4.7. Embryos were electroporated at HH stage 10-11, and left to develop for 12 hours (until HH stage 13-14) or 24 hours (HH stage 15-16). In order to visualise the success of electroporation, the reporter construct was co-electroporated with pMES at 3:1 ratio, such that the reporter plasmid is in excess. Live embryos that had developed normally were harvested, and immediately visualised by fluorescence microscopy. Embryos showing good levels of GFP expression were immediately fixed and processed for lacZ staining. Examples of the expression pattern observed for each construct by lacZ staining are shown in Figure 5.4.

Embryos were initially scored on whether they showed up regulation relative to the control plasmid in wholemount. Since some control embryos contain lacZ activity, only constructs showing consistent absence of expression were deemed to be down-regulating reporter gene activity. Results are shown in Table 5.1.

	12hr		24hr	
Construct	Number of embryos expressing the transgene/ Total number of embryos counted	Overall expression relative to control	Number of embryos expressing the transgene/ Total number of embryos counted	Overall expression relative to control
P1230	13/38	No expression	9/36	No expression
Reg1consA	45/53	Upregulation	11/13	Upregulation
Reg2	10/10	Upregulation	5/7	Upregulation
Reg3	1/6	Down regulation	0/8	Downregulation
Reg4	17/23	Upregulation	9/28	Upregulation
Reg5	10/14	Upregulation	6/21	Slight upregulation
Reg6	12/20	Upregulation	4/7	Slight upregulation
Reg7	5/9	Upregulation	3/7	Upregulation
Reg7b	6/8	Slight upregulation	1/9	Down-regulation
Reg8	13/34	Upregulation	8/21	Upregulation
Reg9	1/6	No effect	14/17	Slight upregulation

Table 5.1: *Effect of conserved regulatory elements on lacZ expression.* Embryos were scored 12 hours and 24 hours post-electroporation for lacZ expression levels. Down regulation is reported where lacZ expression was consistently absent.

5.2.3.1   Regions 4, 5, 6, 7, 7b and 9 have a temporal effect on reporter gene expression

As shown in Table 5.1, several of the conserved regions investigated display variable expression of the transgene between 12 and 24 hours post-electroporation. Regions 4, 5, 6, and 7b produced fewer embryos that express the transgene at the later timepoint. This is interesting considering that the expression of Gli3 within the neural tube is downregulated by the later stages investigated, becoming restricted to dorsal regions. Thus, these regions may be controlled by signalling components that are involved in the restriction of Gli3 expression in neuronal tissue as development proceeds. In contrast, constructs carrying Regions 7 and 9 drive higher levels of reporter gene expression at the later timepoint. This may be because the factors required for their activity are not present at earlier stages. Alternatively, Regions 7 and 9 might have a maintenance role in regulating transgene activity. In the system used here the increased expression observed at later stages could be due to the cumulative maintenance of basal activity. In support of a role in the maintenance of Gli3 expression, Regions 7 and 9 each contain clusters of conserved Smad binding sites that are conserved between mammals and avians (Fig. 3.4Aii). These regions might normally function in conjunction with other enhancer elements that can initiate transcription. Indeed, the absence of an element required for maintaining transcription could explain the decreased activity observed for other regions at the later timepoint.

5.2.3.2   Regions 3 and 7b possess repressor activity

Region 3 is the only conserved element that was found to consistently down-regulate reporter gene expression at both timepoints (Table 5.1; Fig. 5.4 D, O). Reporter gene expression was only driven above basal levels in one embryo harvested 12 hours post-electroporation (Table 5.1). Indeed, transverse sections of embryos electroporated with the Region 3 construct show no expression of the transgene (Fig. 5.4 Di-iii, Oi-iii). This expression profile suggests that Region 3 might contain a repressor activity.

Region 7b also displays repressor activity, but only at the later timepoint. At 12 hours post electroporation I observed a weak expression in the lateral neural tube, which was restricted to a caudal domain (n=6, Fig. 5.4 I). Transverse sections confirm that lateral expression occurs in the mantle zone (n=3, Fig. 5.4 Ii-iii). At 24 hours post-electroporation, reporter gene expression under the control of Region 7b was essentially nil, apart from one or two cells (Fig 5.4 T). These data suggest that Region 7b may control a transient expression at early stages of neural differentiation. This offers a possible mechanism for the initiation of Gli3 expression in newly differentiated cells, or cells in the process of cell cycle exit, which reside in the mantle zone (Leber and Sanes, 1995). This could be further tested by Immunofluorescence with a marker of early differentiation, such as NeuN (Mullen *et al.*, 1992). Initiation of Gli3 transcription has previously been associated with the Wnt pathway (Borycki *et al.*, 2000; Mullor *et al.*, 2001). However, no conserved Lef/Tcf binding sites were found in Region 7b (Fig. 3.4), suggesting that if the Wnt pathway is responsible for transcriptional initiation via Region 7b it acts indirectly. Further binding site analysis using MatInspector revealed two putative binding sites for fibronectin-wnt RE (xtcf), as well as a consensus tcf/sox binding site. A Msx2-fgf8 binding site, and two nkx2.5 binding sites were also found. Each of these binding sites could be investigated in turn to establish whether they influence the activity of the conserved element.

The low levels of expression of the nP1230 control vector made analysis of repressor activity in this system difficult. To further analyse repressor activity an alternative approach is required. One possibility is to fuse potential repressors upstream of the Tk promoter in pSTKnlacZPa, as suggested previously. To quantify expression levels an in vitro luciferase reporter assay could be used (Guo *et al.*, 1995), indeed this system has recently been adapted to quantify reporter gene expression in whole embryos (Das *et al.*, 2006). Investigation of repressor function was not pursued at this stage.

5.2.3.3 Most of the conserved elements identified upregulate reporter gene expression

As shown in Table 5.1 most of the conserved regions investigated drive upregulation of reporter gene expression. However, levels of upregulation and domains of reporter gene expression varied, and conserved regions were often found to drive expression preferentially in posterior domains. To investigate this further, elements driving upregulation of reporter gene expression were scored according to the levels and domains of expression observed at each timepoint (Table 5.2).

5.2.3.4 Several conserved elements drive increased expression in posterior domains

Embryos electroporated with reporter constructs carrying Region 1consA, 2, 5, 6, 7 or 7b displayed increased expression preferentially in the posterior (Fig. 5.4 B, C, F-I). For Regions 5 and 6 this expression profile is readily observed at both timepoints (Table 5.2), although few embryos express the transgene at older stages. This suggests a temporal effect whereby expression levels are reduced in mature regions of the neural tube, such that transgene expression is progressively restricted towards the posterior, and is eventually lost all together. A stronger temporal effect is observed for Region 1consA and Region 2 which initially show high levels of expression throughout the neural tube of young embryos, that becomes progressively restricted towarde the posterior as development proceeds (Table 5.2, Fig. 5.4 B, C, M, N). For Region 1consA, most of the youngest embryos harvested (HH stage 12-13), displayed strong expression at all axial levels (Fig 5.4B), with a small proportion displaying stronger expression in the anterior (Table 5.2, n=5/17). At slightly later developmental stages (HH stages 13-14) activity was highest in the posterior, forming a gradient of expression. In most cases this gradient persisted throughout the electroporated region. Vibratome sections of stained embryos show that in posterior regions where staining is strong, the reporter gene is expressed throughout the DV axis (Fig. 5.4 Bi-ii). As expression decreases along the AP gradient transverse sections show that expression is initially excluded from ventral regions (Fig. 5.4 Biii, n=4). At the 24 hour timepoint expression is further restricted, and staining persists only in the posterior neural tube (Table 5.2, Fig. 5.4M). Furthermore, expression in embryos harvested 24 hours post electroporation varies according to their developmental stage, the strongest expression is observed in embryos that developed slowly. Of the HH stage 16 embryos harvested, 2/7 lacked expression, and 3/7 displayed confined to the tailbud (Table 5.2).

Embryos electroporated with constructs carrying Region 7 displayed weak reporter gene expression 12 hours post electroporation, but strong posterior expression is observed in embryos harvested at the 24 hour timepoint (n=3/7, Table 5.2). I have already postulated that

Region 7 may have a maintenance role. Alternatively, Region 7 might be involved in a second wave of induction, which could account for some of the endogenous Gli3 expression observed in posterior neural tube at later developmental stages.

5.2.3.5   Region 4 upregulates reporter gene expression throughout the neural tube

Region 4 showed consistent expression throughout the A/P axis. At the 12 hour timepoint moderate-strong reporter gene expression was observed in most embryos (Table 5.2, n=14/23), and did not appear to vary with developmental stage. Vibratome sectioning showed that reporter gene expression is uniform throughout the DV axis (Fig. 5.4 Ei-iii, n=5). However, fewer embryos expressed the transgene at the 24 hour timepoint (9/28). Those embryos expressing the transgene at later stages displayed lower expression levels relative embryos harvested 12 hours post electroporation (Fig. 5.4 E, P).

5.2.3.6   Region 8 induces variable levels of transgene expression

Embryos electroporated with the Region 8 construct displayed variable levels of expression, ranging from very weak to strong staining at both timepoints (Table 5.2). Expression levels were consistently weaker at the 24 hour timepoint (19/21 embryos showed little or no expression). In embryos expressing the transgene the expression domain appears to move progressively from anterior to posterior as development proceeds. Approximately one third of the embryos expressing the transgene showed greatest expression in the posterior at the 24 hour timepoint (n=3/8). In younger embryos staining was reproducibly observed in lateral regions (Fig. 5.4 J i-iii, n=4 embryos). In posterior regions, two distinct domains of expression are observed (Fig. 5.4Ji double arrow). In more anterior sections, expression extends throughout the DV axis (Fig. 5.4Jii) later becoming restricted from the ventral side (Fig. 5.4Jiii).

		12hr				24hr			
Region	**Description**	**N**	**Further details**	**Total**		**Description**	**N**	**Further details**	**Total**
Np1230	No staining	25		38		No staining	27		36
	Few cells stained	6				Few cells stained	4		
	Moderate expression levels	7				Moderate expression levels	5		
1consA	Moderate staining, stronger in post	28	Stage 13-14	53		Moderate staining, post only	5	St 14-15	13
	Moderate-strong staining throughout	12	Stage 12-13			Low-mod tail bud only	3	St 16	
	Moderate-strong staining stronger in ant	5	Stage 12-13			Moderate staining	1	St 14	
	No expression	8	Stage 13			moderate staining, higher in ant	1	St 16	
						Very low levels throughout	1	St 16	
						No expression	2	St 16	
2	Moderate staining, stronger in post	5	Stage 12-13	10		Moderate-strong in post only	5	Staining extends more anteriorly in younger embryos	7
	Moderate staining in ant only	1	Stage 14			No expression	2		
	Moderate-strong expression throughout	4	Stage 13-14						
4	Very strong staining	2	No stage effect	23		Weak-moderate staining	8	No stage effect	28
	Moderate-strong staining	12				Moderate staining in gradient from post	1		
	weak staining	3				No staining	19		
	No staining	6							
5	Moderate staining, graded from post	6	Stage 13	14		Moderate staining, graded from post	1	Stage 15	21
	Very little staining, few cells in post	3	Stage 13-14			Very little staining, few cells in post	4	Stage 15	
	Low levels throughout	1	Stage 13			Low levels throughout	1	Stage 16	
	No expression	4	Stage 13-14			No expression	15	Stages 14-16	
6	Strong staining throughout	2	Stage 13	20		Weak staining in post	3	Stage 16	7
	Few cells expressing in post	1	Stage 14			Strong staining, higher in posterior	1	Stage 15	
	Strong staining, stronger in post	6	Stage 13-14			No expression	3	Stage 15-16	
	Low-moderate expression throughout	3	Stage 13-14						
	No expression	8	Stage 14						
7	Moderate staining in middle region along AP axis	3	Stage 13	9		Strong staining in posterior	3		7
	No expression	4	Stage13-14			No expression	4		
	Moderate expression in ant	2	Stage 13						
7b	Very little expression, low levels of staining in lateral posterior neural tube	6	Stage 13	8		No staining	8	Stage 15	9
	No expression	2				Few cells in tail bud only	1		
8	Low-moderate staining	8		34		Low expression	5		21
	Strong staining	3				Moderate expression in tail bud only	2		
	Stronger in posterior	4				Low expression, graded from post	1		
	None	18				No expression	13		
9	Strong lateral	1		6		Very little staining, few cells	14		17
	No expression	5				No expression	3		

Table 5.2: Characterisation of β-Galactosidase expression driven by conserved regions. N numbers are displayed for embryos expressing good levels of GFP throughout the electroporated side of the neural tube.

Figure 5.4: *Comparison of β-Galactosidase activity from plasmids carrying each of the upstream conserved regions identified in Chapter 3.* Conserved regions were cloned upstream of the β-Globin minimal promoter in the β-Galactosidase reporter plasmid nP1230. Embryos were co-electroporated with the reporter construct and a control GFP plasmid. Those displaying strong GFP expression were fixed 12 hours (A-K) or 24 hours (L-V) post electroporation and processed for lacZ staining. The number of embryos showing the displayed expression profile in wholemount is shown. Panels i-iii show 80μm transverse sections of the embryos shown in wholemount, taken at the positions indicated. Success of electroporation along the AP axis can be deduced from GFP expression levels. Green brackets are used to mark the GFP expression domain, whilst blue brackets indicate the extent of lacZ expression. Red arrows are used to indicate specific expression domains referred to in the text.

	GFP wholemount	β-gal wholemount		i	ii	iii
nP1230	A	25/38				
Region 1	B	28/53				
Region 2	C	5/10				
Region 3	D	5/6				
Region 4	E	14/23				
Region 5	F	9/14				
Region 6	G	7/20				
Region 7	H	3/9				
Region 7b	I	6/8				
Region 8	J	13/34				
Region 9	K	1/6				
		25x	50x	400x		

	GFP wholemount	β-gal wholemount		i	ii	iii
P1230	L	iii ii i 27/36				
Region 1	M	i ii 8/13				
Region 2	N	i ii iii 5/7				
Region 3	O	iii ii i 8/8				
Region 4	P	iii ii i 9/28				
Region 5	Q	iii ii i 5/21				
Region 6	R	iii ii i 4/7				
Region 7	S	ii iii i 3/7				
Region 7b	T	iii ii i 8/9				
Region 8	U	iii ii i 3/21				
Region 9	V	ii iii i 14/17				
	25x		50x	400x		

5.2.3.7 Immunofluorescence supports lacZ staining

To further analyse DV patterning of reporter gene expression, Regions 1consA, 4, and 8 were investigated by Immunofluorescence, using an antibody raised against β-Galactosidase. This allows direct comparison of β-Galactosidase and GFP expression levels (Fig. 5.5). 12 hours post-electroporation β-Galactosidase expression from the control plasmid (nP1230) is observed in a few cells (Fig. 5.5A). At the 24 hour timepoint no expression is observed (Fig. 5.5B). In embryos electroporated with the reporter construct carrying Region 4, high levels of reporter gene expression were observed throughout the DV axis in all GFP expressing domains 12 hours post electroportation (Fig. 5.5C; n=3). These data support the results obtained by lacZ staining. Consistent with the data I have reported in Figure 5.4 and in Table 5.2, Immunofluorescence analysis of Region1consA showed that at early stages it drives increased reporter gene expression throughout the electroporated domain (Fig. 5.5 D). As the neural tube matures, expression becomes dorsally restricted (Fig. 5.5Diii-iv, E). However, at the 24 hour timepoint β-Galactosidase expression persists, which is not expected in light of the results from lacZ staining (Fig. 5.5F). This apparent discrepancy could be due to different sensitivity between the two techniques. Consistent with this a previous report showed that Immunofluorescence is more sensitive than lacZ staining for the detection of β-Galactosidase expression (Timmer *et al.*, 2001).
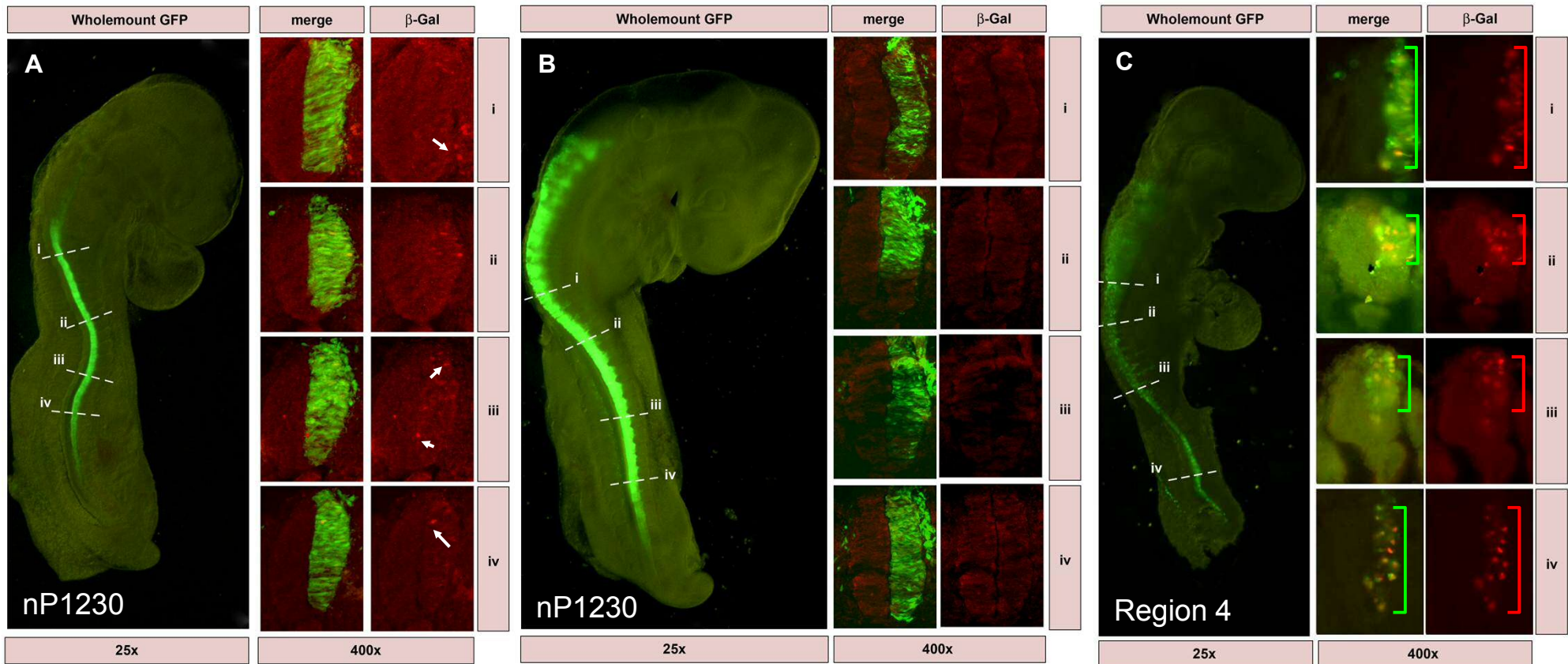
Figure 5.5 *Immunofluorescence detection of β-Galactosidase protein.* 15uM cryo-sections were taken at the position indicated on the embryo (i-iv). Sections were processed for Immunofluorescence using an antibody specific to β-Galactosidase (β-Gal; red), GFP expression shows the extent of successful electroporation (green). Green and red brackets are used to mark the expression domains of GFP and β-Galactosidase, respectively. A and B show embryos electroporated with the nP1230 control plasmid and harvested 12 hours or 24 hours post electroporation, respectively. At 12 hours post electroporation, few lacZ expressing cells are apparent in control embryos, and these are randomly distributed through the GFP positive region. 24 hours post electroporation, no expressing cell is observed. C shows an embryo electroporated with the Region 4 construct. Embryos electroporated with the Region 4 containing plasmid display high levels of reporter gene expression throughout the electroporated domain
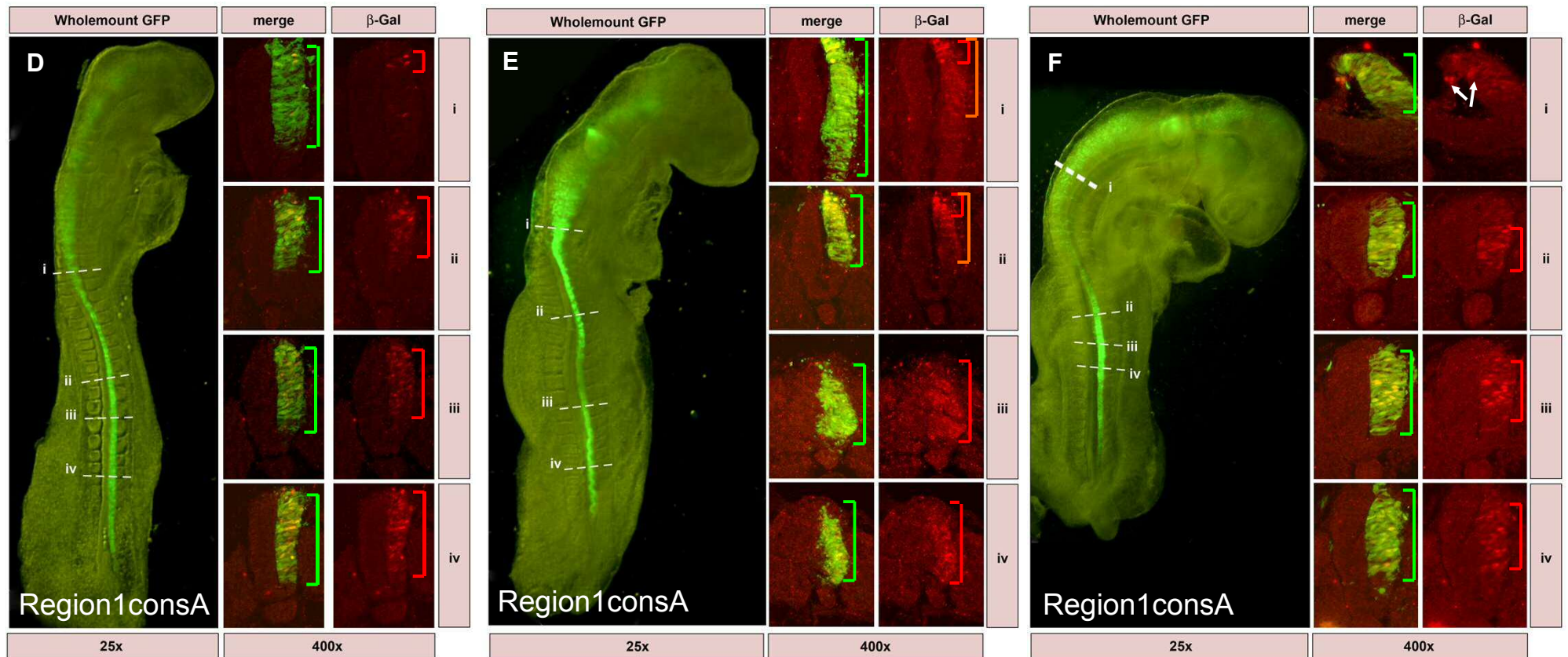
Figure 5.5 Cont... *Immunofluorescence detection of β-Galactosidase protein.* 15uM cryo-sections were taken at the position indicated on the embryo (i-iv). Sections were processed for Immunofluorescence using an antibody specific to β-Galactosidase (β-Gal; red), GFP expression shows the extent of successful electroporation (green). Green and red brackets are used to mark the expression domains of GFP and β-Galactosidase, respectively. D, E and F show embryos electroporated with the Region1consA construct and harvested 12 hours, 18 hours or 24 hours post electroporation, respectively. For the 18 hour embryo, orange brackets indicate the full extent of the lacZ expressing domain, which extends beyond a region of strong expression in the dorsal neural tube. In embryos electroporated with the Region1consaA plasmid, strong expression is observed throughout the posterior neural tube. In more anterior regions expression is progressively confined to a dorsal domain.
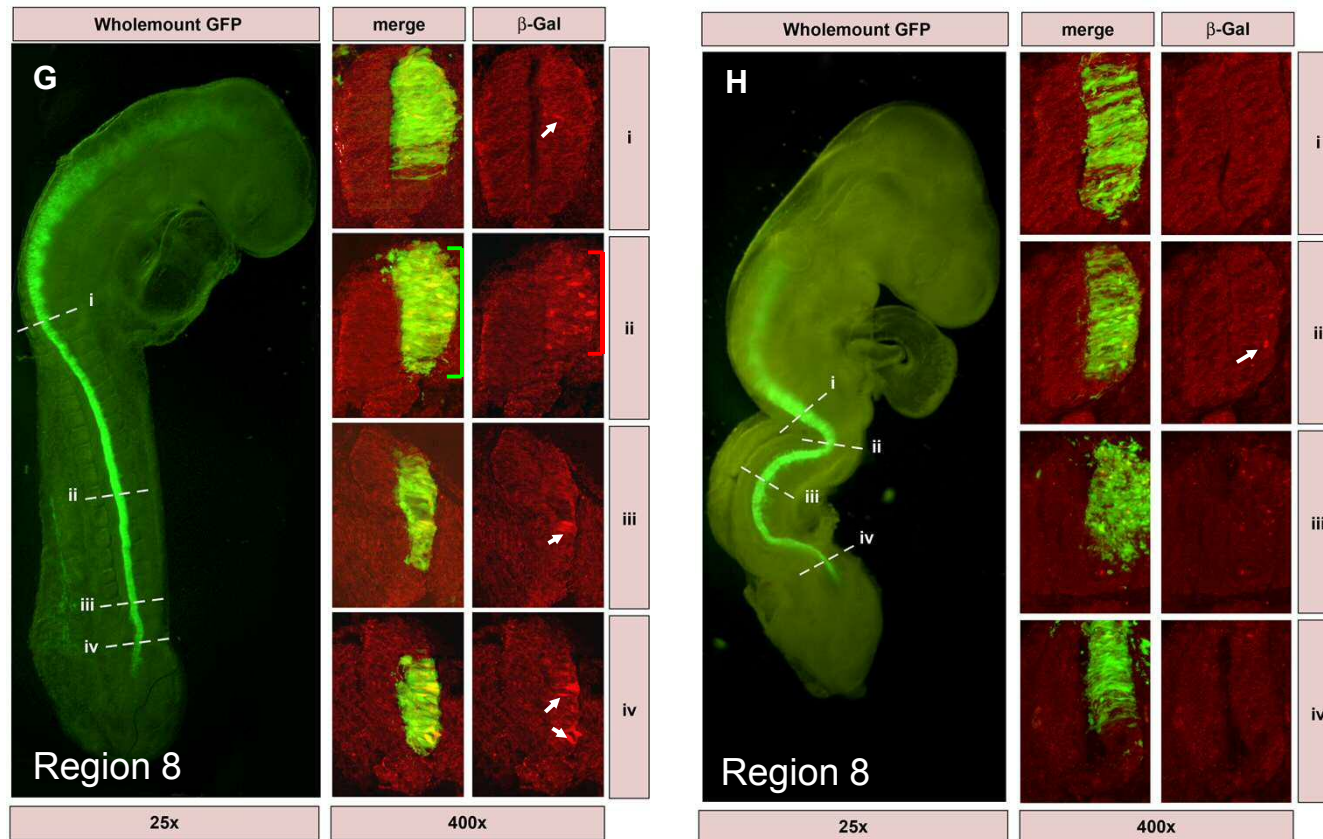
Figure 5.5 Cont... *Immunofluorescence detection of β-Galactosidase protein.* 15uM cryo-sections were taken at the position indicated on the embryo (i-iv). Sections were processed for Immunofluorescence using an antibody specific to β-Galactosidase (β-Gal; red), GFP expression shows the extent of successful electroporation (green). Green and red brackets are used to mark the expression domains of GFP and β-Galactosidase, respectively. F and G show embryos electroporated with the Region 8 construct and harvested 12 hours or 24 hours post electroporation, respectively. At 12 hours post electroporation, high levels of lacZ expression are observed in the caudal half of the neural tube, and lower levels of expression are observed in the rostral half. In the caudal-most section shown, two discrete domains of expression are observed (iv), at the level of the first epithelialised somite only one domain is observed (iii). Expression extends throughout the electroporated domain in intermediate regions of the neural tube (ii). 24 hours post electroporation, very few lacZ expressing cells are observed.

5.2.4   Detailed analysis of Region 8

Immunofluorescence of embryos electroporated with the Region 8 construct show a similar expression pattern to lacZ stained embryos (Fig. 5.5 G, H). The element drives high levels of expression in the caudal half of the neural tube (Fig 5.5 Gii-iv) and low levels of expression in the rostral half of the neural tube at HH stage 14 (Fig. 5.5 Gi). By HH stage 17, only very low levels of β-galactosidase are expressed (Fig. 5.5 H). Interestingly, in the posterior neural tube of HH stage 14 embryos two distinct domains of expression can be observed, (Fig. 5.5 Giv, arrows), which appear to correspond to the domains observed in lacZ staining (Figure 5.4 Ji, arrows). At the position of the first epithelialised somite, only the most dorsal domain expressed the transgene (Fig. 5.5 Giii, n=4). Together these observations suggest that Region 8 drives a highly regulated transient expression of the reporter gene.

To further characterise the domains in which Region 8 drives reporter gene expression in the posterior neural tube, double labelling with transcription factors against factors expressed in similar domains of the neural tube were performed (Fig. 5.6). Mnr2 is expressed in motor neuron (MN) progenitor cells and in newly differentiated motor neurons (Tanabe *et al.*, 1998), and broadly overlaps with β-Galactosidase expression in the posterior neural tube of embryos electroporated with Region 8 (Fig. 5.6A). The expression domain of Nkx6.1, which spans ventral progenitors that give rise to MN, V2 and V3 neurons (Sander *et al.*, 2000), further supports that the discrete expression driven by Region 8 in posterior regions lies within the motor-neuron domain (Fig. 5.6A). The lateral expression of β-Galactosidase in posterior regions suggests that Region 8 might have a role in the initiation of Gli3 expression in differentiating cells. Consistent with this possibility, the initial transgene expression is within the MN pool, which are amongst the first neurons to differentiate (Nornes and Carry, 1978). To test this, I performed double labeling with antibodies against En1, which labels V1 neurons (Ericson *et al.*, 1997b), and Isl1, which is expressed in differentiated motor neurons. However, as shown in Figure 5.6B the reporter gene is expressed prior to the onset of En1 or Isl1 expression, suggesting that the cells expressing the reporter gene in this posterior domain are in a progenitor state, or in an early stage of differentiation, rather than terminally differentiated neurons.
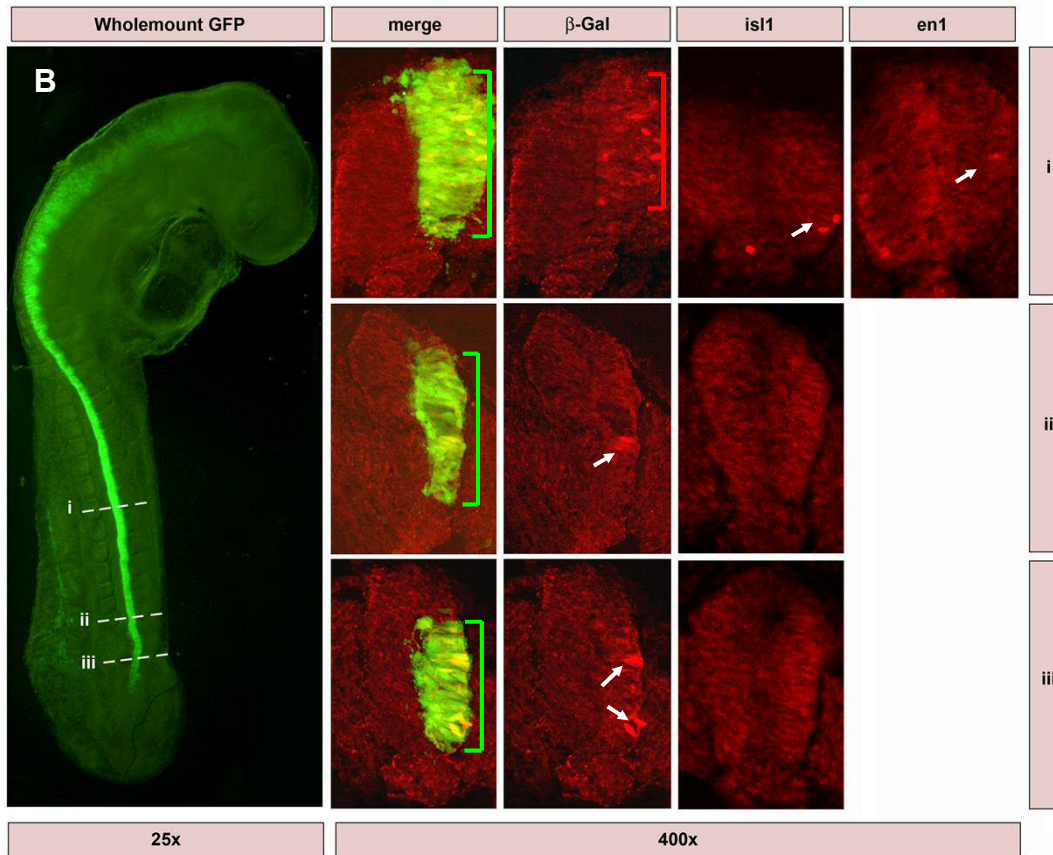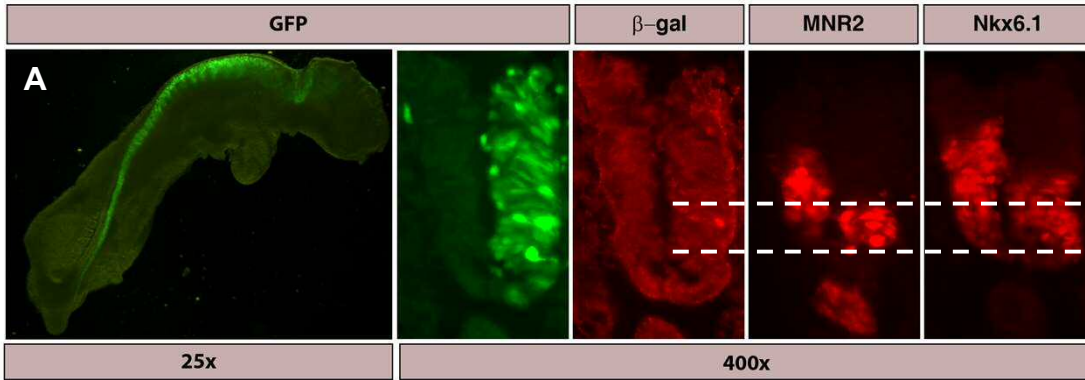
Figure 5.6 *Characterisation of β-Galactosidase expression domains 12hours post electroporation with Region8-nP1230.* 15uM sections at the positions indicated are shown. Serial sections were processed for immunofluorescence using antibodies against β-Galactosidase (β-Gal), Islet1 (Isl1), Engrailed1 (En1), Mnr1 and Nkx6.1. The distinct domain of reporter gene expression observed in the ventral most region of the posterior neural tube occurs in the motor neuron domain prior to Isl1 expression. The second domain appears to correlate in position with the V1 domain, but expression is only observed prior to differentiation, as indicated by the absence of En1 expression.

5.2.4.1 Binding site analysis

To further characterise the expression profiles of reporter gene expression for Region 8, I performed a Transfac search for transcription factor binding sites. In particular I focussed on sites for transcription factors mediating Shh, Wnt and BMP signalling, as they have previously been shown to control Gli3 expression, as well as transcription factors known to play a role in CNS development. In Chapter 3 a single conserved Smad binding site was identified in Region 8 (Fig. 3.4Aii). Further sequence analysis revealed a number of binding sites within Region 8 for transcription factors expressed within the neural tube, these are shown in Figure 5.7. Binding sites for En1 (expressed in V1 neurons) and Nkx2.5 (expressed in cardiac progenitors), a hox/Pbx site and six overlapping NRSF (Neuronal restrictive silencing factor; also known as REST) sites were identified. None of these binding sites are conserved between mammals and avians. Indeed, NRSF and En1 binding sites only occur in the murine sequence. Therefore Region 8 might drive murine specific expression patterns of the reporter gene. NRSF is a zinc finger transcription factor that represses neuronal genes in non-neuronal tissues (Schoenherr and Anderson, 1995a, 1995b; Chen *et al.*, 1998). It is expressed highly in the ventricular zone and is repressed in the mantle zone. Binding of NRSF to Region 8 would provide a possible mechanism for establishing the lateral domain of reporter gene expression driven by Region 8, through repression of transgene expression in medial cells.
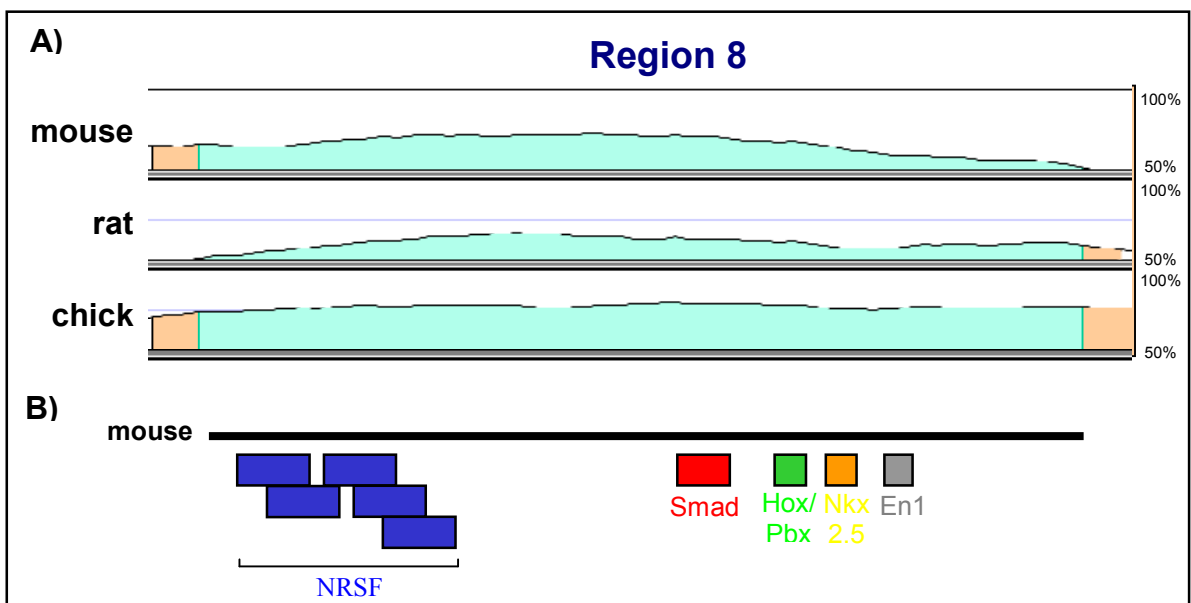
Figure 5.7: *Graphical representation of Region 8.* A) VISTA sequence comparison of mouse, rat and chick against human Region 8. The height of the graph indicates the level of conservation, the region shaded green represents the sequence analysed. B) The position of transcription factor binding sites in mouse Region 8. Coloured boxes represent binding sites for Smad (red), En1 (grey), Hox/Pbx (green), Nkx2.5 (orange) and NRSF (blue).

The Smad binding site identified by rVISTA in Region 8 is conserved in all the species I investigated. To investigate whether it is functional, embryos were co-electroporated with a plasmid driving the expression of a repressor Smad, Smad6 (pCAB Smad6; Linker and Stern, 2004). Smad6 is a potent antagonist of the BMP pathway that acts by blocking BMPR1, as well as competing with Smad4 to bind phosphorylated Smad 1/5/8 (Linker and Stern, 2004). Of four embryos harvested at 12 hours post-electroporation, three showed an upregulation of reporter gene expression relative to P1230 alone (Fig. 5.8Ai). At 24 hours post-electroporation, reporter gene expression was observed in half of the electroporated embryos (n=4/8). Of the embryos expressing the transgene, 3/4 showed an increased expression in the posterior neural tube (Fig. 5.8Aii), whilst one showed increased expression in the anterior where GFP expression was highest (not shown). Initial analysis of β-Galactosidase protein distribution by Immunofluorescence (Fig. 5.9) reveals that Smad6 does not have a strong effect on the activity of Region 8, although there is a slight down regulation of reporter gene expression in the anterior neural tube (Fig. 5.9 Ai-ii). Unfortunately, in the embryo shown in Figure 5.9A, GFP expression does not extend to the ventral neural tube, thus we cannot determine whether BMP signalling affects the ventral-most domain of reporter gene expression observed in the caudal neural tube of embryos electroporated with the Region 8 construct. However, at the axial level at which a single discrete domain of transgene expression is observed in embryos electroporated with Region 8 alone (Fig. 5.5Hiv), Smad6 expression causes reporter gene expression to become more widespread (Fig.5.9Aiv). Further analysis is undoubtably necessary, but this observation suggests that BMP signalling may be required to restrict the domain of Region 8 activity and could regulate transgene expression directly via the Smad binding site. This contrasts with previous reports implicating BMP signalling in the maintenance of Gli3 expression. BMP signalling opposes the ventralising effect of Shh signalling on neuronal patterning (Liem *et al.*, 2000). Upregulation of BMP signalling results in the expansion of dorsal cell types (specifically dl1 and dl3) (Timmer *et al.*, 2002; Chizhikov and Millen, 2004; Liu *et al.*, 2004), and down regulation of BMP signalling by RNAi knockdown of Smad4 results in the ventralisation of the dorsal neural tube (Chesnutt *et al.*, 2004). Therefore, by blocking BMP activity throughout the electroporated domain, dorsal cell types will adopt a more ventral fate, which could account for the expanded domain of reporter gene expression in the posterior neural tube. Immunofluorescence of Smad6 electroporated embryos with the markers used previously to define neuronal regions expressing the Region 8 transgene could be used to verify whether Smad6 electroporation causes ventralisation of the neural tube.
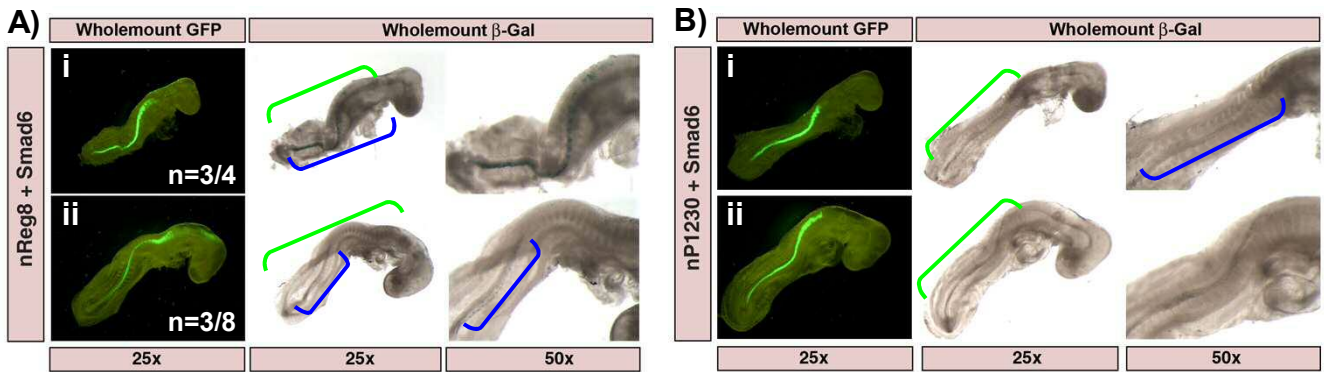
Figure 5.8, *Comparison of Region 8 and nP1230 transcriptional activities in the presence of Smad6.* Embryos were co-electroporated with the reporter construct, a Smad 6 expressing construct and the pMES GFP contruct at a 10:6:3 ratio. Panel A shows embryos electroporated with the Region 8 reporter construct, Panel B shows embryos electroporated with the nP1230 reporter construct. Embryos displaying strong GFP were fixed 12 hours (i) or 24 hours (ii) post electroporation and processed for lacZ staining. Brackets are used to indicate the GFP expressing domain (green) and the extent of lacZ expression (blue).
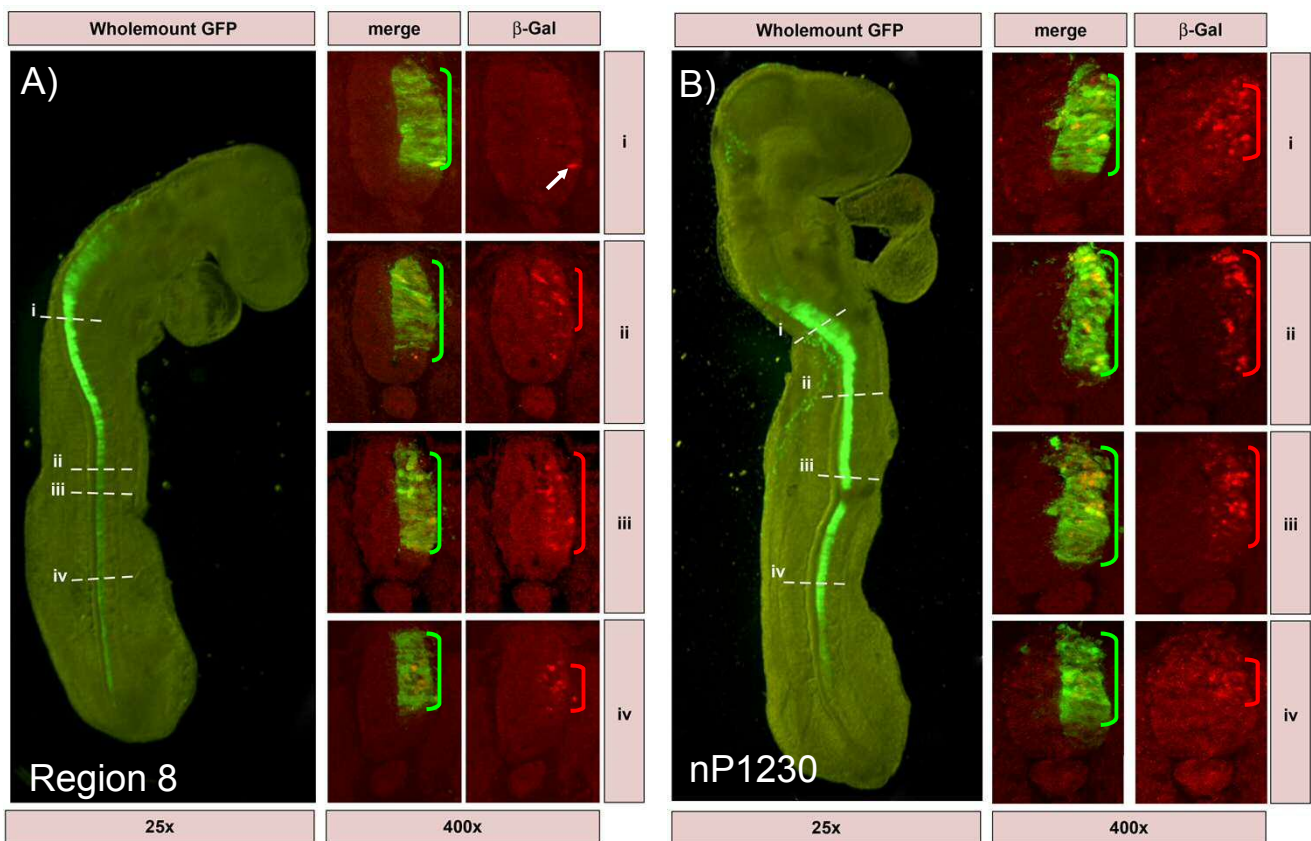
Figure 5.9, *Immunofluorescence imaging showing the effect of Smad 6 on β-Galactosidase expression.* Embryos were co-electroporated with the reporter construct, a Smad 6 expressing construct and the pMES GFP contruct at a 10:6:3 ratio. Panel A shows an embryo electroporated with the Region 8 reporter construct, Panel B shows an embryo electroporated with the nP1230 reporter construct. Embryos displaying strong GFP were fixed 12 hours post electroporation. 15uM sections at the positions indicated are shown. Sections were processed for immunofluorescence using an antibody specific to β-Galactosidase (β-Gal), GFP expression shows the extent of successful electroporation. In electroporated with reporter constructs carrying Region 8, β-Galactosidase expression is observed throughout the electroporated domain. Embryos co-electroporated with the control nP1230 plasmid and the Smad6 construct express higher levels of reporter gene expression than is observed in the absence of Smad6.

It should be noted that the Smad6 construct also appears to upregulate expression of the transgene from the control plasmid (Fig. 5.8B, 5.9B). At the 12 hour timepoint 3/5 embryos processed for lacZ staining expressed the reporter gene (1 embryo had strong expression throughout, one had low levels of expression in a central position along the AP axis, and one had low expression levels throughout-as shown in Fig. 5.8B). At the 24 hour timepoint 3 out of 4 embryos expressed the transgene, although two of these displayed exceptionally high levels of GFP expression. The constitutive upregulation of reporter gene expression might be due to a carrier effect of the DNA. Preparations used for co-electroporation of the reporter construct with the Smad6 expression plasmid contained a higher overall concentration of DNA, and thus will have a greater negative charge.

5.3 Discussion

In this chapter I describe the use of electroporation in the chick neural tube to perform high-throughput analysis of enhancer activity in the chicken embryo. Assessing the activity of a conserved region using this approach allows one to rapidly assess the activating potential of a region of interest, and to establish the spatio-temporal activity of an enhancer. The initial screen presented allowed me to identify several elements that consistently upregulate reporter gene expression in the chick neural tube. Although my approach is not optimised for the analysis of elements with maintenance or repressor activity, several elements were identified that may contain these functions. The use of co-electroporation approaches with activator and repressor forms of putative regulators, combined with co-Immunofluorescence studies, is a powerful method to quickly identify spatio-temporal domains of enhancer activity and putative transcription factors that regulate them.

5.3.1   Combined enhancer activity mimics Gli3 expression in the neural tube

Figure 5.10 shows a schematic representation of the transgene expression domains associated with each conserved region. The combination of these expression profiles would result in high levels of transcription in the posterior neural tube, which become progressively restricted as development proceeds. This is similar to the transcriptional profile of Gli3, which is initially expressed throughout the neural tube before becoming dorsally restricted, and suggests that the elements identified may regulate Gli3 expression in vivo. This is further supported by the profile of transgene expression along the DV axis, which for Region 1consA and Region 2 becomes progressively dorsalised as development proceeds.

Interestingly, Region 4 drives reporter gene expression throughout the AP and DV axis in younger, but not older embryos. This suggests that Region 4 might contain elements

responsible for the initiation of Gli3 expression. Analysis of transcription factor binding sites in Chapter 3 revealed several conserved Lef/Tcf binding site within Region 4 (Fig. 3.4Aii), consistent with the possibility that Region 4 controls Gli3 initiation via the Wnt pathway.

Region 7 also displays a strong activity in the posterior neural tube. However, in contrast to Region 4, this activity is strongest at later stages, suggesting that Region 7 could mediate a second wave of initiation, allowing Gli3 expression at stages where other regulatory elements are no longer active. Alternatively, Region 7 may have a role in the maintenance of transcriptional activity. Indeed, the expression pattern observed of Gli3 in the neural tube is likely to be the result of the combined activity of a number of regulatory elements acting in synchrony. This hypothesis could be tested by analysing reporter gene activity from constructs carrying various combinations of putative enhancer elements, to investigate whether they have an additive effect in recapitulating the endogenous expression pattern of Gli3. As an example, Region 4 and Region 7 could be combined in a reporter construct to establish whether Region 7 is able to maintain the strong transgene expression induced by Region 4.

Two putative repressor elements were also identified, which should be tested further in another system to establish whether they can mediate a strong repression of transgene expression. I have shown that the thymidine kinase promoter drives high basal levels of transcription in the chick neural tube, and offers a potential means to test the repressive activity these elements. Interestingly, the repressive effect of Region 7b was only observed at the later timepoint, indicating that it may have a role in repressing Gli3 expression as development proceeds.

Although the elements identified reproduce to some extent the endogenous profile of Gli3 expression, none of the reporter constructs drove reporter gene expression reproducibly in dorsal domains of the rostral neural tube. Consistent with this, Gli3 expression in the dorsal neural tube appears to be maintained by BMP signals, and by analysing putative enhancers in isolation elements required for this maintenance may have been missed in my study. Alternatively, the regulatory elements responsible for Gli3 expression in the dorsal neural tube may be located elsewhere.
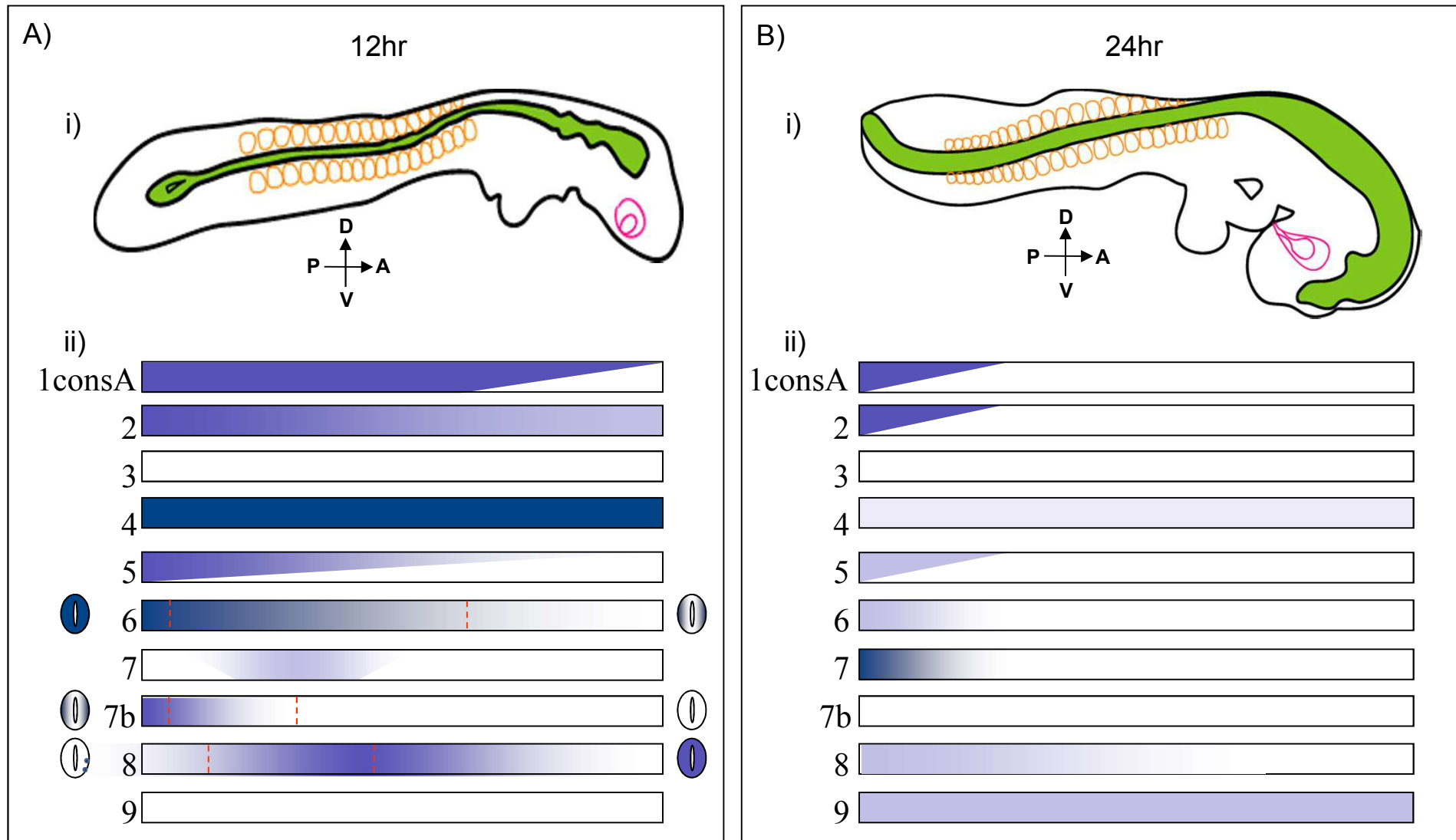
Figure 5.10: *Schematic representation of reporter gene expression patterns in the chick neural tube.* Embryos were electroporated with a reporter construct carrying a lacZ reporter gene fused to each of the conserved elements. β-Galactosidase activity 12 hours (A) and 24 hours (B) post-electroporation is summarised. i) Cartoon showing embryo morphology at each timepoint. Somites are shown in orange, the developing eye is pink, and the neural tube is shown in green. A= anterior, P= posterior, D= dorsal, V= ventral. ii) Expression patterns are indicated in the form of a block which corresponds to a saggital cross section through the neural tube. Darker shading is used to represent higher levels of β-Galactosidase activity. For regions 6, 7b and 8, expression patterns in transverse sections are also shown for the 12 hour timepoint, representing representing the expression pattern observed at the positions indicated by red lines. Combined reporter gene activity is strongest at the earlier timepoint. At the 24 hour timepoint, combined reporter gene activity activity is greatest in the posterior, although Region 4 and Region 9 drive low levels of expression throughout the neural tube

5.3.2 Comparison with previous studies of Gli3 regulatory elements

Two other groups have also studied the regulatory potential of conserved elements in the Gli3 locus, combining in vivo and in vitro data. Table 5.3 shows a comparison of the conserved regions investigated in these studies.

Abbasi *et al.*, (2007); Paparidis *et al.* (2007)				This study	Alvarez-Medina *et al.* (2008)	
Reporter gene expression			Name	Name	Name	Reporter gene expression
H661 cells (Gli3⁺)	H441 cells (Gli3⁻)	Transfected zebrafish embryos				Electroporated chicken neural tubes
-	-	-	-	4	HCNR1	Slight upregulation
Activator	Repressor	Not reported	CNE 12	5		
-	-	-	-	6		
-	-	-	-	7		
Activator	Repressor	CNS (forebrain, midbrain, hindbrain), cardiac chambers, blood cells, skin, median fin fold	CNE 1	I6a	HCNR2	Strong upregulation, particularly in dorsal regions
				I6b		
-	-	-	-	I5		
Activator	No effect	CNS (mainly forebrain), dorsal fin	CNE2	I2	HCNR3	Strong activation
-	-	-	-	I1		
Repressor	Repressor	No expression	CNE3	N/A	HCNR4	Upregulation throughout DV axis

Table 5.3: *Comparison of reported transcriptional activity of conserved regions surrounding the Gli3 locus.* The elements corresponding to a similar genomic region are shown side by side. Tissues and cell lines shown in red do not endogenously express Gli3 (Abbasi *et al.*, 2007; Paparidis *et al.*, 2007; Alvarez-Medina *et al.*, 2008).

Screening potential regulatory elements in the zebrafish embryo resulted in expression patterns that were highly mosaic, as is characteristic of the reporter system used (Muller *et al.*, 1997; Abbasi *et al.*, 2007). Conserved regions were PCR amplified and injected into embryos at the 1-8 cell stage, along with a PCR amplified promoter element. Concatemerisation of the DNA molecules ensures that the elements are in sufficient proximity to produce tissue-specific expression. However, the copy number and integration site will affect promoter activity (Muller *et al.*, 1997). Although some conserved elements were reported to drive reporter gene expression in tissues that express Gli3, such as the CNS, eye and muscle fibres, expression was also observed in numerous sites in which Gli3 expression is absent, such as those shown in red in Table 5.3 (Tyurina *et al.*, 2005; Abbasi *et al.*, 2007). Indeed, the strongest pattern of expression observed was directed to the notochord, which does not endogenously express Gli3. This suggests that Gli3 expression in these sites may normally be repressed by other enhancer elements. Indeed, several of the elements investigated were shown to repress reporter gene activity in cell lines that do not express Gli3 (Table 5.3). Therefore this data supports the hypothesis that Gli3 expression is regulated by multiple regulatory elements. Interestingly, none of the elements investigated

by Abbasi *et al.* reproducibly drives reporter gene expression in the zebrafish neural tube, suggesting that enhancer elements required for this aspect of Gli3 expression were not identified by human:fugu alignments (Abbasi *et al.*, 2007). Furthermore, none of the elements I identified upstream of exon 1 have been investigated in the zebrafish embryo. Thus some of the elements I have investigated may be specifically involved in neural expression of Gli3. Importantly, the technique used by Abbasi and colleagues has previously been used to identify enhancer elements with specific activity in the neural tube. Thus, the lack of neural expression is unlikely to be an artefact of the experimental approach used (Muller *et al.*, 1997). In a more recent study, conservation parameters similar to those used by Abbasi and colleagues were used to define a set of conserved elements in the chicken genome (Alvarez-Medina *et al.*, 2008). The activity of these elements was assessed by their ability to regulate EGFP expression levels from a reporter construct electroporated into the neural tube of HH stage 12 chicken embryos. In contrast to the zebrafish study, several elements were identified that drive strong reporter gene expression in the neural tube (Table 5.3). However, it is important to note that this study utilised the thymidine-kinase promoter, which in my hands drives high levels of reporter gene expression even in the absence of an enhancer element (Fig.5.2). Therefore the transgene expression observed in-ovo may be the result of repressing high basal levels of transcription. Indeed, this is supported by the repressive activity of the homologous human elements in cell lines that do not express Gli3 (Table 5.3), as well as the absence of activity in the zebrafish neural tube. Furthermore, Paparidis showed that transgenic mice carrying a CNE3 reporter did not express the transgene (15 embryos carrying the construct were analysed), whereas the corresponding chick sequence was reported by Alvarez-Medina and colleagues to upregulate reporter gene expression in the chick neural tube (Paparidis, 2005; Abbasi *et al.*, 2007; Alvarez-Medina *et al.*, 2008).

There are several important differences between the two studies that could account for differences in the activity reported for conserved elements: First, Abbasi and colleagues investigated the activity of human enhancers, whilst Alverez-Medina and colleagues investigated the activity of chick elements. Differences in reporter gene expression could be due to sequence divergence. Alternatively, the transcriptional machinery required to regulate transgene expression via the sequence elements investigated could have diverged between zebrafish and mammals, such that mammalian enhancers are unable to function in the zebrafish. However, tightly conserved expression patterns of Gli3 amongst vertebrates suggests that the core regulatory elements required for Gli3 expression are conserved. The majority of regulatory elements are therefore expected to have similar functions in different species. Second, Abbasi and colleagues used shorter regions of homology than Alvarez-

Medina and colleagues (Table 5.4). Unfortunately, sequence information was not provided in the Alvarez-Medina *et al*. publication. However, by analysis of a figure presented in the publication I have determined roughly how the regions identified relate to those investigated by Abbasi and colleagues, and in this thesis (Fig. 3.6, Table 5.3). Comparison of the two studies with my own data can provide an insight into the combinatorial effects of different conserved elements on transcriptional activity.

5.3.2.1 Regions 4, 5, 6 and 7 have been investigated in previous studies

CNE12 studied by Paparidis *et al*. corresponds to Region 5, although sequence limits vary. In vitro analysis demonstrated that CNE12 drives a slight upregulation of reporter gene expression in cells that endogenously express Gli3, and causes a strong down-regulation of reporter gene activity in cells that do not express Gli3 (Paparidis *et al*., 2007). This suggests that Region 5 has a repressor activity in tissues that do not endogenously express Gli3. Indeed, the work presented here suggests that Region 5 does not drive reporter gene expression in domains of the neural tube that are not associated with Gli3 expression, whilst expression is upregulated in Gli3-positive domains.

HCNR1, studied by Alvarez-Medina *et al,* encompasses Regions 4-7. Electroporation of this construct resulted in a slight upregulation of reporter gene expression in the dorsal aspect of the neural tube. Interestingly, transcriptional activity was increased by co-electroporation with constitutively active β-catenin (Alvarez-Medina *et al.*, 2008), which is proposed to act directly via three conserved Tcf binding sites, all within Region 5. I have demonstrated that Tcf binding sites are also present in Regions 4, 6 and 7, although only Region 4 contains a binding site that is conserved between chicken and mammalian genomes (Table 3.2). Of particular interest, I have shown that Region 4 directs strong reporter gene expression at early stages, as is expected if Wnt signalling initiates transcription. My study has also demonstrated a potential maintenance role of Region 7 that could account for the specific activity of HCNR1 in the dorsal neural tube. Thus my results suggest that the activity of HCNR1 is the result of the combined action of a number of discrete regulatory elements and demonstrate the importance of distinguishing between individual regions of conservation. Enhancer elements present in Region 4 might be responsible for increasing transcription in the neural tube at early stages of development, which is maintained by elements in Region 7, and repressed by those in Region 5 to give the dorsal-high gradient observed when these elements are combined. Unfortunately, differences in the activity of HCNR1 at different axial levels, and at different stages of development have not been reported, whereas my study demonstrates that transgene activity varies considerably along the AP axis. Further analysis of the combined

activity of putative enhancers is necessary. Analyses such as these will determine the importance of individual enhancer elements in the regulation of Gli3 expression.

5.3.2.2 Short enhancer regions have been missed in previous studies

Many of the conserved elements investigated in my study were not found in previous studies of Gli3 regulation. These include elements that appear to possess repressor activity (Regions 3 and 7b) and regions that upregulate reporter gene expression (Regions 1consA, 2 and 8). They were not identified in previous studies because they do not meet the stringent selection criteria of >50% identity over 60bp between human and fugu. Indeed, Figure 3.4A shows that these elements are not conserved in Xenopus. However, the selection criteria used here (>60% identity over 100bp between human and chicken alignments) have allowed identification of additional elements that influence reporter gene expression. Furthermore, Regions 1consA, 2 and 3 provide the most consistent expression patterns observed, and Regions 1consA and 2 mimic Gli3 expression in posterior regions of the chick neural tube. It will be interesting to determine whether these elements produce avian/mammal specific patterns of reporter gene activity. This could be determined by assessing each elements activity in zebrafish embryos, using a similar approach to that adopted by Abbasi and colleagues (Abbasi *et al.*, 2007).

5.3.3 Location of enhancer elements

The other studies of Gli3 transcriptional regulation have found conserved elements throughout the genomic locus that contain enhancer activity (Paparidis, 2005; Abbasi *et al.*, 2007; Paparidis *et al.*, 2007; Alvarez-Medina *et al.*, 2008). Furthermore, I showed in Chapter 3 that several regions of high conservation occur upstream of those I have studied. Thus I have only assessed a small proportion of regulatory elements that might influence Gli3 expression in the developing CNS. Indeed, Alvarez-Medina and colleagues (2008) show that elements contained in other intronic regions also influence reporter gene expression in the neural tube. Further analysis is required to establish the full repertoire of enhancer elements that regulate Gli3 expression in the neural tube. The high throughput analysis described here offers a potential approach.

Although I only identified a small proportion of potential Gli3 regulatory sequences, a number of those investigated regulate reporter gene expression within the neural tube. Consistent with this, enhancer elements associated with expression in a particular tissue are often clustered. At the Pax6 locus, upstream elements drive expression in the lens, pancreas and parts of the neural tube, intronic regions drive expression in the retina, forebrain and hindbrain, and 3' regions direct expression to the developing pretectum, retina, and olfactory

regions (Schwarz *et al.*, 2000; Griffin *et al.*, 2002; Kleinjan *et al.*, 2004). At the Sox2 locus elements required for neural expression are clustered downstream of the coding sequence. Hence it is possible that the 15kb region analysed here is specifically associated with regulating neural expression of Gli3. This is further supported by the study of Abbasi and colleagues, that failed to identify neural enhancers. The neural tube is a well characterised site of Gli3 expression both in zebrafish, and in higher vertebrates (Lee *et al.*, 1997; Sasaki *et al.*, 1997; Borycki *et al.*, 2000; Tyurina *et al.*, 2005). This suggests that the elements required for Gli3 expression in the neural tube were not investigated by Abbasi and colleagues. Interestingly, they did not investigate in vivo any of the conserved regions investigated in my screen (Abbasi *et al.*, 2007). This suggests that elements within the locus I studied might be specifically required for Gli3 expression in the neural tube.

5.3.4   Adapting the electroporation approach for other tissues

Electroporation of the chick neural tube is a useful approach for screening conserved regions for elements that regulate Gli3 activity. In the neural tube Gli3 is expressed in a temporally and spatially specific manner, and the three pathways previously implicated in the regulation of Gli3 are known to affect neural patterning. However, it should be noted that the elements I have studied might contain enhancer functions in other tissues, or at a different stage in development. The technique used here could be readily adapted to screen for enhancer activity in the limb bud, and in the somites, both of which are well-characterised sites of Gli3 expression (Lee *et al.*, 1997; Sasaki *et al.*, 1997; Borycki *et al.*, 2000; Schweitzer *et al.*, 2000). Additionally, electroporation at various different stages of development could be used to further establish the temporal effects on enhancer activity. Analysis of enhancer activity in other tissues, and in later stages of development could also be carried out using mouse transgenesis.

Chapter 6

Detailed analysis of Region 1

6.1 Introduction

Until HH stage 13, Region1consA reporter gene expression is observed throughout the neural tube (Fig. 6.1, A-F). As development proceeds, expression becomes progressively restricted to the posterior (Fig. 6.1, G-L). Furthermore, transverse sections show that whereas reporter gene expression is observed along the entire DV axis in the nascent neural tube, dorsal restriction progressively occurs as the neural tube matures (Fig. 6.1, a-l). This expression pattern overlaps in part with that of Gli3 in the neural tube, raising the possibility that Region1consA may control aspects of neuronal Gli3 expression. To investigate this hypothesis, I analysed in detail Region1consA activity. Here I report on the transcription factor binding sites present in Region1consA. Deletion analysis combined with neuronal expression studies reveals a role for TALE family proteins in regulating Region1consA transcriptional activity. Furthermore, I show that Pbx/Meinox binding sites within Region1consA are occupied in vitro, indicating that TALE family proteins may directly control Gli3 expression in the neural tube. In-situ analysis shows that transcription factors belonging to the TALE family are differentially expressed in tissues that express Gli3. I propose that TALE family transcription factors are key regulators of Gli3 expression in the developing embryo.
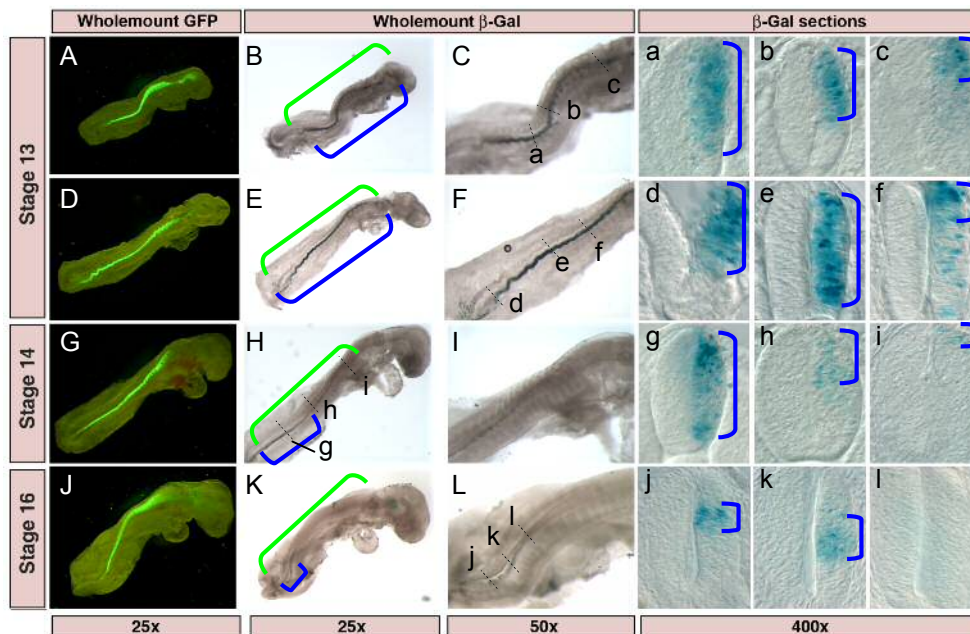
Figure 6.1: *Region1consA transcriptional activity in the neural tube during embryonic development.* Embryos were co-electroporated with the reporter construct and the pMES control plasmid in a 3:1 ratio. Embryos displaying strong GFP expression were fixed 12-24 hours post electroporation and processed for lacZ staining. Wholemount images (A-L) and transverse sections (a-l) taken at the positions indicated by dashed lines are shown. Brackets are used to mark the GFP expression domain (green), and the extent of lacZ staining (blue). Initially reporter gene expression is observed throughout the neural tube. As development proceeds, expression becomes restricted to the posterior and dorsal aspects of the neural tube.

6.2 Results

6.2.1 A detailed map of Region1consA

To further investigate Region1consA, I reviewed the genomic environment of this conserved element to ensure that it doesn't contain sequences with other functions. Interestingly, of all conserved regions analysed in this investigation, Region1consA lies closest to the first exon of Gli3 identified in Chapter 4 (exon 0b). Figure 6.2 shows a detailed map of the region surrounding Region1consA, including promoter predictions, exon positions and ESTs. Two regions of particularly high sequence conservation are indicated by red asterisks in Figure 6.2. Region1consA is most strongly conserved at the 3' end, in which a 77 nucleotide region contains only 5 mismatches between human, mouse and rat sequences (indicated by an orange bar in Fig. 6.2). This highly conserved block is preceded by an area of lower conservation, spanning a further 109 nucleotides in the mouse sequence (indicated by a green bar in Fig. 6.2). Downstream of Region1consA, a second region of high conservation is observed 3' of exon 0, which I refer to as Region1consB. The absence of promoter elements Regions1consA, together with its ability to upregulate reporter gene expression in vivo, suggests that it contains an enhancer element.

To investigate the putative mechanisms regulating Region1consA activity, the sequence was searched for transcription factor binding sites. Initial investigations were performed on Region1consA and Region1consB, transcription factor binding sites identified using Matinspector (Genomatix inc.; Cartharius *et al.*, 2005) are shown in Figure 6.3. A striking observation is the high frequency of binding sites corresponding to the Myt, Pax, and TALE families of transcription factors.

The Myt1 (Myeloid transcription factor 1) family of transcription factors is composed of three zinc finger genes of the CCHHC class (Myt1, Myt1-Like and NZF3) that are expressed predominantly in the developing CNS, in neural progenitors and in the glial lineage (Nielsen *et al.*, 2004; Romm *et al.*, 2005). They have been shown to interact with Sin3B, a protein that mediates transcriptional repression by binding to histone deacetylases (HDACs), and are thought to function in the silencing of genes during development (Romm *et al.*, 2005). Myt1 modulates the proliferation and differentiation of oligodendrocytes, and overexpression in Xenopus promotes neuronal cell fate (Bellefroid *et al.*, 1996; Nielsen *et al.*, 2004).

Figure 6.2: *Genomic environment of Region1.* The sequence spanning 30kb upstream of Gli3 exon 1 in the human sequence was aligned with the corresponding sequence of the mouse, chicken and rat genomes using MAVID (Bray and Pachter, 2004). The alignment encompassing Region1 is shown along with 1kb of mouse genomic sequence immediately upstream. Constructs used in this chapter are represented beneath the alignment as coloured lines, they are: Region1consA (green), Region1consA-myt (orange) and Region1consB-myt (blue). In the aligned region asterisks are used to mark nucleotides conserved across all four species, where these are particularly dense they are shown in red. Two regions of particularly strong conservation are observed. Region1consA was designed to terminate 3' of the first region of high conservation, whereas the second region of high conservation is also included in Region1consB-myt. Regions highlighted in yellow represent exon 0b (in the mouse sequence upstream of the alignment), and exon 0 (within the alignment). Promoter elements I and II identified in Chapter 4 are shown. The region highlighted in green upstream of the alignment, together with exon0b was used by Paparidis and colleagues as a promoter and corresponds to Element I (Paparidis *et al.*, 2007). Element II contains exon 0 and most of Region 1 consB. Nucleotides underlined represent transcriptional start sites identified in EST databases (black) or by 5'RACE (red). All transcriptional start sites identified in Element I produce transcripts that splice from the end of exon 0b to exon 1 (not shown). A single EST identified within Region 1 was identified that splices from the end of exon 0 to exon 1, but does not appear to be expressed in the mouse embryo at E 9.5.

```
              10        20        30        40        50        60        70        80        90       100
  1 GGTACCGGAGCTCAACTTTTGCAACCGGGGCGCTCGGCCTGGACGGGCCCTTAGTCGTTGAAAAGTCGACTTGACGGTCGGGGGGCTCCTTTCTTTCTCT 100
             * **** ** * ** * *** ** *   *       *** *  ** ** **** ** *
            <<<<<<<<<<<                        >>>>>>>>>>>>           <<<<<<<<<<<<<<<
            V$Myt1/Myt1.02                     V$Myt1/Myt1.02         V$GABF/GAGA.01
                                               <<<<<<<<<<<<           <<<<<<<<<<<<<<<
                                               V$Myt1/Myt1.02         V$PRDF.BLIMP1.01
                 >>>>>>>>>>>>>>>>>>>>>                  >>>>>>>>>>>>>>>>>>>  <<<
                 V$MOKF/MOK2.01                         V$DEAF/NUDR.01      V$EGRF/CKROX.01
                                                        <<<<<<<<<<<<<<<      <
                                                        V$GLIF/ZIC2.01      V$EGRF/WT1.01
                                                                            <<<<<<<<<<<<<
                                                                            V$IRFF/IRF3.01
                                                                             <
                                                                            V$MAZF/MAZ.01

             110       120       130       140       150       160       170       180       190       200
101 TTCTCCTCCCACCCGCGGTTGAACCAAATCAGAACTGTCACTCAGGGCCGGTGACTCAGAGGGATCAGATTACAGGATCCCTGAAACTGATGCGCGGGCT 200
    **** **** * *** ** ** *************** **** **** ** **** *********** ****** **** * * * ***** * *
                         <<<<<<<<<<<<<<<<<<                                                          >
                         V$PAX6.PAX6.02                                             V$PAX9/PAX9.01
                       <<<<<<<<<<<<<<<<<<<<<<<<<<<<<<
                       V$PAX5.PAX5.03                                    >>>>>>>>>>>>>>>>>>>>>
                        <<<<<<<<<<<<<                                    V$IRFF/IRF4.02
                        V$HOXH/MEIS1A HOXA9.01                                    >>>>>>>>>>>
                             <<<<<<<<<<<<                                         V$CDEF/CDE.01
                             V$PBXC/PBX1 MEIS1.01
                             >>>>>>>>>>
                             V$TALE/MEIS1.01
    <<<<<<<<         >>>>>>>>>>>     <<<<<<<<<<<<<<<<<<<<<<<<<<
    V$GABF/GAGA.01   V$CAAT/NFY.03   V$AP1R/BACH2.01
    <<<               >>>>>>>>>>>>>>       >>>>>>>>>>>
    V$PRDF.BLIMP1.01  V$GFI1/GFI1.02       V$AP1F/AP1.01
    <<<<<<            <<<<<<<<<<<<<         <<<<<<<<<<
    V$IRFF/IRF3.01    V$MYT1/MYT1L.01      V$AP1F/AP1.01
    <<<<<<<<<<<<<
    V$EGRF/CKROX.01
    <<<<<<<<<<<<<<<
    V$EGRF/WT1.01
    <<<<<<<<<<<<
    V$MAZF/MAZ.01
     <<<<<<<<<<<<<<<<
     V$EGRF/NGFIC.01
       >>>>>>>>>>>>>>>>>>
       V$OAZF/ROAZ.01
         <<<<<<<<<<<<
         V$MYBL/CMYB.01

             210       220       230       240       250       260       270       280       290       300
201 GCAGCGGCGGGAGGGAGCGGACCGGGCCGCACCGCCGGCGCCCGCACACTCGTCCGCCCGCCCGCCGCTCCCCTCCTCGCCCGCTCGCCCCGTCGCCGGC 300
    **            ** *    ** ***  *****  ** * **   ****** ** * *    ** * *               *** *
    >>>>>>>>>>>>>>>>>>>>       <<<<<<<<<<<<<<<       <<<<<<<<<<<<<<<       <<<<<<<<<<<<<<<
    V$PAX9/PAX9.01             V$NRF1/NRF1.01        V$SP1F/SP1.01         V$EKLF/KKLF.01           <<<<<<<<
    <<<<<<<<<<<<<<<<<<<<<<<<<<<<<<                                        <<<<<<<<<<<<<<<<<       V$AHRR/AHRARNT.02
    V$PAX5.PAX5.03                                   V$EKLF/KKLF.01        <<<<<<<<
    >                  >>>>>>>>>>>>>>>>>>>>>>        <<<<<<<<<<<<<<<       V$MAZF/MAZ.01           V$EGRF/EGR3.01
    V$CDEF/CDE.01      V$ZBPF/ZF9.01                 V$MAZF/MAZ.01        <<<<<<<<                >>>>
    >>>>>>>>>>>>>>>>   >>>>>>>>>>>>>>>>>>>>          <<<<<<<<              V$MZF1/MZF1.01          V$P53F/P53.02
    V$E2FF/E2F1.DP1.01 V$NRF1/NRF1.01                V$MZF1/MZF1.01       <<<<<<<<<<<<<           <<<
    >>>>>>>>>>>>>>>>   <<<<<<<<<<<<<<<<              V$NRF1/NRF1.01       V$MAZF/MAZ.01           V$P53F/P53.04
    V$EGRF/CKROX.01    V$NRF1/NRF1.01
                      <<<<<<<<<<<
                      V$ZF5F/ZF5.01

             310       320       330       340       350       360       370       380       390       400
301 CCACGCCACACCATGCCCGCCGGGGCCGCACGCTGGGATCAGTCAGGCCATCCACGGTAGCCCGAGTCCGCCTGCCGCTCCGCCGGCGCAGGTCTGTGGA 400
    **   ** ** ****** *  *  ** *     **  *** ** ****   * **** * ****  * ** ***  *
                                                                                          >>>>>>>>>
                                                                                          V$FAST/FAST1.01
    <<<<<<<<<<<<<<<     >>>>>>>>>>>>>>>>>>>                          <<<<<<<<<<<<<<<
    V$AHRR/AHRARNT.02   V$PLAG/PLAG1.01                              V$NRF1/NRF1.01
    <<<<<<<<<<<<<       <<<<<<<<<<<<<<<<<<<<<<<<<<                    <<<<<<<<<<<<<<<<
    V$EGRF/EGR3.01      V$AHRR.AHRARNT.02                            V$EGRF/EGR1.01
    >>>>>>>>>>>>>>>>    >>>>>>>>>>>>>>>>>>>>>>>>>>>                   >>>>>>>>>>>>>>>>
    V$P53F/P53.02       V$RXRF/VDR.RXR.04                            V$NRF1/NRF1.01
    <<<<<<<<<<<<<<<     <<<<<<<<<<<<<<<<<<<<<<<<<<<<
    V$P53F/P53.04       V$AHRR/AHRARNT.02

             410       420       430       440       450       460       470       480       490       500
401 TTTGGGACCTGACAGCCAAAGCCGCAGGTACGTCCCGCCTGAAGTTCTTTGTTTTGCGGGGCCCCGCAGCCCGCCGCCGCTTCGAGCTTCCCGGCGGCGTCGG 500
                    ** ********* **** ************* **** * * ** ** * ***** ** * ** * *
    >>>>>             <<<<<<<<<<<<<       <<<<<<<<<<<<<<<<         <<<<<<<<<<<<<<<<<<
    V$FAST/FAST1.01   V$ZFHX              V$LEFF                   V$ETSF
    <<<<<<<<<<<       <<<<<<<<<<<<        <<<<<<<<<<
    V$TALE/MEIS1.01   V$NFKB              V$CABL                               >>>>>>>>>>>>
    <<<<<<<<<<<                                                                V$EGRF
    V$TALE            <<<<<<<<<<<<<<<<<<<<<<<<<<
                      V$PAX5
    >>>>>>>>>>>>>>>>
    V$BARB                               <<<<<<<<<<<<<<<<<
                                         V$E2FF
                      <<<<<<<<<<<<<<
                      V$SP1F
                      >>>>>>>>>>>>>>>>>>>>>
                      V$PAX6

             510
501 AGCGGGCTTGTGTGC 514
    * * * *
    >>>>>>
    V$EGRF
```

Figure 6.3: *Binding sites within mouse Region1.* The sequence spanning Region1consA and Region1consB in the mouse genome was inputted to MatInspector (Genomatix inc.; Cartharius *et al.*, 2005) and searched for the binding sites of vertebrate transcription factors. Binding sites are identified by their Transfac accession number (Wingender *et al.*, 1996), their position and orientation is indicated. Asterisks are shown where the corresponding nucleotide is conserved amongst the amniotes studied, and are shown in red where conservation is particularly strong (see Fig. 6.2). Constructs used in this chapter are represented above the DNA sequence as coloured lines, they are: Region1consA (green), Region1consA-myt (orange) and Region1consB-myt (blue). Exon 0 is shown in bold italics and is underlined. Binding sites associated with three protein families appear to be over-represented in the region, the Myt family (yellow), TALE family (blue) and Pax family (grey). A Tcf binding site and a binding site for Fast-1 (which is known to interact with Smad family proteins) are also highlighted (green).

The TALE (three amino acid loop extension) family of homeodomain (HD) containing transcription factors consists of Pbx and Meis proteins (for which binding sites were identified in Fig. 6.3) together with Prep, Iroquois and TGIF proteins (Burglin, 1997; Berthelsen *et al.*, 1998b). Members of this family are emerging as central developmental factors that interact with a range of proteins, including other HD proteins and basic Helix-Loop-Helix (bHLH) transcription factors (Chang *et al.*, 1995; Peltenburg and Murre, 1996; Jacobs *et al.*, 1999; Knoepfler *et al.*, 1999; Liu *et al.*, 2001c; Laurent *et al.*, 2008). They are thought to regulate transcription through the binding of co-factors that affect the specificity and affinity of transcription factor binding, or by modifying the chromatin environment (Laurent *et al.*, 2008). Protein complexes containing TALE family members have been shown to regulate the expression of EphA2, p21, Hoxb1, Hoxb2, Dcn, Bmp4, vhnf1, Flt3, Emx2, non-muscle myosin II heavy chain B (NMHCB), Follicle stimulating hormone β (FSHβ), myogenin, malic enzyme, Pax3, Pax6 and Shh (Chen and Ruley, 1998; Jacobs *et al.*, 1999; Bromleigh and Freedman, 2000; Ferretti *et al.*, 2000; Wang *et al.*, 2001; Huang *et al.*, 2003; Qin *et al.*, 2004; Choe and Sagerstrom, 2005; Sarno *et al.*, 2005; Capellini *et al.*, 2006; Wang *et al.*, 2006; diIorio *et al.*, 2007). Interestingly, the FSHβ gene has recently been shown to be regulated by a complex containing TALE family proteins and Smad 4 (Bailey *et al.*, 2004). In this specific case, TALE family proteins are thought to bind to the gene and recruit Smad proteins or stabilise their binding.

Pax (Paired box) proteins are tissue-specific transcription factors that contain a highly conserved 128 residue DNA-binding 'Paired domain', Pax 3, 4, 6 and 7 also contain a DNA binding homeodomain (Wilson *et al.*, 1993; Epstein *et al.*, 1994). The Pax family consists of 9 members in mammals which are differentially expressed during development in tissues including the CNS, skeleton, B-cells, thyroid, kidney, pancreas and skeletal muscle, and are associated with regulating differentiation (reviewed in Lang *et al.*, 2007). In the context of this study, Pax6 and Pax3 are of particular interest because they are expressed in progenitor cells of the intermediate and dorsal neural tube respectively, in a domain overlapping with that of Gli3. Of particular interest, Pax3 and Pax6 expression in the neural tube is known to be regulated by Shh from the underlying notochord (Ericson *et al.*, 1997; Hynes *et al.*, 1997; Greene *et al.*, 1998; Meyer and Roelink, 2003; Petropoulos *et al.*, 2004).

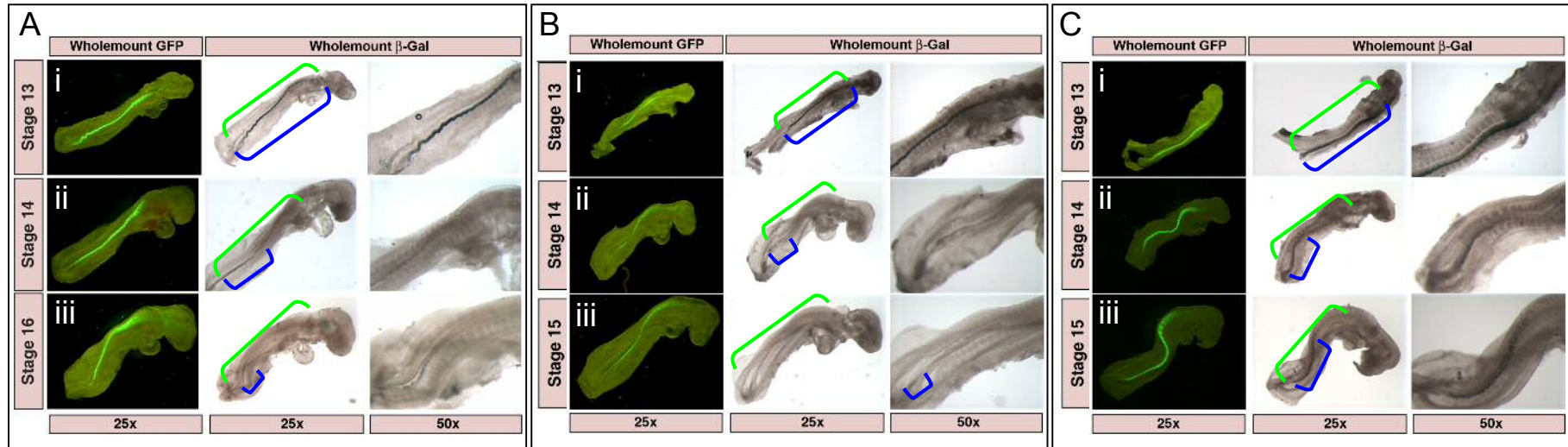### 6.2.2 Defining the minimal sequence required for Region1 enhancer activity

To investigate whether the Myt binding sites located at the 5'end of Region1consA are required in vivo, they were deleted from the Reg1consA construct described in Chapter 5 to produce Region1consA-myt (Fig. 6.2, 6.3 and 6.4A, orange bars). LacZ staining of embryos

electroporated with this construct and harvested 12-18 hours post-electroporation showed a similar expression profile to that of embryos electroporated with the original Region1consA construct (Table 6.1, Fig. 6.5 A and B). Up to HH stage 13, transgene expression was normally observed throughout the electroporated domain, and was later restricted to the posterior. Thus, I concluded that deletion of the Myt binding sites does affect Region1consA transcriptional activity. This is consistent with my observation that the Myt sites are poorly conserved in vertebrates.

Since a second region of homology located 3' of Region1consA was identified (Region1consB, Chapter 3), I prepared a construct including this region of conservation but lacking the Myt sites (Region1consB-myt; Fig. 6.2, 6.3 and 6.4A, blue bars). When electroporated into the neural tube of chick embryos, reporter gene expression driven from this construct was similar to that observed in embryos electroporated with the original Region1consA construct, except that expression levels appeared higher (Table 6.1, Fig. 6.5 A and C). Region1consB contains a putative promoter upstream of exon 0 (Fig. 6.2), which could be responsible for the reporter gene expression observed. However, it also contains binding sites for Lef/Tcf proteins (V\$LEFF) and Fast proteins (V\$FAST/FAST1.01; Fig. 6.3), which could integrate signals from the Wnt and Bmp families respectively, and cause up-regulation. Since Region1consB-myt contains a putative promoter and an exon, this construct was not studied further.

Region	Description	N	HH stage	Pattern	no	Total
	12hr					
1consA	Moderate-strong staining throughout	12	12-13	Higher post	28	53
	Moderate-strong staining stronger in anterior	5	12-13	Throughout	12	
	Moderate staining, stronger in posterior	28	13-14	ant	5	
	No expression	7	13	none	7	
1consA-myt	Moderate- strong staining throughout	5	12-13	Higher post	7	18
	Moderate staining in posterior only	3	13	Throughout	6	
	Low expression throughout	1	14	Ant	0	
	Few cells stained in posterior	4	14	none	5	
	No staining	5	13-14			
1consB-myt	Moderate-strong staining throughout	4	13	Higher post	12	17
	Moderate-strong staining stronger in posterior	2	13	throughout	3	
	Weak staining, stronger in posterior	2	13	ant	2	
	Few cells stained in posterior	2	13	none	0	
	Weak staining, stronger in anterior	1	13			
	Moderate staining, stronger in posterior	6	14			
	Few cells stained in anterior	1	14			

Table 6.1, *Comparison of transgene expression driven from reporter constructs carrying Region1consA, Region1consA-myt, and Region1consB-myt.* Embryos were electroporated with the reporter construct and the pMES control plasmid at a 3:1 ratio. 12-18 hours post electroporation embryos displaying strong GFP expression throughout the AP axis were harvested and processed for lacZ staining. Transgene expression levels were recorded relative to the GFP levels throughout the AP axis. N indicates the number of embryos examined.

Figure 6.4: *Map of Region 1.* A) Map of the different constructs used to investigate Region 1. Binding sites identified using MatInspector are represented by coloured blocks. Binding sites for Myt1 (yellow), TALE family proteins (blue), Pax proteins (grey), and binding sites previously associated with the regulation of Gli3 (green) are shown. The numbers shown in red represent the position of the region relative to human exon 1. Numbers shown in black represent the size of the respective region in the murine sequence. Of particular interest is a cluster of Meis/Pbx and Pax binding sites located within Region 1consA. B) Multi-species alignment of a fragment of Region1consA containing the cluster of binding sites shown in A, alongside the consensus binding sites for Meis, Pbx and Hox proteins.

Sites 1 and 2 show the greatest homology to consensus binding sites. C) Mutations introduced into the Region1consA construct (mouse sequence). Bases shown in red are those mutated in each construct, the complementary strand of the non-mutated DNA is shown in grey. Pax6 consensus binding sites are also indicated.

Figure 6.5: *Comparison Region1consA, Region1consA-myt, and Region1consB-myt transcriptional activities.* Embryos were co-electroporated with the reporter construct and the pMES control plasmid in a 3:1 ratio. Reporter constructs used were Reg1consA (A), Reg1consA-myt (B) and Reg1consB-myt (C). Embryos displaying strong GFP expression were fixed 12-18 hours post electroporation and processed for lacZ staining. Success of electroporation along the AP axis can be visualised with the GFP expression levels (Green). Brackets are used to indicate the GFP expression domain (green), and the extent of lacZ expression (blue).

6.2.3 Meis proteins regulate Region1consA transcriptional activity

As already mentioned, Region1consA-myt contains a cluster of overlapping TALE family binding sites together with a putative Pax binding site and a single Myt binding site (Fig. 6.3, 6.4A). Sequence comparisons revealed that only the Meis and Pbx consensus binding sites have a high degree of conservation among vertebrate genomes (Fig. 6.4B). The Myt binding site was found in human and rat genomes but not in the chicken genome. Likewise, the Pax6 binding site is better conserved in human and rat genomes than in the chicken genome, although in all cases a number of mismatches were found (Fig. 6.4C). The Pax5 binding site was only identified in the mouse and rat sequences. This suggests that Region1consA activity could be due to the presence of Meis/Pbx binding sites.

To establish whether TALE family members regulate reporter gene expression via Region1consA, constitutively active and repressing forms of Meis1a were co-electroporated with the Region1consA-nP1230 reporter construct in the neural tube of HH stage 10-11 chicken embryos. Meis1a-Vp16 contains the entire ORF of Meis1a fused to the activation domain of herpes simplex viral protein Vp16, and thus acts as a constitutive activator. In contrast, Meis1a-En1 contains the entire ORF of Meis1a fused to the repressor domain of Drosophila Engrailed (En1), and functions as a constitutive repressor (Zhang *et al.*, 2002). These constructs have previously been used to investigate the regulation of a Pax6 enhancer by Meis1 in the lens placode (Zhang *et al.*, 2002).

Although at first glance, embryos co-electroporated with the Meis1a-En1 construct appeared similar to embryos electroporated with Reg1consA alone (Fig. 6.6 A-B, Fig. 6.7 A-B, Table 6.2), it became clear that this is only true for young embryos (HH stage 12-13) when the transgene is expressed throughout the AP axis (Fig. 6.6B i-ii, Table 6.2). At HH stage 14, when transgene expression is normally higher in the posterior neural tube, I observed that in the presence of Meis1a-En1 reporter gene expression was maintained throughout the AP axis (Fig. 6.6 Biii, Table 6.2). In addition, anterior sections of HH stage 13 embryos reveal that transgene expression is no longer dorsally restricted as in control embryos (Fig. 6.6 b-c). These data are consistent with the conclusion that the repressor Meis1a-En1 causes a maintenance of reporter gene expression driven by Region1consA in the rostral neural tube.

Embryos electroporated with Meis1a-VP16 displayed an up-regulation of reporter gene expression throughout the DV axis at early stages (up to HH stage 14; Fig. 6.6C i-ii, Fig. 6.7 C, Table 6.2). In more developed embryos, expression followed the same trend as observed for Region1consA alone, with expression progressively restricted towards the posterior (Fig. 6.6 C iii-iv, Table 6.2). Interestingly expression levels appear to become reduced in the anterior at earlier stages than is observed in the absence of the activator construct (Fig. 6.6

Aiii-iv and Ciii-iv, Table 6.2). Indeed, expression levels were highest in the posterior of several of the youngest embryos harvested. This suggests that transactivation of Meis1a allows initiation of reporter gene expression from Region1consA, but prevents its maintenance.

	12hr					
Region	Description	N	HH stage	Pattern	no	Total
1consA	Moderate-strong staining throughout	12	12-13	Higher post	28	53
	Moderate-strong staining stronger in anterior	5	12-13	Throughout	12	
	Moderate staining, stronger in posterior	28	13-14	ant	5	
	No expression	7	13	none	7	
1consA + meis1a-En1	Strong staining, higher in posterior	5	13	Higher post	8	20
	Moderate staining higher in posterior	2	13	throughout	6	
	Few cells stained in posterior	1	13	ant	1	
	Strong staining throughout	5	13-15	none	5	
	Few cells stained throughout	1	14			
	Few cells stained in anterior	1	15			
	No expression	5	13			
1consA + meis1a-VP16	strong staining throughout	5	13	Higher post	18	29
	Weak expression in posterior	12	13-16	Throughout	8	
	Weak expression throughout	1	13	Ant	0	
	Few cells stained throughout	1	13	none	3	
	Moderate staining higher in posterior	2	14			
	Low expression throughout	1	14			
	Strong staining higher in posterior	3	14			
	Moderate expression in posterior only	1	14			
	No staining	3	14, 16			

Table 6.2, *Comparison of transgene expression in embryos co-electroporated with Region1consA and Meis1a-En1 or Meis1a-VP16.* Embryos were electroporated at HH stage 10-11 with the reporter construct, the relevant Meis construct and the pMES control plasmid at a 10:6:3 ratio. 12-18 hours post electroporation embryos displaying strong GFP expression throughout the AP axis were harvested and processed for lacZ staining. Transgene expression levels were quantified relative to the GFP levels throughout the AP axis. Developmental stage was also recorded.

These data suggest that Meis proteins influence reporter gene in two ways. They may control the initiation of transgene expression, as well as the AP patterning of expression. The question remains whether Meis1a acts directly or indirectly on Region1consA activity. Meis proteins have previously been shown to regulate Pax6 expression in the lens and pancreas (Zhang *et al.*, 2002; Zhang *et al.*, 2006). Indeed, Meis1a-En1 represses Pax6 expression in the optic placode in vivo (Zhang *et al.*, 2002). Since Pax6 is expressed in the neural tube, and Pax binding sites were identified by Genomatix within Region1consA, it is possible that Meis1a-En1 acts on Region1consA activity via its control of Pax6 expression. To test this I examined Pax6 expression following co-electroporation of Region1consA and Meis1a-En1. In the electroporated side a slight down-regulation of Pax6 is apparent (Fig. 6.8ii, iii). However, the domain of Pax6 repression does not correlate with the domain of transgene down-regulation (Fig.6.8, arrows), suggesting that Meis1a-En1 does not control Region1consA activity via Pax6.
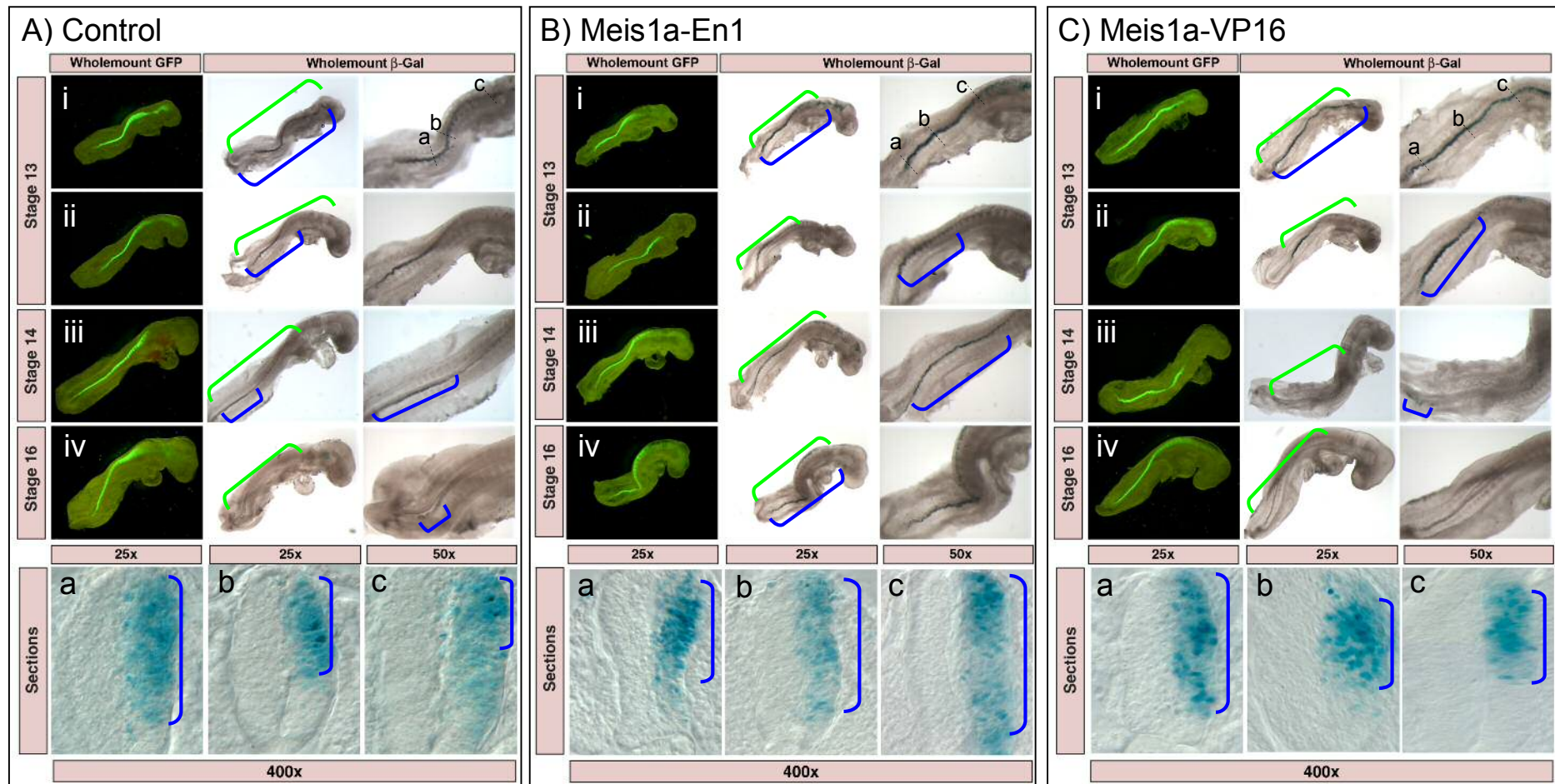
Figure 6.6: *Comparison of β-Galactosidase expression in embryos co-electroporated with Region1consA and Meis1a-En1 or Meis1a-VP16.* Embryos were co-electroporated at HH stage 10-11 with the reporter construct together with the pMES control plasmid and PBS (A), Meis1a-En1 (B), or Meis1a-VP16 (C). Embryos displaying strong GFP expression were fixed 12-18 hours post electroporation and processed for lacZ staining. Panels i-iv show wholemount images of embryos harvested at various stages (HH). Panels a-c show transverse sections of the HH stage 13 embryo in panel i, taken at the positions indicated. Brackets are used to indicate the GFP expression domain (green), and the extent of lacZ expression (blue).
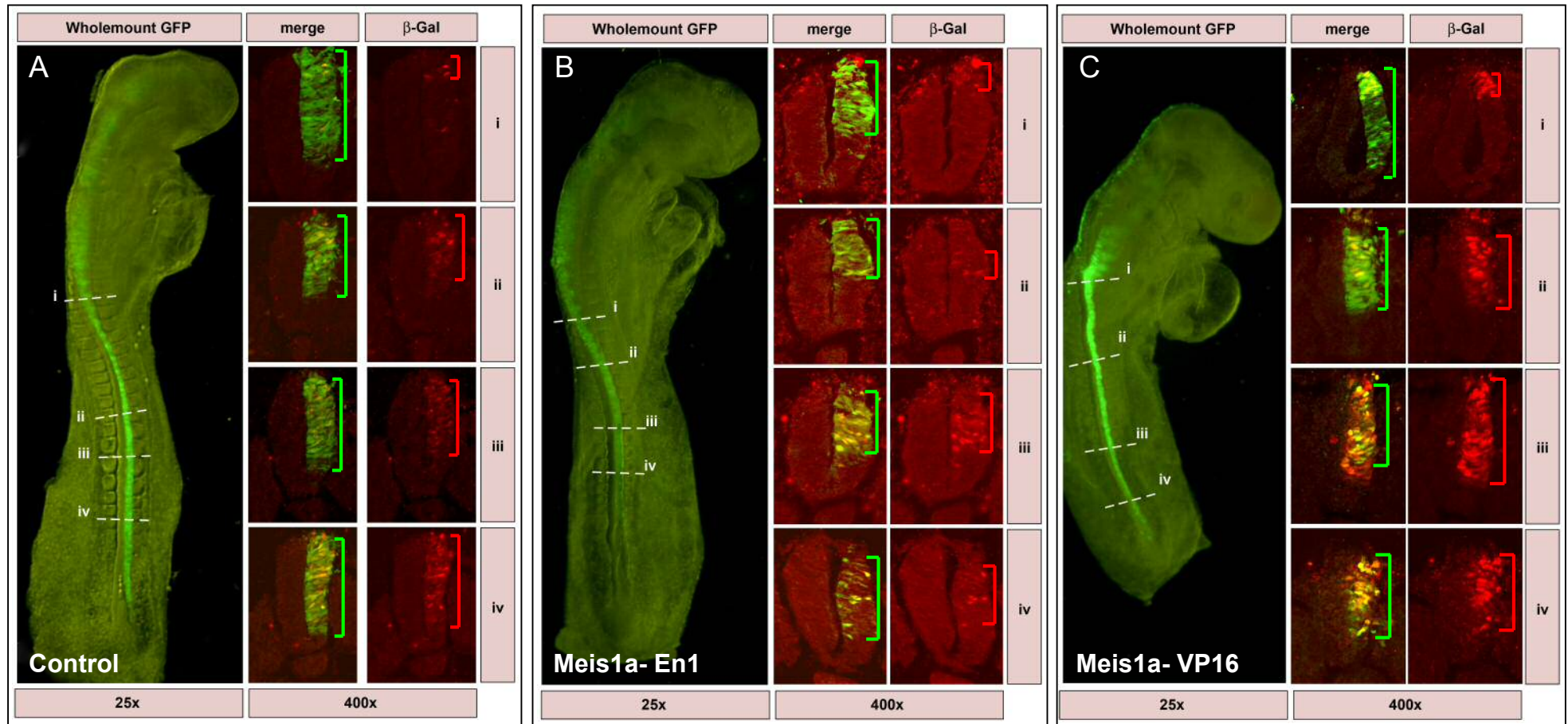
Figure 6.7: *Immunofluorescence showing β-Galactosidase expression following co-electroporation of Region1consA-nP1230 with or without Meis1a-En1 or Meis1a-VP16.* Embryos were co-electroporated at HH stage 10-11 with the reporter construct together with the pMES control plasmid (A), and plamids carrying Meis1a-En1 (B), or Meis1a-VP16 (C) in a 10:3:6 ratio. 12 hours post-electroporation embryos were fixed and processed for cryostat sectioning. 15μM transverse sections were processed for immunofluorescence using an antibody specific to β-Galactosidase (β-Gal; red). Wholemount images show the extent of GFP expression (green) along the AP axis, and indicate successfully electroporated regions. The sections shown correspond to the positions indicated (i-iv). Brackets are used to mark the GFP expression domain (green), and the extent of lacZ expression (red). At this stage of development, Meis1a-En1 has no effect on reporter gene expression. In contrast, Meis1a-VP16 up-regulates reporter gene expression in the posterior neural tube.
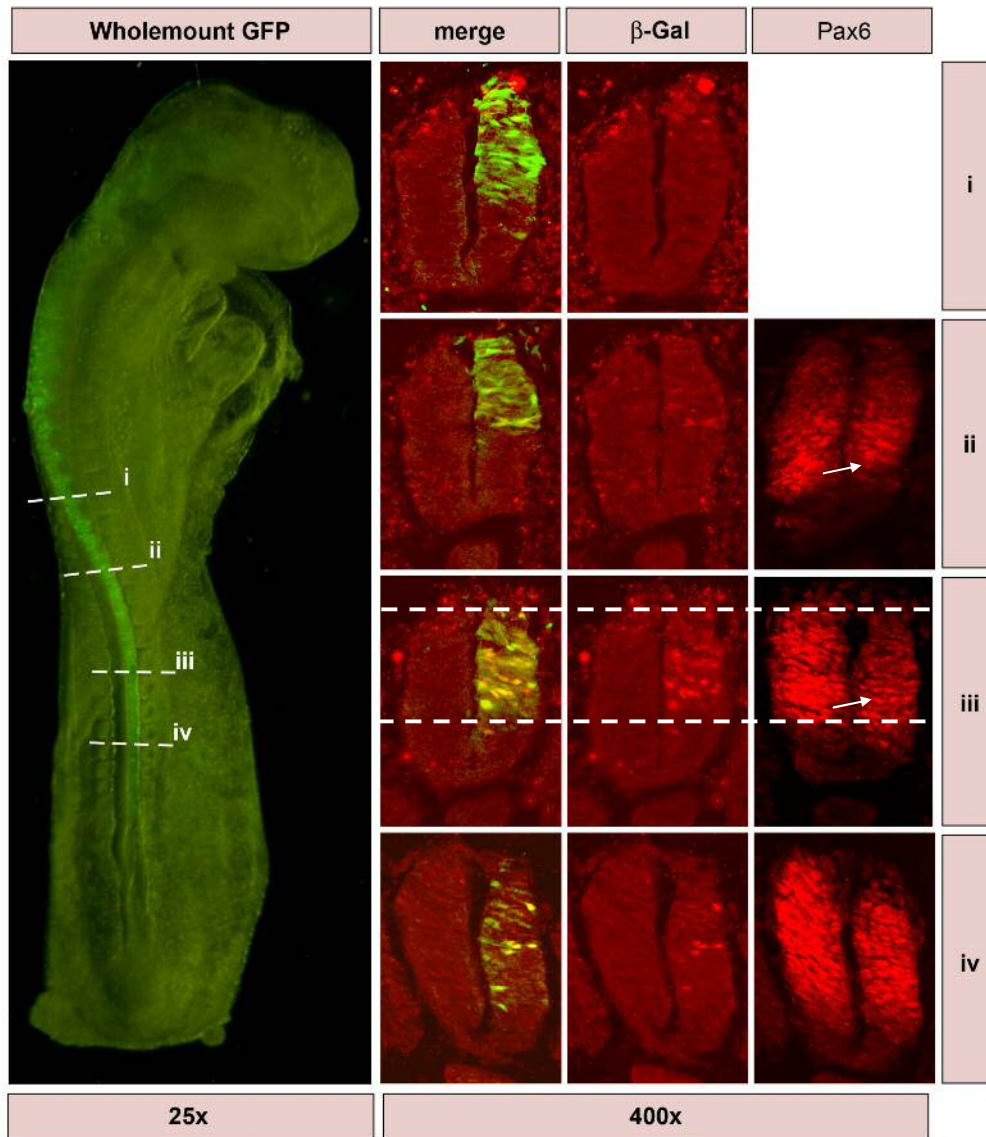
Figure 6.8: *Immunofluorescence showing β-Galactosidase and Pax6 distribution following co-electroporation of Region1consA-nP1230 with Meis1a-En1.* Embryos were co-electroporated at HH stage 10-11 with the reporter construct, the pMES control plasmid and plasmids carrying Meis1a-En1 in a 10:3:6 ratio. 12 hours post-electroporation, embryos were fixed and processed for cryostat sectioning. 15μM adjacent transverse sections were processed for immunofluorescence using antibodies specific to β-Galactosidase (β-Gal, red) or Pax6 (also in red). The sections shown correspond to the positions indicated (i-iv). The merge image shows the extent of GFP expression in sections processed for β-galactosidase expression. White dashed lines are used to mark the domain of GFP expression. A slight down-regulation of Pax6 expression is observed in the electroporated side (arrows).

6.2.4 Further characterisation of TALE binding sites

The identification of conserved TALE family binding sites within Region1consA suggests that these proteins might act directly to regulate Gli3 expression. Structural and functional properties have been used to characterise two sub-families of TALE proteins, which both bind DNA via their homeodomains, but differ in their preferred binding sites. Pbx proteins are members of the PBC family, whereas Meis, Prep and TGIF proteins form the Meinox sub-family (Burglin, 1997). Weak binding is observed for some of these proteins as monomers but binding affinity is greatly enhanced upon the formation of heteromeric complexes, formed with other HD proteins that may or may not be members of the TALE family (Neuteboom and Murre, 1997).

Pbx monomers recognise the consensus sequence 'AATCA' with low affinity (Chang *et al.*, 1996; Shen *et al.*, 1997a). Region1consA contains one perfect site (Fig. 6.4B Site 1), and three sites with one mismatch (Fig. 6.4B Sites 2, 3 and 4). Region1consA also contains the Meinox consensus 'CTGTCA' motif (Fig. 6.4B Site 2; Bertolino *et al.*, 1995; Chang *et al.*, 1997; Shen *et al.*, 1997a; Berthelsen *et al.*, 1998b). This binding site can be bound with high affinity by a heterodimer composed of Pbx1 and Prep1/2 or Meis1 (Berthelsen *et al.*, 1998b; Shanmugam *et al.*, 1999; Fognani *et al.*, 2002; Haller *et al.*, 2002). Furthermore, Site 2 in Region1consA differs by only one nucleotide from a consensus composite binding site consisting of adjacent Pbx and Meinox motifs, 'CTGTCA**A**TCA' (the meinox half-site is underlined and the non-consensus nucleotide in Site 2 is shown in bold; Chang *et al.,* 1997; Knoepfler and Kamps, 1997; Berthelsen *et al.,* 1998a). Iroquois proteins recognise the consensus sequence 'ACACGTGT', which is not present in Region1consA (Bilioni *et al.*, 2005).

Matinspector also identified a composite Meis1/HoxH Hoxa binding site in Region1consA (Fig. 6.4A). Hox proteins are common binding partners of Pbx and Meis proteins and are well characterised as regulators of anterior-posterior identity during development. Pbx proteins, but not Meis proteins form DNA binding complexes with Hox proteins (Chang *et al.*, 1995; Knoepfler and Kamps, 1995; Phelan *et al.*, 1995; Lu and Kamps, 1996; Shen *et al.*, 1996; Chang *et al.*, 1997; Shanmugam *et al.*, 1997; Piper *et al.*, 1999). Pbx:Hox complexes recognise a generalised sequence with the structure 'TGATTNNT', in which the Pbx binding half-site is underlined (Chang *et al.*, 1996). Site 1 overlaps with several Hox half-sites (numbered 1-4 in Fig. 6.4B), but no consensus Hox:Pbx composite site is present.

6.2.5 TALE family binding sites regulate transgene expression levels

Site directed mutagenesis was used to disrupt the TALE protein binding sites identified in Region1consA and assess the effect on Region1consA activity (Fig. 6.4C). I mutated the

Meinox half site in Site 2 (Fig. 6.4B) based on the report of Zhang and colleagues (2004), showing a significant inhibition of protein binding to a disrupted Meis binding site in the Pax6 lens enhancer (Zhang *et al.*, 2002). A mutation used by Andersen and colleagues (1999) was also introduced to the Pbx halfsite, to generate MeisPbxmut (MP) (Fig. 6.4C; Andersen *et al.*, 1999). To prevent Pbx binding to Site 1, mutations used by Andersen and colleagues (Andersen *et al.*, 1999), and Chang and colleagues (Chang *et al.*, 1996) were introduced (Fig. 6.4C). The sequence at this position, 'ACCAAATCA' is similar to the consensus Pbx:Hox binding site of '**ANNA**ATCA', in which the Pbx half site is underlined, and the Hox half site is shown in bold. Chang *et al*. 1996 showed that separation of the two half sites by 1, or 3 nucleotides abrogated all binding (Chang *et al.*, 1996). Therefore the Hox half site adjacent to Site 1 differs from the consensus by presence of a 'C' residue rather than an 'A' residue at the 5' end. To ensure that this putative Hox binding site is non-permissive to binding, an additional A to G mutation was introduced based on a mutation shown to prevent the binding of any Hox protein (Pbxmut, Fig. 6.4 C; Chang *et al.*, 1996; Chang *et al.*, 1997; Shen *et al.*, 1997b), the residue targeted for mutation is shown in red.

The targeted mutations were introduced into Region1consA-nP1230. A third construct, Pbx;MeisPbxmut (MPP), carries mutations both in the MeisPbx composite site and in the individual Pbx site (Fig. 6.4C). Reporter gene expression in embryos electroporated with each of these constructs was analysed. Region1consAPbxmut maintains the AP gradient of expression described for Region1consA (Fig. 6.9C, Table 6.3). However, it drives increased expression levels (Fig. 6.10C, Table 6.3). The MPP construct has a similar effect to Region1consAPbxmut, causing an upregulation of reporter gene expression whilst maintaining the AP patterning (Fig. 6.9B, 6.10B, Table 6.3). In contrast, the MeisPbx construct produces weaker levels of transgene expression than the original construct (Fig. 6.9D, 6.10D, Table 6.3).

A summary of the expression pattern observed for each of the mutated constructs is shown in Figure 6.11. Mutation of Site 2 (MeisPbxmut, see Fig. 6.4B) suggests that it is required for high levels of reporter gene expression. This is interesting considering that Meis1a-VP16 caused an upregulation of reporter gene expression at early stages, and suggests that an activator protein might associate with Site 2 to activate transcription. Since Meinox proteins each recognise the same consensus, this activity could be mediated by other Meinox family members. Mutation of Site 1 causes a strong upregulation of reporter gene expression, suggesting that Site 1 normally functions to repress transcription. This suggests that the activity of Region1consA may be determined by competition between an activator complex associated with Site 2, and a repressor complex associated with Site 1. Since both sites are

predicted to bind TALE family proteins, the transcriptional activity of Region1consA may be determined by differential expression of TALE family proteins during development. A mechanism for this could be that a protein binding to Site 1 precludes the binding of an activator complex to Site 2, such that activation is facilitated upon mutation of Site 1, resulting in an upregulation of reporter gene expression. However, when Sites 1 and 2 are disrupted in conjunction, a strong upregulation of reporter gene expression is observed, similar to that seen upon mutation of Site 1 alone. One explanation for this could be that mutation of Site 1 creates a novel transcription factor binding site for Ets transcription factors, that could be responsible for the increased activity.

| Region | 12hr | | | | | | |
	Description	N	HH stage	Pattern	no	Total
1consA	Moderate-strong staining throughout	12	12-13	Higher post	28	53
	Moderate-strong staining stronger in anterior	5	12-13	Throughout	12	
	No expression	7	13	ant	5	
	Moderate staining, stronger in posterior	28	13-14	none	7	
1consA pbx mut	Strong expressioin throughout	6	13	Higher post	9	21
	Strong expression, strongest in posterior	4	13	Throughout	12	
	Strong expressioin throughout	5	14	Ant	0	
	Strong expression in posterior only	2	14	None	0	
	Strong expression, strongest in posterior	2	14			
	Strong expression, strongest in posterior	1	16			
	Stong expression throughout	1	16			
1consA meispbx mut	Strong expression, strongest in posterior	2	13	Higher post	12	23
	Strong expression throughout	1	13	Throughout	10	
	Moderate expression, stronger in posterior	4	13	ant	1	
	Moderate expression throughout	3	13	none	0	
	Weak expression, stronger in posterior	1	13			
	Weak expression throughout	2	13			
	Strong expression, strongest in posterior	2	14			
	Strong throughout	3	14			
	Moderate expression, stronger in posterior	3	14			
	Weak expression in anterior only	1	14			
	Strong expression throughout	1	15			
1consA MPPmut	Strong expression, strongest in posterior	1	13	Higher post	4	13
	Strong expression throughout	3	13	Throughout	9	
	Strong expression, strongest in post	2	14	ant	0	
	Strong expression throughout	6	14	none	0	
	Strong expression in posterior only	1	15			

Table 6.3: *Comparison of transgene expression driven from Region1consA-p1230 carrying mutations in TALE family binding sites.* Embryos were electroporated with the reporter construct, and the pMES control plasmid at a 3:1 ratio. 12 hours post electroporation embryos displaying strong GFP expression throughout the AP axis were harvested and processed for lacZ staining. Transgene expression levels were quantified relative to the GFP levels throughout the AP axis. Developmental stage was also recorded.
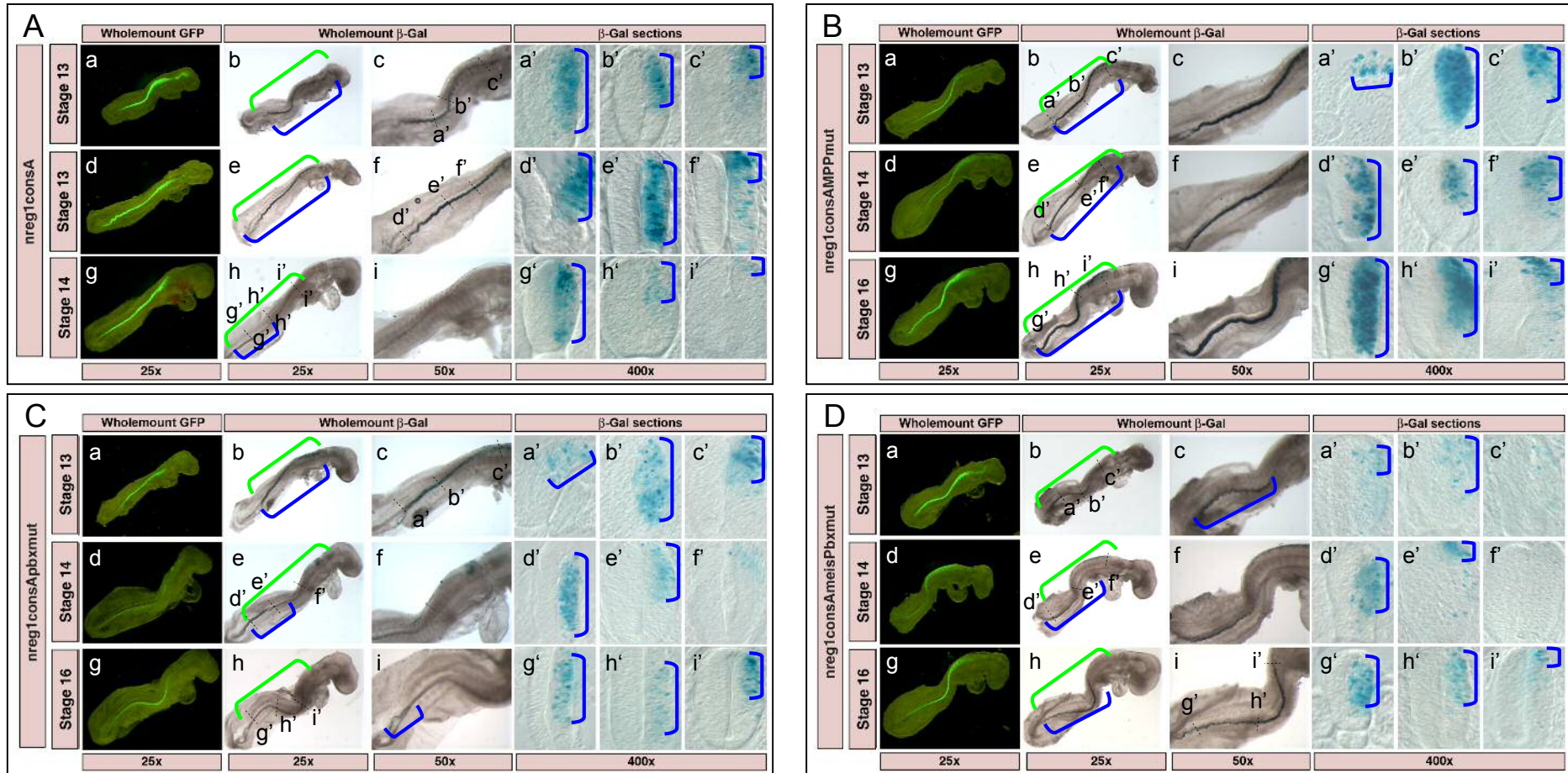
Figure 6.9: *Comparison of b-Galactosidase expression from plasmids containing Region1consA with mutated Meis/Pbx binding sites.* Embryos were co-electroporated with Region1consA (A), Region1consAMPPmut (B), Region1consAPbxmut (C) or Region1consAMeisPbxmut (D) and the pMES control plasmid at a 3:1 ratio. Embryos displaying strong GFP expression were fixed 12-18 hours post electroporation and processed for lacZ staining. Panels a-i show wholemount images of embryos harvested at various stages (HH). Panels a'-i' show transverse sections taken at the positions indicated. Brackets are used to mark the GFP expression domain (green), and the extent of lacZ expression (blue).
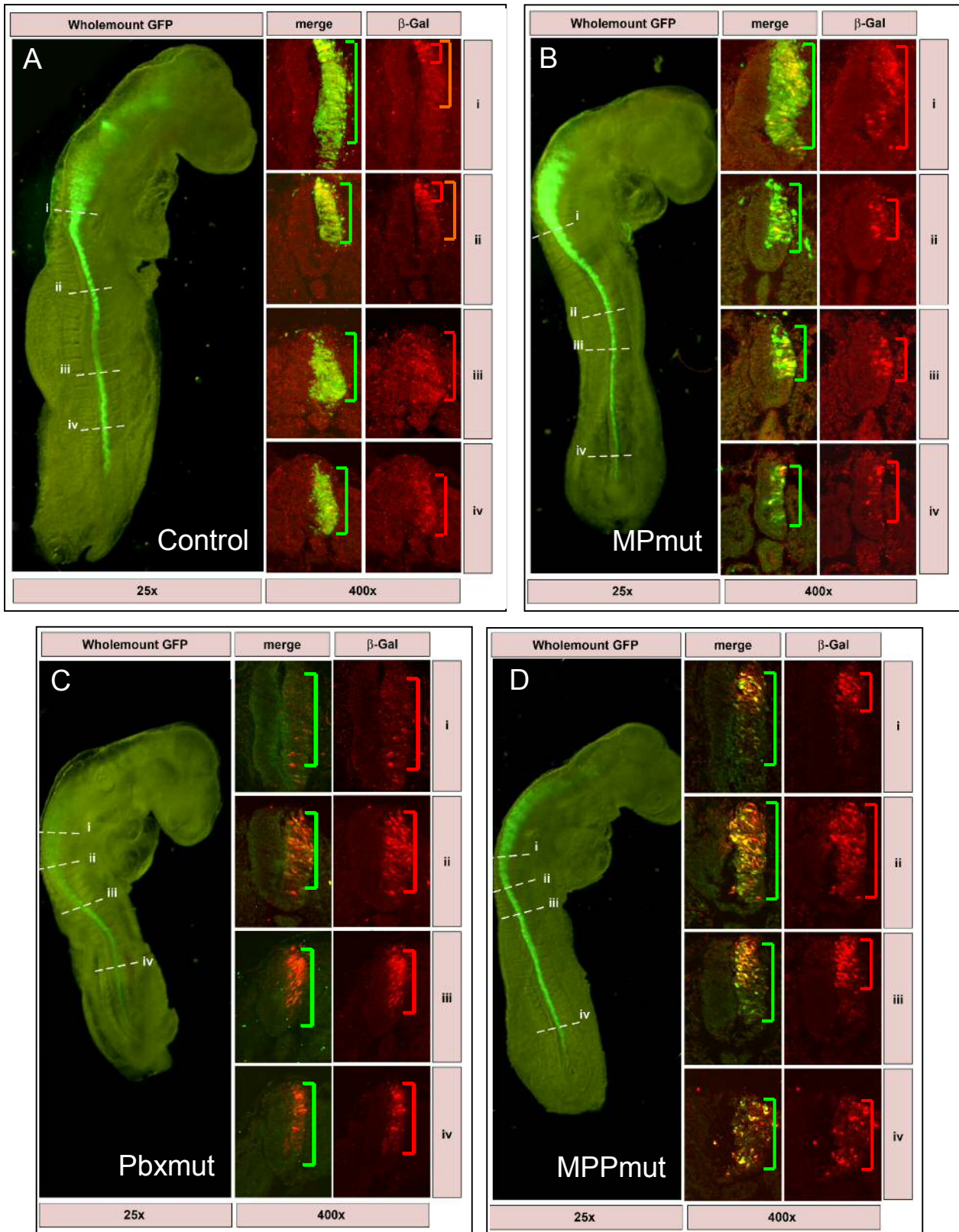
Figure 6.10: *Immunofluorescence showing β-Galactosidase expression following electroporation with Region1consA-nP1230 carrying mutations in Meis/Pbx binding sites.* Embryos were co-electroporated at stage 10-11 with wildtype Region1consA (A), Region1consAMPmut (B), Region1consApbxmut (C) or Region1consAMPPmut (D) and a control GFP reporter at a 3:1 ratio. 18 hours post-electroporation embryos were fixed and processed for cryostat sectioning. 15μM adjacent transverse sections were processed for immunofluorescence using an antibody specific to β-Galactosidase (β-Gal; red). Wholemount images show the extent of GFP expression (green) along the AP axis, and indicate successfully electroporated regions. The sections shown correspond to the positions indicated (i-iv). Brackets are used to mark the GFP expression domain (green), and the extent of lacZ expression (red).
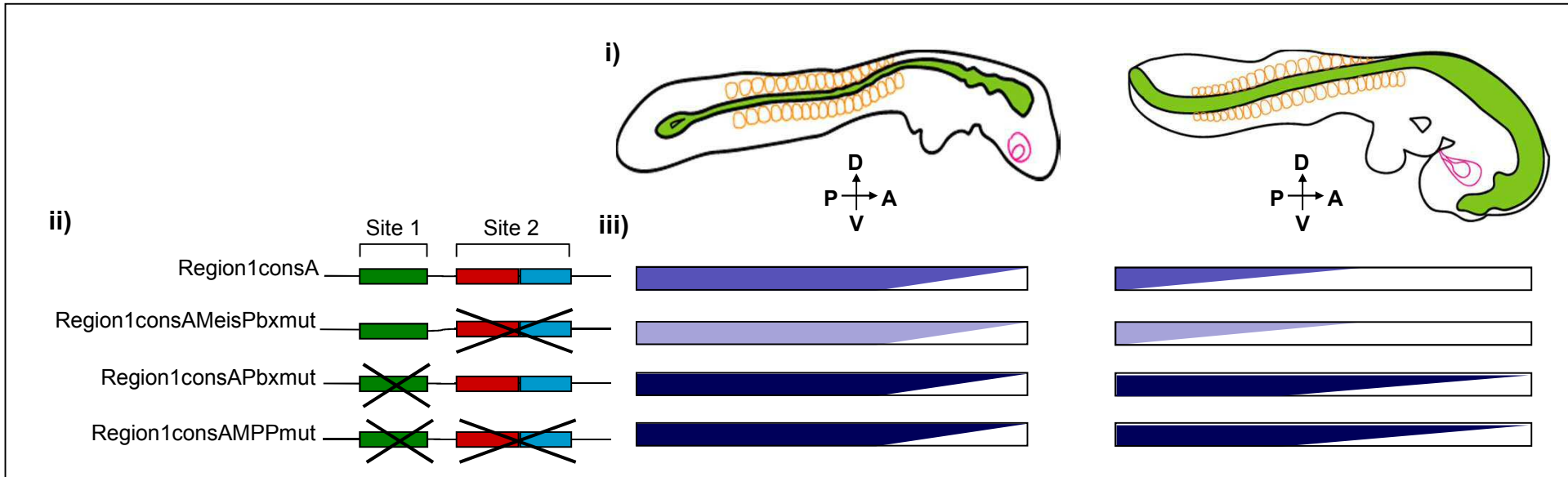
Figure 6.11: *Schematic representation showing the effect of mutating of TALE family binding sites in Region1consA on reporter gene activity.* β-Galactosidase activity 12 hours (A) and 18 hours (B) post-electroporation is summarised. i) Cartoon showing embryo morphology at each timepoint. Somites are shown in orange, the developing eye is pink, and the neural tube is shown in green. A= anterior, P= posterior, D= dorsal, V= ventral. ii) Cartoon summarising the mutations introduced to each construct. Site 1 contains a Pbx binding site (green), Site 2 represents a composite binding site composed of Meis (red) and Pbx (blue) half sites. Binding sites disrupted in each construct are marked by a cross. iii) LacZ expression in the neural tube is summarised. Expression patterns are indicated in the form of a block which corresponds to a saggital cross section through the neural tube. Darker shading is used to represent higher levels of β-Galactosidase activity. For Region1consA transgene activity extends throughout the AP axis at early stages, but is progressively restricted to the dorsal neural tube in anterior regions. As embryos mature, expression is progressively restricted to the posterior neural tube. Constructs carrying mutations in Site 1 drive higher levels of reporter gene expression than the wildtype construct. Expression is maintained at high levels throughout the AP axis in older embryos, although the DV gradient of expression is still observed in anterior regions. Conversely, the Region1consAMeisPbx construct, that carries mutations in Site 2 only, drives lower levels of reporter gene activity than are observed for the wildtype construct.

6.2.6   Investigation of Meis/Pbx expression

6.2.6.1   Meis/Pbx proteins are expressed in HH stage 12-13 chick nuclear extract

To further characterise whether the activity of Region1consA is due to the binding of Meinox/Pbx proteins, I used an electrophotetic mobility shift assay (EMSA) to establish whether the TALE binding sites identified are occupied in vitro by nuclear proteins. I first needed to establish whether Meinox/Pbx proteins are expressed at the appropriate stages of development.

In vertebrates there are four distinct Pbx proteins, three Meis proteins, two Prep proteins, and two TGIF proteins (Nakamura *et al.*, 1996; Imoto *et al.*, 2000; Wagner *et al.*, 2001; Fognani *et al.*, 2002). Additionally, alternative splicing creates multiple isoforms: Pbx1 has two splice variants (a and b), and Pbx3 has four splice variants (a-d; Monica *et al.*, 1991; Milech *et al.*, 2001). 'a' isoforms represent the full length protein. 'b' and 'd' isoforms lack a coding exon at the 3' end of the transcript and are thus truncated at the C-terminus. 'c' and 'd' isoforms are truncated at the N-terminus due to the absence of a coding exon near the 5' end (Monica *et al.*, 1991; Milech *et al.*, 2001). Meis1 has two splice variants (a and b), whilst four splice variants have been identified for Meis2 (a-d). The 'a' isoforms represent the full length protein, whilst 'b', 'c' and 'd' isoforms represent alternatively spliced forms, lacking one or more coding exons (Moskow *et al.*, 1995; Oulad-Abdelghani *et al.*, 1997). Expression levels were analysed in HH stage 12-13 chick embryo, and in two neuronal cell lines, DAOY (human melludoblastoma) and PC12 (rat pheochromocytoma) (Greene and Tischler, 1976). Expression levels were determined both at the mRNA level by RT-PCR, and at the protein level by Western blot.

Western blotting was performed on nuclear extracts using antibodies raised against Pbx or Meis proteins. The pan-Pbx antibody used (Pbx1/2/3) recognises a C-terminal domain of Pbx proteins that is not present in Pbx1b, Pbx3b or Pbx3d. The antibody also does not recognise Pbx4. The Meis antibody used (Meis1/2) recognises an N-terminal motif of Meis1 and 2, but not Meis3. In embryonic and Daoy cell nuclear extract a single Pbx isoform was identified, migrating at approximately 42kDa (Fig. 6.12A, band a). This isoform is not identified in PC12 cells, but 2 other isoforms are detected, migrating at approximately 49 and 50kDa, which are most likely to represent Pbx2 and Pbx3a respectively (Fig. 6.12B bands b and c; Okada *et al.*, 2003). Two Meis isoforms are identified in embryonic tissue, migrating at 42 and 47kDa (Fig. 6.12 A, bands a' and b'). The more prominent 47kDa band is most likely to represent Meis1a, and is also expressed in PC12 cells (Fig. 6.12 B), but not in Daoy cells (Fig. 6.12A) (Haller *et al.*, 2002; Okada *et al.*, 2003; Huang *et al.*, 2005). A 42

kDa isoform of Meis1/2, or Pbx1-3 has not previously been characterised, thus bands a and a' may represent non-specific binding. Similarly an isoform corresponding to molecular weight represented by band c' (~45kDa) has not been previously reported for Meis protins, this too may represent non-specific binding. The results demonstrate that Pbx and Meis proteins are differentially expressed in nuclear extracts. PC12 cell nuclear extract contains Pbx2, Pbx3a and Meis 1a, whilst none of these are present in Daoy cell nuclei. Chick embryo nuclear extract contains Meis1a, but not Pbx2 or Pbx3a.
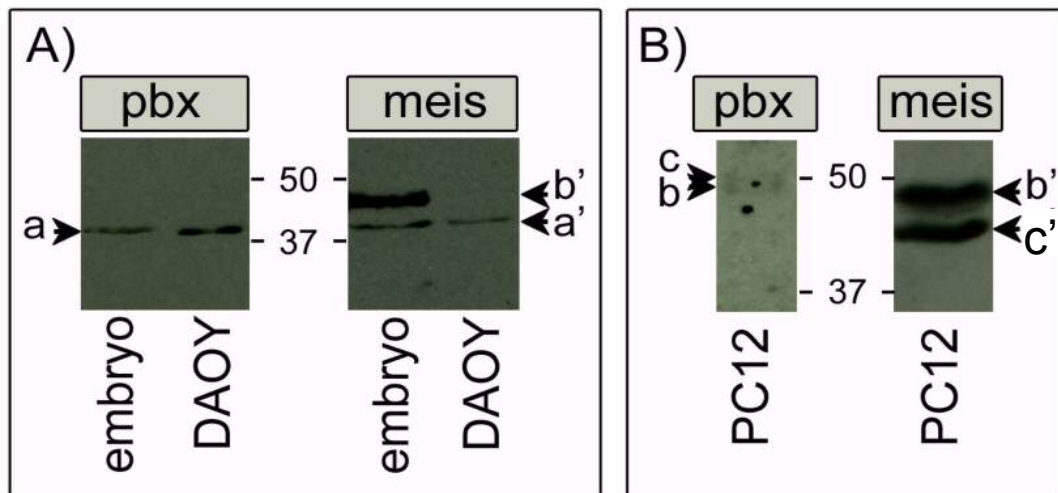

Figure 6.12: *Western Blot analysis of Meis and Pbx protein expression.* Nuclear extracts from HH stage 12-13 embryos and Daoy cells (A), and PC12 cells (B) were processed for Western blot analysis using αPbx1/2/3 (pbx) or αMeis1/2 (meis) antibodies. Protein sizes were estimated using a protein ladder, sizes are shown in kDa. A) In embryonic and Daoy cell nuclear extract a single band is detected by αPbx 1/2/3, migrating at approximately 42kDa (a). The Meis1/2 antibody detects two bands in nuclear extract (a' and b') migrating at 42 and 47kDa respectively. The more prominant band (b') is not detected in Daoy cells. B) PC12 cells express two Pbx isoforms that do not appear to be present in Daoy cells or embryonic tissue (b and c). Both migrate at approximately 50kDa. Band b' identified in embryo extract is detected in PC12 cells, but the band labelled c' appears to be of a higher molecular weight than a', migrating at approximately 45kDa.

Chicken cDNA sequences of Pbx1 (ENSGALT00000021887), (Pbx3 (ENSGALT00000001426), Pbx4 (ENSGALT00000021886), Prep1 (ENSGALT00000026100) and Prep2 (ENSGALT00000000422) were extracted from the Ensembl database, and used to design specific PCR primers. The primers were used in RT PCR analysis of cDNAs generated from cell lines and from chicken embryos. Primers specific for Gli3 were also used. Sequence analysis revealed that chick Pbx4 carries an additional coding exon at the 3' end of the transcript that has not been reported in other species, I refer to the chicken isoform as cPbx4. Pbx1 and Pbx3 primers were designed in a region common to all isoforms and therefore will detect all splice isoforms.

RT-PCR products corresponding to the expected size for Pbx1, Pbx3, cPbx4, Prep1, Prep2 are detected in cDNA generated from HH stage 13 chicken embryos, which express Gli3 (Fig. 5.1). All isoforms of Pbx1 and 3 were detected by RT-PCR, although the b isoform is most

prominent (Fig. 6.13A). The identity of various products were confirmed by sequencing, with the exception of Pbx3a, Pbx3c and Pbx3d which were not cloned. Similarly, all splice variants of Pbx1 and 3 were identified in DAOY cells, that express Gli3 (Fig. 6.13B). Prep genes are not expressed. In addition to Pbx isoforms, PC12 cells that do not express Gli3, express Prep genes at high levels (Fig. 6.13C). Pbx4 primers, designed by homology to the chicken sequence, produce non-specific transcripts when used to amplify the human and rat cDNA present in Daoy and PC12 cells respectively (Fig. 6.13B, C). However, a rat transcript of a similar size to cPbx4 suggests that this isoform might exist in mammals (Fig. 6.13C).
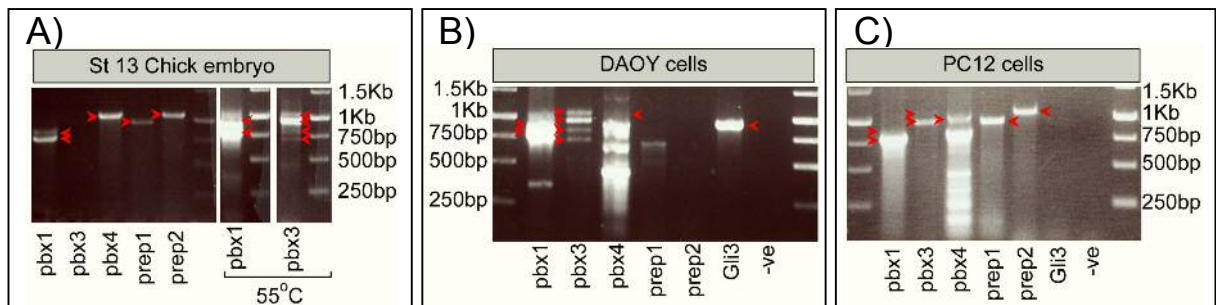

Figure 6.13: TALE gene expression in HH stage 13 chick embryos and in DAOY and PC12 neuronal cell lines. RT-PCR was carried out on cDNAs from HH stage 13 chick embryos (A), DAOY cells (B) and PC12 cells (C). Unlabelled lanes contain a DNA ladder, bands correlate to the fragment sizes shown. Pbx1a (901bp), Pbx1b (788bp), Pbx3a (1073bp), Pbx3b (960bp), Pbx3c (831bp), Pbx3d (718bp), cPbx4 (1064bp), Pbx4 (983bp), Prep1 (959bp), Prep2 (1059bp) and Gli3 (929bp) were analysed using standard PCR conditions with an annealing temperature of 56ºC unless otherwise stated, expected product sizes are shown in brackets and are labelled by red arrowheads. Primers used for Pbx1 and 3 recognise both a and b isoforms. Primers used for Pbx4 amplification will produce a larger product when chicken cDNA is used because of an additional exon in cPbx4. A) At HH stage 13, all Pbx genes investigated are expressed in the chicken embryo. Pbx1b and 3b isoforms appear to be the dominant isoforms, although Pbx1a was also successfully cloned from PCR products. Although Pbx3 was not initially detected in the embryo, both isoforms were clearly visible when the PCR was repeated a slightly lower annealing temperature (as shown). Gli3 expression was not examined in embryo extract, but is well documented. B) DAOY cells express Gli3, along with Pbx 1 and 3, but do not appear to express Prep genes. The faint band observed for Prep1 is of a smaller size than the expected product. C) PC12 cells express Prep1, Prep2, Pbx1b and Pbx3b, Gli3 is only weakly expressed. Primers used for Pbx4 amplification work poorly on cDNAs generated from human and rat genomes represented in Daoy and PC12 cell lines.

Expression analysis demonstrates that all TALE family proteins investigated are expressed in the developing chick embryo, including multiple Pbx isoforms. However, Western analysis suggests that not all protein products are present in nuclear extract. This is supported by previous reports which have shown that transcriptional regulation of TALE family proteins can be regulated by subcellular localisation (Berthelsen *et al.*, 1999; Saleh *et al.*, 2000a; Fognani *et al.*, 2002; Haller *et al.*, 2002; Kilstrup-Nielsen *et al.*, 2003). In support of a possible function of Meis proteins in the regulation of Region1consA, I have demonstrated that Meis1a is expressed in nuclear extracts. Furthermore I have demonstrated that concomitant with a change in Gli3 expression between Daoy and PC12 neuronal cell lines, Meis and Prep genes are differentially expressed. Daoy cells, which express Gli3 do

not express Prep or Meis proteins. Conversely, PC12 cells which do not express Gli3, express high levels of Prep 1 and Prep2 mRNA, as well as high levels of Meis1a. Thus these cell lines provide useful tools for discriminating whether the change in Gli3 expression might be regulated by the binding of different TALE family complexes to Region1consA. By comparing binding of nuclear proteins present in each extract I hope to establish whether Meis/Pbx sites are occupied in vitro by proteins present in the chick embryo. By comparison of nuclear extracts from DAOY and PC12 cells it will be possible to determine whether changes in Gli3 expression are associated with differential binding to Region1consA.

6.2.7 A 35bp fragment of Region1consA containing Meis/Pbx sites is occupied in vitro

To investigate whether the Meis/Pbx sites are occupied, the electrophoretic mobility shift assay (EMSA) was used. A 35bp Oligo-nucleotide including the MeisPbx binding site and the individual Pbx binding site was designed for EMSA and is shown in Figure 6.14. To establish whether this oligonucleotide binds nuclear proteins, the WT labelled oligo was incubated in the presence of nuclear extracts (NE) prepared from HH stage 12-13 chicken embryos, DAOY cells or PC12 cells. Two separate nuclear nuclear extract preparations were made from chick embryos and Daoy cells (named A and B). Species binding to the probe were separated by electrophoresis on a 5.5% polyacrylamide:bis gel (Fig. 6.15).
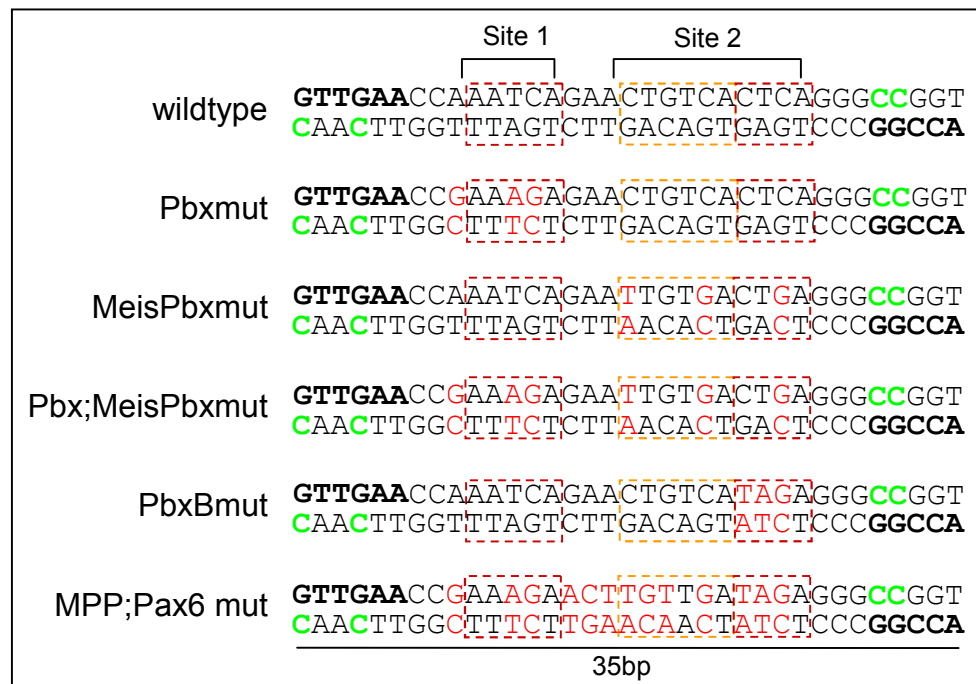
Figure 6.14: *Oligonucleotide design for Electrophoretic mobility shift assay (EMSA).* Oligonucleotides used for EMSA were 35 nucleotides long and incorporated the Meis, Pbx and Pax6 binding sites identified by MatInspector. Complimentary primers were designed with 5' overhangs (bold) that included two G residues at each end. Fill-in reactions were carried out in the presence of P[32]-dCTP resulting in radioactively labelled oligos. Radioactively labelled bases are shown in green. Binding sites were mutated to sequences previously characterised as non-permissive to binding. Substituted bases are shown in red.

NE preparation	A		B		A		B		A	
chick st12	+	+	+	+	-	-	-	-	-	-
DAOY	-	-	-	-	+	+	+	+	-	-
PC12	-	-	-	-	-	-	-	-	+	+
WT* oligo	+	+	+	+	+	+	+	+	+	+
1400x WT cold oligo	-	+	-	+	-	+	-	+	-	+

C	++	-	++	-	++	-	++	-	+	-
B	++	-	+	-	+	-	++	-	+	-
A	++	-	++	-	-	-	-	-	-	-
	1	2	3	4	5	6	7	8	9	10

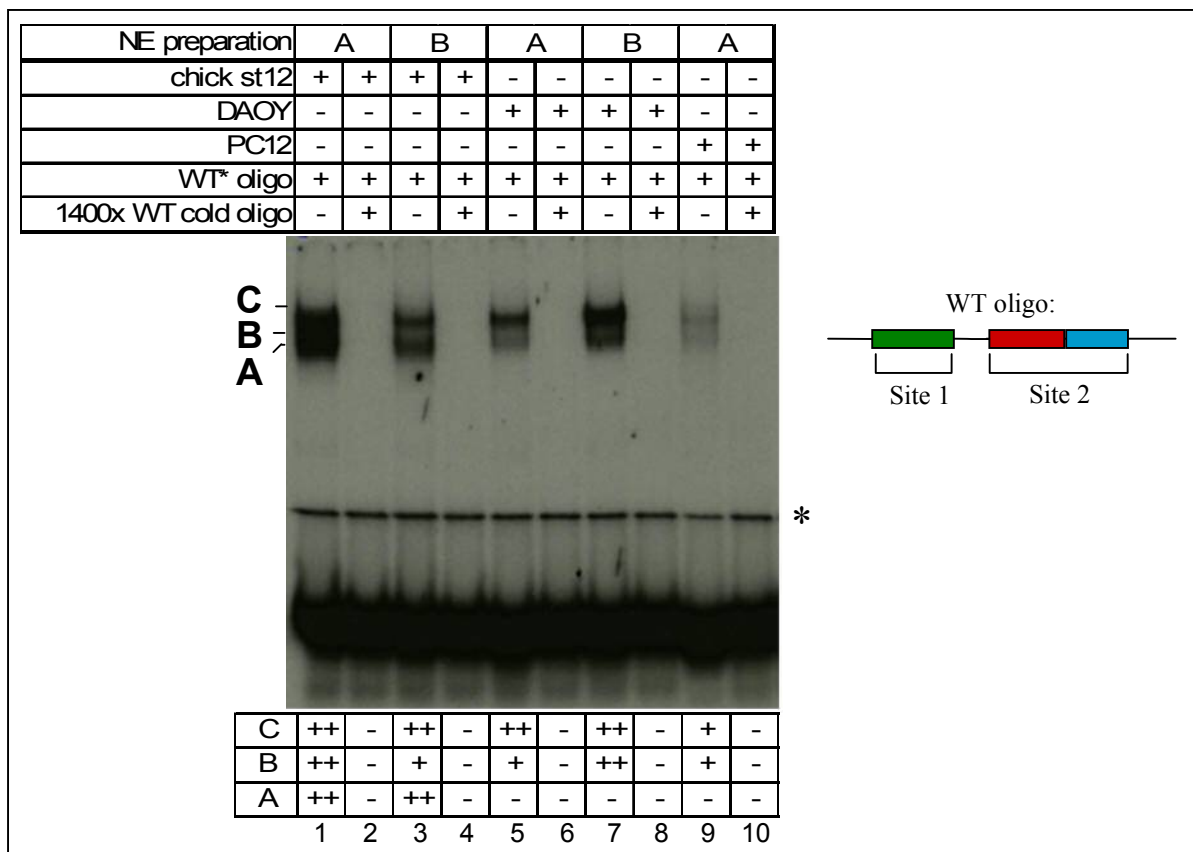***Figure 6.15:*** *Binding of Nuclear proteins to Region1consA.* EMSA experiments were performed using nuclear extracts derived from HH stage 12 chicken embryos, DAOY cells, and PC12 cells and the wild-type oligo radiolabelled containing Meis and Pbx binding sites (WT*). Three bands (A-C) are observed, as summarised in the table below. Each of these bands is competed upon the addition of excess cold wildtype oligo. * designates non-specific binding.

Figure 6.15 shows three bands (labelled A-C) forming in the presence of nuclear extracts from chick embryo and two bands (B and C) forming in the presence of DAOY or PC12 nuclear extracts, indicating that two or three distinct complexes can form on this oligo. To confirm that the bands seen are specific, binding was competed with excess cold (non-radioactive) competitor. In the presence of 1400x cold competitor bands A-C are lost, indicating that they bind specifically to the oligonucleotide. The band labelled '*' in Figure 6.15 is non-specific, since binding is not competed by excess cold probe. Comparing the intensity of each band reveals that band 'C' is most abundant in each sample suggesting that this larger complex is primarily formed. Using PC12 cell nuclear extracts, bands B and C are weak, although similar amounts of proteins were loaded as judged by the intensity of band *, furthermore, band A is absent. Since PC12 cells do not express Gli3 the data are consistent with complexes A-C being required for transcriptional activity of Region1consA.

There are several possible explanations for the complexes observed. The three complexes could represent the binding of three proteins to distinct sites in the oligo. In this case the bands A, B and C observed would represent the binding of 1, 2 and 3 proteins respectively (Fig. 6.16, Model A). The absence of one protein would be expected to reduce the molecular

weight of the resulting complex, such that bands migrate at a faster rate. Alternatively the different bands could represent the binding of three distinct complexes to the oligo, each associating with the same binding site (Fig. 6.16, Model B). The ability of such a complex to associate with DNA will be determined by the abundance of constituent proteins. Individual complexes may share proteins in common, or may each be composed of a unique combination of proteins. Finally it is possible that bands A-C represent a combination of models A and B. The presence of the high molecular complex in PC12 cells, but not of lower molecular weight complexes favours Model B, in which the absence of a particular factor impedes the formation of complexes A and B, but allows binding of complex C.
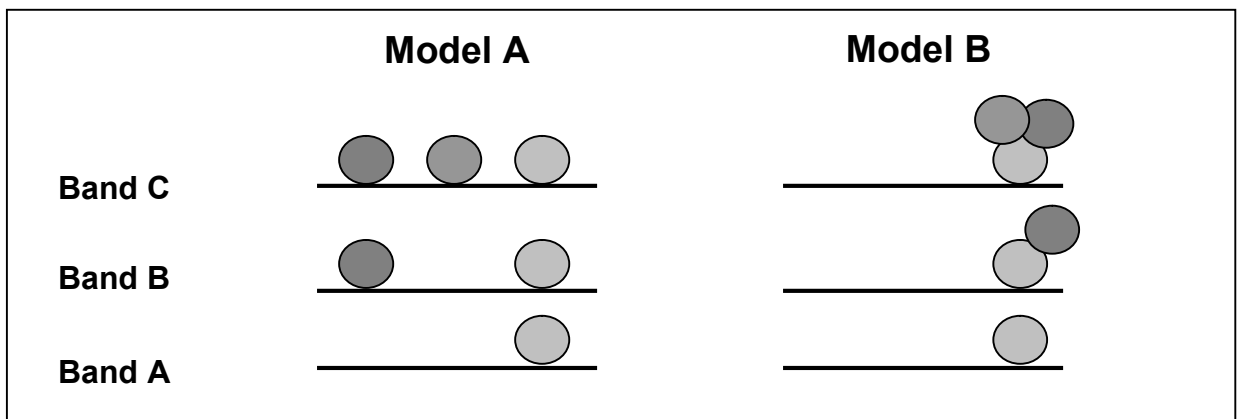

Figure 6.16: *Models for protein binding to Region1consA. Model A) Bands A, B and C observed in EMSA experiments represent the binding of one, two or three proteins respectively, each associating with a different site in the oligo. Model B) Bands A, B and C observed in EMSA experiments represent the binding of different sized protein complexes, each associating with the same binding site. It is also feasible that the binding observed in EMSA experiments results from a combination of the two models proposed here.*

6.2.8 TALE family binding sites are required for protein binding

To establish whether the complexes binding to the labelled oligonucleotide interact directly with the TALE family binding sites, oligonucleotides containing the mutations in the Pbx and Meis core elements used previously in electroporation approaches, as well as new mutations were generated (Fig. 6.4C and 6.15). The mutated oligonucleotides were first used as competitors in EMSA experiments using a radioactively labelled wildtype oligo to assess their ability to compete binding of proteins present in embryonic nuclear extract (Fig. 6.17). Results in Figure 6.17 show that mutations in the individual Pbx site (Site 1) do not disrupt protein binding on the oligo, as the WT and mutated oligo have equal competing ability (Fig. 6.17, compare lanes 2-5 with lanes 10-13). However, oligonucleotides carrying mutations in the MeisPbx binding site showed a dramatically reduced ability to compete binding to the wildtype probe (Fig. 6.17, compare lanes 2-5 with lanes 6-9 and 14-17), although high levels of mutated competitor partially abolish binding to the WT oligo (Fig. 6.17 lanes 9 and 17, 1000x competitor).
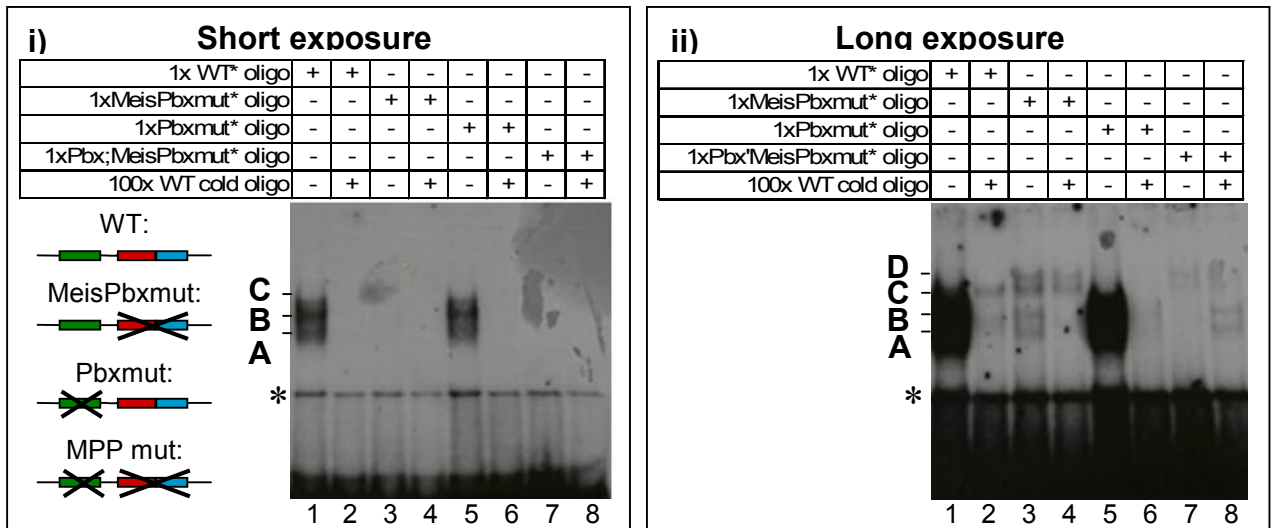
competitor probe	-	Wildtype				MeisPbxmut				PbxBmut				MPPmut				-	-	Pbxbmut				Pax6mut			
1x WT* oligo	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
5x cold competitor	-	+	-	-	-	+	-	-	-	+	-	-	-	+	-	-	-	-	-	+	-	-	-	+	-	-	-
50x cold competitor	-	-	+	-	-	-	+	-	-	-	+	-	-	-	+	-	-	-	-	-	+	-	-	-	+	-	-
100xcold competitor	-	-	-	+	-	-	-	+	-	-	-	+	-	-	-	+	-	-	-	-	-	+	-	-	-	+	-
1000x cold competitor	-	-	-	-	+	-	-	-	+	-	-	-	+	-	-	-	+	-	-	-	-	-	+	-	-	-	+

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27
A	++	-	-	-	-	++	++	++	+	+	-	-	-	++	++	++	+	+	++	++	++	++	-	++	++	++	+
B	++	+	-	-	-	++	++	++	+	+	-	-	-	++	++	++	+	+	++	++	++	++	-	++	++	++	+
C	++	-	-	-	-	++	++	++	+	+	-	-	-	++	++	++	+	+	++	++	++	++	-	++	++	++	+

Figure 6.17: *Oligonucleotides carrying a mutated MeisPbx binding site have reduced binding affinity.* EMSA experiments were performed using nuclear extracts derived from HH stage 12 chicken embryos, and a 35bp radioactively labelled oligonucleotide encoding a segment of Region1consA encompassing Meis and Pbx binding sites (WT). Binding was competed using an excess of unlabelled cold competitor oligonucleotide carrying each of the mutations shown in Figure 6.14. Three bands (A-C) bind to WT, as summarised in the table below. Each of these bands is competed upon the addition of excess cold oligo, but competition is reduced upon mutation of the MeisPbx site indicating that the mutated sequence has a lower affinity for the species bound to the WT oligo. * designates non-specific binding.

These data are consistent with a model whereby distinct complexes form on the Meis/Pbx site (Fig 6.16, Model B). To test this further I performed EMSA experiments using radioactively labelled mutant oligonucleotides (Fig. 6.18). Consistent with my hypothesis, oligonucleotides carrying mutations in the MeisPbx site failed to bind nuclear proteins (Fig. 6.18, lanes 3 and 7), whereas the Pbxmut oligo forms similar complexes to the wildtype oligo (Fig. 6.18, lanes 1 and 5). However, upon long radiography exposures faint binding to the MeisPbx mutant oligo is observed (Fig. 6.18ii lanes 3 and 7). Interestingly, three of the bands are competed by wildtype oligonucleotide, confirming that they correspond to species A-C observed previously (lane 4). However, a fourth band (D) of higher molecular weight is created that is not competed with WT oligo, suggesting that it represents the binding to a novel site introduced by the mutation. Consistent with this I found that the mutations introduced in the MeisPbx mutant probe create binding sites for Ets, Pax3 and GCM (Glial Cells Missing) transcription factors. Alternatively, the novel band could represent the binding of a large complex, the formation of which is inhibited in the presence of a functional Meinox:Pbx binding site.

Figure 6.18: *Oligonucleotides carrying a mutated MeisPbx binding site fail to bind nuclear proteins.* EMSA experiments were performed using nuclear extracts from HH stage 12 chicken embryos, and a 35bp radioactively labelled oligonucleotide carrying mutations within Meis and Pbx sites identified in Region1consA. Binding was competed using an excess of the wildtype sequence. i) Three bands (A-C) bind to the wildtype sequence, and a similar degree of binding is observed to oligos carrying a mutation in the individual Pbx binding site. ii) In long radiography exposures it is apparent that oligos carrying the MeisPbx mutation bind very weakly to complexes of a similar size to those that bind the wildtype sequence. A novel complex forms upon mutation of the MeisPbx sequence, as indicated by the presence band D, which is not competed by the wildtype probe.

Two additional mutated oligonucleotides were designed to further characterise the binding sites recognised by nuclear proteins. PbxBmut contains a mutation in the Pbx half site of Site 2 that does not affect the Meis half site. The mutation was designed based on previous mutations used by Andersen and colleagues that were shown to abolish Pbx and Pdx binding (Andersen *et al.*, 1999). PbxBmut failed to compete binding of nuclear proteins to the WT oligo in a similar manner than MP (Fig. 6.17, lanes 20-23), suggesting that the Pbx half site is critical for protein binding to the MeisPbx site. However, high concentrations of cold PbxBmut oligo fully competed binding to the WT oligo (Fig. 6.17, lane 23) whereas a small amount of binding remained in the presence of a similar concentration of cold MeisPbx mutated oligo (Fig. 6.17, lane 9), confirming that the MeisPbx mutation is more severe. This suggests that both Meis and Pbx half sites are critical to complex formation. Interestingly, mutations introduced to the 35bp oligo do not preferentially disrupt the binding of any single species bound to the wildtype probe, since the intensity of each band is affected to the same degree at a given concentration of competitor. This suggests that the distinct complexes formed each bind to the same site (Fig. 6.16, Model B).

The second mutated oligonucleotide generated, called MPP:Pax6mut contains a mutation in the putative Pax6 binding site. The consensus binding site for the paired domain is (G/T)T(T/C)(C/A)(C/T)(G/C)(G/C), but different family members have different nucleotide preferences (Epstein *et al.*, 1994; Chalepakis and Gruss, 1995; Czerny and Busslinger, 1995;

Jun and Desplan, 1996; Lang *et al.*, 2007). Pax homeodomains recognize palindromic binding sites of TAAT(N)$_{2-3}$ATTA (Wilson *et al.*, 1993), which could not be identified within Region1consA. The Pax6 consensus YNMKTNASTWCGCACTTNA (Epstein *et al.*, 1994) aligns weakly at two positions with Region1consA (Fig. 6.4C). MPP;Pax6mut was designed to prevent Pax6 binding to either of these sites and test the possibility that band D in Figure 6.18 is due to a protein binding at this site. However the MPP;Pax6mut cold oligos gave a similar pattern on gel shift assays to the cold MPP oligos when used as a competitor. This data suggest that the Pax6 binding site may not be active in Region1consA, although further EMSA analyses using radiolabelled MPP;Pax6mut oligo are required.

6.2.8.1 Identifying the proteins bound to the Pbx/Meis sites in Region1consA

To identify the proteins present in nuclear extract that bind to the MeisPbx site I used antibodies against Pbx and Meis, since the PbxMeis site is expected to bind a heterodimeric Pbx:Meinox complex. Antibodies can have two effects on EMSA experiments: disrupting the binding of nuclear proteins or altering the migration rate of the Protein/DNA complex in the gel. The respective consequences can be a loss of the band in the gel, or a 'supershift' of the band due to a complex with a higher molecular weight forming upon antibody binding. Thus, Pbx1/2/3 (C20) or Meis1/2 (N17) antibodies described previously were added to the binding reaction before the addition of labelled oligo. A third antibody, Meis1/2* (H-80) is similar to the N17 antibody but recognises a longer region at the N-terminus, and is concentrated to facilitate its use in EMSA experiments. However, I did not observe significant loss of a radioactively labelled band, nor a supershift, when antibodies were added separately or in combination to the EMSA reaction (Fig. 6.19). A possible explanation for this is that the Meis or Pbx proteins binding to Reg1consA are not recognised by the antibody used. Consistent with this the Pbx1/2/3 antibody performed poorly in Western blot analysis, only one band was detected in embryonic nuclear extract, which appears to be non-specific. The Pbx antibody used does not recognise b or d isoforms of Pbx 1 or 3, nor will it recognise Pbx4. Likewise, the Meis1/2 antibodies do not recognise Meis3, nor the Prep or TGIF members of the Meinox family. Whilst Meis1a appears to be expressed in nuclear extract of chick embryos, its binding to Region1consA may be dependant on the expression of co-factors, such as Pbx proteins, or might be competed by high levels of other Meinox family members. EMSA experiments using other antibodies could address whether other TALE family proteins bind to the wildtype oligo. My PCR data indicate that Pbx1a, 1b, 3a, 3b, 3c, 3d and cPbx4, together with Prep1 and Prep2 are expressed in HH stage 13 chick embryos, additional RT-PCR experiments are required to

test whether Meis3 and TGIF genes are expressed. The expression profile of each protein associating with Region1consA will determine its activity. Importantly the RT-PCR analysis presented here will not reveal tissue specificity of RNA transcripts, nor does not establish the cellular localisation of proteins produced. Furthermore, complexes forming in EMSA experiments are representative of those binding throughout the embryo, and cannot be associated with activation or repression of transgene activity. This prompted me to investigate further the expression pattern of TALE family proteins during embryonic development, in particular, in relation to Region1consA activity and Gli3 expression.

	DAOY A				Embryo B				PC12				Embryo B					
meis1/2	-	+	-	+	-	+	-	+	-	+	-	+	-	+	-	+	-	-
pbx1/2/3	-	-	+	+	-	-	+	+	-	-	+	+	-	-	+	+	-	-
meis1/2* (H80)	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	+	-
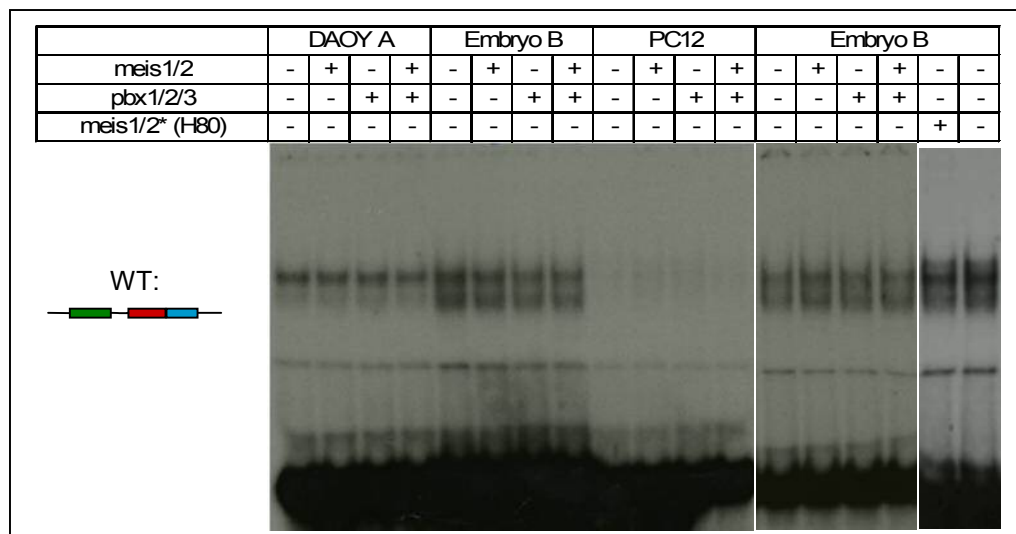
Figure 6.19: Binding to the Region1consA oligo is not disrupted by α-Meis/Pbx antibodies. EMSA analysis was performed using nuclear extracts from HH stage 12 chicken embryos, DAOY cells, and PC12 cells and a 35bp radioactively labelled oligonucleotide encoding a segment of Region1consA encompassing Meis and Pbx binding sites (WT). Binding was competed using antibodies raised against Meis or Pbx proteins. Pbx1/2/3 recognises Pbx 1-3 but does not recognise Pbx1b. Meis 1/2 recognises the N-terminus of Meis1 and 2 but will not recognise Meis 3. Meis1.2* recognises the same proteins as Meis1/2 but is at a higher concentration. None of these antibodies appear to interfere with the binding of nuclear proteins to the wildtype oligonucleotide. Note that the concentration of NE used in this assay was half of that used in previous EMSAs.

6.2.9 Meinox/Pbx family proteins are expressed in tissues associated with Gli3 expression

cDNAs for Pbx1a, 3b, 4, Prep1 and Prep2 were cloned and sequenced to confirm their identity. In addition, cMeis1 and cMeis2 cDNAs were included, although their expression pattern has been described previously (Mercader *et al.*, 1999). The sequence for cMeis3 and cPbx2 could not be found in Ensembl, Unigene, or NCBI databases and have not been investigated at this stage. Expression patterns of Pbx1a, Pbx3b, cPbx4, Prep 1, Prep 2, Meis 1 and Meis 2 were characterised by wholemount in-situ hybridisation using DIG-labelled RNA probes on chick embryos at HH stage 10 to 18. TALE proteins display overlapping yet distinct expression patterns in the paraxial, intermediate and lateral mesoderm, in the CNS and in the limb buds (Fig. 6.20-6.26).
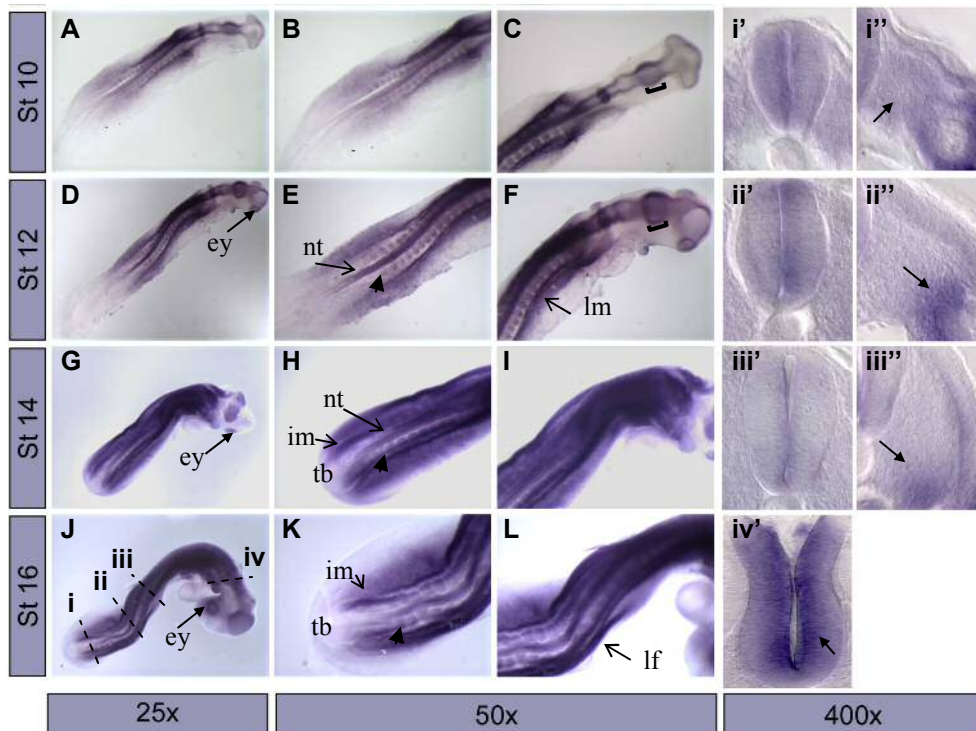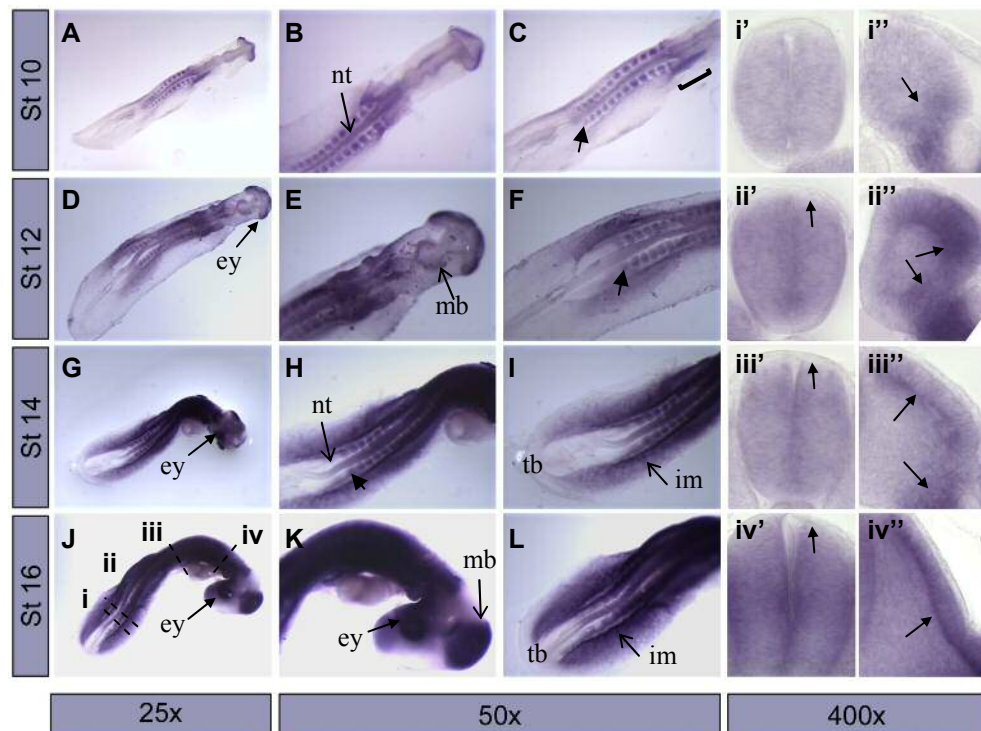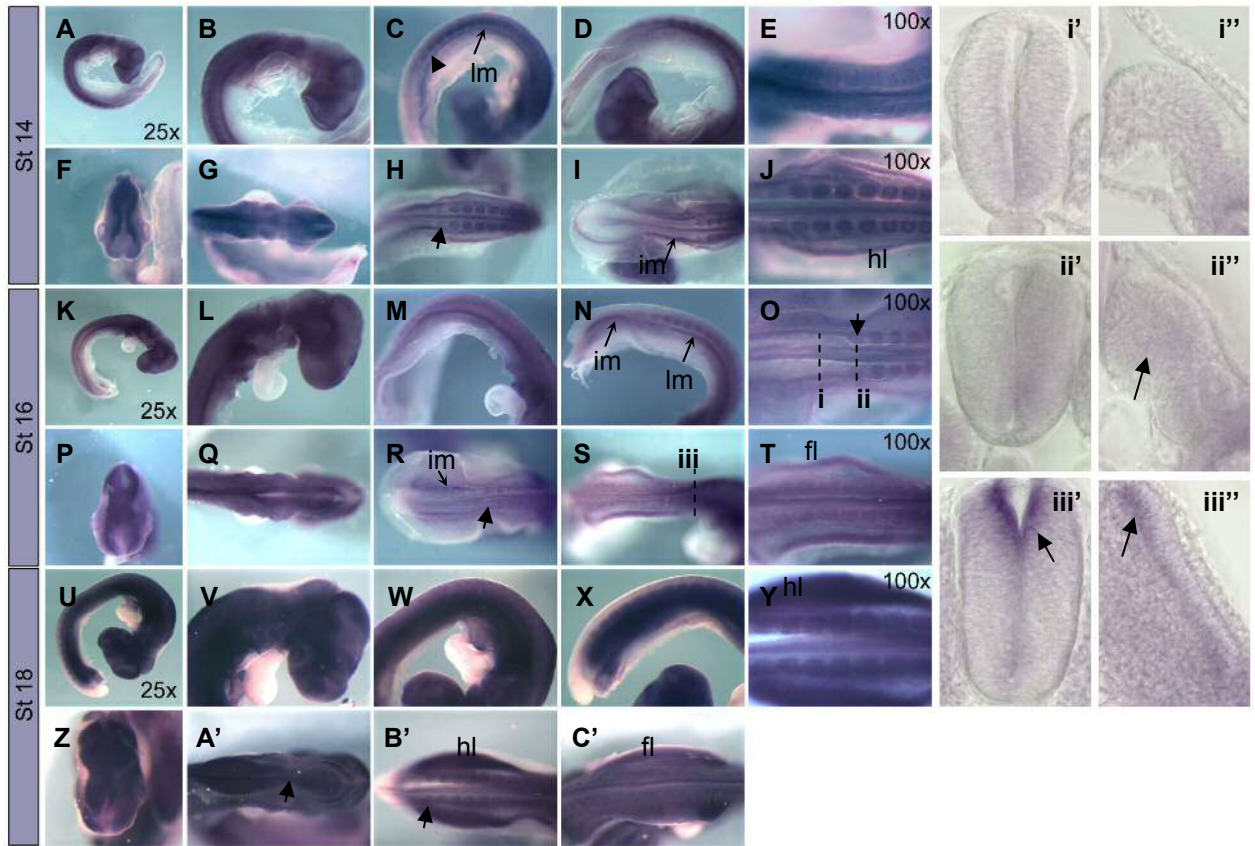
Figure 6.20: *Meis1 expression during development of the chicken embryo*

Wholemount RNA in-situ hybridisation of Meis1 in wildtype chicken embryos at HH stages 10 (A-C), 12 (D-F), 14 (G-I) and 16 (J-L). i-iv show transverse sections of the HH stage 16 embryo shown, taken at the positions indicated in J, 'denotes neural tube, '' denotes paraxial mesoderm. Meis1 is expressed in the paraxial mesoderm upon somite formation (black arrowheads) and in somite 0 (arrow, i''). Upon somite maturation expression is restricted to a ventrolateral sclerotome (Arrows, ii'', iii''). Expression in the neural tube (nt) also initiates around the time of somite formation (E), becoming stronger as development proceeds. In anterior regions expression is restricted to the ventricular zone (iv'). Elsewhere in the CNS Meis1 is expressed in the midbrain (brackets in C and F), and in the developing eye (ey). Expression is also observed in the hindlimb field (lf) and in the intermediate mesoderm (im). No expression is observed in the tail bud (tb).

Figure 6.21: *Meis2 expression during development of the chicken embryo.*
Wholemount RNA in-situ hybridisation of Meis2 in wildtype chicken embryos at HH stages 10 (A-C), 12 (D-F), 14 (G-I) and 16 (J-L). i-iv show transverse sections of the HH stage 16 embryo shown, taken at the positions indicated in J, 'denotes neural tube, '' denotes paraxial mesoderm. Meis2 is expressed in the paraxial mesoderm upon somite formation (black arrowheads), and becomes stronger as somites mature due to upregulated expression in the myotome (C, iii'', iv''). Expression in the neural tube (nt) also initiates around the time of somite formation (C, F, H, L), becoming stronger as development proceeds (i'-iv'). Expression is also observed the intermediate mesoderm (im), and is absent in the tailbud (tb). Other sites of expression include the developing eye (ey), and the midbrain (mb).
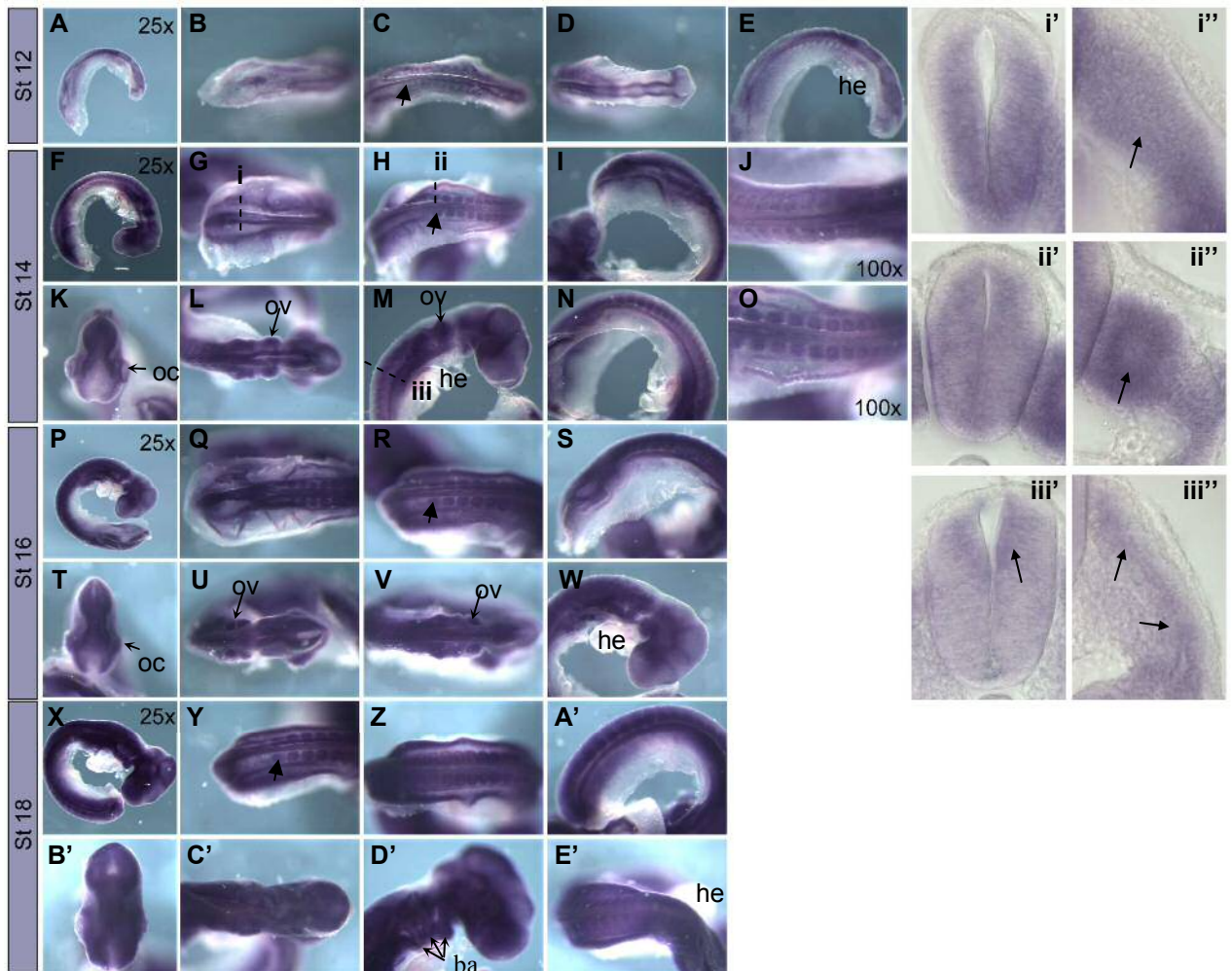
Figure 6.22: _Pbx1a expression during development of the chicken embryo._
Wholemount RNA in-situ hybridisation of Pbx1a in wildtype chicken embryos at HH stages 14 (A-J), 16 (K-T), and 18 (U-C'). Images were taken at 50x magnification unless otherwise stated. i-iii show transverse sections of the stage 16 embryo shown, taken at the positions indicated in O and S, ' denotes neural tube, '' denotes paraxial mesoderm, images were taken at 400x magnification. Pbx1 is expressed in the paraxial mesoderm, posterior intermediate mesoderm (im), lateral mesoderm (lm) and in the brain. In the paraxial mesoderm Pbx1 is expressed upon somite formation (black arrowheads). Initially expression is broad throughout the somite (ii''), but becomes restricted to the dorsal myotome upon differentiation (arrow, iii''), with weak expression in the sclerotome. At HH stage 18 expression is upregulated in the limb buds. In the forelimb (fl) expression is strongest in the anterior, whereas transcripts are detected throughout the hindlimb (hl). Expression is absent in the posterior neural tube, but expressed dorsally in anterior regions (iii').
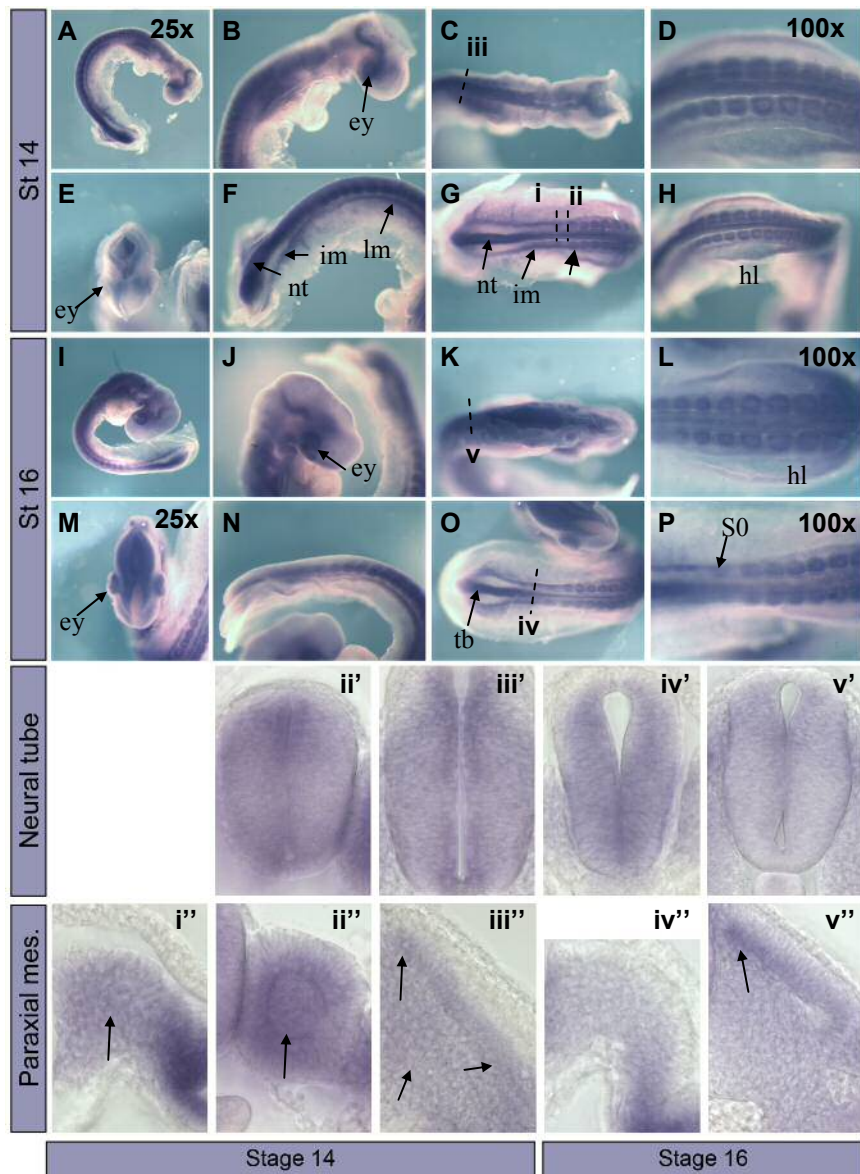
Figure 6.2: *Pbx3b expression during development of the chicken embryo.*
Wholemount RNA in-situ hybridisation of Pbx3 in wildtype chicken embryos at HH stages 10 (A-D), 12 (E-H), 14 (I-O) and 16 (P-W). Images were taken at 50x magnification unless otherwise stated. At HH stages 10-12, Pbx3 is expressed in the lateral mesoderm (lm), in the midbrain and in the forebrain (D and H). At HH stage 14 expression is also observed in the posterior intermediate mesoderm and in the anterior neural tube (nt). At HH stage 16, further expression is observed in the forebrain, the developing eye (ey), the septum transversum (st) and branchial arches. In the branchial arches expression is particularly strong in the anterior region of branchial arch 1 (ba1).

Figure 6.24: *cPbx4 expression during development of the chicken embryo.*
A-E', Wholemount RNA in-situ hybridisation of cPbx4 in wildtype chicken embryos at HH stages 12 (A-E), 14 (F-O), 16 (P-W) and 18 (X-E'). Images shown were taken at 50x magnification unless otherwise stated. i-iii show transverse sections of the HH stage 14 embryo shown, taken at the positions indicated in G, H and M, ' denotes neural tube, '' denotes paraxial mesoderm. Images were taken at 400x magnification. cPbx4 is expressed broadly in the early neural ectoderm and paraxial mesoderm. It is expressed in the presomitic mesoderm (i'') but is upregulated upon somite formation (black arrowheads in C, R and H). Upon somite maturation expression is restricted to the myotome (iii'''). In the lateral mesoderm expression is strong in the limb fields. Intitially, Pbx4 is expressed throughout the neural ectoderm, but becomes restricted as the neural tube matures. In anterior regions expression is confined to the dorsal ventricular zone (iii'). Strong posterior expression is observed in the intermediate mesoderm. In the head expression is observed in the forebrain, midbrain and hindbrain. In the otic vesicles (ov) expression is strongest at HH stage 16. Expression is also observed in the optic cup (oc) and in the branchial arches (ba). Notably expression is excluded from the developoing heart (he) at all stages.
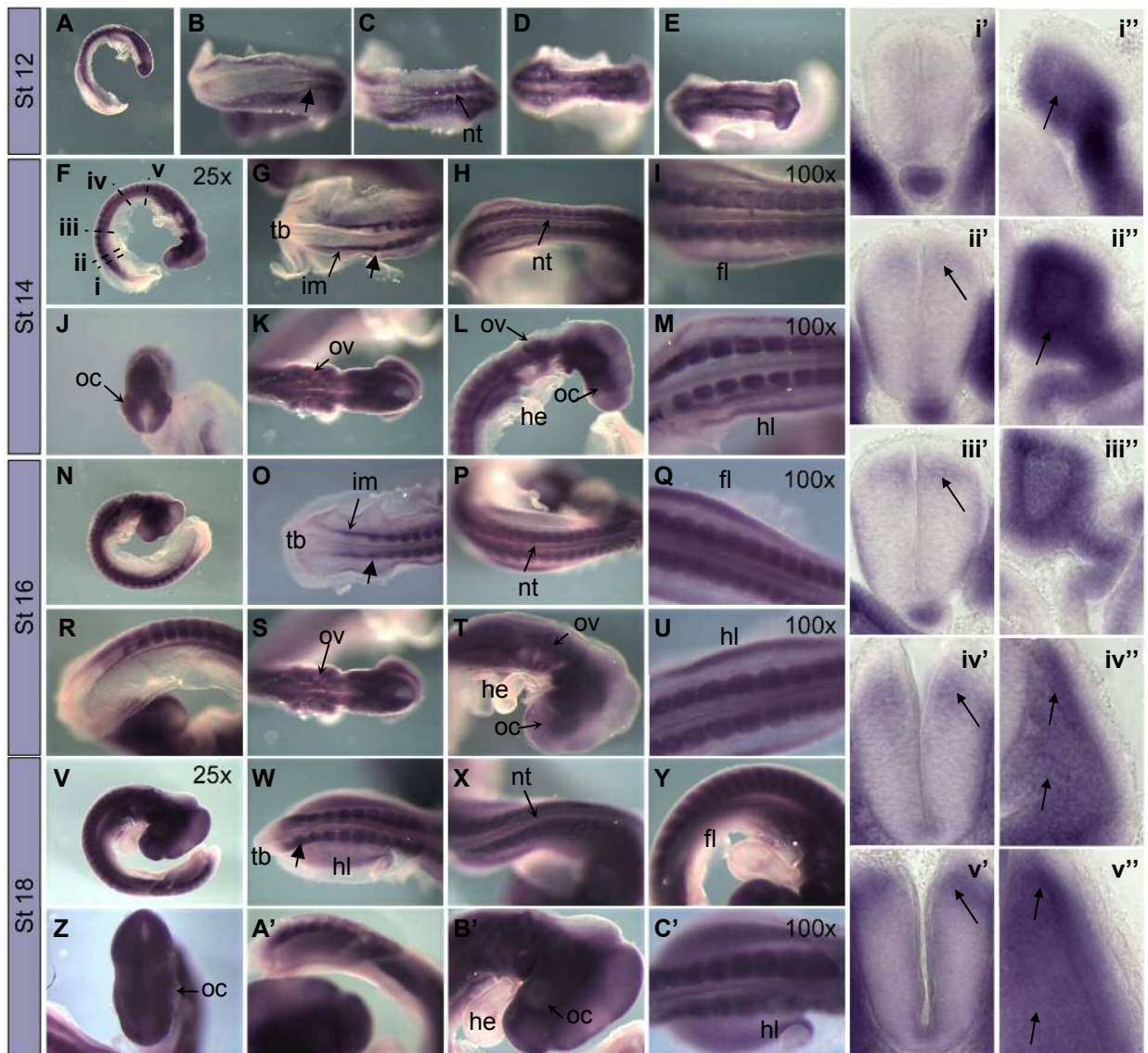
Figure 6.25: *Prep1 expression during development of the Chicken embryo.*
Wholemount RNA in-situ hybridisation of Prep1 in wildtype chicken embryos at HH stages 14 (A-H) and 16 (I-P). Images shown were taken at 50x magnification unless otherwise stated. i-v show transverse sections of the HH stage 14 (i-iii) and HH stage 16 (iv-v) embryos shown, taken at the positions indicated in C, G, K and O. Images were taken at 400x magnification. Prep1 is expressed in the neural tube (nt), intermediate mesoderm (im), and lateral mesoderm (lm) and laterally in the limb buds. Paraxial mesoderm expression initiates upon somite formation (black arrowhead, G). Expression is also observed in the posterior half of somite 0 (S0). Expression is upregulated throughout somite I (ii''). Upon differentiation, expression is maintained in the sclerotome and myotome (iii''). At HH stage 16 anterior somites express Prep1 predominantly in the dorsal myotome (v''). Prep1 is initially expressed throughout the neural tube, regressing towards the posterior as development proceeds. By HH stage 16 expression is strongest in the tail bud (tb), forming a gradient until the first somite. Expression is also maintained in the brain (K), and anterior neural tube (v'). Expression is observed in the lateral boundary of the hindlimb (hl, H), and is later confined to the anterior hindlimb (L). In the head expression is observed mainly in the midbrain (B, C, E, J, K, N). Expression is also strong in the develoing eye (ey).

Figure 6.26: *Prep2 expression during development of the Chicken embryo.*
A-C', Wholemount RNA in-situ hybridisation of Prep2 in wildtype chicken embryos at HH stages 12 (A-E), 14 (F-M), 16 (N-U) and 18 (V-C'). Images shown were taken at 50x magnification unless otherwise stated. i-v, Transverse sections of the HH stage 14 embryo shown, taken at the positions indicated in F. Images were taken at 400x magnification. Prep2 is expressed strongly in the intermediate, paraxial and lateral mesoderm and in the notochord. Paraxial mesoderm expression initiates in somite 0 (black arrowheads and i'). Prep2 is expressed strongly throughout the first somite (ii''). Upon differentiation, expression is maintained in the sclerotome and myotome, but is excluded from the dermomyotome (iv''). In more anterior somites expression is strongest in the dorsal mytome (v''). Transverse sections also reveal that Prep2 is expressed in the dorsal neural tube, initiating at the time of somite formation (ii'). Expression is also observed in the posterior intermediate mesoderm (im). In the lateral mesoderm expression is observed in the limb fields of both the forelimb (fl) and hindlimb (hl). In the head expression is observed in the forebrain, midbrain and hindbrain (E, J, L, T, Z, B'), in the otic vesicles (ov) and optic cup (oc). Notably expression is absent in the tail bud (tb) and heart (he).

Meis1 is expressed broadly in the lateral mesoderm and the intermediate mesoderm (Fig. 6.20 E, F, K). Expression is upregulated at HH stage 16 in the limb field (Fig. 6.20L). In the paraxial mesoderm expression is not detected in the presomitic mesoderm and is activated prior to somite formation (Fig. 6.20 B, E, H). Although initially broadly expressed throughout the somites, expression is ultimately restricted to the ventrolateral sclerotome (Fig. 6.20 ii'', iii''). In the CNS, Meis1 mRNA transcripts are observed in the neural tube, midbrain, and in the eye (Fig. 6.16 A-L). Transverse sections show that although initially expressed throughout the neural tube, Meis1 transcripts become restricted to the ventricular zone (Fig. 6.20 iv').

Meis2 is expressed broadly in the intermediate mesoderm, as well as in the neural tube and paraxial mesoderm (Fig. 6.21 C, F, H, L). As for Meis1, expression in the paraxial mesoderm initiates upon somite formation, and becomes stronger as tissues mature (Fig. 6.21 C, H). Transverse sections show that Meis2 is initially expressed throughout the mesoderm, but is higher in the ventro-lateral domain (Fig. 6.21 i''). Upon differentiation, expression is upregulated in the myotome, but is maintained at low levels throughout the sclerotome (Fig. 6.21 iv'). In the CNS, expression is observed in the neural tube, midbrain and in the developing eye (Fig. 6. 21 E, G, K). In the neural tube, expression initiates at a position adjacent to somite 0 (Fig. 6. 21 H). Transverse sections show that Meis2 is expressed throughout the DV axis, with the exception of the roofplate (Fig. 6.21 i'-iv').

The results presented here suggest that Meis1 and Meis 2 are co-expressed in the neural tube, intermediate mesoderm, paraxial mesoderm and midbrain. Co-expression has also been reported in proximal domains of mouse and chick limb buds (Capdevila *et al.*, 1999; Mercader *et al.*, 1999). Similar expression patterns have been reported mouse embryos (Cecconi *et al.*, 1997; Mercader *et al.*, 2000). However, neural expression of the two proteins appears to have diverged between chicken and mouse. Meis1 does not appear to be strongly expressed in the mouse neural tube (Mercader *et al.*, 2000), and Meis2 mRNA is confined to the most dorsal part of the neural tube in mouse embryos (Cecconi *et al.*, 1997).

Pbx1a is expressed in the paraxial, intermediate and lateral mesoderm. Expression is up-regulated at HH stage 18 in the hind-limb bud and in the anterior forelimb bud (Fig. 6.22 Y, B', C'). In the paraxial mesoderm expression initiates upon somite formation (Fig. 6.22 H, R, Y, i''. ii''). Although initially expressed at low levels throughout the somite (ii''), upon differentiation expression is upregulated in the myotome (iii''). Transverse sections show dorsomedial expression of Pbx1a mRNA in the anterior neural tube (iii'). In the CNS expression is also observed in the brain vesicles, and in the otic vesicle (Fig. 6.22, F, G, P, Q, Z, A'). This expression pattern is similar to that described of murine Pbx1b, which was identified as the main isoform expressed during mouse development (Schnabel *et al.*, 2001).

Pbx1a expression in the mouse has not been described in detail, but is primarily restricted to developing neural tissues. This suggests a possible divergence between Pbx1a expression profiles between mouse and chicken embryos.

Pbx3 is expressed strongly in the lateral mesoderm and rostral regions of the CNS at HH stages 10 – 16 (Fig. 6.23). From HH stage 14 Pbx3 mRNA is also detected in the intermediate mesoderm (Fig. 6.23 K, M), and at HH stage 16 several anterior regions express the transcript, including the septum transversum, branchial arches (Fig. 6.23 V), and developing eye (Fig. 6.23 P, T). In the lateral mesoderm expression is particularly strong in tissues that will give rise to the forelimb (Fig. 6.18 R, U), but is absent in regions that will give rise to the hindlimb (Fig. 6.23 Q, W). In the CNS Pbx3 is expressed in the hindbrain from early stages (Fig. 6.23 C, D), midbrain expression initiates at stage 12 (Fig. 6.23 H). Both spatially and temporally expression domains are similar to those reported in mouse (Di Giacomo *et al.*, 2006).

Chick Pbx4 is broadly expressed in the paraxial mesoderm, CNS, lateral mesoderm and posterior intermediate mesoderm (Fig. 6.24). Expression is upregulated at HH stage 16 in the limb field (Fig. 6.24 R, Z, E'). In the paraxial mesoderm expression is activated prior to somite formation (Fig. 6.24 G, H, Q, R, Y, i''). Although initially broadly expressed throughout the mesoderm, expression is upregulated upon epithelialisation, and is ultimately restricted to the myotome (Fig. 6.24 ii'', iii''). In the CNS cPbx4 is initially expressed throughout the neural ectoderm, progressively diminishing upon maturation of the neural tube (Fig. 6.24 G, Q, O, Y). Expression remains strong in the tail bud, in brain vesicles and in the optic capsule (Fig. 6.24 K, Q, T, C'). Transverse sections show that upon maturation of the neural tube cPbx4 mRNA transcripts become restricted to the dorsal ventricular zone (iii'). Expression is strong in the otic vesicle from HH stage 14 (Fig. 6.24 L, M), and at HH stage 18 expression is also observed in the branchial arches, particularly in anterior regions (Fig. 6.24 D'). As described previously cPbx4 contains an additional coding exon that results in extension of the C-terminus. The expression profile of cPbx4 differs from murine Pbx4, which was not detected by in-hybridisation on mouse embryos at E7-E10.5, and is testes specific in the adult (Wagner *et al.*, 2001). cPbx4 could represent a common Pbx4 isoform that has not yet been identified in mammals, or could be unique to avians. It remains to be seen whether an isoform orthologous to murine Pbx4 is expressed at later stages of chick development.

These data confirm that Pbx proteins are expressed in distinct yet overlapping patterns in the developing chick embryo. RT-PCR analysis has shown that various isoform of Pbx1 and 3 are expressed, and Pbx4 may also encode multiple isoforms. In mouse embryos Pbx1a and 1b are differentially expressed (Schnabel *et al.*, 2001), as are the various Pbx3 isoforms (Milech *et al.*,

2001; Di Giacomo *et al.*, 2006). Thus dynamic expression profiles of Pbx family members, and of their various isoforms may contribute to differences in the expression of target genes.

Prep1 is expressed strongly in the CNS and paraxial mesoderm (Fig. 6.25). In the neural tube, expression initially spans the AP axis (Fig. 6.25 C, G), but regresses towards the posterior as development proceeds (Fig. 6.25 L, O, P, also compare iii' to v'). By HH stage 16 expression is strongest in the tail bud, forming a gradient up to the first somite (Fig. 6.25 O, P). Expression is also maintained in the brain and is particularly strong in the midbrain (Fig. 6.25 K, M). In the lateral mesoderm expression is observed in the lateral boundary of the hindlimb, and by HH stage 16 is confined to the anterior hindlimb (Fig. 6.25 H, L). Paraxial mesoderm expression is apparent upon somite formation (Fig. 6.25 G, O), but is also observed in the posterior half of somite 0 (Fig. 6.25 P, i'). Upon differentiation Prep1 mRNA expression is maintained in the sclerotome and myotome (Fig. 6.25 iii''). Ultimately expression is restricted to the dorsal myotome (Fig. 6.25 v''). Prep1 is also expressed in the developing eye (Fig. 6.25 E, J, M). This dynamic expression pattern is in contrast to the ubiquitous expression of Prep1 mRNA transcripts reported in mouse embryos (Ferretti *et al.*, 1999). However, a detailed analysis of Prep1 expression has not been reported previously.

Prep2 is expressed strongly in the paraxial mesoderm, intermediate mesoderm and limb fields (Fig. 6.26 G, I, M, O, Q, U). At HH stage 16 expression is upregulated in the limb field (Fig. 6.26 Q, U). Expression is excluded from the presomitic paraxial mesoderm, but is activated prior to somite formation, in somite 0 (Fig. 6.26 G, O, W, i''). Transverse sections show that Prep2 mRNA transcripts are upregulated throughout somite I (Fig. 6.26 ii''). Upon differentiation, expression is maintained in the sclerotome and myotome, and is particularly strong in the dorsal myotome (Fig 6.26 iv'' and v''). In the CNS Prep2 mRNA transcripts are expressed in ventral regions of the midbrain, forebrain, and hinbrain, in the otic vesicle, and in the optic cup (Fig. 6.26 J-L, S, T, Z-B'). Transverse sections reveal that Prep2 is expressed at low levels in the dorsal neural tube, initiating adjacent to the first somite (ii'). In the mouse embryo expression has previously been described at E10.5 in the brain, spinal cord, tail somites, forelimb bud, heart and pharyngeal arch (Haller *et al.*, 2002). Notably in the chicken embryo expression is absent in the heart (Fig. 6.26 L, T, B').

The results presented here demonstrate that Prep1 and Prep2 are differentially expressed in the chick embryo. Whilst similar expression patterns are observed in the paraxial and intermediate mesoderm, mRNA levels differ in the neural tube and lateral mesoderm. Prep2 is expressed throughout the limb fields, whilst Prep1 is confined to the anterior hindlimb at HH stage 16. In the neural tube Prep1 is expressed predominantly in the posterior, whilst Prep2 is expressed at low levels in the dorsal neural tube anterior to the first somite.

The expression of Meis, Pbx and Prep proteins in the trunk of the chicken embryo is summarised in Figure 6.27. In the CNS, TALE family proteins are differentially expressed along the AP axis. Pbx4 together with Prep1 is expressed strongly in the tail bud, with lower levels expressed towards the anterior. Prep2 is absent in the caudal neural tube, but low levels of expression are observed in the dorsal third of the neural tube anterior to somite I. Similarly, Pbx1a is expressed in dorsal regions of the anterior neural tube. Meis1 and Meis2 are expressed throughout the neural tube, anterior to somite 0. In the brain vesicles Pbx and Prep proteins are broadly expressed, although Prep1 is confined to ventral aspects. Meis1 and Meis 2 are strongly expressed in the midbrain, but are absent in the forebrain and hindbrain.
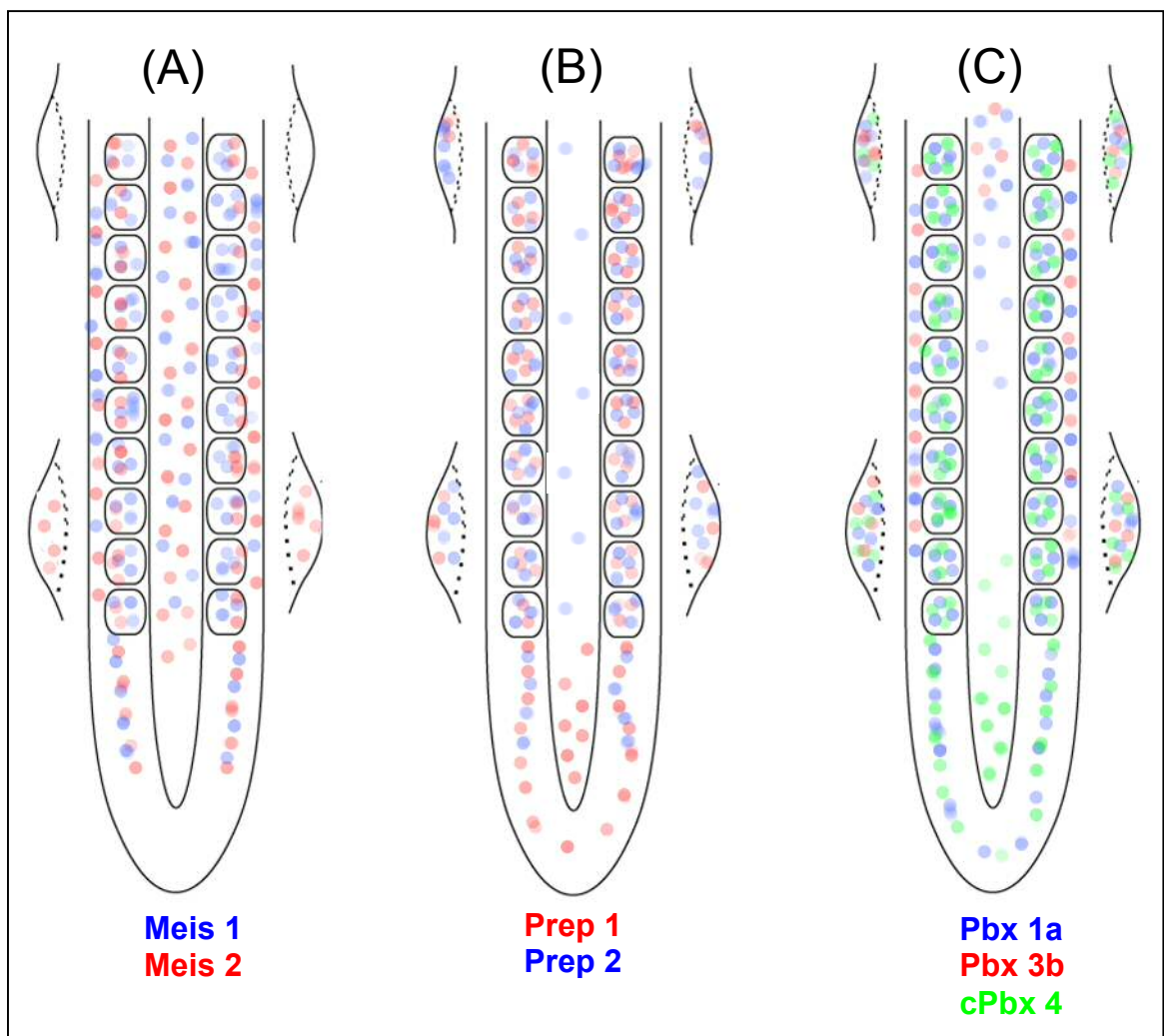
Figure 6.27: *Summary of TALE protein expression patterns in the developing chicken embryo.* Expression patterns in the trunk of the embryo, and in the limb buds is shown. Coloured dots are used to mark regions expressing the respective protein. Expression of members of the Meis (A), Prep (B) and Pbx (C) protein families as shown in Figures 6.20-Fig. 6.26 are indicated. cPbx 4 and Prep2 are of particular interest since they are the only proteins expressed in the posterior neural tube where Region1consA is upregulates expression. Meis proteins are expressed in the neural tube from around the time of somite formation, and could control the Anterior/Posterior pattern of expression due to Region1consA transcriptional activity.

6.3 Discussion

In this chapter I have investigated a possible molecular mechanism underlying Region1consA transcriptional activity. The conserved region contains binding sites for TALE family transcription factors, and consistent with this observation I found that manipulating Meis activity alters reporter gene expression under the control of Region1consA. Mutations of these TALE family binding sites that prevent protein binding in vitro, cause also an up-regulation of reporter gene expression in vivo.

In particular I have also shown that nuclear proteins bind both the Meinox and Pbx halfsites of the composite site. Antibodies specific to Pbx1a, 2 and 3a, or Meis1/2 did not disrupt the complexes formed, suggesting that other TALE family proteins are associated with the MeisPbx binding site. Preliminary evidence implicates cPbx4 and Prep1 as potential binding partners of Region1consA, because they are expressed in the domains of reporter gene expression.

6.3.1 Meis proteins regulate the activity of Region1consA

Electroporation of activator and repressor forms of Meis1a indicates that Meis1 affects both the initiation of Region1consA expression and its maintenance. Consistent with this, mutation of the MeisPbx 'Site 2' causes a down-regulation of reporter gene expression, which could be caused by reduced initiation. Indeed, EMSA experiments show that mutation of Site 2 abrogates almost all protein binding to the 35bp oligo, suggesting that binding is required for high levels of reporter gene expression. It should be noted that since all Meinox proteins recognise the same consensus sequence, the Meis activator and repressor constructs used could reflect other Meinox protein activities.

Conversely, the presence of a second mutation in the Pbx consensus sequence 'Site 1' results in an upregulation of reporter gene expression, although it does not appear to disrupt protein binding in vitro. Despite this, mutation of Site 1 behaves in a manner dominant to that of Site 2, since the MPP double mutant construct also causes an upregulation of reporter gene expression in vivo. Although puzzling at first, the observations have several possible explanations. First, the mutations introduced in the Region1consA construct may create novel binding sites for factors that influence transcriptional activity. Second, abrogation of one of the two binding sites investigated here could disrupt chromatin remodelling, which has been associated with TALE protein activity. Finally, Site 1 could inhibit the activity of Site 2 by preventing its association with transcriptional complexes required for activation, whilst not itself involved in protein binding.

6.3.2 Binding sites created or lost in both Pbx and MPP constructs cannot account for up-regulation of reporter gene expression

Table 6.4 summarises the binding sites predicted by Genomatix to be created and disrupted in each of the mutant constructs investigated by EMSA. Pbx and MPP constructs share in common a novel binding site 'V$ETSF' created, and the V$CAAT and V$GFII binding sites disrupted by the mutation introduced in the Pbx site. There is a possibility that the gain or loss of these sites could account for the up-regulation of transcriptional activity observed.

V$ETSF is a binding site for human and murine Ets1 factor, which regulates the expression of oncogenes and tumour suppressors by activating transcription (Sharrocks *et al.*, 1997). However, neither Ets1 nor other members of the Ets family are expressed in the neural tube, suggesting that they are not responsible for upregulating reporter gene activity (Fafeur *et al.*, 1997). V$CAAT and GFI1 are binding sites for NF-Y and GFI1 respectively, and are disrupted in the Pbx and MPP mutant constructs. However, no loss of complex was observed in EMSA experiments performed with labelled oligos carrying the Pbx mutation, suggesting that these sites are not responsible for the up-regulation observed.

6.3.3 TALE binding could cause changes in the chromatin structure of Region1consA

One mechanism by which TALE family proteins are known to function is by accessing regions of condensed chromatin, where they mark genes for activation by allowing other factors to access DNA (Berkes *et al.*, 2004; Maves *et al.*, 2007). In the case of the myogenin promoter, binding of Meis and Pbx proteins to adjacent sites is required for activation upon MyoD expression (Berkes *et al.*, 2004). MyoD recruits SWI/SNF (Switch/Sucrose NonFermentable) members of the chromatin remodelling complex, causing conformational changes in the DNA structure that are necessary for activation of the myogenin locus prior to muscle differentiation. However, the authors also speculate that the complex represses myogenin expression in some cellular contexts, owing to a change in the expression of Meis/Pbx isoforms (Berkes *et al.*, 2004). Indeed, Pbx isoforms display different transcriptional activities depending on their association with co-factors. For instance, when complexed to Pdx1, Pbx1a acts as a transcription repressor through its association with the co-repressors NcoR (Nuclear receptor corepressor) and SMRT (silencing mediator for retinoid and thyroid hormone receptors). In the absence of this interaction, the complex functions as a transcriptional activator via an activation domain carried by Pdx1 (Asahara *et al.*, 1999; Goudet *et al.*, 1999).

NcoR1 and SMRT appear to establish a protein-protein link between DNA bound transcription factors and histone deacelylases (HDACs) (Pazin and Kadonaga, 1997). The SWI/SNF enzymes recruited by MyoD have the potential to interact with both HDACs and histone actltransferases

(HACs), and thus have the potential to co-ordinate chromatin-remodelling activities (Martens and Winston, 2003; Sif, 2004). Additionally, one of the repressor domains in the N-terminus of Pbx proteins has been shown to associate directly with class 1 HDACs (Saleh *et al.*, 2000b), as have C-terminal motifs of TGIF1 and 2 (Wotton *et al.*, 1999; Melhuish *et al.*, 2001). Thus, TALE family proteins are capable of regulating transcriptional activity by recruiting components of the chromatin remodelling machinery.

This leads me to propose a two step model for the regulation of Region1consA transcriptional activity (Fig. 6.28). In the first step (step A), association of TALE family proteins with Region1consA is required to repress transcription, mediated by their association with HDAC complexes (Fig. 6.28 A). In the second step (step B), TALE proteins bind to Site 2 and associate with other co-factors, the abundance of which will determine transcriptional activity (Fig. 6.28 B). Proteins that are known to interact with TALE family proteins, and could influence transcriptional activity include other homeodomain proteins and basic Helix-loop-helix transcription factors, which are widely expressed in the developing neural tube (Chang *et al.*, 1995; Peltenburg and Murre, 1996; Jacobs *et al.*, 1999; Knoepfler *et al.*, 1999; Liu *et al.*, 2001c; Laurent *et al.*, 2008).
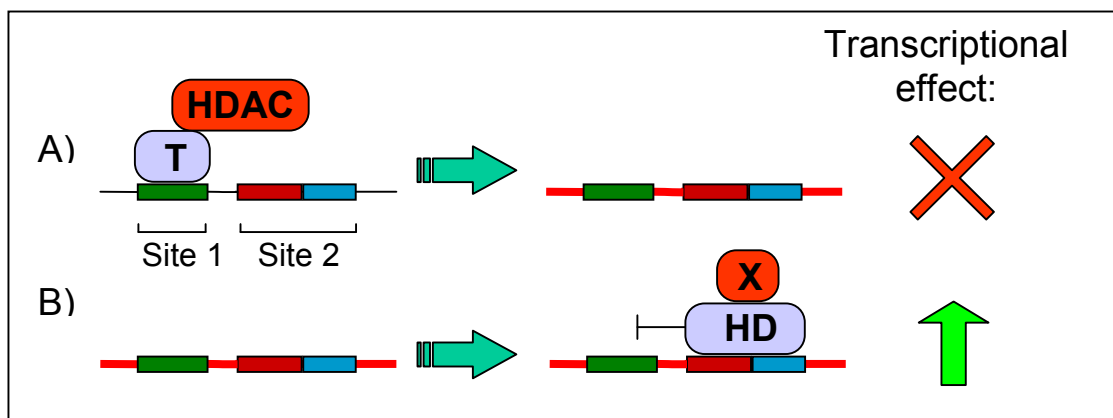
Figure 6.28: *Model for regulation of Region1consA transcriptional activity.* Binding sites within the 35bp region analysed by EMSA are represented by coloured boxes. Site 1 contains a Pbx binding site (green), Site 2 contains a composite binding site composed of a Meinox half-site (red) and a Pbx half-site (blue). A two step model is proposed. In step A, Site 1 is bound by a TALE family complex (T) that associates with histone deacetylase complexes (HDAC). HDACs mediate transcriptional repression by inducing a conformational change in chromatin structure, condensed chromatin is represented in red. In step B, TALE family proteins associate with Site 2, and in doing so bloc k access to Site 1. TALE family proteins are represented as heterodimers (HD), although EMSA experiments suggest that monomers can also bind. TALE family heterodimers associate with co-factors (X) and upregulate transcriptional activity.

Site 1 may be required transiently to induce the formation of a condensed chromatin environment (Fig. 6.28). The transient nature of this binding could account for the apparent absence of protein complexes associated with Site 1 in EMSA assays, whilst a requirement for repressing the transcriptional activity of Region1consA would cause upregulation of reporter gene expression in its absence. EMSA experiments provide two results that support this model.

First, I showed that Site 1 mutation has a minimal effect on complex formation. Second, a novel complex with a high molecular weight forms on radioactively labelled oligos upon mutation of Site 2. This novel complex could represent the association of HDACs with a complex binding to Site 1. Indeed, binding is extremely weak in the absence of functional Site 1. Binding of the high molecular weight complex is inhibited by the presence of a functional Site 2. Thus, in the second step of my model complex formation on Site 2 is proposed to inhibit the association of HDACs with Site 1 (Fig. 6.28 B). This could also account for the inability of wildtype oligo to compete binding of complexes C and D on MPmut oligos.

According to this model, mutation of Site 1 would block the repression exerted on the chromatin surrounding Region1consA (step A), facilitating complex formation on Site 2. In vivo, this up-regulates transcriptional activity. Mutation of Site 2 allows condensation of the chromatin environment, but prevents binding of factors required for activation (step B). Consequently, transcriptional activity is down regulated.

6.3.4   Up-regulation of transcriptional activity in MPP mutants could be due to the creation of a Pax3 binding site

My model predicts that the MPP mutation (in which both Site 1 and Site 2 are disrupted) would result in an open chromatin environment, permissive to protein binding (Fig.6.28 A). However, in the absence of Site 2, TALE family proteins would no-longer bind, and thus their co-factors would be unable to regulate transcriptional activity. Other factors must therefore account for the increased transcription observed. Binding site analysis reveals the formation of Pax3 and GCM binding sites in both PbxMut and MPP mutant constructs. Pax3 is expressed in the dorsal neural tube from HH stage 8 (Goulding *et al.*, 1991). Therefore, Pax3 could mediate transcriptional activity in the absence of Sites 1 and 2. Conversely, the model proposes that in the MPmut construct chromatin will remain in a condensed state, which can account for the down-regulation of reporter gene expression observed.

6.3.5   AP1 sites may convey activator function to Region1consA

An alternative explanation for the upregulation of reporter gene activity upon mutation of Site 1 is that the open chromatin environment extends beyond the 35bp highly conserved region, such that the mutation facilitates protein binding in a broader region. AP1, CDE and IRF4 binding sites are present in Region1consA outside of the 35 bp oligo. The (C/G)GCGG consensus of the core CDE binding site (cell-cycle dependant element) is not contained in the 3' end of Region1consA (Lucibello *et al.*, 1997), whilst IRF4 (Interferon regulatory factor 4) is predominantly expressed in the immune system and has not been characterised as a

regulator of neuronal gene expression (Mamane *et al.*, 1999). However, three overlapping AP1 sites are of interest.

The AP1 (activator protein 1) transcription factor consists of dimers of the Fos (Fos, Fra1, Fra2 and FosB) and Jun (Jun, JunB and JunD) families of basic leucine zipper domain proteins, together with the more recently discovered Maf family proteins (Karreth *et al.*, 2004). AP1 has been implicated in the regulation of differentiation, proliferation, apoptosis and oncogenic transformation (Jochum *et al.*, 2001). Numerous AP1 family proteins are expressed in the developing neural tube, including MafA, c-Jun and Fra2 (Katsuoka *et al.*, 2003; Karreth *et al.*, 2004; Lecoin *et al.*, 2004). In particular, the expression pattern of MafA in the developing CNS resembles that of Gli3 (Lecoin *et al.*, 2004). Thus, in an open chromatin environment, AP1 could bind to Region1consA and drive transcription.

Indeed, in the reporter construct used in which the β-globin promoter is located in close proximity to Region1consA, it is feasible that the open chromatin induced by mutation of Site 1 could spread to the promoter, and facilitate assembly of the pre-initiation complex.

Binding site	WT	Pbx	MeisPbx	MPP	PbxB	MPPPax6
V$CAAT	✓	X	✓	X	✓	X
V$GFII	✓	X	✓	X	✓	X
V$HOXH	✓	✓	X	X	✓	X
V$MYT1	✓	✓	X	X	✓	X
V$PAX6	✓	X	X	X	X	X
V$PBXC	✓	✓	X	X	X	X
V$TALE	✓	✓	X	X	✓	X
V$ETSF	X	✓	X	✓	X	✓
V$BRNF	X	X	✓	X	X	✓
V$PAX3	X	X	✓	✓	X	X
V$GCMF	X	X	✓	✓	X	X
V$HOXC	X	X	X	X	X	✓
Transcriptional activity	moderate	high	low	High	N/A	N/A
Species A	+	+	(v.weak)	(v.weak)	weak	v.weak
Species B	+	+	(v.weak)	(v.weak)	weak	v.weak
Species C	+	+	(v.weak)	(v.weak)	weak	v.weak
Species D	-	-	(v.weak)	(v.weak)	?	?

Table 6.4: *Comparison of binding sites identified by MatInspector in each of the mutated oligonucleotides used in EMSA experiments.* Each construct investigated by EMSA was searched for transcription factor binding sites using MatInspector (Cartharius *et al.*, 2005). Binding sites present in the unmutated oligo are shown in bold. Note that MatInspector does not recognise the presence of a Pbx binding site in Site 1, although the sequence perfectly matches the consensus (Chang *et al.*, 1996). Bands observed in EMSA experiments are shown (Species A-D), as is the overall level of reporter gene activity for constructs investigated in vivo. Pbx and MPP mutations each cause upregulation of reporter gene expression in the chick neural tube. They share a binding site for V$ETSF that does not exist in the WT sequence. The MeisPbx mutation results in a down regulation of reporter gene expression, which could be mediated by the creation of a BRNF binding site.

6.3.6 Down regulation of reporter gene expression in MeisPbxmut could be due to the creation of a Brn-2 binding site

This alternative model, where factors outside of the 35bp oligo studied control Region1consA activity, presents the caveat that differences between the MeisPbx and MPP constructs must account for the downregulation observed of MeisPbx mutant constructs. Noticeably, the MeisPbx construct carries a Brn-2 binding site that is not present in the MPP construct (Table 6.4). Brn proteins are members of the Pou III family of transcription factors, which bind DNA via a central Pou domain, composed of a Pou-specific region and a Pou homeodomain (Verrijzer and Van der Vliet, 1993). These proteins are expressed in the embryonic and adult CNS, and function in the proliferation and specification of neuronal cell types. Brn-2 Expression is highest in the anterior CNS, and by HH stage 10 forms a gradient that decreases towards the posterior neural tube (Pruitt *et al.*, 2004; Lan *et al.*, 2006). Brn-2 alone is a transcriptional activator, whereas in association with co-factors it is converted into a repressor (Bert *et al.*, 2000). Thus, Brn-2 repressor activity may mediate the down-regulation of transcriptional activity observed for the MPmut construct. In particular, protein binding to a newly generated site in the MeisPbx construct could also explain the novel complex observed when a radioactively labelled MeisPbx mutant probe is used in EMSA experiments (Fig. 6.18).

6.3.7 How can TALE family proteins account for the spatial and temporal expression pattern regulated by Region1consA?

Figure 6.27 shows a summary of Pbx, Meis and Prep expression patterns in the developing chicken embryo. TALE family proteins are expressed in domains overlapping that of Gli3, including the developing limb, somites and neural tube, suggesting that they may be involved in regulating Gli3 expression in these tissues. In the posterior neural tube, where Region1consA activity is highest, Prep1 and cPbx4 are co-expressed, and thus are potential candidates for controlling the onset of Region1consA activity. In contrast, Meis1, Meis2 and Prep2 are activated in the neural tube at the level of somite formation, when reporter gene expression from Region1consA becomes restricted to the dorsal neural tube. This expression pattern suggests that Meis1/2 or Prep2 are involved in Gli3 dorsalisation. Thus, it is tempting to speculate that Gli3 expression may be regulated by complexes that contain distinct combinations of individual TALE family proteins, that form at different positions along the AP and DV axis of the developing embryo. I propose that complexes containing Pbxa and/or TGIF proteins, which are known to interact with HDAC, bind to Site 1 and cause repression my modifying the chromatin environment (Wotton *et al.*, 1999; Saleh *et al.*, 2000b; Melhuish *et al.*, 2001). In the posterior neural tube TALE family proteins associate with Site 2 and

upregulate transcriptional activity by association with other co-factors. TALE proteins usually associate with homeodomain and bHLH proteins (reviewed in Laurent *et al.,* 2008), although a recent report shows that TALE proteins associate with Smad proteins to regulate the expression of the follicle stimulating hormone β (Fsh-β) gene (Bailey *et al.*, 2004). Thus, dorsalisation of Gli3 expression could be mediated by the transcriptional activity of Region1consA in the dorsal neural tube and controlled by Smad proteins associated with TALE family members.

Meis proteins are expressed in regions of the neural tube in which Region1consA reporter gene activity is repressed. Thus, a mechanism through which A/P patterning of Region1consA activity could occur is via the sequestration of proteins required for activity. It will be interesting to examine the expression patterns of other TALE family genes to gain further insight into the mechanism controlling Region1consA activity. Of particular interest are the Pbx b isoforms which cannot bind NcoR and SMRT, although they can compete for binding with Pbx a isoforms in tissues where they are co-expressed (Asahara *et al.*, 1999; Goudet *et al.*, 1999). Data presented here suggest that Pbx3b is not expressed in the chick neural tube. However, Pbx1b expression has not been investigated. The b isoforms of Pbx proteins are also not recognised by the pan-Pbx antibody used in EMSA assays, and thus they remain likely candidates for binding to Site 2.

Figure 6.29 summarises my views and hypotheses on the proteins binding to, and regulating the activity of the 35bp segment of Region1consA. EMSA assays have demonstrated that three complexes bind to Site 2, which differ by a single nucleotide from the MeisPbx consensus. Upon mutation of Meinox or Pbx half sites none of these complexes are able to form (Fig. 6.29, c-f). Although the identity of the proteins binding to Region1consA could not be determined precisely, my evidence support the possibility that Pbx1b or Pbx4 together with Prep proteins control the transcriptional activity of Region1consA by the mechanism proposed in Figure 6.28. The smallest complex observed (A) may represent the binding of a monomeric protein (M). EMSA results suggest that of the complexes binding Site 2, this complex has the lowest affinity, as is expected of monomeric TALE proteins (Lu and Kamps, 1996). Band B is likely to represent a Pbx:Meinox heterodimer (HD), whilst Band C represents a higher order complex consisting of a heterodimer and other co-factor(s) (X). Band D, which is produced only upon mutation of Site 2, is likely to represent the association of Band C with histone deacetylase machinery (HDAC; Fig. 6.29 a, b, c).

The change in reporter gene activity mediated by Region 1consA along the neural tube corrolates with a change in the expression profiles of Meis, Pbx and Prep proteins. The various TALE family members display differences in their binding affinity, both with protein

partners and with DNA (Discussed later in Chapter 7). Thus, the combination of TALE family proteins expressed in a particular tissue and at a particular stage in development will determine the nature of TALE family complexes formed. In addition to determining the binding affinity, this will affect the ability of a complex to associate with co-factors, and will determine the size of the complex formed. It should be noted that each band observed in EMSA could also represent distinct complexes migrating at a similar rate.
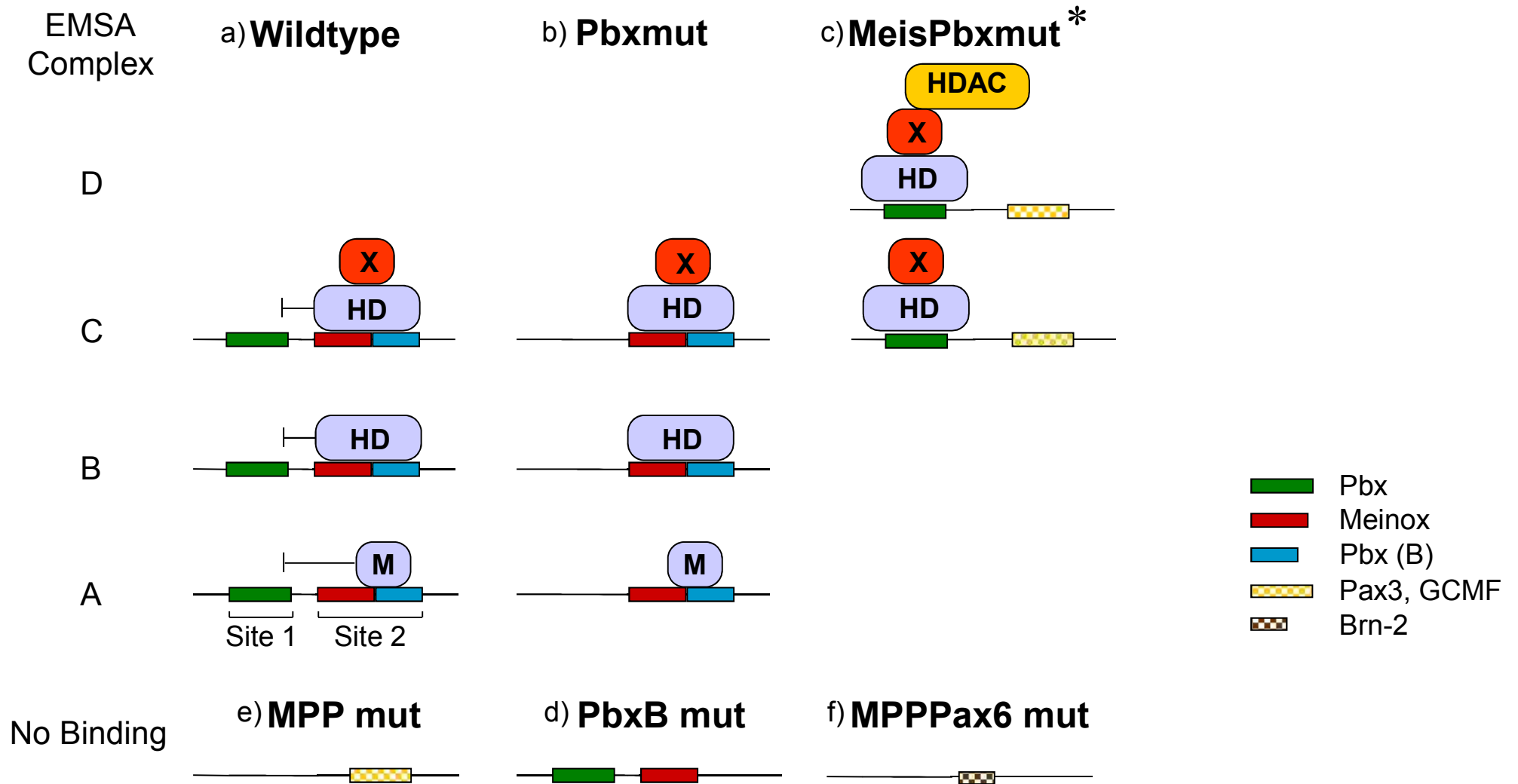
Figure 6.29: *Model for the transcriptional regulation of Region1consA.* Binding sites within the 35bp region analysed by EMSA are represented by coloured boxes, and are predicted to bind the factors shown. Site 1 contains a consensus Pbx binding site, Site 2 contains a composite binding site composed of Pbx and Meinox half sites. Brn-2, Pax3 and GCMF binding sites referred to in the text that are created by the mutations introduced are shown by chequered boxes. The complexes binding to each oligo used in EMSA assays are shown (a-f). Four different complexes (A-D) associate with the oligos. Complex A represents monomeric binding of TALE family transcription factors (M). Complex B is of a higher molecular weight and is likely to represent the binding of a TALE protein heterodimer (HD). Complex C could represent the association of a TALE protein heterodimer with one or more co-factors (X). Complexes A-C associate with both Pbx and Meinox half sites of Site 2, and in doing so inhibit binding to Site 1. Radioactively labelled MeisPbx mut oligos bind nuclear extracts to produce low levels (*) of bands C and D. Band D is likely to represent the association of the band C complex with HDAC proteins.

Chapter 7

Final Discussion

7.1 Identification of putative enhancer elements

Using phylogenetic footprinting, I have identified several putative enhancer elements in the Gli3 locus. A percentage identity cut-off of 60% over 100bp between human and chicken genomes yielded 18 non-coding elements likely to be involved in Gli3 transcriptional regulation. 5' RACE analysis established that in mammals all putative enhancer elements are intragenic, owing to the identification of a novel untranslated exon upstream of the annotated mammalian transcript. No further conserved non-coding elements were identified in 85Kb of genomic DNA upstream of this exon. An in-vivo screening technique was devised to analyse the transcriptional activity of each element in the developing chick neural tube. Several putative elements can upregulate reporter gene expression in a spatially and temporally restricted manner within the developing neural tube. Thus, I propose that these regulatory elements act in synergy to orchestrate the expression profile of Gli3 in the developing embryo. This could be achieved via the differential expression of transcription factors that control the activity of enhancers both temporally and spatially. To illustrate this hypothesis I have further characterised Region1consA, one element that drives reporter gene expression in the developing neural tube in a pattern that mimics that of endogenous Gli3. Further analysis of this region established that it contains binding sites for, and is regulated by TALE family proteins. Expression pattern analysis of the members of this protein family revealed that they are expressed in regions of the embryo overlapping with Gli3 expression. Their known function as transcriptional co-factors offers a mechanism for the integration of various developmental pathways, including Shh and BMP pathways, in the regulation of Gli3 expression. For the first time this implicates TALE family transcription factors as key regulators of Gli3 expression.

7.2 Conserved elements upstream of Gli3 regulate reporter gene expression in the chick neural tube

Several conserved regions, including Region1consA, drive reporter gene expression in the chick neural tube. It remains possible that the lack of transcriptional activity for other conserved elements is due to the assay chosen, which is likely to miss combinatorial activities between two or more conserved elements. Consistent with this possibility, other groups have investigated larger elements, encompassing several of the regions investigated here, and found different effects on reporter gene expression (Abbasi *et al.*, 2007; Paparidis *et al.*, 2007; Alvarez-Medina *et al.*, 2008). To address a putative cumulative transcriptional activity on the control of Gli3 expression, each enhancer element should be studied in combination with other elements. Alternatively, the progressive deletion of putative enhancers identified

here in the context of a large BAC that faithfully reproduces Gli3 expression will reveal their individual contribution to the transcriptional output.

It is also worth noting that enhancer activity may vary depending on the tissue studied. Thus, absence of transcriptional activity in the neural tube may indicate that the element is either not active, or its activity is over-ridden by the repressive activity of other elements. A lack of activity might be due to the absence of transcription factors essential for enhancer activity, or increased expression of repressors. To address this, each conserved element should also be investigated in other sites of Gli3 expression, such as limb buds and paraxial mesoderm, in which electroporation in the chicken embryo is feasible. Alternatively, mouse transgenesis could be used to investigate the effect of each enhancer element on transgene expression in the whole embryo.

Finally, it should be noted that the conserved non-coding elements investigated in this study represent only a small proportion of conserved elements in the Gli3 locus that may control transcriptional activity. Indeed, Abbasi and colleagues have identified additional conserved non-coding regions throughout the Gli3 locus (Abbasi *et al.*, 2007). Furthermore, my genomic alignments have revealed the presence of non-coding conserved elements up to 700kb upstream of the Gli3 locus, suggesting a high level of complexity in the regulation of Gli3.

7.3 Promoter effects on transcriptional activity

In addition to the activity of enhancer elements, the methylation status of the promoter may also affect transcriptional activity. Gli3 has at least three alternative promoters, allowing differential inclusion of exon 0, exon 0b and exon 1b. I have not investigated which promoter the enhancer elements identified are likely to associate with. However, at the developmental stage investigated exon 0b is likely to be the prominent start site, since it was identified by 5' RACE on cDNAs generated from E9.5 mouse embryos. A similar transcript has also been amplified from human placental cDNA (Paparidis, 2005). Chick cDNA corresponding to this transcript could not be identified. In the chick embryo an alternative transcript initiates downstream of exon 0, suggesting the use of a further promoter that may be species specific.

To investigate which promoter elements are active in vivo, Chip on chip technology could be employed. A micro-array 'chip' could be generated of the genomic locus surrounding Gli3, and promoter elements could be isolated using antibodies raised against components of the pre-initiation complex such as RNAP and TFIID. This approach has led to the identification of more than 10,000 active promoters in human fibroblast cells (Kim *et al.*, 2005). However, the nearest promoter to the Gli3 locus was approximately 700kb upstream

of the transcript. This may be due to the fact that Gli3 is not transcribed in the cell line used to generate the microarray investigated. Likewise, no Gli3 promoter was identified in a Chip on chip screen for promoter elements within CpG islands (Weinmann *et al.*, 2002).

RNA-TRAP could be used to investigate which enhancer elements come into close proximity with each of the transcriptional start sites identified. The technique relies on hybridising a digoxigenin-labelled intronic probe to the primary transcript of a gene, which is bound by a Fab fragment conjugated to horse-radish peroxidase. Biotin tyramide is added, which will bind chromatin in close proximity to HRP, allowing the labelling of distal elements that loop into close proximity with the promoter of interest (Carter *et al.*, 2002).

Methylation of one promoter might cause transcription to initiate at another, and could therefore affect the contribution of various enhancer elements to transcriptional control. The methylation status of a promoter can be investigated by DnaseI sensitivity or by comparing the activity of methylation-sensitive and methylation-insensitive restriction endonucleases on the genomic DNA in various tissues, or at different developmental time-points (Levy-Wilson and Fortier, 1989).

To verify that the conserved elements identified here do not function as promoters, a reporter system could be used to determine whether the element is able to drive basal transcription independently of another promoter. Alternatively enhancer function could be verified by inverting their orientation, since enhancer elements can function in either direction.

7.4 TALE family transcription factors regulate Gli3 expression

Of particular interest in this study is the identification of the TALE transcription factors as regulators of Gli3 expression. TALE family proteins are emerging as key regulators of developmental gene expression. They have been associated with the regulation of several developmentally regulated genes such as Pax3, Pax6, Shh, myogenin and several Hox family genes (Popperl *et al.*, 1995; Maconochie *et al.*, 1997; Jacobs *et al.*, 1999; Ferretti *et al.*, 2000; Zhang *et al.*, 2002; Pruitt *et al.*, 2004; Capellini *et al.*, 2006; Zhang *et al.*, 2006; diIorio *et al.*, 2007). I have demonstrated that a composite Pbx:Meinox binding site identified in Region1consA is occupied in vitro. Furthermore, I showed that binding requires both half-sites, and disruption of either of these sites or of a Pbx binding site nearby affects reporter gene expression in vivo. Although the sites are predicted to bind TALE family proteins, the composition of the bound complexes could not be determined.

In mammals, the TALE family consists of Pbx proteins of the PBC subfamily, together with Meis, Prep and TGIF proteins of the Meinox subfamily, and Iroquois proteins (Burglin, 1997; Berthelsen *et al.*, 1998b). The proteins are characterised by an atypical homeodomain

that contains three additional residues between helix 1 and 2 (three amino acid loop extension; Bertolino *et al.,* 1995). In addition, the Meinox proteins share an N-terminal Meinox domain and the PBC proteins carry two conserved PBC domains (Burglin, 1997). Several paralogs of each protein exist. To date four Pbx proteins, three Meis proteins, two Prep proteins, three TGIF proteins and six Iroquois proteins have been identified (Nakamura *et al.*, 1996; Cohen *et al.*, 2000; Wagner *et al.*, 2001; Haller *et al.*, 2002). Members of each subfamily recognise a similar consensus binding site. Binding sites for PBC (Chang *et al.*, 1996; Shen *et al.*, 1997a) and Meinox proteins (Bertolino *et al.*, 1995; Chang *et al.*, 1997; Shen *et al.*, 1997a; Berthelsen *et al.*, 1998b), but not Iroquois proteins (Bilioni *et al.*, 2005) were identified in Region1consA.

TALE family proteins regulate transcriptional activity by associating with co-factors. In some cases, this interaction increases the binding specificity or affinity of a co-factor with activating potential (Knoepfler *et al.*, 1999). Alternatively, TALE proteins can prevent the binding of other complexes, either by sequestering their components or by occluding their binding sites (Bertolino *et al.*, 1995; Shen *et al.*, 1997a; Berthelsen *et al.*, 1998a; Jacobs *et al.*, 1999; Shanmugam *et al.*, 1999). In several instances, TALE proteins have been shown to interact with chromatin remodelling machinery, and can thus regulate transcriptional activity by inducing conformational changes in the DNA (Pazin and Kadonaga, 1997; Asahara *et al.*, 1999; Wotton *et al.*, 1999a; Wotton *et al.*, 1999b). Below, I describe interactions of TALE family proteins that have been described, in an attempt to understand the complexes that could regulate Region1consA activity, and thus Gli3 expression. EMSA experiments demonstrated that at least three protein complexes associate with the composite Meinox:Pbx binding site. Upon mutation of this site, weak binding of a larger complex was detected, which is predicted to associate with the Pbx binding site.

7.4.1 TALE proteins function through protein:protein interactions

Some TALE family proteins, including Pbx2 and Pbx3, but not Pbx1 can bind DNA as monomers (Neuteboom and Murre, 1997; Calvo *et al.*, 1999). Although Pbx1 and Pbx3 have been shown to homodimerise in vitro, stable heterodimers composed of TALE family proteins are formed most readily by the binding of a PBC protein to a Meinox family protein (Neuteboom and Murre, 1997; Calvo *et al.*, 1999). Meinox proteins bind to the PBC-A domain in the N-terminus of PBC proteins, via their own N-terminal Meinox domain. Pbx3c and 3d isoforms which lack a large portion of the PBC-A domain are unable to interact with Prep1, and interact only weakly with Meis1 (Berthelsen *et al.*, 1999).

Subtle sequence differences also affect Protein:Protein interactions between TALE proteins. For example, Pbx1:Prep1 and Pbx1:Prep2 heterodimers display differences in their dissociation rates (Fognani *et al.*, 2002). Thus, the expression levels of various TALE proteins in a tissue, will affect the occupancy of the binding sites identified in Region1consA, whilst the combination of different TALE proteins will determine the identity of the bound complex, and its stability. The binding of monomeric and dimeric complexes to Region1consA can account for two of the bands observed in EMSA experiments, suggesting that the third band observed could represent a higher order complex. However, TALE proteins themselves do not appear to be strong activators or repressors of transcription. Instead their activity is determined by the factors with which they associate.

PBC proteins contain a pocket in their homeodomain that is able to bind a conserved motif known as the PBC interacting domain (PID, also known as the 'hexapeptide', 'pentapeptide' or 'YPWM' motif) (Chang *et al.*, 1995; Knoepfler and Kamps, 1995; Phelan *et al.*, 1995; Lu and Kamps, 1996a; Shen *et al.*, 1996; Chang *et al.*, 1997; Shanmugam *et al.*, 1997; Piper *et al.*, 1999). This domain was initially identified in Hox proteins but has since been identified in multiple homeodomain and basic helix-loop-helix (bHLH) transcription factors (Chang *et al.*, 1995; Knoepfler and Kamps, 1995; Phelan *et al.*, 1995; Lu and Kamps, 1996a; Shen *et al.*, 1996; Chang *et al.*, 1997; Shanmugam *et al.*, 1997; Piper *et al.*, 1999). Pbx/Meinox heterodimers interact with multiple transcription factors including the pancreatic factor Pdx1, Engrailed, and the myogenic bHLH proteins MyoD, myogenin, Mrf-4 and Myf-5 (Peers *et al.*, 1995; Peltenburg and Murre, 1996; Goudet *et al.*, 1999; Knoepfler *et al.*, 1999; Knudsen, 1999; Liu *et al.*, 2001; Berkes *et al.*, 2004; Erickson *et al.*, 2007). Similar motifs have been detected in Nk, Lim and Pax family proteins, although functional interactions have not yet been identified (In der Rieden *et al.*, 2004). Thus Pbx:Meinox heterodimers can facilitate the recruitment, or augment the function of other transcription factors (Laurent *et al.*, 2008). Since homeodomain and bHLH proteins are each widely expressed in the neural tube, it is reasonable to assume that the activity of Region1consA might be regulated by such interactions. Indeed, my data are consistent with the hypothesis that the three bands observed in EMSA experiments represent monomeric and heteromeric TALE family proteins, together with a trimeric complex composed of a Pbx:Meinox heterodimer associated with an accessory protein, each competing for the same binding site (the composite Pbx/Meinox site referred to as Site 2 in Chapter 6). In this case transcriptional activity would be determined by the expression levels of homeodomain and bHLH co-factors, together with the co-expression of sufficient TALE family proteins to facilitate/compete binding.

Another mechanism by which the TALE proteins might function is by increasing the affinity of other transcription factors in conditions where they show little or no inherent DNA binding affinity. Indeed, binding of TALE family proteins to the myogenin promoter permits the binding of MyoD to E-box motifs that poorly match the consensus sequence (Berkes *et al.*, 2004). Within Region1consA, TALE family binding sites overlap with non-consensus binding sites for Myt1, Pax5, Pax6, CAAT and GFT1. TALE family binding could allow proteins to bind to these sites, that would not otherwise be occupied.

Recently, dimerisation of TALE family proteins with non-homeobox proteins has also emerged. In a two hybrid screen only 6% of putative Pbx partners contained a homeodomain domain, and only 18% were transcription factors (Laurent *et al.*, 2008). Non-homeodomain proteins identified included non-muscle myosin II heavy chain B (NMHCB), and a novel zinc finger containing protein, ZFPIP (Zinc-finger PBX1 interacting protein) (Huang *et al.*, 2003; Laurent *et al.*, 2007). Such interactions regulate the activity of transcriptional complexes containing TALE proteins. Binding of ZFPIP or HPIP to Pbx1 prevents Hox:Pbx complexes from binding to their consensus DNA site in vitro (Abramovich *et al.*, 2002; Laurent *et al.*, 2007), whilst up-regulation of NMHCB causes cytoplasmic accumulation of Pbx and Meis in vitro (Huang *et al.*, 2003). Therefore the activity of Region1consA may also be influenced by the expression of multiple proteins that do not associate with the DNA-bound complexes.

7.4.2 TALE family proteins regulate chromatin remodelling

Upon mutation of the Pbx binding site in Region1consA, reporter gene expression is up-regulated in vivo, suggesting that the region is required for repressing transcription. Consistent with this, Pbx1 has been shown to associate with the chromatin remodelling machinery. In its N-terminal domain, Pbx1a contains a repressor domain that interacts with co-repressors NcoR (nuclear receptor co-repressor) and SMRT (Silencing Mediator for Retinoid and Thyroid-hormone receptors)(Asahara *et al.*, 1999). These proteins establish a protein-protein link between Pbx1 and histone deacelylases (HDACs) (Pazin and Kadonaga, 1997). Additionally, one of the repressor domains in the N-terminus of Pbx proteins has been shown to associate directly with class 1 HDACs (Saleh *et al.*, 2000b). Repression could also be imposed by TGIF binding to the Pbx:Meinox site, since TGIF proteins contain HDAC interacting domains in their C-termini. HDAC activity results in condensation of the chromatin environment, which is non-permissive to transcription (Pazin and Kadonaga, 1997). The condensed chromatin structure could repress enhancer activity by preventing the binding of other proteins, or by directly affecting transcriptional initiation. Indeed, TGIF is able to repress transcription from a distance (Wotton *et al.*, 1999b).

TALE family proteins have previously been shown to regulate the chromatin environment at the myogenin locus (Berkes *et al.*, 2004). Pbx1 binds to the myogenin promoter prior to gene expression, and may mediate the condensed status of the chromatin. The promoter is activated by an association of MyoD with Pbx1. MyoD recruits SWI/SNF members of the chromatin remodelling complex, which have the potential to interact with both HDACs and histone acetyltransferases (HACs), and co-ordinate chromatin-remodelling activities (Martens and Winston, 2003; Sif, 2004). The authors propose that TALE family proteins can access regions of condensed chromatin, and mark genes for activation by allowing other factors to access DNA (Berkes *et al.*, 2004; Maves *et al.*, 2007). Such a mechanism could account for the activity of Region1consA. The two step model proposed in Chapter 6 suggests that the binding of Pbx proteins to Region1consA causes repression of gene expression by association with HDAC proteins. Expression could be activated in a context dependant manner by the association of Pbx:Meinox heterodimers with bHLH or homeodomain transcription factors that contain a PID domain. Alternatively, another uncharacterised protein with similar properties to MyoD could associate with Region1consA and co-ordinate chromatin remodelling, resulting in transcriptional activation.

Alternative splicing of Meis and Pbx proteins adds to their complexity by regulating their ability to interact with co-factors. Pbx1b and Pbx3b/d lack the C-terminal interaction domain and is unable to bind NcoR and SMRT co-repressors (Monica *et al.*, 1991; Oulad-Abdelghani *et al.*, 1997; Asahara *et al.*, 1999; Goudet *et al.*, 1999; Milech *et al.*, 2001). Similarly Pbx3c and Pbx3d, which lack a large portion of the N-terminal PBC domain, have a reduced ability to form heterodimers. Therefore, the ability of Pbx1 to repress transcription of Region1consA will depend on the distribution of various splice forms.

7.4.3   TALE family proteins are Smad co-factors

It has previously been proposed that the maintenance of Gli3 expression is regulated by BMP signalling. An interesting observation is that Smad family proteins have been shown to interact with TALE family proteins. Thus, the TALE family binding sites in Region1consA offer a means of integrating BMP signals in the regulation of Gli3 expression.

Smad family transcription factors (including Smad 2, 3 and 4) interact with Pbx1:Prep1 heterodimers, via a Smad interaction domain in the N-terminus of Prep1 (Bailey *et al.*, 2004). At the Follicle-Stimulating Hormone β–subunit (FSHβ) gene promoter, Pbx1/Prep1 binding is required to recruit Smad proteins or stabilise their binding following activation of TGFβ signalling, resulting in an upregulation of reporter gene expression. Smad 4 is also responsible for transducing BMP signalling. To determine whether Smad binding regulates transcriptional

activity of Region1consA, the reporter construct could be co-electroporated in the chick neural tube with activators or repressors of BMP signalling (such as Alk3 and Smad6 constructs respectively; Imamura *et al.,* 1997; James and Schultheiss, 2005; Linker and Stern, 2004). Interestingly, TGIF also associates with Smad proteins, although the resulting complex represses genes regulated by TGFβ signalling (Wotton *et al.*, 1999a; Wotton *et al.*, 1999b). Thus, the activator function of Pbx1:Prep1:Smad complexes could be competed by a repressor function of TGIF:Smad complexes. Both are expected to compete for binding to the composite Meinox:Pbx site in Region1consA.

7.4.4   Differential expression of TALE family proteins

Since the binding site identified in Region1consA is predicted to bind a number of TALE family complexes, occupancy by a particular complex will depend on expression levels of specific TALE proteins. In-situ hybridisation analysis has demonstrated that various TALE proteins are differentially expressed in the developing neural tube.

The species observed in EMSA experiments to bind Region1consA do not appear to contain Pbx1a, 2, 3a, 3c or Meis1/2, as shown by the inability of antibodies that recognise these proteins to compete binding. However, my RT-PCR results demonstrate that Pbx1b, 3b, 3d, cPbx4, Prep1 and Prep2 are expressed in the chick embryo. These remain candidates for binding to Region1consA, along with TGIF proteins whose expression I have not analysed. Murine TGIF is expressed in the brain and in tail-bud regions of the neural ectoderm, and a dorsal-high gradient of expression has been reported for TGIF in the chick neural tube (Bertolino *et al.*, 1995; Shen and Walsh, 2005; Jin *et al.*, 2006; Knepper *et al.*, 2006). My situ hybridisation analysis demonstrates that Prep1 and cPbx4 are expressed in the posterior neural tube, where transgene expression is at its highest. In this domain these proteins, together with TGIF, may compete for occupancy of the Meinox/Pbx binding site, and allow activation by association with bHLH and homeobox transcription factors. In more anterior regions similar mechanisms might exist, but expression patterns are restricted by low levels of TALE protein expression, together with altered expression of co-factors. The maintenance of reporter gene expression in the dorsal neural tube can be accounted for by the integration of BMP signalling on the activity of Region1consA. Indeed, Prep2 is expressed in the dorsal neural tube anterior to the first somite. The high degree of homology between Prep1 and Prep2 suggests that both will contain similar protein interaction domains, allowing Prep2 to associate with Smad proteins (Fognani *et al.*, 2002).

Interestingly my in-situ hybridisation analysis demonstrated that Meis proteins are widely expressed anterior to the first somite. This coincides with the axial level at which

Region1consA activity is repressed. Meis proteins might therefore have a repressive role on Region1consA activity. Since Meis proteins do not appear to associate with the TALE binding sites identified, their function may be to sequester other proteins required for transcriptional activity. Indeed, in several cases a Pbx:Hox:Meinox trimer has been found as a stable complex in the cell in the absence of DNA binding (Shen *et al.*, 1997a; Berthelsen *et al.*, 1998a; Jacobs *et al.*, 1999; Shanmugam *et al.*, 1999).

7.4.5 The function of TALE family proteins is regulated by subcellular localisation

In addition to their levels of expression and co-factors they associate with, TALE protein activity is regulated by subcellular localisation. Pbx1 contains two nuclear export signals located in its interaction domain with Meinox proteins (the PBC-A domain; Berthelsen *et al.*, 1999; Kilstrup-Nielsen *et al.*, 2003). Formation of Pbx:Meinox complexes blocks these sites and nuclear export, resulting in the accumulation of active complexes in the nucleus. In addition, Pbx1 contains two co-operative nuclear localisation sequences (not present in Meinox proteins) that cause increased nuclear localisation of Meinox proteins when associated with Pbx1 (Berthelsen *et al.*, 1999; Saleh *et al.*, 2000a; Fognani *et al.*, 2002; Haller *et al.*, 2002; Kilstrup-Nielsen *et al.*, 2003). These NLSs are also inhibited by an intra-molecular interaction between the N-terminus and homeodomain of Pbx1, such that they are only exposed upon a conformational change that results from Meinox binding (Saleh *et al.*, 2000a). Indeed, in the mouse limb bud Pbx1 is only nuclear where Meis1 is co-expressed (Saleh *et al.*, 2000a). Thus, Meis1 could help moderate the repressive function of Pbx1 on Region1consA activity. Additionally, increased nuclear localisation of Meinox proteins such as Prep1 protects Pbx proteins from degradation (Longobardi and Blasi, 2003). Thus, Meinox protein association with Pbx proteins affects both their nuclear localisation and stability. The importance of Pbx subcellular localisation is clearly illustrated in the analysis of Pbx1-deficient mice, which display a phenotype only in proximal limb skeletal elements where Pbx1 is nuclear (Selleri *et al.*, 2001).

Nuclear export of PBC proteins can also be antagonised by PKA mediated phosphorylation of the PBC-B domain (Kilstrup-Nielsen *et al.*, 2003). Indeed, the authors propose that in most cell types PKA basal activity is sufficient for nuclear import of PBC proteins independently of their association with Meinox proteins. However, in some tissues PKA activity is blocked, or counteracted by a phosphatase, resulting in increased nuclear export. Interestingly Shh is proposed to induce the nuclear export of PBC proteins by counteracting PKA mediated phosphorylation of the PBC-B domain, independantly of Pbx:Meinox interaction (Kilstrup-Nielsen *et al.*, 2003). Thus in some instances, TALE family proteins

can be regulated by Shh signalling, providing a putative link between Shh expression and the regulation of Region1consA activity. Indeed, Shh is believed to prevent PBC nuclear localisation in distal limb elements by repressing distal expression of Meinox genes (Kilstrup-Nielsen *et al.*, 2003).

7.5 Evidence for a link between TALE family proteins and Shh signalling

7.5.1 Mutant phenotypes suggest a link between TGIF and Shh signalling

Holoprosencephaly (HPE) is the most common congenital anomaly in the human forebrain. Mutations in seven genes have been associated with the condition, including Gli2, Ptc and Shh. Mutations in TGIF have also been shown to cause HPE, suggesting a possible link between TGIF and the Shh pathway (Knepper *et al.*, 2006).

If TGIF is able to repress Gli3 transcription, mutation in this gene would be expected to upregulate Gli3 expression in a similar manner to that observed in Shh-/- and Ptc-/- embryos (Litingtung and Chiang, 2000; Motoyama *et al.*, 2003). Indeed, overexpression of TGIF in the chick neural tube results in repression of dorsal neural tube markers, which could be attributed to a loss of Gli3 expression (Persson *et al.*, 2002; Knepper *et al.*, 2006). Expression of the ventral neural tube markers Pax3, FoxD3, Nkx2.2 and Isl1 was unaffected, suggesting that TGIF specifically represses genes expressed in the dorsal neural tube, where Gli3 is more strongly expressed (Knepper *et al.*, 2006).

However, mouse embryos homozygous for a mutation in TGIF appear to have normal DV and AP patterning in the neural tube and brain, as judged by expression profiles of Shh, Foxa2, Pax2, Pax7, Nkx2.2, Nkx2.1, Six3, Fgf6 and Otx3 (Jin *et al.*, 2006), and do not present holoprosencephaly (Shen and Walsh, 2005). Gli expression has not been investigated in these mutants. Motayama and colleagues (2003) demonstrated that in the context of the Ptc-/- mouse, Gli2 and Gli3 are functionally redundant (Motoyama *et al.*, 2003). They are required for Foxa2, Nkx2.2 and Isl1/2 expression, and repress Pax7 expression, but the effect of Gli3 on expression of these proteins was only observed in the absence of Gli2 (Motoyama *et al.*, 2003). Functional redundancy between Gli2 and Gli3 could account for the maintenance of expression of these proteins in TGIF mutant mice.

It has been proposed that in humans additional genetic abnormalities might contribute to the HPE phenotype in addition to TGIF depletion, or that other TGFβ antagonists may compensate for the loss of TGIF (Shen and Walsh, 2005; Jin *et al.*, 2006). However, phenotypic differences between human and mouse could also be due to a divergence in TGIF function. For instance, TGIF could repress transcription of both Gli2 and Gli3 in humans, whereas in mice only Gli3 would be affected, resulting in a more severe phenotype in humans.

7.5.2   Evidence for TALE family regulation of Shh signalling

A further relationship between Shh signalling and TALE family proteins has been uncovered in the developing limb. It is known that down-regulation of Meis2 in the distal limb bud is required for normal limb outgrowth (Capdevila *et al.*, 1999; Mercader *et al.*, 1999). Misexpression of Meis2 in the distal limb results in a strong repression of genes involved in the Shh/FGF regulatory loop (Capdevila *et al.*, 1999). Furthermore, Pbx1$^{-/-}$ Pbx2$^{+/-}$ mice have a similar hindlimb phenotype to embryos carrying a mutation in the distal Shh limb enhancer, that is consistent with a loss of Shh in the hindlimb. Indeed, Pbx2 binds the Shh limb enhancer in vitro, although in-vivo binding is not essential (Capellini *et al.*, 2006). It has been proposed that limb outgrowth is mediated by Pbx1/2 proteins interacting with Hox proteins and controlling Shh expression (Capellini *et al.*, 2006). Interestingly, Pbx1$^{-/-}$Pbx2$^{+/-}$ hindlimbs display an expansion of Gli3 expression throughout the distal limb bud, a phenotype similar to, though less severe than that of Pbx1 depleted embryos (Selleri *et al.*, 2001). This is consistent with a role of Pbx1 in repressing Gli3 expression, whilst is also suggestive that Pbx2 upregulates Gli3 expression. Zebrafish Meis3 has also been shown to regulate Shh expression in the anterior endoderm (diIorio *et al.*, 2007).

7.5.3   TALE proteins may have a more general role in regulating Gli3 expression

Interestingly, other TALE family binding sites were identified by Abbasi and colleagues in various conserved non-coding elements of Gli3 (Abbasi *et al.*, 2007). One of these regions, which functions as an activator in Gli3-expressing cell lines, contains a 50bp module with Pbx1, Pax2 and Meis1 binding sites that was shown to be necessary but not sufficient for reporter gene expression in vitro. However, this putative enhancer was unable to drive reporter gene expression in vivo in transiently transfected zebrafish embryos. To establish whether multiple TALE binding sites are active during Gli3 transcription, a chip on chip approach could be used.

7.6 The role of TALE family proteins in Gli3 regulation

Together, the data suggest that TALE family complexes can function as activators or repressors in a context dependant manner. In the case of Region1consA they appear to preferentially invoke repressor activity, since mutation of the Pbx binding site causes an upregulation of transgene expression. Repression is alleviated in specific domains of the neural tube, perhaps through the association of TALE family proteins with bHLH and homeodomain transcription factors. The association of TALE family proteins with Smad transcription factors offers a means of integrating BMP signalling on the regulation of Gli3 expression. Furthermore, regulation of TALE protein activity by Shh signalling could also

influence Gli3 expression. Phenotypic similarities between Pbx and Shh mutant mice, together with their misregulation of embryonic Gli gene expression is consistent with TALE family proteins regulating the expression of Gli3.

7.7 Concluding remarks

Within a 170 Kb region surrounding the Gli3 transcriptional start site, numerous non-coding DNA regions are highly conserved amongst vertebrates. Investigation of ten of these regions in the chick neural tube established that many have enhancer activity. One element was identified as being of particular interest in the regulation of Gli3 in the chick neural tube, since it drove reporter gene expression in a pattern that mimics that of the endogenous gene. Within this region, two highly conserved binding sites for TALE family proteins were identified. Furthermore, I have demonstrated that this element is protein-bound in vitro, and mutations in the binding sites disrupt transgene expression and DNA binding.

Further characterisation of how TALE family proteins control Region1consA activity, and how this influences Gli3 expression is undoubtedly necessary. However, the identification of a functional binding site within an element that drives reporter gene expression in a manner that mimics that of endogenous Gli3 in the neural tube, presents a novel mechanism for the regulation of this developmentally important gene. TALE family proteins have not previously been directly associated with the regulation of Gli3 expression, although their expression patterns correlate with that of Gli3 in the developing embryo. Transcriptional regulation by these factors offers a means to integrate signals from the BMP and Shh pathways, which have previously been implicated in Gli3 regulation. In summary, the work presented here identifies TALE family proteins as novel regulators of Gli3 expression.

References:

Abbasi, A. A., Paparidis, Z., Malik, S., Goode, D. K., Callaway, H., Elgar, G. and Grzeschik, K. H. (2007). Human GLI3 intragenic conserved non-coding sequences are tissue-specific enhancers. PLoS ONE 2 (4): e366.

Abramovich, C., Chavez, E. A., Lansdorp, P. M. and Humphries, R. K. (2002). Functional characterization of multiple domains involved in the subcellular localization of the hematopoietic Pbx interacting protein (HPIP). Oncogene 21 (44): 6766-71.

Agren, M., Kogerman, P., Kleman, M. I., Wessling, M. and Toftgard, R. (2004). Expression of the PTCH1 tumor suppressor gene is regulated by alternative promoters and a single functional Gli-binding site. Gene 330: 101-14.

Alcedo, J., Ayzenzon, M., Von Ohlen, T., Noll, M. and Hooper, J. E. (1996). The Drosophila smoothened gene encodes a seven-pass membrane protein, a putative receptor for the hedgehog signal. Cell 86 (2): 221-32.

Alcedo, J., Zou, Y. and Noll, M. (2000). Posttranscriptional regulation of smoothened is part of a self-correcting mechanism in the Hedgehog signaling system. Mol Cell 6 (2): 457-65.

Alexandre, C., Jacinto, A. and Ingham, P. W. (1996). Transcriptional activation of hedgehog target genes in Drosophila is mediated directly by the cubitus interruptus protein, a member of the GLI family of zinc finger DNA-binding proteins. Genes Dev 10 (16): 2003-13.

Altschul, S., Gish, W., Miller, W., Myers, E. and Lipman., D. (1990). Basic local alignment search tool. J Mol Biol. 215: 403-10.

Alvarez-Medina, R., Cayuso, J., Okubo, T., Takada, S. and Marti, E. (2008). Wnt canonical pathway restricts graded Shh/Gli patterning activity through the regulation of Gli3 expression. Development 135 (2): 237-247.

Andersen, F. G., Jensen, J., Heller, R. S., Petersen, H. V., Larsson, L. I., Madsen, O. D. and Serup, P. (1999). Pax6 and Pdx1 form a functional complex on the rat somatostatin gene upstream enhancer. FEBS Lett 445 (2-3): 315-20.

Antequera, F. and Bird, A. (1993). Number of CpG islands and genes in human and mouse. Proc Natl Acad Sci U S A 90 (24): 11995-9.

Aoto, K., Nishimura, T., Eto, K. and Motoyama, J. (2002). Mouse GLI3 regulates Fgf8 expression and apoptosis in the developing neural tube, face, and limb bud. Dev Biol 251 (2): 320-32.

Aparicio, S., Morrison, A., Gould, A., Gilthorpe, J., Chaudhuri, C., Rigby, P., Krumlauf, R. and Brenner, S. (1995). Detecting conserved regulatory elements with the model genome of the Japanese puffer fish, Fugu rubripes. Proc Natl Acad Sci U S A 92 (5): 1684-8.

Arkell, R. and Beddington, R. S. (1997). BMP-7 influences pattern and growth of the developing hindbrain of mouse embryos. Development 124 (1): 1-12.

Arnone, M. and Davidson, E. (1997). The hardwiring of development: organization and function of genomic regulatory systems. Development 124 (10): 1851-1864.

Aruga, J., Mizugishi, K., Koseki, H., Imai, K., Balling, R., Noda, T. and Mikoshiba, K. (1999). Zic1 regulates the patterning of vertebral arches in cooperation with Gli3. Mech Dev 89 (1-2): 141-50.

Asahara, H., Dutta, S., Kao, H. Y., Evans, R. M. and Montminy, M. (1999). Pbx-Hox heterodimers recruit coactivator-corepressor complexes in an isoform-specific manner. Mol Cell Biol 19 (12): 8219-25.

Ashburner, M., Ball, C. A., Blake, J. A., Botstein, D., Butler, H., Cherry, J. M., Davis, A. P., Dolinski, K., Dwight, S. S., Eppig, J. T., Harris, M. A., Hill, D. P., Issel-Tarver, L., Kasarskis, A., Lewis, S., Matese, J. C., Richardson, J. E., Ringwald, M., Rubin, G. M. and Sherlock, G. (2000). Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. Nat Genet 25 (1): 25-9.

Aza-Blanc, P., Lin, H. Y., Ruiz I Altaba, A. and Kornberg, T. B. (2000). Expression of the vertebrate Gli proteins in Drosophila reveals a distribution of activator and repressor activities. Development 127 (19): 4293-301.

Bagheri-Fam, S., Ferraz, C., Demaille, J., Scherer, G. and Pfeifer, D. (2001). Comparative genomics of the SOX9 region in human and Fugu rubripes: conservation of short regulatory sequence elements within large intergenic regions. Genomics 78 (1-2): 73-82.

Bai, C. B. and Joyner, A. L. (2001). Gli1 can rescue the in vivo function of Gli2. Development 128 (24): 5161-72.

Bai, C. B., Auerbach, W., Lee, J. S., Stephen, D. and Joyner, A. L. (2002). Gli2, but not Gli1, is required for initial Shh signaling and ectopic activation of the Shh pathway. Development 129 (20): 4753-61.

Bai, C. B., Stephen, D. and Joyner, A. L. (2004). All mouse ventral spinal cord patterning by hedgehog is Gli dependent and involves an activator function of Gli3. Dev Cell 6 (1): 103-15.

Bailey, J. S., Rave-Harel, N., Mcgillivray, S. M., Coss, D. and Mellon, P. L. (2004). Activin regulation of the follicle-stimulating hormone beta-subunit gene involves Smads and the TALE homeodomain proteins Pbx1 and Prep1. Mol Endocrinol 18 (5): 1158-70.

Bailey, P. J., Klos, J. M., Andersson, E., Karlen, M., Kallstrom, M., Ponjavic, J., Muhr, J., Lenhard, B., Sandelin, A. and Ericson, J. (2006). A global genomic transcriptional code associated with CNS-expressed genes. Exp Cell Res 312 (16): 3108-19.

Bajanca, F., Luz, M., Duxson, M. J. and Thorsteinsdottir, S. (2004). Integrins in the mouse myotome: developmental changes and differences between the epaxial and hypaxial lineage. Dev Dyn 231 (2): 402-15.

Bajic, V. B. and Seah, S. H. (2003). Dragon gene start finder: an advanced system for finding approximate locations of the start of gene transcriptional units. Genome Res 13 (8): 1923-9.

Bajic, V. B., Seah, S. H., Chong, A., Zhang, G., Koh, J. L. and Brusic, V. (2002). Dragon Promoter Finder: recognition of vertebrate RNA polymerase II promoters. Bioinformatics 18 (1): 198-9.

Bajic, V. B., Tan, S. L., Christoffels, A., Schonbach, C., Lipovich, L., Yang, L., Hofmann, O., Kruger, A., Hide, W., Kai, C., Kawai, J., Hume, D. A., Carninci, P. and Hayashizaki, Y. (2006). Mice and men: their promoter properties. PLoS Genet 2 (4): e54.

Bajic, V. B., Tan, S. L., Suzuki, Y. and Sugano, S. (2004). Promoter prediction analysis on the whole human genome. Nat Biotechnol 22 (11): 1467-73.

Barnes, W. (1994). PCR Amplification of up to 35-kb DNA with High Fidelity and High Yield from {lambda} Bacteriophage Templates. Proceedings of the National Academy of Sciences 91 (6): 2216-2220.

Barth, K. A., Kishimoto, Y., Rohr, K. B., Seydler, C., Schulte-Merker, S. and Wilson, S. W. (1999). Bmp activity establishes a gradient of positional information throughout the entire neural plate. Development 126 (22): 4977-87.

Bastida, M. F., Delgado, M. D., Wang, B., Fallon, J. F., Fernandez-Teran, M. and Ros, M. A. (2004). Levels of Gli3 repressor correlate with Bmp4 expression and apoptosis during limb development. Dev Dyn 231 (1): 148-60.

Behrens, J., Von Kries, J. P., Kuhl, M., Bruhn, L., Wedlich, D., Grosschedl, R. and Birchmeier, W. (1996). Functional interaction of beta-catenin with the transcription factor LEF-1. Nature 382 (6592): 638-42.

Bejerano, G., Haussler, D. and Blanchette, M. (2004). Into the heart of darkness: large-scale clustering of human non-coding DNA. Bioinformatics 20 (suppl_1): i40-48.

Bellefroid, E. J., Bourguignon, C., Hollemann, T., Ma, Q., Anderson, D. J., Kintner, C. and Pieler, T. (1996). X-MyT1, a Xenopus C2HC-type zinc finger protein with a regulatory function in neuronal differentiation. Cell 87 (7): 1191-202.

Ben-Arie, N., Hassan, B. A., Bermingham, N. A., Malicki, D. M., Armstrong, D., Matzuk, M., Bellen, H. J. and Zoghbi, H. Y. (2000). Functional conservation of atonal and Math1 in the CNS and PNS. Development 127 (5): 1039-48.

Bergeron, S. A., Milla, L. A., Villegas, R., Shen, M. C., Burgess, S. M., Allende, M. L., Karlstrom, R. O. and Palma, V. (2008). Expression profiling identifies novel Hh/Gli-regulated genes in developing zebrafish embryos. Genomics 91 (2): 165-77.

Berkes, C. A., Bergstrom, D. A., Penn, B. H., Seaver, K. J., Knoepfler, P. S. and Tapscott, S. J. (2004). Pbx marks genes for activation by MyoD indicating a role for a homeodomain protein in establishing myogenic potential. Mol Cell 14 (4): 465-77.

Berman, B. P., Pfeiffer, B. D., Laverty, T. R., Salzberg, S. L., Rubin, G. M., Eisen, M. B. and Celniker, S. E. (2004a). Computational identification of developmental enhancers: conservation and function of transcription factor binding-site clusters in Drosophila melanogaster and Drosophila pseudoobscura. Genome Biol 5 (9): R61.

Berman, D. M., Desai, N., Wang, X., Karhadkar, S. S., Reynon, M., Abate-Shen, C., Beachy, P. A. and Shen, M. M. (2004b). Roles for Hedgehog signaling in androgen production and prostate ductal morphogenesis. Dev Biol 267 (2): 387-98.

Berman, D. M., Karhadkar, S. S., Hallahan, A. R., Pritchard, J. I., Eberhart, C. G., Watkins, D. N., Chen, J. K., Cooper, M. K., Taipale, J., Olson, J. M. and Beachy, P. A. (2002). Medulloblastoma growth inhibition by hedgehog pathway blockade. Science 297 (5586): 1559-61.

Bert, A. G., Burrows, J., Hawwari, A., Vadas, M. A. and Cockerill, P. N. (2000). Reconstitution of T cell-specific transcription directed by composite NFAT/Oct elements. J Immunol 165 (10): 5646-55.

Berthelsen, J., Kilstrup-Nielsen, C., Blasi, F., Mavilio, F. and Zappavigna, V. (1999). The subcellular localization of PBX1 and EXD proteins depends on nuclear import and export signals and is modulated by association with PREP1 and HTH. Genes Dev 13 (8): 946-53.

Berthelsen, J., Zappavigna, V., Ferretti, E., Mavilio, F. and Blasi, F. (1998a). The novel homeoprotein Prep1 modulates Pbx-Hox protein cooperativity. Embo J 17 (5): 1434-45.

Berthelsen, J., Zappavigna, V., Mavilio, F. and Blasi, F. (1998b). Prep1, a novel functional partner of Pbx proteins. Embo J 17 (5): 1423-33.

Bertolino, E., Reimund, B., Wildt-Perinic, D. and Clerc, R. G. (1995). A novel homeobox protein which recognizes a TGT core and functionally interferes with a retinoid-responsive motif. J Biol Chem 270 (52): 31178-88.

Bien-Willner, G. A., Stankiewicz, P. and Lupski, J. R. (2007). SOX9cre1, a cis-acting regulatory element located 1.1 Mb upstream of SOX9, mediates its enhancement through the SHH pathway. Hum. Mol. Genet.: ddm061.

Bilioni, A., Craig, G., Hill, C. and Mcneill, H. (2005). Iroquois transcription factors recognize a unique motif to mediate transcriptional repression in vivo. Proc Natl Acad Sci U S A 102 (41): 14671-6.

Birney, E., Stamatoyannopoulos, J. A., Dutta, A., Guigo, R., Gingeras, T. R., Margulies, E. H., Weng, Z., Snyder, M., Dermitzakis, E. T., Thurman, R. E., Kuehn, M. S., Taylor, C. M., Neph, S., Koch, C. M., Asthana, S., Malhotra, A., Adzhubei, I., Greenbaum, J. A., Andrews, R. M., Flicek, P., Boyle, P. J., Cao, H., Carter, N. P., Clelland, G. K., Davis, S., Day, N., Dhami, P., Dillon, S. C., Dorschner, M. O., Fiegler, H., Giresi, P. G., Goldy, J., Hawrylycz, M., Haydock, A., Humbert, R., James, K. D., Johnson, B. E., Johnson, E. M., Frum, T. T., Rosenzweig, E. R., Karnani, N., Lee, K., Lefebvre, G. C., Navas, P. A., Neri, F., Parker, S. C., Sabo, P. J., Sandstrom, R., Shafer, A., Vetrie, D., Weaver, M., Wilcox, S., Yu, M., Collins, F. S., Dekker, J., Lieb, J. D., Tullius, T. D., Crawford, G. E., Sunyaev, S., Noble, W. S., Dunham, I., Denoeud, F., Reymond, A., Kapranov, P., Rozowsky, J., Zheng, D., Castelo, R., Frankish, A., Harrow, J., Ghosh, S., Sandelin, A., Hofacker, I. L., Baertsch, R., Keefe, D., Dike, S., Cheng, J., Hirsch, H. A., Sekinger, E. A., Lagarde, J., Abril, J. F., Shahab, A., Flamm, C., Fried, C., Hackermuller, J., Hertel, J., Lindemeyer, M., Missal, K., Tanzer, A., Washietl, S., Korbel, J., Emanuelsson, O., Pedersen, J. S., Holroyd, N., Taylor, R., Swarbreck, D., Matthews, N., Dickson, M. C., Thomas, D. J., Weirauch, M. T., Gilbert, J., Drenkow, J., Bell, I., Zhao, X., Srinivasan, K. G., Sung, W. K., Ooi, H. S., Chiu, K. P., Foissac, S., Alioto, T., Brent, M., Pachter, L., Tress, M. L., Valencia, A., Choo, S. W., Choo, C. Y., Ucla, C., Manzano, C., Wyss, C., Cheung, E., Clark, T. G., Brown, J. B., Ganesh, M., Patel, S., Tammana, H., Chrast, J., Henrichsen, C. N., Kai, C., Kawai, J., Nagalakshmi, U., Wu, J., Lian, Z., Lian, J., Newburger, P., Zhang, X., Bickel, P., Mattick, J. S., Carninci, P., Hayashizaki, Y., Weissman, S., Hubbard, T., Myers, R. M., Rogers, J., Stadler, P. F., Lowe, T. M., Wei, C. L., Ruan, Y., Struhl, K., Gerstein, M., Antonarakis, S. E., Fu, Y., Green, E. D., Karaoz, U., Siepel, A., Taylor, J., Liefer, L. A., Wetterstrand, K. A., Good, P. J., Feingold, E. A., Guyer, M. S., Cooper, G. M., Asimenos, G., Dewey, C. N., Hou, M., Nikolaev, S., Montoya-Burgos, J. I., Loytynoja, A., Whelan, S., Pardi, F., Massingham, T., Huang, H., Zhang, N. R., Holmes, I., Mullikin, J. C., Ureta-Vidal, A., Paten, B., Seringhaus, M., Church, D., Rosenbloom, K., Kent, W. J., Stone, E. A., Batzoglou, S., Goldman, N., Hardison, R. C., Haussler, D., Miller, W., Sidow, A., Trinklein, N. D., Zhang, Z. D., Barrera, L., Stuart, R., King, D. C., Ameur, A., Enroth, S., Bieda, M. C., Kim, J., Bhinge, A. A., Jiang, N., Liu, J., Yao, F., Vega, V. B., Lee, C. W., Ng, P., Yang, A., Moqtaderi, Z., Zhu, Z., Xu, X., Squazzo, S., Oberley, M. J., Inman, D., Singer, M. A., Richmond, T. A., Munn, K. J., Rada-Iglesias, A., Wallerman, O., Komorowski, J., Fowler, J. C., Couttet, P., Bruce, A. W., Dovey, O. M., Ellis, P. D., Langford, C. F., Nix, D. A., Euskirchen, G., Hartman, S., Urban, A. E., Kraus, P., Van Calcar, S., Heintzman, N., Kim, T. H., Wang, K., Qu, C., Hon, G., Luna, R., Glass, C. K., Rosenfeld, M. G., Aldred, S. F., Cooper, S. J., Halees, A., Lin, J. M., Shulha, H. P., Xu, M., Haidar, J. N., Yu, Y., Iyer, V. R., Green, R. D., Wadelius, C., Farnham, P. J., Ren, B., Harte, R. A., Hinrichs, A. S., Trumbower, H., Clawson, H., Hillman-Jackson, J., Zweig, A. S., Smith, K., Thakkapallayil, A., Barber, G., Kuhn, R. M., Karolchik, D., Armengol, L., Bird, C. P., De Bakker, P. I., Kern, A. D., Lopez-Bigas, N., Martin, J. D., Stranger, B. E., Woodroffe, A., Davydov, E., Dimas, A., Eyras, E., Hallgrimsdottir, I. B., Huppert, J., Zody,

M. C., Abecasis, G. R., Estivill, X., Bouffard, G. G., Guan, X., Hansen, N. F., Idol, J. R., Maduro, V. V., Maskeri, B., Mcdowell, J. C., Park, M., Thomas, P. J., Young, A. C., Blakesley, R. W., Muzny, D. M., Sodergren, E., Wheeler, D. A., Worley, K. C., Jiang, H., Weinstock, G. M., Gibbs, R. A., Graves, T., Fulton, R., Mardis, E. R., Wilson, R. K., Clamp, M., Cuff, J., Gnerre, S., Jaffe, D. B., Chang, J. L., Lindblad-Toh, K., Lander, E. S., Koriabine, M., Nefedov, M., Osoegawa, K., Yoshinaga, Y., Zhu, B. and De Jong, P. J. (2007). Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. Nature 447 (7146): 799-816.

Bitgood, M. J. and Mcmahon, A. P. (1995). Hedgehog and Bmp genes are coexpressed at many diverse sites of cell-cell interaction in the mouse embryo. Dev Biol 172 (1): 126-38.

Bitgood, M. J., Shen, L. and Mcmahon, A. P. (1996). Sertoli cell signaling by Desert hedgehog regulates the male germline. Curr Biol 6 (3): 298-304.

Boffelli, D., Mcauliffe, J., Ovcharenko, D., Lewis, K. D., Ovcharenko, I., Pachter, L. and Rubin, E. M. (2003). Phylogenetic Shadowing of Primate Sequences to Find Functional Regions of the Human Genome. Science 299 (5611): 1391-1394.

Borycki, A. G., Brunk, B., Tajbakhsh, S., Buckingham, M., Chiang, C. and Emerson, C. P., Jr. (1999). Sonic hedgehog controls epaxial muscle determination through Myf5 activation. Development 126 (18): 4053-63.

Borycki, A. G., Mendham, L. and Emerson, C. P., Jr. (1998). Control of somite patterning by Sonic hedgehog and its downstream signal response genes. Development 125 (4): 777-90.

Borycki, A., Brown, A. M. and Emerson, C. P., Jr. (2000). Shh and Wnt signaling pathways converge to control Gli gene activation in avian somites. Development 127 (10): 2075-87.

Bose, J., Grotewold, L. and Ruther, U. (2002). Pallister-Hall syndrome phenotype in mice mutant for Gli3. Hum Mol Genet 11 (9): 1129-35.

Bradac, J. A., Gruber, C. E., Forry-Schaudies, S. and Hughes, S. H. (1989). Isolation and characterization of related cDNA clones encoding skeletal muscle beta-tropomyosin and a low-molecular-weight nonmuscle tropomyosin isoform. Mol Cell Biol 9 (1): 185-92.

Brannon, M., Gomperts, M., Sumoy, L., Moon, R. T. and Kimelman, D. (1997). A beta-catenin/XTcf-3 complex binds to the siamois promoter to regulate dorsal axis specification in Xenopus. Genes Dev 11 (18): 2359-70.

Brasset, E. and Vaury, C. (2005). Insulators are fundamental components of the eukaryotic genomes. Heredity 94 (6): 571-6.

Bray, N. and Pachter, L. (2004). MAVID: Constrained Ancestral Alignment of Multiple Sequences. Genome Res. 14 (4): 693-699.

Bray, N., Dubchak, I. and Pachter, L. (2003). AVID: A global alignment program. Genome Res 13 (1): 97-102.

Breathnach, R. and Chambon, P. (1981). Organization and expression of eucaryotic split genes coding for proteins. Annu Rev Biochem 50: 349-83.

Briscoe, J. and Novitch, B. G. (2008). Regulatory pathways linking progenitor patterning, cell fates and neurogenesis in the ventral neural tube. Philos Trans R Soc Lond B Biol Sci 363 (1489): 57-70.

Briscoe, J., Pierani, A., Jessell, T. M. and Ericson, J. (2000). A homeodomain protein code specifies progenitor cell identity and neuronal fate in the ventral neural tube. Cell 101 (4): 435-45.

Briscoe, J., Sussel, L., Serup, P., Hartigan-O'connor, D., Jessell, T. M., Rubenstein, J. L. and Ericson, J. (1999). Homeobox gene Nkx2.2 and specification of neuronal identity by graded Sonic hedgehog signalling. Nature 398 (6728): 622-7.

Bromleigh, V. C. and Freedman, L. P. (2000). p21 is a transcriptional target of HOXA10 in differentiating myelomonocytic cells. Genes Dev 14 (20): 2581-6.

Brudno, M., Do, C. B., Cooper, G. M., Kim, M. F., Davydov, E., Program, N. C. S., Green, E. D., Sidow, A. and Batzoglou, S. (2003a). LAGAN and Multi-LAGAN: Efficient Tools for Large-Scale Multiple Alignment of Genomic DNA. Genome Res. 13 (4): 721-731.

Brudno, M., Malde, S., Poliakov, A., Do, C. B., Couronne, O., Dubchak, I. and Batzoglou, S. (2003b). Glocal alignment: finding rearrangements during alignment. Bioinformatics 19 (suppl_1): i54-62.

Bucher, P. (1990). Weight matrix descriptions of four eukaryotic RNA polymerase II promoter elements derived from 502 unrelated promoter sequences. J Mol Biol 212 (4): 563-78.

Bulfone, A., Puelles, L., Porteus, M. H., Frohman, M. A., Martin, G. R. and Rubenstein, J. L. (1993). Spatially restricted expression of Dlx-1, Dlx-2 (Tes-1), Gbx-2, and Wnt-3 in the embryonic day 12.5 mouse forebrain defines potential transverse and longitudinal segmental boundaries. J Neurosci 13 (7): 3155-72.

Bumcrot, D. A., Takada, R. and Mcmahon, A. P. (1995). Proteolytic processing yields two secreted forms of sonic hedgehog. Mol Cell Biol 15 (4): 2294-303.

Burglin, T. R. (1997). Analysis of TALE superclass homeobox genes (MEIS, PBC, KNOX, Iroquois, TGIF) reveals a novel domain conserved between plants and animals. Nucleic Acids Res 25 (21): 4173-80.

Burke, R., Nellen, D., Bellotto, M., Hafen, E., Senti, K. A., Dickson, B. J. and Basler, K. (1999). Dispatched, a novel sterol-sensing domain protein dedicated to the release of cholesterol-modified hedgehog from signaling cells. Cell 99 (7): 803-15.

Buscher, D. and Ruther, U. (1998). Expression profile of Gli family members and Shh in normal and mutant mouse limb development. Dev Dyn 211 (1): 88-96.

Buscher, D., Bosse, B., Heymer, J. and Ruther, U. (1997). Evidence for genetic control of Sonic hedgehog by Gli3 in mouse limb development. Mech Dev 62 (2): 175-82.

Buscher, D., Grotewold, L. and Ruther, U. (1998). The XtJ allele generates a Gli3 fusion transcript. Mamm Genome 9 (8): 676-8.

Buttitta, L., Mo, R., Hui, C. C. and Fan, C. M. (2003). Interplays of Gli2 and Gli3 and their requirement in mediating Shh-dependent sclerotome induction. Development 130 (25): 6233-43.

Calvo, K. R., Knoepfler, P., Mcgrath, S. and Kamps, M. P. (1999). An inhibitory switch derepressed by pbx, hox, and Meis/Prep1 partners regulates DNA-binding by pbx1 and E2a-pbx1 and is dispensable for myeloid immortalization by E2a-pbx1. Oncogene 18 (56): 8033-43.

Capdevila, J., Tabin, C. and Johnson, R. L. (1998). Control of dorsoventral somite patterning by Wnt-1 and beta-catenin. Dev Biol 193 (2): 182-94.

Capdevila, J., Tsukui, T., Rodriquez Esteban, C., Zappavigna, V. and Izpisua Belmonte, J. C. (1999). Control of vertebrate limb outgrowth by the proximal factor Meis2 and distal antagonism of BMPs by Gremlin. Mol Cell 4 (5): 839-49.

Capellini, T. D., Di Giacomo, G., Salsi, V., Brendolan, A., Ferretti, E., Srivastava, D., Zappavigna, V. and Selleri, L. (2006). Pbx1/Pbx2 requirement for distal limb patterning is mediated by the hierarchical control of Hox gene spatial distribution and Shh expression. Development 133 (11): 2263-73.

Carninci, P., Sandelin, A., Lenhard, B., Katayama, S., Shimokawa, K., Ponjavic, J., Semple, C. A., Taylor, M. S., Engstrom, P. G., Frith, M. C., Forrest, A. R., Alkema, W. B., Tan, S. L., Plessy, C., Kodzius, R., Ravasi, T., Kasukawa, T., Fukuda, S., Kanamori-Katayama, M., Kitazume, Y., Kawaji, H., Kai, C., Nakamura, M., Konno, H., Nakano, K., Mottagui-Tabar, S., Arner, P., Chesi, A., Gustincich, S., Persichetti, F., Suzuki, H., Grimmond, S. M., Wells, C. A., Orlando, V., Wahlestedt, C., Liu, E. T., Harbers, M., Kawai, J., Bajic, V. B., Hume, D. A. and Hayashizaki, Y. (2006). Genome-wide analysis of mammalian promoter architecture and evolution. Nat Genet 38 (6): 626-35.

Carpenter, D., Stone, D. M., Brush, J., Ryan, A., Armanini, M., Frantz, G., Rosenthal, A. and De Sauvage, F. J. (1998). Characterization of two patched receptors for the vertebrate hedgehog protein family. Proc Natl Acad Sci U S A 95 (23): 13630-4.

Carter, D., Chakalova, L., Osborne, C. S., Dai, Y. F. and Fraser, P. (2002). Long-range chromatin regulatory interactions in vivo. Nat Genet 32 (4): 623-6.

Cartharius, K., Frech, K., Grote, K., Klocke, B., Haltmeier, M., Klingenhoff, A., Frisch, M., Bayerlein, M. and Werner, T. (2005). MatInspector and beyond: promoter analysis based on transcription factor binding sites. Bioinformatics 21 (13): 2933-42.

Caspary, T., Garcia-Garcia, M. J., Huangfu, D., Eggenschwiler, J. T., Wyler, M. R., Rakeman, A. S., Alcorn, H. L. and Anderson, K. V. (2002). Mouse Dispatched homolog1 is required for long-range, but not juxtacrine, Hh signaling. Curr Biol 12 (18): 1628-32.

Caspary, T., Larkins, C. E. and Anderson, K. V. (2007). The graded response to Sonic Hedgehog depends on cilia architecture. Dev Cell 12 (5): 767-78.

Cauthen, C. A., Berdougo, E., Sandler, J. and Burrus, L. W. (2001). Comparative analysis of the expression patterns of Wnts and Frizzleds during early myogenesis in chick embryos. Mech Dev 104 (1-2): 133-8.

Cavallo, R. A., Cox, R. T., Moline, M. M., Roose, J., Polevoy, G. A., Clevers, H., Peifer, M. and Bejsovec, A. (1998). Drosophila Tcf and Groucho interact to repress Wingless signalling activity. Nature 395 (6702): 604-8.

Cawley, S., Bekiranov, S., Ng, H. H., Kapranov, P., Sekinger, E. A., Kampa, D., Piccolboni, A., Sementchenko, V., Cheng, J., Williams, A. J., Wheeler, R., Wong, B., Drenkow, J., Yamanaka, M., Patel, S., Brubaker, S., Tammana, H., Helt, G., Struhl, K. and Gingeras, T. R. (2004). Unbiased mapping of transcription factor binding sites along human chromosomes 21 and 22 points to widespread regulation of noncoding RNAs. Cell 116 (4): 499-509.

Cecconi, F., Proetzel, G., Alvarez-Bolado, G., Jay, D. and Gruss, P. (1997). Expression of Meis2, a Knotted-related murine homeobox gene, indicates a role in the differentiation of the forebrain and the somitic mesoderm. Dev Dyn 210 (2): 184-90.

Cepko, C., Ryder, E., Fekete, D. and Bruhn, S. Detection of ß-Galactosidase and alkaline phosphatase activities in tissue. http://axon.med.harvard.edu/~cepko/protocol/xgalplap-stain.htm.

Chalepakis, G. and Gruss, P. (1995). Identification of DNA recognition sequences for the Pax3 paired domain. Gene 162 (2): 267-70.

Chang, C. P., Brocchieri, L., Shen, W. F., Largman, C. and Cleary, M. L. (1996). Pbx modulation of Hox homeodomain amino-terminal arms establishes different DNA-binding specificities across the Hox locus. Mol Cell Biol 16 (4): 1734-45.

Chang, C. P., Jacobs, Y., Nakamura, T., Jenkins, N. A., Copeland, N. G. and Cleary, M. L. (1997). Meis proteins are major in vivo DNA binding partners for wild-type but not chimeric Pbx proteins. Mol Cell Biol 17 (10): 5679-87.

Chang, C. P., Shen, W. F., Rozenfeld, S., Lawrence, H. J., Largman, C. and Cleary, M. L. (1995). Pbx proteins display hexapeptide-dependent cooperative DNA binding with a subset of Hox proteins. Genes Dev 9 (6): 663-74.

Charrier, J. B., Lapointe, F., Le Douarin, N. M. and Teillet, M. A. (2001). Anti-apoptotic role of Sonic hedgehog protein at the early stages of nervous system organogenesis. Development 128 (20): 4011-20.

Chen, F. C. and Li, W. H. (2001). Genomic divergences between humans and other hominoids and the effective population size of the common ancestor of humans and chimpanzees. Am J Hum Genet 68 (2): 444-56.

Chen, G., Fernandez, J., Mische, S. and Courey, A. J. (1999). A functional interaction between the histone deacetylase Rpd3 and the corepressor groucho in Drosophila development. Genes Dev 13 (17): 2218-30.

Chen, J. and Ruley, H. E. (1998). An enhancer element in the EphA2 (Eck) gene sufficient for rhombomere-specific expression is activated by HOXA1 and HOXB1 homeobox proteins. J Biol Chem 273 (38): 24670-5.

Chen, W., Burgess, S. and Hopkins, N. (2001). Analysis of the zebrafish smoothened mutant reveals conserved and divergent functions of hedgehog activity. Development 128 (12): 2385-96.

Chen, X., Weisberg, E., Fridmacher, V., Watanabe, M., Naco, G. and Whitman, M. (1997). Smad4 and FAST-1 in the assembly of activin-responsive factor. Nature 389 (6646): 85-9.

Chen, Y. and Struhl, G. (1996). Dual roles for patched in sequestering and transducing Hedgehog. Cell 87 (3): 553-63.

Chen, Z. F., Paquette, A. J. and Anderson, D. J. (1998). NRSF/REST is required in vivo for repression of multiple neuronal target genes during embryogenesis. Nat Genet 20 (2): 136-42.

Cheng, S., Fockler, C., Barnes, W. and Higuchi, R. (1994). Effective Amplification of Long Targets from Cloned Inserts and Human Genomic DNA. Proceedings of the National Academy of Sciences 91 (12): 5695-5699.

Chesnutt, C., Burrus, L. W., Brown, A. M. and Niswander, L. (2004). Coordinate regulation of neural tube patterning and proliferation by TGFbeta and WNT activity. Dev Biol 274 (2): 334-47.

Chiang, C., Litingtung, Y., Lee, E., Young, K. E., Corden, J. L., Westphal, H. and Beachy, P. A. (1996). Cyclopia and defective axial patterning in mice lacking Sonic hedgehog gene function. Nature 383 (6599): 407-13.

Chiang, C., Swan, R. Z., Grachtchouk, M., Bolinger, M., Litingtung, Y., Robertson, E. K., Cooper, M. K., Gaffield, W., Westphal, H., Beachy, P. A. and Dlugosz, A. A. (1999). Essential role for Sonic hedgehog during hair follicle morphogenesis. Dev Biol 205 (1): 1-9.

Chicken Genome Sequencing Consortium (2004). Sequence and comparative analysis of the chicken genome provide unique perspectives on vertebrate evolution. 432 (7018): 695-716.

Chizhikov, V. V. and Millen, K. J. (2004). Control of roof plate development and signaling by Lmx1b in the caudal vertebrate CNS. J Neurosci 24 (25): 5694-703.

Chizhikov, V. V. and Millen, K. J. (2005). Roof plate-dependent patterning of the vertebrate dorsal central nervous system. Dev Biol 277 (2): 287-95.

Choe, S. K. and Sagerstrom, C. G. (2005). Variable Meis-dependence among paralog group-1 Hox proteins. Biochem Biophys Res Commun 331 (4): 1384-91.

Chuang, P. T. and Mcmahon, A. P. (1999). Vertebrate Hedgehog signalling modulated by induction of a Hedgehog-binding protein. Nature 397 (6720): 617-21.

Chuong, C. M., Patel, N., Lin, J., Jung, H. S. and Widelitz, R. B. (2000). Sonic hedgehog signaling pathway in vertebrate epithelial appendage morphogenesis: perspectives in development and evolution. Cell Mol Life Sci 57 (12): 1672-81.

Clevers, H. C. and Grosschedl, R. (1996). Transcriptional control of lymphoid development: lessons from gene targeting. Immunol Today 17 (7): 336-43.

Cohen, D. R., Cheng, C. W., Cheng, S. H. and Hui, C. C. (2000). Expression of two novel mouse Iroquois homeobox genes during neurogenesis. Mech Dev 91 (1-2): 317-21.

Cohen, M. M., Jr. (1989). Perspectives on holoprosencephaly: Part I. Epidemiology, genetics, and syndromology. Teratology 40 (3): 211-35.

Cohen-Kaminsky, S., Maouche-Chretien, L., Vitelli, L., Vinit, M. A., Blanchard, I., Yamamoto, M., Peschle, C. and Romeo, P. H. (1998). Chromatin immunoselection defines a TAL-1 target gene. Embo J 17 (17): 5151-60.

Cooper, G. M. and Sidow, A. (2003). Genomic regulatory regions: insights from comparative sequence analysis. Curr Opin Genet Dev 13 (6): 604-10.

Cooper, G. M., Brudno, M., Stone, E. A., Dubchak, I., Batzoglou, S. and Sidow, A. (2004). Characterization of evolutionary rates and constraints in three Mammalian genomes. Genome Res 14 (4): 539-48.

Corbit, K. C., Aanstad, P., Singla, V., Norman, A. R., Stainier, D. Y. and Reiter, J. F. (2005). Vertebrate Smoothened functions at the primary cilium. Nature 437 (7061): 1018-21.

Couronne, O., Poliakov, A., Bray, N., Ishkhanov, T., Ryaboy, D., Rubin, E., Pachter, L. and Dubchak, I. (2003). Strategies and tools for whole-genome alignments. Genome Res 13 (1): 73-80.

Cremer, T. and Cremer, C. (2001). Chromosome territories, nuclear architecture and gene regulation in mammalian cells. Nat Rev Genet 2 (4): 292-301.

Cui, C., Elsam, T., Tian, Q., Seykora, J. T., Grachtchouk, M., Dlugosz, A. and Tseng, H. (2004). Gli proteins up-regulate the expression of basonuclin in Basal cell carcinoma. Cancer Res 64 (16): 5651-8.

Czerny, T. and Busslinger, M. (1995). DNA-binding and transactivation properties of Pax-6: three amino acids in the paired domain are responsible for the different sequence recognition of Pax-6 and BSAP (Pax-5). Mol Cell Biol 15 (5): 2858-71.

Dai, P., Akimaru, H., Tanaka, Y., Maekawa, T., Nakafuku, M. and Ishii, S. (1999). Sonic Hedgehog-induced activation of the Gli1 promoter is mediated by GLI3. J Biol Chem 274 (12): 8143-52.

Das, R. M., Van Hateren, N. J., Howell, G. R., Farrell, E. R., Bangs, F. K., Porteous, V. C., Manning, E. M., Mcgrew, M. J., Ohyama, K., Sacco, M. A., Halley, P. A., Sang, H. M., Storey, K. G., Placzek, M., Tickle, C., Nair, V. K. and Wilson, S. A. (2006). A robust system for RNA interference in the chicken using a modified microRNA operon. Dev Biol 294 (2): 554-63.

Davidson, E. H., Mcclay, D. R. and Hood, L. (2003). Regulatory gene networks and the properties of the developmental process. Proc Natl Acad Sci U S A 100 (4): 1475-80.

Davuluri, R. V., Grosse, I. and Zhang, M. Q. (2001). Computational identification of promoters and first exons in the human genome. Nat Genet 29 (4): 412-7.

Derynck, R., Gelbart, W. M., Harland, R. M., Heldin, C. H., Kern, S. E., Massague, J., Melton, D. A., Mlodzik, M., Padgett, R. W., Roberts, A. B., Smith, J., Thomsen, G. H., Vogelstein, B. and Wang, X. F. (1996). Nomenclature: vertebrate mediators of TGFbeta family signals. Cell 87 (2): 173.

Di Giacomo, G., Koss, M., Capellini, T. D., Brendolan, A., Popperl, H. and Selleri, L. (2006). Spatio-temporal expression of Pbx3 during mouse organogenesis. Gene Expr Patterns 6 (7): 747-57.

Dickinson, M. E., Krumlauf, R. and Mcmahon, A. P. (1994). Evidence for a mitogenic effect of Wnt-1 in the developing mammalian central nervous system. Development 120 (6): 1453-71.

Dickmeis, T. and Muller, F. (2005). The identification and functional characterisation of conserved regulatory elements in developmental genes. Brief Funct Genomic Proteomic 3 (4): 332-50.

Dickmeis, T., Plessy, C., Rastegar, S., Aanstad, P., Herwig, R., Chalmel, F., Fischer, N. and Strahle, U. (2004). Expression profiling and comparative genomics identify a conserved regulatory region controlling midline expression in the zebrafish embryo. Genome Res 14 (2): 228-38.

Diez Del Corral, R., Olivera-Martinez, I., Goriely, A., Gale, E., Maden, M. and Storey, K. (2003). Opposing FGF and retinoid pathways control ventral neural pattern, neuronal differentiation, and segmentation during body axis extension. Neuron 40 (1): 65-79.

Diiorio, P., Alexa, K., Choe, S. K., Etheridge, L. and Sagerstrom, C. G. (2007). TALE-family homeodomain proteins regulate endodermal sonic hedgehog expression and pattern the anterior endoderm. Dev Biol 304 (1): 221-31.

Dillon, R., Gadgil, C. and Othmer, H. G. (2003). Short- and long-range effects of Sonic hedgehog in limb development. Proc Natl Acad Sci U S A 100 (18): 10152-7.

Ding, Q., Motoyama, J., Gasca, S., Mo, R., Sasaki, H., Rossant, J. and Hui, C. C. (1998). Diminished Sonic hedgehog signaling and lack of floor plate differentiation in Gli2 mutant mice. Development 125 (14): 2533-43.

Dorsky, R. I., Sheldahl, L. C. and Moon, R. T. (2002). A transgenic Lef1/beta-catenin-dependent reporter is expressed in spatially restricted domains throughout zebrafish development. Dev Biol 241 (2): 229-37.

Down, T. A. and Hubbard, T. J. (2002). Computational detection and location of transcription start sites in mammalian genomic DNA. Genome Res 12 (3): 458-61.

Du, S. J., Purcell, S. M., Christian, J. L., Mcgrew, L. L. and Moon, R. T. (1995). Identification of distinct classes and functional domains of Wnts through expression of wild-type and chimeric proteins in Xenopus embryos. Mol Cell Biol 15 (5): 2625-34.

Dudley, A. T. and Robertson, E. J. (1997). Overlapping expression domains of bone morphogenetic protein family members potentially account for limited tissue defects in BMP7 deficient embryos. Dev Dyn 208 (3): 349-62.

Dudley, A. T., Lyons, K. M. and Robertson, E. J. (1995). A requirement for bone morphogenetic protein-7 during development of the mammalian kidney and eye. Genes Dev 9 (22): 2795-807.

Dunn, N. R., Winnier, G. E., Hargett, L. K., Schrick, J. J., Fogo, A. B. and Hogan, B. L. (1997). Haploinsufficient phenotypes in Bmp4 heterozygous null mice and modification by mutations in Gli3 and Alx4. Dev Biol 188 (2): 235-47.

Eastman, Q. and Grosschedl, R. (1999). Regulation of LEF-1/TCF transcription factors by Wnt and other signals. Curr Opin Cell Biol 11 (2): 233-40.

Echelard, Y., Epstein, D. J., St-Jacques, B., Shen, L., Mohler, J., Mcmahon, J. A. and Mcmahon, A. P. (1993). Sonic hedgehog, a member of a family of putative signaling molecules, is implicated in the regulation of CNS polarity. Cell 75 (7): 1417-30.

Eddy, S. R. (2002). Computational genomics of noncoding RNA genes. Cell 109 (2): 137-40.

Eichberger, T., Regl, G., Ikram, M. S., Neill, G. W., Philpott, M. P., Aberger, F. and Frischauf, A. M. (2004). FOXE1, a new transcriptional target of GLI2 is expressed in human epidermis and basal cell carcinoma. J Invest Dermatol 122 (5): 1180-7.

Epstein, D., Mcmahon, A. and Joyner, A. (1999). Regionalization of Sonic hedgehog transcription along the anteroposterior axis of the mouse central nervous system is regulated by Hnf3-dependent and -independent mechanisms. Development 126 (2): 281-292.

Epstein, J., Cai, J., Glaser, T., Jepeal, L. and Maas, R. (1994). Identification of a Pax paired domain recognition sequence and evidence for DNA-dependent conformational changes. J Biol Chem 269 (11): 8355-61.

Erickson, T., Scholpp, S., Brand, M., Moens, C. B. and Waskiewicz, A. J. (2007). Pbx proteins cooperate with Engrailed to pattern the midbrain-hindbrain and diencephalic-mesencephalic boundaries. Dev Biol 301 (2): 504-17.

Ericson, J., Briscoe, J., Rashbass, P., Van Heyningen, V. and Jessell, T. M. (1997a). Graded sonic hedgehog signaling and the specification of cell fate in the ventral neural tube. Cold Spring Harb Symp Quant Biol 62: 451-66.

Ericson, J., Morton, S., Kawakami, A., Roelink, H. and Jessell, T. M. (1996). Two critical periods of Sonic Hedgehog signaling required for the specification of motor neuron identity. Cell 87 (4): 661-73.

Ericson, J., Muhr, J., Placzek, M., Lints, T., Jessell, T. M. and Edlund, T. (1995). Sonic hedgehog induces the differentiation of ventral forebrain neurons: a common signal for ventral patterning within the neural tube. Cell 81 (5): 747-56.

Ericson, J., Rashbass, P., Schedl, A., Brenner-Morton, S., Kawakami, A., Van Heyningen, V., Jessell, T. M. and Briscoe, J. (1997b). Pax6 controls progenitor cell identity and neuronal fate in response to graded Shh signaling. Cell 90 (1): 169-80.

Evans, R., Fairley, J. A. and Roberts, S. G. (2001). Activator-mediated disruption of sequence-specific DNA contacts by the general transcription factor TFIIB. Genes Dev 15 (22): 2945-9.

Fafeur, V., Tulasne, D., Queva, C., Vercamer, C., Dimster, V., Mattot, V., Stehelin, D., Desbiens, X. and Vandenbunder, B. (1997). The ETS1 transcription factor is expressed during epithelial-mesenchymal transitions in the chick embryo and is activated in scatter factor-stimulated MDCK epithelial cells. Cell Growth Differ 8 (6): 655-65.

Fan, C. M. and Tessier-Lavigne, M. (1994). Patterning of mammalian somites by surface ectoderm and notochord: evidence for sclerotome induction by a hedgehog homolog. Cell 79 (7): 1175-86.

Fan, C. M., Lee, C. S. and Tessier-Lavigne, M. (1997a). A role for WNT proteins in induction of dermomyotome. Dev Biol 191 (1): 160-5.

Fan, H. and Khavari, P. A. (1999). Sonic hedgehog opposes epithelial cell cycle arrest. J Cell Biol 147 (1): 71-6.

Fan, H., Oro, A. E., Scott, M. P. and Khavari, P. A. (1997b). Induction of basal cell carcinoma features in transgenic human skin expressing Sonic Hedgehog. Nat Med 3 (7): 788-92.

Faure, S., De Santa Barbara, P., Roberts, D. J. and Whitman, M. (2002). Endogenous patterns of BMP signaling during early chick development. Dev Biol 244 (1): 44-65.

Ferretti, E., Marshall, H., Popperl, H., Maconochie, M., Krumlauf, R. and Blasi, F. (2000). Segmental expression of Hoxb2 in r4 requires two separate sites that integrate cooperative interactions between Prep1, Pbx and Hox proteins. Development 127 (1): 155-66.

Ferretti, E., Schulz, H., Talarico, D., Blasi, F. and Berthelsen, J. (1999). The PBX-regulating protein PREP1 is present in different PBX-complexed forms in mouse. Mech Dev 83 (1-2): 53-64.

Fognani, C., Kilstrup-Nielsen, C., Berthelsen, J., Ferretti, E., Zappavigna, V. and Blasi, F. (2002). Characterization of PREP2, a paralog of PREP1, which defines a novel sub-family of the MEINOX TALE homeodomain transcription factors. Nucleic Acids Res 30 (9): 2043-51.

Fougerousse, F., Bullen, P., Herasse, M., Lindsay, S., Richard, I., Wilson, D., Suel, L., Durand, M., Robson, S., Abitbol, M., Beckmann, J. S. and Strachan, T. (2000). Human-mouse differences in the embryonic expression patterns of developmental control genes and disease genes. Hum Mol Genet 9 (2): 165-73.

Frazer, K. A., Sheehan, J. B., Stokowski, R. P., Chen, X., Hosseini, R., Cheng, J. F., Fodor, S. P., Cox, D. R. and Patil, N. (2001). Evolutionarily conserved sequences on human chromosome 21. Genome Res 11 (10): 1651-9.

Frazer, K., Pachter, L., Poliakov, A., Rubin, E. and Dubchak, I. (2004). VISTA: computational tools for comparative genomics. Nucleic Acids Res. 32 (Web server issue): W273-9.

Frith, M. C., Hansen, U. and Weng, Z. (2001). Detection of cis-element clusters in higher eukaryotic DNA. Bioinformatics 17 (10): 878-89.

Frohman, M. A., Dush, M. K. and Martin, G. R. (1988). Rapid production of full-length cDNAs from rare transcripts: amplification using a single gene-specific oligonucleotide primer. Proc Natl Acad Sci U S A 85 (23): 8998-9002.

Fukue, Y., Sumida, N., Nishikawa, J. and Ohyama, T. (2004). Core promoter elements of eukaryotic genes have a highly distinctive mechanical property. Nucleic Acids Res 32 (19): 5834-40.

Gajovic, S., St-Onge, L., Yokota, Y. and Gruss, P. (1997). Retinoic acid mediates Pax6 expression during in vitro differentiation of embryonic stem cells. Differentiation 62 (4): 187-92.

Galas, D. J. and Schmitz, A. (1978). DNAse footprinting: a simple method for the detection of protein-DNA binding specificity. Nucleic Acids Res 5 (9): 3157-70.

Galceran, J., Farinas, I., Depew, M. J., Clevers, H. and Grosschedl, R. (1999). Wnt3a-/--like phenotype and limb deficiency in Lef1(-/-)Tcf1(-/-) mice. Genes Dev 13 (6): 709-17.

Galceran, J., Sustmann, C., Hsu, S. C., Folberth, S. and Grosschedl, R. (2004). LEF1-mediated regulation of Delta-like1 links Wnt and Notch signaling in somitogenesis. Genes Dev 18 (22): 2718-23.

Gardiner-Garden, M. and Frommer, M. (1987). CpG islands in vertebrate genomes. J Mol Biol 196 (2): 261-82.

Garner, M. M. and Revzin, A. (1981). A gel electrophoresis method for quantifying the binding of proteins to specific DNA regions: application to components of the Escherichia coli lactose operon regulatory system. Nucl. Acids Res. 9 (13): 3047-3060.

Gershenzon, N. I. and Ioshikhes, I. P. (2005). Synergy of human Pol II core promoter elements revealed by statistical sequence analysis. Bioinformatics 21 (8): 1295-300.

Giese, K., Cox, J. and Grosschedl, R. (1992). The HMG domain of lymphoid enhancer factor 1 bends DNA and facilitates assembly of functional nucleoprotein structures. Cell 69 (1): 185-95.

Gillemans, N., Mcmorrow, T., Tewari, R., Wai, A. W., Burgtorf, C., Drabek, D., Ventress, N., Langeveld, A., Higgs, D., Tan-Un, K., Grosveld, F. and Philipsen, S. (2003). Functional and comparative analysis of globin loci in pufferfish and humans. Blood 101 (7): 2842-9.

Glazko, G. V., Koonin, E. V., Rogozin, I. B. and Shabalina, S. A. (2003). A significant fraction of conserved noncoding DNA in human and mouse consists of predicted matrix attachment regions. Trends Genet 19 (3): 119-24.

Goodman, M., Porter, C. A., Czelusniak, J., Page, S. L., Schneider, H., Shoshani, J., Gunnell, G. and Groves, C. P. (1998). Toward a phylogenetic classification of Primates based on DNA evidence complemented by fossil evidence. Mol Phylogenet Evol 9 (3): 585-98.

Goodrich, L. V., Milenkovic, L., Higgins, K. M. and Scott, M. P. (1997). Altered neural cell fates and medulloblastoma in mouse patched mutants. Science 277 (5329): 1109-13.

Goto, K., Kamiya, Y., Imamura, T., Miyazono, K. and Miyazawa, K. (2007). Selective inhibitory effects of Smad6 on bone morphogenetic protein type I receptors. J Biol Chem 282 (28): 20603-11.

Gottgens, B., Barton, L. M., Gilbert, J. G., Bench, A. J., Sanchez, M. J., Bahn, S., Mistry, S., Grafham, D., Mcmurray, A., Vaudin, M., Amaya, E., Bentley, D. R., Green, A. R. and Sinclair, A. M. (2000). Analysis of vertebrate SCL loci identifies conserved enhancers. Nat Biotechnol 18 (2): 181-6.

Goudet, G., Delhalle, S., Biemar, F., Martial, J. A. and Peers, B. (1999). Functional and cooperative interactions between the homeodomain PDX1, Pbx, and Prep1 factors on the somatostatin promoter. J Biol Chem 274 (7): 4067-73.

Goulding, M. D., Chalepakis, G., Deutsch, U., Erselius, J. R. and Gruss, P. (1991). Pax-3, a novel murine DNA binding protein expressed during early neurogenesis. Embo J 10 (5): 1135-47.

Goumans, M. J., Valdimarsdottir, G., Itoh, S., Rosendahl, A., Sideras, P. and Ten Dijke, P. (2002). Balancing the activation state of the endothelium via two distinct TGF-beta type I receptors. Embo J 21 (7): 1743-53.

Gowan, K., Helms, A. W., Hunsaker, T. L., Collisson, T., Ebert, P. J., Odom, R. and Johnson, J. E. (2001). Crossinhibitory activities of Ngn1 and Math1 allow specification of distinct dorsal interneurons. Neuron 31 (2): 219-32.

Greene, L. A. and Tischler, A. S. (1976). Establishment of a noradrenergic clonal line of rat adrenal pheochromocytoma cells which respond to nerve growth factor. Proc Natl Acad Sci U S A 73 (7): 2424-8.

Greene, N. D., Gerrelli, D., Van Straaten, H. W. and Copp, A. J. (1998). Abnormalities of floor plate, notochord and somite differentiation in the loop-tail (Lp) mouse: a model of severe neural tube defects. Mech Dev 73 (1): 59-72.

Griffin, C., Kleinjan, D. A., Doe, B. and Van Heyningen, V. (2002). New 3' elements control Pax6 expression in the developing pretectum, neural retina and olfactory region. Mech Dev 112 (1-2): 89-100.

Gross, D. S. and Garrard, W. T. (1988). Nuclease hypersensitive sites in chromatin. Annu Rev Biochem 57: 159-97.

Gross, M. K., Dottori, M. and Goulding, M. (2002). Lbx1 specifies somatosensory association interneurons in the dorsal spinal cord. Neuron 34 (4): 535-49.

Gross, P. and Oelgeschlager, T. (2006). Core promoter-selective RNA polymerase II transcription. Biochem Soc Symp (73): 225-36.

Gumucio, D. L., Heilstedt-Williamson, H., Gray, T. A., Tarle, S. A., Shelton, D. A., Tagle, D. A., Slightom, J. L., Goodman, M. and Collins, F. S. (1992). Phylogenetic footprinting reveals a nuclear protein which binds to silencer sequences in the human gamma and epsilon globin genes. Mol. Cell. Biol. 12 (11): 4919-4929.

Guo, D.-F., Uno, S., Ishihata, A., Nakamura, N. and Inagami, T. (1995). Identification of a cis-Acting Glucocorticoid Responsive Element in the Rat Angiotensin II Type 1A Promoter. Circ Res 77 (2): 249-257.

Hadchouel, J., Carvajal, J. J., Daubas, P., Bajard, L., Chang, T., Rocancourt, D., Cox, D., Summerbell, D., Tajbakhsh, S., Rigby, P. W. and Buckingham, M. (2003). Analysis of a key regulatory region upstream of the Myf5 gene reveals multiple phases of myogenesis, orchestrated at each site by a combination of elements dispersed throughout the locus. Development 130 (15): 3415-26.

Hahn, S. A., Schutte, M., Hoque, A. T., Moskaluk, C. A., Da Costa, L. T., Rozenblum, E., Weinstein, C. L., Fischer, A., Yeo, C. J., Hruban, R. H. and Kern, S. E. (1996). DPC4, a candidate tumor suppressor gene at human chromosome 18q21.1. Science 271 (5247): 350-3.

Haller, K., Rambaldi, I., Kovacs, E. N., Daniels, E. and Featherstone, M. (2002). Prep2: cloning and expression of a new prep family member. Dev Dyn 225 (3): 358-64.

Hallikas, O., Palin, K., Sinjushina, N., Rautiainen, R., Partanen, J., Ukkonen, E. and Taipale, J. (2006). Genome-wide prediction of mammalian enhancers based on analysis of transcription-factor binding affinity. Cell 124 (1): 47-59.

Hamburger, V. and Hamilton, H. L. (1992). A series of normal stages in the development of the chick embryo. 1951. Dev Dyn 195 (4): 231-72.

Hanyu, A., Ishidou, Y., Ebisawa, T., Shimanuki, T., Imamura, T. and Miyazono, K. (2001). The N domain of Smad7 is essential for specific inhibition of transforming growth factor-beta signaling. J Cell Biol 155 (6): 1017-27.

Hardcastle, Z., Mo, R., Hui, C. C. and Sharpe, P. T. (1998). The Shh signalling pathway in tooth development: defects in Gli2 and Gli3 mutants. Development 125 (15): 2803-11.

Hatzis, P. and Talianidis, I. (2002). Dynamics of enhancer-promoter communication during differentiation-induced gene activation. Mol Cell 10 (6): 1467-77.

Hayashi, H., Abdollah, S., Qiu, Y., Cai, J., Xu, Y. Y., Grinnell, B. W., Richardson, M. A., Topper, J. N., Gimbrone, M. A., Jr., Wrana, J. L. and Falb, D. (1997). The MAD-related protein Smad7 associates with the TGFbeta receptor and functions as an antagonist of TGFbeta signaling. Cell 89 (7): 1165-73.

Haycraft, C. J., Banizs, B., Aydin-Son, Y., Zhang, Q., Michaud, E. J. and Yoder, B. K. (2005). Gli2 and Gli3 localize to cilia and require the intraflagellar transport protein polaris for processing and function. PLoS Genet 1 (4): e53.

Hedgepeth, C. M., Conrad, L. J., Zhang, J., Huang, H. C., Lee, V. M. and Klein, P. S. (1997). Activation of the Wnt signaling pathway: a molecular mechanism for lithium action. Dev Biol 185 (1): 82-91.

Hedges, S. B. (2002). The origin and evolution of model organisms. Nat Rev Genet 3 (11): 838-49.

Helms, A. W. and Johnson, J. E. (1998). Progenitors of dorsal commissural interneurons are defined by MATH1 expression. Development 125 (5): 919-28.

Helms, A. W., Battiste, J., Henke, R. M., Nakada, Y., Simplicio, N., Guillemot, F. and Johnson, J. E. (2005). Sequential roles for Mash1 and Ngn2 in the generation of dorsal spinal cord interneurons. Development 132 (12): 2709-19.

Hollyday, M., Mcmahon, J. A. and Mcmahon, A. P. (1995). Wnt expression patterns in chick embryo nervous system. Mech Dev 52 (1): 9-25.

Hu, M. C. and Rosenblum, N. D. (2005). Smad1, beta-catenin and Tcf4 associate in a molecular complex with the Myc promoter in dysplastic renal tissue and cooperate to control Myc transcription. Development 132 (1): 215-25.

Huang, H., Paliouras, M., Rambaldi, I., Lasko, P. and Featherstone, M. (2003). Nonmuscle myosin promotes cytoplasmic localization of PBX. Mol Cell Biol 23 (10): 3636-45.

Huang, H., Rastegar, M., Bodner, C., Goh, S. L., Rambaldi, I. and Featherstone, M. (2005). MEIS C termini harbor transcriptional activation domains that respond to cell signaling. J Biol Chem 280 (11): 10119-27.

Hubbard, T. J. P., Aken, B. L., Beal, K., Ballester, B., Caccamo, M., Chen, Y., Clarke, L., Coates, G., Cunningham, F., Cutts, T., Down, T., Dyer, S. C., Fitzgerald, S., Fernandez-Banet, J., Graf, S., Haider, S., Hammond, M., Herrero, J., Holland, R., Howe, K., Johnson, N., Kahari, A., Keefe, D., Kokocinski, F., Kulesha, E., Lawson, D., Longden, I., Melsopp, C., Megy, K., Meidl, P., Ouverdin, B., Parker, A., Prlic, A., Rice, S., Rios, D., Schuster, M., Sealy, I., Severin, J., Slater, G., Smedley, D., Spudich, G., Trevanion, S., Vilella, A., Vogel, J., White, S., Wood, M., Cox, T., Curwen, V., Durbin, R., Fernandez-Suarez, X. M., Flicek, P., Kasprzyk, A., Proctor, G., Searle, S., Smith, J., Ureta-Vidal, A. and Birney, E. (2007). Ensembl 2007. Nucl. Acids Res. 35 (suppl_1): D610-617.

Huber, O., Korn, R., Mclaughlin, J., Ohsugi, M., Herrmann, B. G. and Kemler, R. (1996). Nuclear localization of beta-catenin by interaction with transcription factor LEF-1. Mech Dev 59 (1): 3-10.

Hui, C. and Joyner, A. (1993). A mouse model of Greig cephalo-polysyndactyly syndrome: the extra-toesJ mutation contains an intragenic deletion of the Gli3 gene. Nature Genetics 3 (3): 241-6.

Hui, C. C., Slusarski, D., Platt, K. A., Holmgren, R. and Joyner, A. L. (1994). Expression of three mouse homologs of the Drosophila segment polarity gene cubitus interruptus, Gli, Gli-2, and Gli-3, in ectoderm- and mesoderm-derived tissues suggests multiple roles during postimplantation development. Dev Biol 162 (2): 402-13.

Hussein, S. M., Duff, E. K. and Sirard, C. (2003). Smad4 and beta-catenin co-activators functionally interact with lymphoid-enhancing factor to regulate graded expression of Msx2. J Biol Chem 278 (49): 48805-14.

Hynes, M., Stone, D. M., Dowd, M., Pitts-Meek, S., Goddard, A., Gurney, A. and Rosenthal, A. (1997). Control of cell pattern in the neural tube by the zinc finger transcription factor and oncogene Gli-1. Neuron 19 (1): 15-26.

Hynes, M., Ye, W., Wang, K., Stone, D., Murone, M., Sauvage, F. and Rosenthal, A. (2000). The seven-transmembrane receptor smoothened cell-autonomously induces multiple ventral cell types. Nat Neurosci 3 (1): 41-6.

Ikeya, M. and Takada, S. (1998). Wnt signaling from the dorsal neural tube is required for the formation of the medial dermomyotome. Development 125 (24): 4969-76.

Ikeya, M., Lee, S. M., Johnson, J. E., Mcmahon, A. P. and Takada, S. (1997). Wnt signalling required for expansion of neural crest and CNS progenitors. Nature 389 (6654): 966-70.

Ikram, M. S., Neill, G. W., Regl, G., Eichberger, T., Frischauf, A. M., Aberger, F., Quinn, A. and Philpott, M. (2004). GLI2 is expressed in normal human epidermis and BCC and induces GLI1 expression by binding to its promoter. J Invest Dermatol 122 (6): 1503-9.

Ille, F., Atanasoski, S., Falk, S., Ittner, L. M., Marki, D., Buchmann-Moller, S., Wurdak, H., Suter, U., Taketo, M. M. and Sommer, L. (2007). Wnt/BMP signal integration regulates the balance between proliferation and differentiation of neuroepithelial cells in the dorsal spinal cord. Dev Biol 304 (1): 394-408.

Imamura, T., Takase, M., Nishihara, A., Oeda, E., Hanai, J., Kawabata, M. and Miyazono, K. (1997). Smad6 inhibits signalling by the TGF-beta superfamily. Nature 389 (6651): 622-6.

Imoto, I., Pimkhaokham, A., Watanabe, T., Saito-Ohara, F., Soeda, E. and Inazawa, J. (2000). Amplification and overexpression of TGIF2, a novel homeobox gene of the TALE superclass, in ovarian cancer cell lines. Biochem Biophys Res Commun 276 (1): 264-70.

In Der Rieden, P. M., Mainguy, G., Woltering, J. M. and Durston, A. J. (2004). Homeodomain to hexapeptide or PBC-interaction-domain distance: size apparently matters. Trends Genet 20 (2): 76-9.

Ingham, P. W. and Mcmahon, A. P. (2001). Hedgehog signaling in animal development: paradigms and principles. Genes Dev 15 (23): 3059-87.

Inman, G. J. and Hill, C. S. (2002). Stoichiometry of active smad-transcription factor complexes on DNA. J Biol Chem 277 (52): 51008-16.

Ishitani, T., Ninomiya-Tsuji, J., Nagai, S., Nishita, M., Meneghini, M., Barker, N., Waterman, M., Bowerman, B., Clevers, H., Shibuya, H. and Matsumoto, K. (1999). The TAK1-NLK-MAPK-related pathway antagonizes signalling between beta-catenin and transcription factor TCF. Nature 399 (6738): 798-802.

Itasaki, N., Bel-Vialar, S. and Krumlauf, R. (1999). 'Shocking' developments in chick embryology: electroporation and in ovo gene expression. Nat Cell Biol 1 (8): E203-7.

Jacob, J. and Briscoe, J. (2003). Gli proteins and the control of spinal-cord patterning. EMBO Rep 4 (8): 761-5.

Jacobs, Y., Schnabel, C. A. and Cleary, M. L. (1999). Trimeric association of Hox and TALE homeodomain proteins mediates Hoxb2 hindbrain enhancer activity. Mol Cell Biol 19 (7): 5134-42.

Javahery, R., Khachi, A., Lo, K., Zenzie-Gregory, B. and Smale, S. T. (1994). DNA sequence requirements for transcriptional initiator activity in mammalian cells. Mol Cell Biol 14 (1): 116-27.

Jessell, T. M., Bovolenta, P., Placzek, M., Tessier-Lavigne, M. and Dodd, J. (1989). Polarity and patterning in the neural tube: the origin and function of the floor plate. Ciba Found Symp 144: 255-76; discussion 276-80, 290-5.

Jia, J., Tong, C. and Jiang, J. (2003). Smoothened transduces Hedgehog signal by physically interacting with Costal2/Fused complex through its C-terminal tail. Genes Dev 17 (21): 2709-20.

Jiang, J. (2002). Degrading Ci: who is Cul-pable? Genes Dev 16 (18): 2315-21.

Jiang, J. and Struhl, G. (1998). Regulation of the Hedgehog and Wingless signalling pathways by the F-box/WD40-repeat protein Slimb. Nature 391 (6666): 493-6.

Jin, J. Z., Gu, S., Mckinney, P. and Ding, J. (2006). Expression and functional analysis of Tgif during mouse midline development. Dev Dyn 235 (2): 547-53.

Jochum, W., Passegue, E. and Wagner, E. F. (2001). AP-1 in mouse development and tumorigenesis. Oncogene 20 (19): 2401-12.

Johnson, D. R. (1967). Extra-toes: anew mutant gene causing multiple abnormalities in the mouse. J Embryol Exp Morphol 17 (3): 543-81.

Johnson, R. L., Laufer, E., Riddle, R. D. and Tabin, C. (1994). Ectopic expression of Sonic hedgehog alters dorsal-ventral patterning of somites. Cell 79 (7): 1165-73.

Joyner, A. L. (2002). Establishment of Anterior-Posterior and Dorsal-Ventral Pattern in the Early Central Nervous System; Mouse Development. J. Rossant and P. P. L. Tam. San Diego, Academic Press: 107-126.

Jun, S. and Desplan, C. (1996). Cooperative interactions between paired domain and homeodomain. Development 122 (9): 2639-50.

Kalff-Suske, M., Wild, A., Topp, J., Wessling, M., Jacobsen, E. M., Bornholdt, D., Engel, H., Heuer, H., Aalfs, C. M., Ausems, M. G., Barone, R., Herzog, A., Heutink, P., Homfray, T., Gillessen-Kaesbach, G., Konig, R., Kunze, J., Meinecke, P., Muller, D., Rizzo, R., Strenge, S., Superti-Furga, A. and Grzeschik, K. H. (1999). Point mutations throughout the GLI3 gene cause Greig cephalopolysyndactyly syndrome. Hum Mol Genet 8 (9): 1769-77.

Kamachi, Y., Uchikawa, M., Tanouchi, A., Sekido, R. and Kondoh, H. (2001). Pax6 and SOX2 form a co-DNA-binding partner complex that regulates initiation of lens development. Genes Dev 15 (10): 1272-86.

Kang, S. H., Vieira, K. and Bungert, J. (2002). Combining chromatin immunoprecipitation and DNA footprinting: a novel method to analyze protein-DNA interactions in vivo. Nucleic Acids Res 30 (10): e44.

Kang, S., Graham, J. M., Jr., Olney, A. H. and Biesecker, L. G. (1997a). GLI3 frameshift mutations cause autosomal dominant Pallister-Hall syndrome. Nat Genet 15 (3): 266-8.

Kang, S., Rosenberg, M., Ko, V. D. and Biesecker, L. G. (1997b). Gene structure and allelic expression assay of the human GLI3 gene. Hum Genet 101 (2): 154-7.

Kapranov, P., Drenkow, J., Cheng, J., Long, J., Helt, G., Dike, S. and Gingeras, T. R. (2005). Examples of the complex architecture of the human transcriptome revealed by RACE and high-density tiling arrays. Genome Res 15 (7): 987-97.

Karlstrom, R. O., Tyurina, O. V., Kawakami, A., Nishioka, N., Talbot, W. S., Sasaki, H. and Schier, A. F. (2003). Genetic analysis of zebrafish gli1 and gli2 reveals divergent requirements for gli genes in vertebrate development. Development 130 (8): 1549-64.

Karolchik, D., Baertsch, R., Diekhans, M., Furey, T. S., Hinrichs, A., Lu, Y. T., Roskin, K. M., Schwartz, M., Sugnet, C. W., Thomas, D. J., Weber, R. J., Haussler, D. and Kent, W. J. (2003). The UCSC Genome Browser Database. Nucleic Acids Res 31 (1): 51-4.

Karreth, F., Hoebertz, A., Scheuch, H., Eferl, R. and Wagner, E. F. (2004). The AP1 transcription factor Fra2 is required for efficient cartilage development. Development 131 (22): 5717-25.

Katagiri, T., Boorla, S., Frendo, J. L., Hogan, B. L. and Karsenty, G. (1998). Skeletal abnormalities in doubly heterozygous Bmp4 and Bmp7 mice. Dev Genet 22 (4): 340-8.

Katoh, Y. and Katoh, M. (2005). Hedgehog signaling pathway and gastric cancer. Cancer Biol Ther 4 (10): 1050-4.

Katsuoka, F., Motohashi, H., Tamagawa, Y., Kure, S., Igarashi, K., Engel, J. D. and Yamamoto, M. (2003). Small Maf compound mutants display central nervous system neuronal degeneration, aberrant transcription, and Bach protein mislocalization coincident with myoclonus and abnormal startle response. Mol Cell Biol 23 (4): 1163-74.

Kawabata, M., Inoue, H., Hanyu, A., Imamura, T. and Miyazono, K. (1998). Smad proteins exist as monomers in vivo and undergo homo- and hetero-oligomerization upon activation by serine/threonine kinase receptors. Embo J 17 (14): 4056-65.

Kawaji, H., Kasukawa, T., Fukuda, S., Katayama, S., Kai, C., Kawai, J., Carninci, P. and Hayashizaki, Y. (2006). CAGE Basic/Analysis Databases: the CAGE resource for comprehensive promoter analysis. Nucleic Acids Res 34 (Database issue): D632-6.

Kawakami, T., Kawcak, T., Li, Y. J., Zhang, W., Hu, Y. and Chuang, P. T. (2002). Mouse dispatched mutants fail to distribute hedgehog proteins and are defective in hedgehog signaling. Development 129 (24): 5753-65.

Kel, A. E., Gossling, E., Reuter, I., Cheremushkin, E., Kel-Margoulis, O. V. and Wingender, E. (2003). MATCH: A tool for searching transcription factor binding sites in DNA sequences. Nucleic Acids Res 31 (13): 3576-9.

Kenney, A. M. and Rowitch, D. H. (2000). Sonic hedgehog promotes G(1) cyclin expression and sustained cell cycle progression in mammalian neuronal precursors. Mol Cell Biol 20 (23): 9055-67.

Kenney, A. M., Cole, M. D. and Rowitch, D. H. (2003). Nmyc upregulation by sonic hedgehog signaling promotes proliferation in developing cerebellar granule neuron precursors. Development 130 (1): 15-28.

Kewley, R. J., Whitelaw, M. L. and Chapman-Smith, A. (2004). The mammalian basic helix-loop-helix/PAS family of transcriptional regulators. Int J Biochem Cell Biol 36 (2): 189-204.

Kilstrup-Nielsen, C., Alessio, M. and Zappavigna, V. (2003). PBX1 nuclear export is regulated independently of PBX-MEINOX interaction by PKA phosphorylation of the PBC-B domain. Embo J 22 (1): 89-99.

Kim, C. H., Oda, T., Itoh, M., Jiang, D., Artinger, K. B., Chandrasekharappa, S. C., Driever, W. and Chitnis, A. B. (2000). Repressor activity of Headless/Tcf3 is essential for vertebrate head formation. Nature 407 (6806): 913-6.

Kim, T. H., Barrera, L. O., Zheng, M., Qu, C., Singer, M. A., Richmond, T. A., Wu, Y., Green, R. D. and Ren, B. (2005). A high-resolution map of active promoters in the human genome. Nature 436 (7052): 876-80.

Kinzler, K. W., Bigner, S. H., Bigner, D. D., Trent, J. M., Law, M. L., O'brien, S. J., Wong, A. J. and Vogelstein, B. (1987). Identification of an amplified, highly expressed gene in a human glioma. Science 236 (4797): 70-3.

Kinzler, K. W., Ruppert, J. M., Bigner, S. H. and Vogelstein, B. (1988). The GLI gene is a member of the Kruppel family of zinc finger proteins. Nature 332 (6162): 371-4.

Klein, P. S. and Melton, D. A. (1996). A molecular mechanism for the effect of lithium on development. Proc Natl Acad Sci U S A 93 (16): 8455-9.

Kleinjan, D. A., Seawright, A., Childs, A. J. and Van Heyningen, V. (2004). Conserved elements in Pax6 intron 7 involved in (auto)regulation and alternative transcription. Dev Biol 265 (2): 462-77.

Knepper, J. L., James, A. C. and Ming, J. E. (2006). TGIF, a gene associated with human brain defects, regulates neuronal development. Dev Dyn 235 (6): 1482-90.

Knoepfler, P. S. and Kamps, M. P. (1995). The pentapeptide motif of Hox proteins is required for cooperative DNA binding with Pbx1, physically contacts Pbx1, and enhances DNA binding by Pbx1. Mol Cell Biol 15 (10): 5811-9.

Knoepfler, P. S., Bergstrom, D. A., Uetsuki, T., Dac-Korytko, I., Sun, Y. H., Wright, W. E., Tapscott, S. J. and Kamps, M. P. (1999). A conserved motif N-terminal to the DNA-binding domains of myogenic bHLH transcription factors mediates cooperative DNA binding with pbx-Meis1/Prep1. Nucleic Acids Res 27 (18): 3752-61.

Knudsen, S. (1999). Promoter2.0: for the recognition of PolII promoter sequences. Bioinformatics 15 (5): 356-61.

Kretzschmar, M., Liu, F., Hata, A., Doody, J. and Massague, J. (1997). The TGF-beta family mediator Smad1 is phosphorylated directly and activated functionally by the BMP receptor kinase. Genes Dev 11 (8): 984-95.

Kriks, S., Lanuza, G. M., Mizuguchi, R., Nakafuku, M. and Goulding, M. (2005). Gsh2 is required for the repression of Ngn1 and specification of dorsal interneuron fate in the spinal cord. Development 132 (13): 2991-3002.

Kruger, M., Mennerich, D., Fees, S., Schafer, R., Mundlos, S. and Braun, T. (2001). Sonic hedgehog is a survival factor for hypaxial muscles during mouse development. Development 128 (5): 743-52.

Kuo, M. H. and Allis, C. D. (1999). In vivo cross-linking and immunoprecipitation for studying dynamic Protein:DNA associations in a chromatin environment. Methods 19 (3): 425-33.

Kuschel, S., Ruther, U. and Theil, T. (2003). A disrupted balance between Bmp/Wnt and Fgf signaling underlies the ventralization of the Gli3 mutant telencephalon. Dev Biol 260 (2): 484-95.

Kutach, A. K. and Kadonaga, J. T. (2000). The downstream promoter element DPE appears to be as widely used as the TATA box in Drosophila core promoters. Mol Cell Biol 20 (13): 4754-64.

Labbe, E., Letamendia, A. and Attisano, L. (2000). Association of Smads with lymphoid enhancer binding factor 1/T cell-specific factor mediates cooperative signaling by the transforming growth factor-beta and wnt pathways. Proc Natl Acad Sci U S A 97 (15): 8358-63.

Labbe, E., Silvestri, C., Hoodless, P. A., Wrana, J. L. and Attisano, L. (1998). Smad2 and Smad3 positively and negatively regulate TGF beta-dependent transcription through the forkhead DNA-binding protein FAST2. Mol Cell 2 (1): 109-20.

Ladomery, M. and Dellaire, G. (2002). Multifunctional zinc finger proteins in development and disease. Ann Hum Genet 66 (Pt 5-6): 331-42.

Laemmli, U. K., Kas, E., Poljak, L. and Adachi, Y. (1992). Scaffold-associated regions: cis-acting determinants of chromatin structural loops and functional domains. Curr Opin Genet Dev 2 (2): 275-85.

Lagna, G., Hata, A., Hemmati-Brivanlou, A. and Massague, J. (1996). Partnership between DPC4 and SMAD proteins in TGF-beta signalling pathways. Nature 383 (6603): 832-6.

Lagrange, T., Kapanidis, A. N., Tang, H., Reinberg, D. and Ebright, R. H. (1998). New core promoter element in RNA polymerase II-dependent transcription: sequence-specific DNA binding by transcription factor IIB. Genes Dev 12 (1): 34-44.

Lan, L., Liu, M., Liu, Y., Zhang, W., Xue, J., Xue, Z. and He, R. (2006). Expression of qBrn-1, a new member of the POU gene family, in the early developing nervous system and embryonic kidney. Dev Dyn 235 (4): 1107-14.

Lang, D., Powell, S. K., Plummer, R. S., Young, K. P. and Ruggeri, B. A. (2007a). PAX genes: Roles in development, pathophysiology, and cancer. Biochemical Pharmacology 73 (1): 1-14.

Latres, E., Chiaur, D. S. and Pagano, M. (1999). The human F box protein beta-Trcp associates with the Cul1/Skp1 complex and regulates the stability of beta-catenin. Oncogene 18 (4): 849-54.

Laurent, A., Bihan, R., Deschamps, S., Guerrier, D., Dupe, V., Omilli, F., Burel, A. and Pellerin, I. (2007). Identification of a new type of PBX1 partner that contains zinc finger motifs and inhibits the binding of HOXA9-PBX1 to DNA. Mech Dev 124 (5): 364-76.

Laurent, A., Bihan, R., Omilli, F., Deschamps, S. and Pellerin, I. (2008). PBX proteins: much more than Hox cofactors. Int J Dev Biol 52 (1): 9-20.

Leber, S. M. and Sanes, J. R. (1995). Migratory paths of neurons and glia in the embryonic chick spinal cord. J Neurosci 15 (2): 1236-48.

Lebrun, J. J., Takabe, K., Chen, Y. and Vale, W. (1999). Roles of pathway-specific and inhibitory Smads in activin receptor signaling. Mol Endocrinol 13 (1): 15-23.

Lecoin, L., Sii-Felice, K., Pouponnot, C., Eychene, A. and Felder-Schmittbuhl, M. P. (2004). Comparison of maf gene expression patterns during chick embryo development. Gene Expr Patterns 4 (1): 35-46.

Lee, C. S., Buttitta, L. and Fan, C. M. (2001). Evidence that the WNT-inducible growth arrest-specific gene 1 encodes an antagonist of sonic hedgehog signaling in the somite. Proc Natl Acad Sci U S A 98 (20): 11347-52.

Lee, J. J., Ekker, S. C., Von Kessler, D. P., Porter, J. A., Sun, B. I. and Beachy, P. A. (1994). Autoproteolysis in hedgehog protein biogenesis. Science 266 (5190): 1528-37.

Lee, J., Platt, K. A., Censullo, P. and Ruiz I Altaba, A. (1997). Gli1 is a target of Sonic hedgehog that induces ventral neural tube development. Development 124 (13): 2537-52.

Lee, K. J., Dietrich, P. and Jessell, T. M. (2000). Genetic ablation reveals that the roof plate is essential for dorsal interneuron specification. Nature 403 (6771): 734-40.

Lee, K. J., Mendelsohn, M. and Jessell, T. M. (1998). Neuronal patterning by BMPs: a requirement for GDF7 in the generation of a discrete class of commissural interneurons in the mouse spinal cord. Genes Dev 12 (21): 3394-407.

Lee, M. P., Howcroft, K., Kotekar, A., Yang, H. H., Buetow, K. H. and Singer, D. S. (2005). ATG deserts define a novel core promoter subclass. Genome Res 15 (9): 1189-97.

Lei, Q., Jeong, Y., Misra, K., Li, S., Zelman, A. K., Epstein, D. J. and Matise, M. P. (2006). Wnt signaling inhibitors regulate the transcriptional response to morphogenetic Shh-Gli signaling in the neural tube. Dev Cell 11 (3): 325-37.

Lei, Q., Zelman, A. K., Kuang, E., Li, S. and Matise, M. P. (2004). Transduction of graded Hedgehog signaling by a combination of Gli2 and Gli3 activator functions in the developing spinal cord. Development 131 (15): 3593-604.

Lemos, B., Yunes, J. A., Vargas, F. R., Moreira, M. A., Cardoso, A. A. and Seuanez, H. N. (2004). Phylogenetic footprinting reveals extensive conservation of Sonic Hedgehog (SHH) regulatory elements. Genomics 84 (3): 511-23.

Letamendia, A., Labbe, E. and Attisano, L. (2001). Transcriptional regulation by Smads: crosstalk between the TGF-beta and Wnt pathways. J Bone Joint Surg Am 83-A Suppl 1 (Pt 1): S31-9.

Lettice, L. A., Heaney, S. J., Purdie, L. A., Li, L., De Beer, P., Oostra, B. A., Goode, D., Elgar, G., Hill, R. E. and De Graaff, E. (2003). A long-range Shh enhancer regulates expression in the developing limb and fin and is associated with preaxial polydactyly. Hum Mol Genet 12 (14): 1725-35.

Lettice, L. A., Horikoshi, T., Heaney, S. J. H., Van Baren, M. J., Van Der Linde, H. C., Breedveld, G. J., Joosse, M., Akarsu, N., Oostra, B. A., Endo, N., Shibata, M., Suzuki, M., Takahashi, E., Shinka, T., Nakahori, Y., Ayusawa, D., Nakabayashi, K., Scherer, S. W., Heutink, P., Hill, R. E. and Noji, S. (2002). Disruption of a long-range cis-acting regulator for Shh causes preaxial polydactyly. Proceedings of the National Academy of Sciences 99 (11): 7548-7553.

Levanon, D., Goldstein, R. E., Bernstein, Y., Tang, H., Goldenberg, D., Stifani, S., Paroush, Z. and Groner, Y. (1998). Transcriptional repression by AML1 and LEF-1 is mediated by the TLE/Groucho corepressors. Proc Natl Acad Sci U S A 95 (20): 11590-5.

Levin, M., Johnson, R. L., Stern, C. D., Kuehn, M. and Tabin, C. (1995). A molecular pathway determining left-right asymmetry in chick embryogenesis. Cell 82 (5): 803-14.

Levy-Wilson, B. and Fortier, C. (1989). Tissue-specific undermethylation of DNA sequences at the 5' end of the human apolipoprotein B gene. J Biol Chem 264 (17): 9891-6.

Lewis, P. M., Dunn, M. P., Mcmahon, J. A., Logan, M., Martin, J. F., St-Jacques, B. and Mcmahon, A. P. (2001). Cholesterol modification of sonic hedgehog is required for long-range signaling activity and effective modulation of signaling by Ptc1. Cell 105 (5): 599-612.

Li, W. H., Ellsworth, D. L., Krushkal, J., Chang, B. H. and Hewett-Emmett, D. (1996). Rates of nucleotide substitution in primates and rodents and the generation-time effect hypothesis. Mol Phylogenet Evol 5 (1): 182-7.

Liberati, N. T., Datto, M. B., Frederick, J. P., Shen, X., Wong, C., Rougier-Chapman, E. M. and Wang, X. F. (1999). Smads bind directly to the Jun family of AP-1 transcription factors. Proc Natl Acad Sci U S A 96 (9): 4844-9.

Liem, K. F., Jr., Jessell, T. M. and Briscoe, J. (2000). Regulation of the neural patterning activity of sonic hedgehog by secreted BMP inhibitors expressed by notochord and somites. Development 127 (22): 4855-66.

Liem, K. F., Jr., Tremml, G. and Jessell, T. M. (1997). A role for the roof plate and its resident TGFbeta-related proteins in neuronal patterning in the dorsal spinal cord. Cell 91 (1): 127-38.

Liem, K. F., Jr., Tremml, G., Roelink, H. and Jessell, T. M. (1995). Dorsal differentiation of neural plate cells induced by BMP-mediated signals from epidermal ectoderm. Cell 82 (6): 969-79.

Lim, L. P., Glasner, M. E., Yekta, S., Burge, C. B. and Bartel, D. P. (2003). Vertebrate microRNA genes. Science 299 (5612): 1540.

Linker, C. and Stern, C. D. (2004). Neural induction requires BMP inhibition only as a late step, and involves signals other than FGF and Wnt antagonists. Development 131 (22): 5671-81.

Litingtung, Y. and Chiang, C. (2000). Specification of ventral neuron types is mediated by an antagonistic interaction between Shh and Gli3. Nat Neurosci 3 (10): 979-85.

Litingtung, Y., Lei, L., Westphal, H. and Chiang, C. (1998). Sonic hedgehog is essential to foregut development. Nat Genet 20 (1): 58-61.

Liu, B., Dou, C. L., Prabhu, L. and Lai, E. (1999a). FAST-2 is a mammalian winged-helix protein which mediates transforming growth factor beta signals. Mol Cell Biol 19 (1): 424-30.

Liu, C., Kato, Y., Zhang, Z., Do, V. M., Yankner, B. A. and He, X. (1999b). beta-Trcp couples beta-catenin phosphorylation-degradation and regulates Xenopus axis formation. Proc Natl Acad Sci U S A 96 (11): 6273-8.

Liu, F., Massague, J. and Altaba., A. R. I. (1998). Carboxy-terminally truncated Gli3 proteins associate with Smads. 20 (4): 325-326.

Liu, F., Pouponnot, C. and Massague, J. (1997). Dual role of the Smad4/DPC4 tumor suppressor in TGFbeta-inducible transcriptional complexes. Genes Dev 11 (23): 3157-67.

Liu, F., Ventura, F., Doody, J. and Massague, J. (1995). Human type II receptor for bone morphogenic proteins (BMPs): extension of the two-kinase receptor model to the BMPs. Mol Cell Biol 15 (7): 3479-86.

Liu, Y., Macdonald, R. J. and Swift, G. H. (2001). DNA binding and transcriptional activation by a PDX1.PBX1b.MEIS2b trimer and cooperation with a pancreas-specific basic helix-loop-helix complex. J Biol Chem 276 (21): 17985-93.

Liu, Y., Titus, L., Barghouthi, M., Viggeswarapu, M., Hair, G. and Boden, S. D. (2004). Glucocorticoid regulation of human BMP-6 transcription. Bone 35 (3): 673-81.

Lo, K. and Smale, S. T. (1996). Generality of a functional initiator consensus sequence. Gene 182 (1-2): 13-22.

Lobo, S. M., Lister, J., Sullivan, M. L. and Hernandez, N. (1991). The cloned RNA polymerase II transcription factor IID selects RNA polymerase III to transcribe the human U6 gene in vitro. Genes Dev 5 (8): 1477-89.

Logan, C. Y. and Nusse, R. (2004). The Wnt signaling pathway in development and disease. Annu Rev Cell Dev Biol 20: 781-810.

Longobardi, E. and Blasi, F. (2003). Overexpression of PREP-1 in F9 teratocarcinoma cells leads to a functionally relevant increase of PBX-2 by preventing its degradation. J Biol Chem 278 (40): 39235-41.

Loots, G., Ovcharenko, I., Pachter, L., Dubchak, I. and Rubin, E. (2002). rVISTA for comparative sequence-based discovery of functional transcription factor binding sites. Genome. Res. 12: 832-839.

Lu, Q. and Kamps, M. P. (1996). Structural determinants within Pbx1 that mediate cooperative DNA binding with pentapeptide-containing Hox proteins: proposal for a model of a Pbx1-Hox-DNA complex. Mol Cell Biol 16 (4): 1632-40.

Lucibello, F. C., Liu, N., Zwicker, J., Gross, C. and Muller, R. (1997). The differential binding of E2F and CDF repressor complexes contributes to the timing of cell cycle-regulated transcription. Nucleic Acids Res 25 (24): 4921-5.

Luger, K., Mader, A. W., Richmond, R. K., Sargent, D. F. and Richmond, T. J. (1997). Crystal structure of the nucleosome core particle at 2.8 A resolution. Nature 389 (6648): 251-60.

Lum, L. and Beachy, P. A. (2004). The Hedgehog response network: sensors, switches, and routers. Science 304 (5678): 1755-9.

Luo, G., Hofmann, C., Bronckers, A. L., Sohocki, M., Bradley, A. and Karsenty, G. (1995). BMP-7 is an inducer of nephrogenesis, and is also required for eye development and skeletal patterning. Genes Dev 9 (22): 2808-20.

Macias-Silva, M., Hoodless, P. A., Tang, S. J., Buchwald, M. and Wrana, J. L. (1998). Specific activation of Smad1 signaling pathways by the BMP7 type I receptor, ALK2. J Biol Chem 273 (40): 25628-36.

Maconochie, M. K., Nonchev, S., Studer, M., Chan, S. K., Popperl, H., Sham, M. H., Mann, R. S. and Krumlauf, R. (1997). Cross-regulation in the mouse HoxB complex: the expression of Hoxb2 in rhombomere 4 is regulated by Hoxb1. Genes Dev 11 (14): 1885-95.

Maeda, N., Kasukawa, T., Oyama, R., Gough, J., Frith, M., Engstrom, P. G., Lenhard, B., Aturaliya, R. N., Batalov, S., Beisel, K. W., Bult, C. J., Fletcher, C. F., Forrest, A. R., Furuno, M., Hill, D., Itoh, M., Kanamori-Katayama, M., Katayama, S., Katoh, M., Kawashima, T., Quackenbush, J., Ravasi, T., Ring, B. Z., Shibata, K., Sugiura, K., Takenaka, Y., Teasdale, R. D., Wells, C. A., Zhu, Y., Kai, C., Kawai, J., Hume, D. A., Carninci, P. and Hayashizaki, Y. (2006). Transcript annotation in FANTOM3: mouse gene catalog based on physical cDNAs. PLoS Genet 2 (4): e62.

Majumder, S., Zhao, Z., Kaneko, K. and Depamphilis, M. L. (1997). Developmental acquisition of enhancer function requires a unique coactivator activity. Embo J 16 (7): 1721-31.

Malecova, B., Gross, P., Boyer-Guittaut, M., Yavuz, S. and Oelgeschlager, T. (2007). The initiator core promoter element antagonizes repression of TATA-directed transcription by negative cofactor NC2. J Biol Chem 282 (34): 24767-76.

Mamane, Y., Heylbroeck, C., Genin, P., Algarte, M., Servant, M. J., Lepage, C., Deluca, C., Kwon, H., Lin, R. and Hiscott, J. (1999). Interferon regulatory factors: the next generation. Gene 237 (1): 1-14.

Mao, J., Wang, J., Liu, B., Pan, W., Farr, G. H., 3rd, Flynn, C., Yuan, H., Takada, S., Kimelman, D., Li, L. and Wu, D. (2001). Low-density lipoprotein receptor-related protein-5 binds to Axin and regulates the canonical Wnt signaling pathway. Mol Cell 7 (4): 801-9.

Margulies, E. H., Blanchette, M., Haussler, D. and Green, E. D. (2003). Identification and characterization of multi-species conserved sequences. Genome Res 13 (12): 2507-18.

Marigo, V., Johnson, R. L., Vortkamp, A. and Tabin, C. J. (1996). Sonic hedgehog differentially regulates expression of GLI and GLI3 during limb development. Dev Biol 180 (1): 273-83.

Marin, M., Karis, A., Visser, P., Grosveld, F. and Philipsen, S. (1997). Transcription factor Sp1 is essential for early embryonic development but dispensable for cell growth and differentiation. Cell 89 (4): 619-28.

Marine, J. C., Bellefroid, E. J., Pendeville, H., Martial, J. A. and Pieler, T. (1997). A role for Xenopus Gli-type zinc finger proteins in the early embryonic patterning of mesoderm and neuroectoderm. Mech Dev 63 (2): 211-25.

Martens, J. A. and Winston, F. (2003). Recent advances in understanding chromatin remodeling by Swi/Snf complexes. Curr Opin Genet Dev 13 (2): 136-42.

Marti, E., Bumcrot, D. A., Takada, R. and Mcmahon, A. P. (1995). Requirement of 19K form of Sonic hedgehog for induction of distinct ventral cell types in CNS explants. Nature 375 (6529): 322-5.

Martin, D. I., Fiering, S. and Groudine, M. (1996). Regulation of beta-globin gene expression: straightening out the locus. Curr Opin Genet Dev 6 (4): 488-95.

Massague, J. (1998). TGF-beta signal transduction. Annu Rev Biochem 67: 753-91.

Masuya, H., Sagai, T., Moriwaki, K. and Shiroishi, T. (1997). Multigenic control of the localization of the zone of polarizing activity in limb morphogenesis in the mouse. Dev Biol 182 (1): 42-51.

Matise, M. P., Epstein, D. J., Park, H. L., Platt, K. A. and Joyner, A. L. (1998). Gli2 is required for induction of floor plate and adjacent cells, but not most ventral neurons in the mouse central nervous system. Development 125 (15): 2759-70.

Maves, L., Waskiewicz, A. J., Paul, B., Cao, Y., Tyler, A., Moens, C. B. and Tapscott, S. J. (2007). Pbx homeodomain proteins direct Myod activity to promote fast-muscle differentiation. Development 134 (18): 3371-82.

Maye, P., Zheng, J., Li, L. and Wu, D. (2004). Multiple mechanisms for Wnt11-mediated repression of the canonical Wnt signaling pathway. J Biol Chem 279 (23): 24659-65.

Mayor, C., Brudno, M., Schwartz, J. R., Poliakov, A., Rubin, E. M., Frazer, K. A., Pachter, L. S. and Dubchak, I. (2000). VISTA : visualizing global DNA sequence alignments of arbitrary length. Bioinformatics 16 (11): 1046-1047.

Mcdermott, A., Gustafsson, M., Elsam, T., Hui, C. C., Emerson, C. P., Jr. and Borycki, A. G. (2005). Gli2 and Gli3 have redundant and context-dependent function in skeletal muscle formation. Development 132 (2): 345-57.

Mcmahon, A. P. and Bradley, A. (1990). The Wnt-1 (int-1) proto-oncogene is required for development of a large region of the mouse brain. Cell 62 (6): 1073-85.

Megason, S. G. and Mcmahon, A. P. (2002). A mitogen gradient of dorsal midline Wnts organizes growth in the CNS. Development 129 (9): 2087-98.

Melhuish, T. A., Gallo, C. M. and Wotton, D. (2001). TGIF2 interacts with histone deacetylase 1 and represses transcription. J Biol Chem 276 (34): 32109-14.

Mercader, N., Leonardo, E., Azpiazu, N., Serrano, A., Morata, G., Martinez-a, C. and Torres, M. (1999). Conserved regulation of proximodistal limb axis development by Meis1/Hth. 402 (6760): 425-429.

Mercader, N., Leonardo, E., Piedra, M. E., Martinez, A. C., Ros, M. A. and Torres, M. (2000). Opposing RA and FGF signals control proximodistal vertebrate limb development through regulation of Meis genes. Development 127 (18): 3961-70.

Merrill, B. J., Pasolli, H. A., Polak, L., Rendl, M., Garcia-Garcia, M. J., Anderson, K. V. and Fuchs, E. (2004). Tcf3: a transcriptional regulator of axis induction in the early embryo. Development 131 (2): 263-74.

Methot, N. and Basler, K. (2000). Suppressor of fused opposes hedgehog signal transduction by impeding nuclear accumulation of the activator form of Cubitus interruptus. Development 127 (18): 4001-10.

Meyer, N. P. and Roelink, H. (2003). The amino-terminal region of Gli3 antagonizes the Shh response and acts in dorsoventral fate specification in the developing spinal cord. Dev Biol 257 (2): 343-55.

Milech, N., Kees, U. R. and Watt, P. M. (2001). Novel alternative PBX3 isoforms in leukemia cells with distinct interaction specificities. Genes Chromosomes Cancer 32 (3): 275-80.

Millonig, J. H., Millen, K. J. and Hatten, M. E. (2000). The mouse Dreher gene Lmx1a controls formation of the roof plate in the vertebrate CNS. Nature 403 (6771): 764-9.

Miyazaki, Y., Oshima, K., Fogo, A., Hogan, B. L. and Ichikawa, I. (2000). Bone morphogenetic protein 4 regulates the budding site and elongation of the mouse ureter. J Clin Invest 105 (7): 863-73.

Miyazawa, K., Shinozaki, M., Hara, T., Furuya, T. and Miyazono, K. (2002). Two major Smad pathways in TGF-beta superfamily signalling. Genes Cells 7 (12): 1191-204.

Miyazono, K. (1999). Signal transduction by bone morphogenetic protein receptors: functional roles of Smad proteins. Bone 25 (1): 91-93.

Miyazono, K. and Miyazawa, K. (2002). Id: a target of BMP signaling. Sci STKE 2002 (151): PE40.

Mo, R., Freer, A. M., Zinyk, D. L., Crackower, M. A., Michaud, J., Heng, H. H., Chik, K. W., Shi, X. M., Tsui, L. C., Cheng, S. H., Joyner, A. L. and Hui, C. (1997). Specific and redundant functions of Gli2 and Gli3 zinc finger genes in skeletal patterning and development. Development 124 (1): 113-23.

Molenaar, M., Van De Wetering, M., Oosterwegel, M., Peterson-Maduro, J., Godsave, S., Korinek, V., Roose, J., Destree, O. and Clevers, H. (1996). XTcf-3 transcription factor mediates beta-catenin-induced axis formation in Xenopus embryos. Cell 86 (3): 391-9.

Monica, K., Galili, N., Nourse, J., Saltman, D. and Cleary, M. L. (1991). PBX2 and PBX3, new homeobox genes with extensive homology to the human proto-oncogene PBX1. Mol Cell Biol 11 (12): 6149-57.

Monnier, V., Ho, K. S., Sanial, M., Scott, M. P. and Plessis, A. (2002). Hedgehog signal transduction proteins: contacts of the Fused kinase and Ci transcription factor with the kinesin-related protein Costal2. BMC Dev Biol 2: 4.

Moskow, J. J., Bullrich, F., Huebner, K., Daar, I. O. and Buchberg, A. M. (1995). Meis1, a PBX1-related homeobox gene involved in myeloid leukemia in BXH-2 mice. Mol Cell Biol 15 (10): 5434-43.

Motoyama, J., Liu, J., Mo, R., Ding, Q., Post, M. and Hui, C. C. (1998). Essential function of Gli2 and Gli3 in the formation of lung, trachea and oesophagus. Nat Genet 20 (1): 54-7.

Motoyama, J., Milenkovic, L., Iwama, M., Shikata, Y., Scott, M. P. and Hui, C. C. (2003). Differential requirement for Gli2 and Gli3 in ventral neural cell fate specification. Dev Biol 259 (1): 150-61.

Mouse Genome Sequencing Consortium (2002). Initial sequencing and comparative analysis of the mouse genome. 420 (6915): 520-562.

Mueller, P. R. and Wold, B. (1989). In vivo footprinting of a muscle specific enhancer by ligation mediated PCR. Science 246 (4931): 780-6.

Muhr, J., Andersson, E., Persson, M., Jessell, T. M. and Ericson, J. (2001). Groucho-mediated transcriptional repression establishes progenitor cell pattern and neuronal fate in the ventral neural tube. Cell 104 (6): 861-73.

Mullen, R. J., Buck, C. R. and Smith, A. M. (1992). NeuN, a neuronal specific nuclear protein in vertebrates. Development 116 (1): 201-11.

Muller, F., Albert, S., Blader, P., Fischer, N., Hallonet, M. and Strahle, U. (2000). Direct action of the nodal-related signal cyclops in induction of sonic hedgehog in the ventral midline of the CNS. Development 127 (18): 3889-97.

Muller, F., Williams, D. W., Kobolak, J., Gauvry, L., Goldspink, G., Orban, L. and Maclean, N. (1997). Activator effect of coinjected enhancers on the muscle-specific expression of promoters in zebrafish embryos. Mol Reprod Dev 47 (4): 404-12.

Muller, T., Anlag, K., Wildner, H., Britsch, S., Treier, M. and Birchmeier, C. (2005). The bHLH factor Olig3 coordinates the specification of dorsal neurons in the spinal cord. Genes Dev 19 (6): 733-43.

Muller, T., Brohmann, H., Pierani, A., Heppenstall, P. A., Lewin, G. R., Jessell, T. M. and Birchmeier, C. (2002). The homeodomain factor lbx1 distinguishes two major programs of neuronal differentiation in the dorsal spinal cord. Neuron 34 (4): 551-62.

Mullor, J. L., Dahmane, N., Sun, T. and Ruiz I Altaba, A. (2001). Wnt signals are targets and mediators of Gli function. Curr Biol 11 (10): 769-73.

Muramatsu, T., Mizutani, Y., Ohmori, Y. and Okumura, J. (1997). Comparison of three nonviral transfection methods for foreign gene expression in early chicken embryos in ovo. Biochem Biophys Res Commun 230 (2): 376-80.

Murone, M., Rosenthal, A. and De Sauvage, F. J. (1999). Sonic hedgehog signaling by the patched-smoothened receptor complex. Curr Biol 9 (2): 76-84.

Muroyama, Y., Fujihara, M., Ikeya, M., Kondoh, H. and Takada, S. (2002). Wnt signaling plays an essential role in neuronal specification of the dorsal spinal cord. Genes Dev 16 (5): 548-53.

Muta, M., Kamachi, Y., Yoshimoto, A., Higashi, Y. and Kondoh, H. (2002). Distinct roles of SOX2, Pax6 and Maf transcription factors in the regulation of lens-specific delta1-crystallin enhancer. Genes Cells 7 (8): 791-805.

Nakamura, T., Jenkins, N. A. and Copeland, N. G. (1996). Identification of a new family of Pbx-related homeobox genes. Oncogene 13 (10): 2235-42.

Nakano, Y., Nystedt, S., Shivdasani, A. A., Strutt, H., Thomas, C. and Ingham, P. W. (2004). Functional domains and sub-cellular distribution of the Hedgehog transducing protein Smoothened in Drosophila. Mech Dev 121 (6): 507-18.

Nakao, A., Afrakhte, M., Moren, A., Nakayama, T., Christian, J. L., Heuchel, R., Itoh, S., Kawabata, M., Heldin, N. E., Heldin, C. H. and Ten Dijke, P. (1997a). Identification of Smad7, a TGFbeta-inducible antagonist of TGF-beta signalling. Nature 389 (6651): 631-5.

Nakao, A., Imamura, T., Souchelnytskyi, S., Kawabata, M., Ishisaki, A., Oeda, E., Tamaki, K., Hanai, J., Heldin, C. H., Miyazono, K. and Ten Dijke, P. (1997b). TGF-beta receptor-mediated signalling through Smad2, Smad3 and Smad4. Embo J 16 (17): 5353-62.

Nakashima, K., Takizawa, T., Ochiai, W., Yanagisawa, M., Hisatsune, T., Nakafuku, M., Miyazono, K., Kishimoto, T., Kageyama, R. and Taga, T. (2001). BMP2-mediated alteration in the developmental pathway of fetal mouse brain cells from neurogenesis to astrocytogenesis. Proceedings of the National Academy of Sciences 98 (10): 5868-5873.

Nakashima, K., Yanagisawa, M., Arakawa, H., Kimura, N., Hisatsune, T., Kawabata, M., Miyazono, K. and Taga, T. (1999). Synergistic Signaling in Fetal Brain by STAT3-Smad1 Complex Bridged by p300. Science 284 (5413): 479-482.

Nardone, J., Lee, D. U., Ansel, K. M. and Rao, A. (2004). Bioinformatics for the 'bench biologist': how to find regulatory regions in genomic DNA. Nat Immunol 5 (8): 768-74.

Neuteboom, S. T. and Murre, C. (1997). Pbx raises the DNA binding specificity but not the selectivity of antennapedia Hox proteins. Mol Cell Biol 17 (8): 4696-706.

Nguyen, V. H., Trout, J., Connors, S. A., Andermann, P., Weinberg, E. and Mullins, M. C. (2000). Dorsal and intermediate neuronal cell types of the spinal cord are established by a BMP signaling pathway. Development 127 (6): 1209-20.

Nielsen, J. A., Berndt, J. A., Hudson, L. D. and Armstrong, R. C. (2004). Myelin transcription factor 1 (Myt1) modulates the proliferation and differentiation of oligodendrocyte lineage cells. Mol Cell Neurosci 25 (1): 111-23.

Nohe, A., Keating, E., Knaus, P. and Petersen, N. O. (2004). Signal transduction of bone morphogenetic protein receptors. Cell Signal 16 (3): 291-9.

Nornes, H. O. and Carry, M. (1978). Neurogenesis in spinal cord of mouse: an autoradiographic analysis. Brain Res 159 (1): 1-6.

Novitch, B. G., Chen, A. I. and Jessell, T. M. (2001). Coordinate regulation of motor neuron subtype identity and pan-neuronal properties by the bHLH repressor Olig2. Neuron 31 (5): 773-89.

Odent, S., Atti-Bitach, T., Blayau, M., Mathieu, M., Aug, J., Delezo De, A. L., Gall, J. Y., Le Marec, B., Munnich, A., David, V. and Vekemans, M. (1999). Expression of the Sonic hedgehog (SHH) gene during early human development and phenotypic expression of new mutations causing holoprosencephaly. Hum Mol Genet 8 (9): 1683-9.

Ogden, S. K., Ascano, M., Jr., Stegman, M. A., Suber, L. M., Hooper, J. E. and Robbins, D. J. (2003). Identification of a functional interaction between the transmembrane protein Smoothened and the kinesin-related protein Costal2. Curr Biol 13 (22): 1998-2003.

O'gorman, W., Kwek, K. Y., Thomas, B. and Akoulitchev, A. (2006). Non-coding RNA in transcription initiation. Biochem Soc Symp (73): 131-40.

Ohler, U., Stemmer, G., Harbeck, S. and Niemann, H. (2000). Stochastic segment models of eukaryotic promoter regions. Pac Symp Biocomput: 380-91.

Okada, Y., Nagai, R., Sato, T., Matsuura, E., Minami, T., Morita, I. and Doi, T. (2003). Homeodomain proteins MEIS1 and PBXs regulate the lineage-specific transcription of the platelet factor 4 gene. Blood 101 (12): 4748-56.

Olivera-Martinez, I. and Storey, K. G. (2007). Wnt signals provide a timing mechanism for the FGF-retinoid differentiation switch during vertebrate body axis extension. Development 134 (11): 2125-35.

O'shea-Greenfield, A. and Smale, S. T. (1992). Roles of TATA and initiator elements in determining the start site location and direction of RNA polymerase II transcription. J Biol Chem 267 (9): 6450.

Osumi, N. and Inoue, T. (2001). Gene transfer into cultured mammalian embryos by electroporation. Methods 24 (1): 35-42.

Oulad-Abdelghani, M., Chazaud, C., Bouillet, P., Sapin, V., Chambon, P. and Dolle, P. (1997). Meis2, a novel mouse Pbx-related homeobox gene induced by retinoic acid during differentiation of P19 embryonal carcinoma cells. Dev Dyn 210 (2): 173-83.

Ovcharenko, I., Stubbs, L. and Loots, G. G. (2004). Interpreting mammalian evolution using Fugu genome comparisons. Genomics 84 (5): 890-5.

Pan, Y. and Wang, B. (2007). A novel protein-processing domain in Gli2 and Gli3 differentially blocks complete protein degradation by the proteasome. J Biol Chem 282 (15): 10846-52.

Paparidis, Z. (2005). Cis-acting elements controlling the expression of the human Gli3 gene. PhD Thesis. Fachbereich Biologie der Philipps-Universität Marburg, Germany.

Paparidis, Z., Abbasi, A. A., Malik, S., Goode, D. K., Callaway, H., Elgar, G., Degraaff, E., Lopez-Rios, J., Zeller, R. and Grzeschik, K. H. (2007). Ultraconserved non-coding sequence element controls a subset of spatiotemporal GLI3 expression. Dev Growth Differ 49 (6): 543-53.

Park, H. L., Bai, C., Platt, K. A., Matise, M. P., Beeghly, A., Hui, C. C., Nakashima, M. and Joyner, A. L. (2000). Mouse Gli1 mutants are viable but have defects in SHH signaling in combination with a Gli2 mutation. Development 127 (8): 1593-605.

Parr, B. A. and Mcmahon, A. P. (1994). Wnt genes and vertebrate development. Curr Opin Genet Dev 4 (4): 523-8.

Parr, B. A., Shea, M. J., Vassileva, G. and Mcmahon, A. P. (1993). Mouse Wnt genes exhibit discrete domains of expression in the early embryonic CNS and limb buds. Development 119 (1): 247-61.

Parras, C. M., Schuurmans, C., Scardigli, R., Kim, J., Anderson, D. J. and Guillemot, F. (2002). Divergent functions of the proneural genes Mash1 and Ngn2 in the specification of neuronal subtype identity. Genes Dev 16 (3): 324-38.

Pavletich, N. P. and Pabo, C. O. (1993). Crystal structure of a five-finger GLI-DNA complex: new perspectives on zinc fingers. Science 261 (5129): 1701-7.

Pazin, M. J. and Kadonaga, J. T. (1997). What's up and down with histone deacetylation and transcription? Cell 89 (3): 325-8.

Pearse, R. V., 2nd, Collier, L. S., Scott, M. P. and Tabin, C. J. (1999). Vertebrate homologs of Drosophila suppressor of fused interact with the gli family of transcriptional regulators. Dev Biol 212 (2): 323-36.

Peers, B., Sharma, S., Johnson, T., Kamps, M. and Montminy, M. (1995). The pancreatic islet factor STF-1 binds cooperatively with Pbx to a regulatory element in the somatostatin promoter: importance of the FPWMK motif and of the homeodomain. Mol Cell Biol 15 (12): 7091-7.

Peltenburg, L. T. and Murre, C. (1996). Engrailed and Hox homeodomain proteins contain a related Pbx interaction motif that recognizes a common structure present in Pbx. Embo J 15 (13): 3385-93.

Pennacchio, L. A. (2003). Insights from human/mouse genome comparisons. Mamm Genome 14 (7): 429-36.

Persson, M., Stamataki, D., Te Welscher, P., Andersson, E., Bose, J., Ruther, U., Ericson, J. and Briscoe, J. (2002). Dorsal-ventral patterning of the spinal cord requires Gli3 transcriptional repressor activity. Genes Dev 16 (22): 2865-78.

Petropoulos, H., Gianakopoulos, P. J., Ridgeway, A. G. and Skerjanc, I. S. (2004). Disruption of Meox or Gli activity ablates skeletal myogenesis in P19 cells. J Biol Chem 279 (23): 23874-81.

Phelan, M. L., Rambaldi, I. and Featherstone, M. S. (1995). Cooperative interactions between HOX and PBX proteins mediated by a conserved peptide motif. Mol Cell Biol 15 (8): 3989-97.

Pierani, A., Brenner-Morton, S., Chiang, C. and Jessell, T. M. (1999). A sonic hedgehog-independent, retinoid-activated pathway of neurogenesis in the ventral spinal cord. Cell 97 (7): 903-15.

Piper, D. E., Batchelor, A. H., Chang, C. P., Cleary, M. L. and Wolberger, C. (1999). Structure of a HoxB1-Pbx1 heterodimer bound to DNA: role of the hexapeptide and a fourth homeodomain helix in complex formation. Cell 96 (4): 587-97.

Placzek, M., Tessier-Lavigne, M., Yamada, T., Jessell, T. and Dodd, J. (1990). Mesodermal control of neural cell identity: floor plate induction by the notochord. Science 250 (4983): 985-8.

Plessy, C., Dickmeis, T., Chalmel, F. and Strahle, U. (2005). Enhancer sequence conservation between vertebrates is favoured in developmental regulator genes. Trends in Genetics 21 (4): 207-210.

Pohl, T., Mattei, M. and Ruther, U. (1990). Evidence for allelism of the recessive insertional mutation add and the dominant mouse mutation extra-toes (Xt). Development 110 (4): 1153-1157.

Ponger, L. and Mouchiroud, D. (2002). CpGProD: identifying CpG islands associated with transcription start sites in large genomic mammalian sequences. Bioinformatics 18 (4): 631-3.

Ponjavic, J., Lenhard, B., Kai, C., Kawai, J., Carninci, P., Hayashizaki, Y. and Sandelin, A. (2006). Transcriptional and structural impact of TATA-initiation site spacing in mammalian core promoters. Genome Biol 7 (8): R78.

Popperl, H., Bienz, M., Studer, M., Chan, S. K., Aparicio, S., Brenner, S., Mann, R. S. and Krumlauf, R. (1995). Segmental expression of Hoxb-1 is controlled by a highly conserved autoregulatory loop dependent upon exd/pbx. Cell 81 (7): 1031-42.

Popperl, H., Schmidt, C., Wilson, V., Hume, C. R., Dodd, J., Krumlauf, R. and Beddington, R. S. (1997). Misexpression of Cwnt8C in the mouse induces an ectopic embryonic axis and causes a truncation of the anterior neuroectoderm. Development 124 (15): 2997-3005.

Porter, J. A., Ekker, S. C., Park, W. J., Von Kessler, D. P., Young, K. E., Chen, C. H., Ma, Y., Woods, A. S., Cotter, R. J., Koonin, E. V. and Beachy, P. A. (1996). Hedgehog patterning activity: role of a lipophilic modification mediated by the carboxy-terminal autoprocessing domain. Cell 86 (1): 21-34.

Prestridge, D. S. (1995). Predicting Pol II promoter sequences using transcription factor binding sites. J Mol Biol 249 (5): 923-32.

Price, M. A. and Kalderon, D. (2002). Proteolysis of the Hedgehog signaling effector Cubitus interruptus requires phosphorylation by Glycogen Synthase Kinase 3 and Casein Kinase 1. Cell 108 (6): 823-35.

Price, S. R. and Briscoe, J. (2004). The generation and diversification of spinal motor neurons: signals and responses. Mech Dev 121 (9): 1103-15.

Pruitt, S. C., Bussman, A., Maslov, A. Y., Natoli, T. A. and Heinaman, R. (2004). Hox/Pbx and Brn binding sites mediate Pax3 expression in vitro and in vivo. Gene Expr Patterns 4 (6): 671-85.

Qin, B. Y., Chacko, B. M., Lam, S. S., De Caestecker, M. P., Correia, J. J. and Lin, K. (2001). Structural basis of Smad1 activation by receptor kinase phosphorylation. Mol Cell 8 (6): 1303-12.

Qin, P., Haberbusch, J. M., Zhang, Z., Soprano, K. J. and Soprano, D. R. (2004). Pre-B cell leukemia transcription factor (PBX) proteins are important mediators for retinoic acid-dependent endodermal and neuronal differentiation of mouse embryonal carcinoma P19 cells. J Biol Chem 279 (16): 16263-71.

Quandt, K., Frech, K., Karas, H., Wingender, E. and Werner, T. (1995). MatInd and MatInspector: new fast and versatile tools for detection of consensus matches in nucleotide sequence data. Nucl. Acids Res. 23 (23): 4878-4884.

Radhakrishna, U., Wild, A., Grzeschik, K. H. and Antonarakis, S. E. (1997). Mutation in GLI3 in postaxial polydactyly type A. Nat Genet 17 (3): 269-71.

Reese, M. G. (2001). Application of a time-delay neural network to promoter annotation in the Drosophila melanogaster genome. Comput Chem 26 (1): 51-6.

Regl, G., Kasper, M., Schnidar, H., Eichberger, T., Neill, G. W., Philpott, M. P., Esterbauer, H., Hauser-Kronberger, C., Frischauf, A. M. and Aberger, F. (2004). Activation of the BCL2 promoter in response to Hedgehog/GLI signal transduction is predominantly mediated by GLI2. Cancer Res 64 (21): 7724-31.

Ren, B., Robert, F., Wyrick, J. J., Aparicio, O., Jennings, E. G., Simon, I., Zeitlinger, J., Schreiber, J., Hannett, N., Kanin, E., Volkert, T. L., Wilson, C. J., Bell, S. P. and Young, R. A. (2000). Genome-wide location and function of DNA binding proteins. Science 290 (5500): 2306-9.

Riddle, R. D., Johnson, R. L., Laufer, E. and Tabin, C. (1993). Sonic hedgehog mediates the polarizing activity of the ZPA. Cell 75 (7): 1401-16.

Robbins, D. J., Nybakken, K. E., Kobayashi, R., Sisson, J. C., Bishop, J. M. and Therond, P. P. (1997). Hedgehog elicits signal transduction by means of a large complex containing the kinesin-related protein costal2. Cell 90 (2): 225-34.

Roberts, D. J., Johnson, R. L., Burke, A. C., Nelson, C. E., Morgan, B. A. and Tabin, C. (1995). Sonic hedgehog is an endodermal signal inducing Bmp-4 and Hox genes during induction and regionalization of the chick hindgut. Development 121 (10): 3163-74.

Robertson, C. P., Braun, M. M. and Roelink, H. (2004). Sonic hedgehog patterning in chick neural plate is antagonized by a Wnt3-like signal. Dev Dyn 229 (3): 510-9.

Roelink, H. and Nusse, R. (1991). Expression of two members of the Wnt family during mouse development--restricted temporal and spatial patterns in the developing neural tube. Genes Dev 5 (3): 381-8.

Roelink, H., Porter, J. A., Chiang, C., Tanabe, Y., Chang, D. T., Beachy, P. A. and Jessell, T. M. (1995). Floor plate and motor neuron induction by different concentrations of the amino-terminal cleavage product of sonic hedgehog autoproteolysis. Cell 81 (3): 445-55.

Romm, E., Nielsen, J. A., Kim, J. G. and Hudson, L. D. (2005). Myt1 family recruits histone deacetylase to regulate neural transcription. J Neurochem 93 (6): 1444-53.

Roose, J. and Clevers, H. (1999). TCF transcription factors: molecular switches in carcinogenesis. Biochim Biophys Acta 1424 (2-3): M23-37.

Roose, J., Molenaar, M., Peterson, J., Hurenkamp, J., Brantjes, H., Moerer, P., Van De Wetering, M., Destree, O. and Clevers, H. (1998). The Xenopus Wnt effector XTcf-3 interacts with Groucho-related transcriptional repressors. Nature 395 (6702): 608-12.

Rosenzweig, B. L., Imamura, T., Okadome, T., Cox, G. N., Yamashita, H., Ten Dijke, P., Heldin, C. H. and Miyazono, K. (1995). Cloning and characterization of a human type II receptor for bone morphogenetic proteins. Proc Natl Acad Sci U S A 92 (17): 7632-6.

Rowitch, D. H., B, S. J., Lee, S. M., Flax, J. D., Snyder, E. Y. and Mcmahon, A. P. (1999). Sonic hedgehog regulates proliferation and inhibits differentiation of CNS precursor cells. J Neurosci 19 (20): 8954-65.

Rowitch, D. H., Echelard, Y., Danielian, P. S., Gellner, K., Brenner, S. and Mcmahon, A. P. (1998). Identification of an evolutionarily conserved 110 base-pair cis-acting regulatory sequence that governs Wnt-1 expression in the murine neural plate. Development 125 (14): 2735-46.

Ruiz I Altaba, A. (1998). Combinatorial Gli gene function in floor plate and neuronal inductions by Sonic hedgehog. Development 125 (12): 2203-12.

Ruiz I Altaba, A. (1999). Gli proteins encode context-dependent positive and negative functions: implications for development and disease. Development 126 (14): 3205-16.

Ruppert, J. M., Kinzler, K. W., Wong, A. J., Bigner, S. H., Kao, F. T., Law, M. L., Seuanez, H. N., O'brien, S. J. and Vogelstein, B. (1988). The GLI-Kruppel family of human genes. Mol Cell Biol 8 (8): 3104-13.

Ruppert, J. M., Vogelstein, B., Arheden, K. and Kinzler, K. W. (1990). GLI3 encodes a 190-kilodalton protein with multiple regions of GLI similarity. Mol Cell Biol 10 (10): 5408-15.

Saleh, M., Huang, H., Green, N. C. and Featherstone, M. S. (2000a). A conformational change in PBX1A is necessary for its nuclear localization. Exp Cell Res 260 (1): 105-15.

Saleh, M., Rambaldi, I., Yang, X. J. and Featherstone, M. S. (2000b). Cell signaling switches HOX-PBX complexes from repressors to activators of transcription mediated by histone deacetylases and histone acetyltransferases. Mol Cell Biol 20 (22): 8623-33.

Salinas, P. C. and Nusse, R. (1992). Regional expression of the Wnt-3 gene in the developing mouse forebrain in relationship to diencephalic neuromeres. Mech Dev 39 (3): 151-60.

Sambrook, J., Fritsch, E. F. and Maniatis, T. (1989). Molecular Cloning - A Laboratory Manual, 2nd Edition. Cold Spring Habour Laboratory Press.

Sampath, K., Rubinstein, A. L., Cheng, A. M., Liang, J. O., Fekany, K., Solnica-Krezel, L., Korzh, V., Halpern, M. E. and Wright, C. V. (1998). Induction of the zebrafish ventral brain and floorplate requires cyclops/nodal signalling. Nature 395 (6698): 185-9.

Sandelin, A., Alkema, W., Engstrom, P., Wasserman, W. W. and Lenhard, B. (2004a). JASPAR: an open-access database for eukaryotic transcription factor binding profiles. Nucleic Acids Res 32 (Database issue): D91-4.

Sandelin, A., Bailey, P., Bruce, S., Engstrom, P., Klos, J., Wasserman, W., Ericson, J. and Lenhard, B. (2004b). Arrays of ultraconserved non-coding regions span the loci of key developmental genes in vertebrate genomes. BMC Genomics 5 (1): 99.

Sandelin, A., Carninci, P., Lenhard, B., Ponjavic, J., Hayashizaki, Y. and Hume, D. A. (2007). Mammalian RNA polymerase II core promoters: insights from genome-wide studies. Nat Rev Genet 8 (6): 424-36.

Sandelin, A., Wasserman, W. W. and Lenhard, B. (2004c). ConSite: web-based prediction of regulatory elements using cross-species comparison. Nucleic Acids Res 32 (Web Server issue): W249-52.

Sander, M., Paydar, S., Ericson, J., Briscoe, J., Berber, E., German, M., Jessell, T. M. and Rubenstein, J. L. (2000). Ventral neural patterning by Nkx homeobox genes: Nkx6.1 controls somatic motor neuron and ventral interneuron fates. Genes Dev 14 (17): 2134-9.

Sarno, J. L., Kliman, H. J. and Taylor, H. S. (2005). HOXA10, Pbx2, and Meis1 protein expression in the human endometrium: formation of multimeric complexes on HOXA10 target genes. J Clin Endocrinol Metab 90 (1): 522-8.

Sasaki, H., Hui, C., Nakafuku, M. and Kondoh, H. (1997). A binding site for Gli proteins is essential for HNF-3beta floor plate enhancer activity in transgenics and can respond to Shh in vitro. Development 124 (7): 1313-22.

Sasaki, H., Nishizaki, Y., Hui, C., Nakafuku, M. and Kondoh, H. (1999). Regulation of Gli2 and Gli3 activities by an amino-terminal repression domain: implication of Gli2 and Gli3 as primary mediators of Shh signaling. Development 126 (17): 3915-24.

Savage, C., Das, P., Finelli, A. L., Townsend, S. R., Sun, C. Y., Baird, S. E. and Padgett, R. W. (1996). Caenorhabditis elegans genes sma-2, sma-3, and sma-4 define a conserved family of transforming growth factor beta pathway components. Proc Natl Acad Sci U S A 93 (2): 790-4.

Scherf, M., Klingenhoff, A. and Werner, T. (2000). Highly specific localization of promoter regions in large genomic sequences by PromoterInspector: a novel context analysis approach. J Mol Biol 297 (3): 599-606.

Schimmang, T., Lemaistre, M., Vortkamp, A. and Ruther, U. (1992). Expression of the zinc finger gene Gli3 is affected in the morphogenetic mouse mutant extra-toes (Xt). Development 116 (3): 799-804.

Schimmang, T., Van Der Hoeven, F. and Ruther, U. (1993). Gli3 expression is affected in the morphogenetic mouse mutants add and Xt. Prog Clin Biol Res 383A: 153-61.

Schmid, C. D., Perier, R., Praz, V. and Bucher, P. (2006). EPD in its twentieth year: towards complete promoter coverage of selected model organisms. Nucleic Acids Res 34 (Database issue): D82-5.

Schmid, C. D., Praz, V., Delorenzi, M., Perier, R. and Bucher, P. (2004). The Eukaryotic Promoter Database EPD: the impact of in silico primer extension. Nucleic Acids Res 32 (Database issue): D82-5.

Schmidt, M., Patterson, M., Farrell, E. and Munsterberg, A. (2004). Dynamic expression of Lef/Tcf family members and beta-catenin during chick gastrulation, neurulation, and early limb development. Dev Dyn 229 (3): 703-7.

Schmidt, M., Tanaka, M. and Munsterberg, A. (2000). Expression of (beta)-catenin in the developing chick myotome is regulated by myogenic signals. Development 127 (19): 4105-13.

Schnabel, C. A., Selleri, L., Jacobs, Y., Warnke, R. and Cleary, M. L. (2001). Expression of Pbx1b during mammalian organogenesis. Mech Dev 100 (1): 131-5.

Schoenherr, C. J. and Anderson, D. J. (1995a). The neuron-restrictive silencer factor (NRSF): a coordinate repressor of multiple neuron-specific genes. Science 267 (5202): 1360-3.

Scholey, J. M. and Anderson, K. V. (2006). Intraflagellar transport and cilium-based signaling. Cell 125 (3): 439-42.

Schwabish, M. A. and Struhl, K. (2004). Evidence for eviction and rapid deposition of histones upon transcriptional elongation by RNA polymerase II. Mol Cell Biol 24 (23): 10111-7.

Schwarz, M., Cecconi, F., Bernier, G., Andrejewski, N., Kammandel, B., Wagner, M. and Gruss, P. (2000). Spatial specification of mammalian eye territories by reciprocal transcriptional repression of Pax2 and Pax6. Development 127 (20): 4325-34.

Schweitzer, R., Vogan, K. J. and Tabin, C. J. (2000). Similar expression and regulation of Gli2 and Gli3 in the chick limb bud. Mech Dev 98 (1-2): 171-4.

Sekelsky, J. J., Newfeld, S. J., Raftery, L. A., Chartoff, E. H. and Gelbart, W. M. (1995). Genetic characterization and cloning of mothers against dpp, a gene required for decapentaplegic function in Drosophila melanogaster. Genetics 139 (3): 1347-58.

Selleri, L., Depew, M. J., Jacobs, Y., Chanda, S. K., Tsang, K. Y., Cheah, K. S., Rubenstein, J. L., O'gorman, S. and Cleary, M. L. (2001). Requirement for Pbx1 in skeletal patterning and programming chondrocyte proliferation and differentiation. Development 128 (18): 3543-57.

Shanmugam, K., Featherstone, M. S. and Saragovi, H. U. (1997). Residues flanking the HOX YPWM motif contribute to cooperative interactions with PBX. J Biol Chem 272 (30): 19081-7.

Shanmugam, K., Green, N. C., Rambaldi, I., Saragovi, H. U. and Featherstone, M. S. (1999). PBX and MEIS as non-DNA-binding partners in trimeric complexes with HOX proteins. Mol Cell Biol 19 (11): 7577-88.

Sharrocks, A. D., Brown, A. L., Ling, Y. and Yates, P. R. (1997). The ETS-domain transcription factor family. Int J Biochem Cell Biol 29 (12): 1371-87.

Shen, J. and Walsh, C. A. (2005). Targeted disruption of Tgif, the mouse ortholog of a human holoprosencephaly gene, does not result in holoprosencephaly in mice. Mol Cell Biol 25 (9): 3639-47.

Shen, W. F., Chang, C. P., Rozenfeld, S., Sauvageau, G., Humphries, R. K., Lu, M., Lawrence, H. J., Cleary, M. L. and Largman, C. (1996). Hox homeodomain proteins exhibit selective complex stabilities with Pbx and DNA. Nucleic Acids Res 24 (5): 898-906.

Shen, W. F., Montgomery, J. C., Rozenfeld, S., Moskow, J. J., Lawrence, H. J., Buchberg, A. M. and Largman, C. (1997a). AbdB-like Hox proteins stabilize DNA binding by the Meis1 homeodomain proteins. Mol Cell Biol 17 (11): 6448-58.

Shen, W. F., Rozenfeld, S., Lawrence, H. J. and Largman, C. (1997b). The Abd-B-like Hox homeodomain proteins can be subdivided by the ability to form complexes with Pbx1a on a novel DNA target. J Biol Chem 272 (13): 8198-206.

Shimeld, S. M., Mckay, I. J. and Sharpe, P. T. (1996). The murine homeobox gene Msx-3 shows highly restricted expression in the developing neural tube. Mech Dev 55 (2): 201-10.

Shimizu, H., Julius, M. A., Giarre, M., Zheng, Z., Brown, A. M. and Kitajewski, J. (1997). Transformation by Wnt family proteins correlates with regulation of beta-catenin. Cell Growth Differ 8 (12): 1349-58.

Shiraki, T., Kondo, S., Katayama, S., Waki, K., Kasukawa, T., Kawaji, H., Kodzius, R., Watahiki, A., Nakamura, M., Arakawa, T., Fukuda, S., Sasaki, D., Podhajska, A., Harbers, M., Kawai, J., Carninci, P. and Hayashizaki, Y. (2003). Cap analysis gene expression for high-throughput analysis of transcriptional starting point and identification of promoter usage. Proc Natl Acad Sci U S A 100 (26): 15776-81.

Sif, S. (2004). ATP-dependent nucleosome remodeling complexes: enzymes tailored to deal with chromatin. J Cell Biochem 91 (6): 1087-98.

Sisson, J. C., Ho, K. S., Suyama, K. and Scott, M. P. (1997). Costal2, a novel kinesin-related protein in the Hedgehog signaling pathway. Cell 90 (2): 235-45.

Smale, S. T. and Baltimore, D. (1989). The "initiator" as a transcription control element. Cell 57 (1): 103-13.

Smale, S. T. and Kadonaga, J. T. (2003). The RNA polymerase II core promoter. Annu Rev Biochem 72: 449-79.

Smale, S. T., Jain, A., Kaufmann, J., Emami, K. H., Lo, K. and Garraway, I. P. (1998). The initiator element: a paradigm for core promoter heterogeneity within metazoan protein-coding genes. Cold Spring Harb Symp Quant Biol 63: 21-31.

Solloway, M. J., Dudley, A. T., Bikoff, E. K., Lyons, K. M., Hogan, B. L. and Robertson, E. J. (1998). Mice lacking Bmp6 function. Dev Genet 22 (4): 321-39.

Solovyev, V. and Salamov, A. (1997). The Gene-Finder computer tools for analysis of human and model organisms genome sequences. Proc Int Conf Intell Syst Mol Biol 5: 294-302.

Sonnenburg, S., Zien, A. and Ratsch, G. (2006). ARTS: accurate recognition of transcription starts in human. Bioinformatics 22 (14): e472-80.

Sorek, R. and Ast, G. (2003). Intronic sequences flanking alternatively spliced exons are conserved between human and mouse. Genome Res 13 (7): 1631-7.

Souchelnytskyi, S., Nakayama, T., Nakao, A., Moren, A., Heldin, C. H., Christian, J. L. and Ten Dijke, P. (1998). Physical and functional interaction of murine and Xenopus Smad7 with bone morphogenetic protein receptors and transforming growth factor-beta receptors. J Biol Chem 273 (39): 25364-70.

Speek, M., Njunkova, O., Pata, I., Valdre, E. and Kogerman, P. (2006). A potential role of alternative splicing in the regulation of the transcriptional activity of human GLI2 in gonadal tissues. BMC Mol Biol 7: 13.

Stamataki, D., Ulloa, F., Tsoni, S. V., Mynett, A. and Briscoe, J. (2005). A gradient of Gli activity mediates graded Sonic Hedgehog signaling in the neural tube. Genes Dev 19 (5): 626-41.

Stojanovic, N., Florea, L., Riemer, C., Gumucio, D., Slightom, J., Goodman, M., Miller, W. and Hardison, R. (1999). Comparison of five methods for finding conserved sequences in multiple alignments of gene regulatory regions. Nucleic Acids Res 27 (19): 3899-910.

Stone, D. M., Hynes, M., Armanini, M., Swanson, T. A., Gu, Q., Johnson, R. L., Scott, M. P., Pennica, D., Goddard, A., Phillips, H., Noll, M., Hooper, J. E., De Sauvage, F. and Rosenthal, A. (1996). The tumour-suppressor gene patched encodes a candidate receptor for Sonic hedgehog. Nature 384 (6605): 129-34.

Stormo, G. D. (2000). DNA binding sites: representation and discovery. Bioinformatics 16 (1): 16-23.

Su, G., Mao, B. and Wang, J. (2006). A web server for transcription factor binding site prediction. Bioinformation 1 (5): 156-7.

Suzuki, Y., Yamashita, R., Nakai, K. and Sugano, S. (2002). DBTSS: DataBase of human Transcriptional Start Sites and full-length cDNAs. Nucleic Acids Res 30 (1): 328-31.

Svard, J., Heby-Henricson, K., Persson-Lek, M., Rozell, B., Lauth, M., Bergstrom, A., Ericson, J., Toftgard, R. and Teglund, S. (2006). Genetic elimination of Suppressor of fused reveals an essential repressor function in the mammalian Hedgehog signaling pathway. Dev Cell 10 (2): 187-97.

Swartz, M., Eberhart, J., Mastick, G. S. and Krull, C. E. (2001). Sparking new frontiers: using in vivo electroporation for genetic manipulations. Dev Biol 233 (1): 13-21.

Tagle, D. A., Koop, B. F., Goodman, M., Slightom, J. L., Hess, D. L. and Jones, R. T. (1988). Embryonic epsilon and gamma globin genes of a prosimian primate (Galago crassicaudatus). Nucleotide and amino acid sequences, developmental regulation and phylogenetic footprints. J Mol Biol 203 (2): 439-55.

Tago, K., Nakamura, T., Nishita, M., Hyodo, J., Nagai, S., Murata, Y., Adachi, S., Ohwada, S., Morishita, Y., Shibuya, H. and Akiyama, T. (2000). Inhibition of Wnt signaling by ICAT, a novel beta-catenin-interacting protein. Genes Dev 14 (14): 1741-9.

Takada, S., Stark, K. L., Shea, M. J., Vassileva, G., Mcmahon, J. A. and Mcmahon, A. P. (1994). Wnt-3a regulates somite and tailbud formation in the mouse embryo. Genes Dev 8 (2): 174-89.

Takebayashi, H., Ohtsuki, T., Uchida, T., Kawamoto, S., Okubo, K., Ikenaka, K., Takeichi, M., Chisaka, O. and Nabeshima, Y. (2002). Non-overlapping expression of Olig3 and Olig2 in the embryonic neural tube. Mech Dev 113 (2): 169-74.

Takemaru, K., Yamaguchi, S., Lee, Y. S., Zhang, Y., Carthew, R. W. and Moon, R. T. (2003). Chibby, a nuclear beta-catenin-associated antagonist of the Wnt/Wingless pathway. Nature 422 (6934): 905-9.

Tamai, K., Zeng, X., Liu, C., Zhang, X., Harada, Y., Chang, Z. and He, X. (2004). A mechanism for Wnt coreceptor activation. Mol Cell 13 (1): 149-56.

Tamaki, K., Souchelnytskyi, S., Itoh, S., Nakao, A., Sampath, K., Heldin, C. H. and Ten Dijke, P. (1998). Intracellular signaling of osteogenic protein-1 through Smad5 activation. J Cell Physiol 177 (2): 355-63.

Tanabe, Y., William, C. and Jessell, T. M. (1998). Specification of Motor Neuron Identity by the MNR2 Homeodomain Protein. Cell 95 (1): 67-80.

Tanimura, A., Dan, S. and Yoshida, M. (1998). Cloning of novel isoforms of the human Gli2 oncogene and their activities to enhance tax-dependent transcription of the human T-cell leukemia virus type 1 genome. J Virol 72 (5): 3958-64.

Tatusova, T. A. and Madden, T. L. (1999). BLAST 2 S, a new tool for comparing protein and nucleotide sequences. FEMS Microbiology Letters 174 (2): 247-250.

Te Welscher, P., Zuniga, A., Kuijper, S., Drenth, T., Goedemans, H. J., Meijlink, F. and Zeller, R. (2002). Progression of vertebrate limb development through SHH-mediated counteraction of GLI3. Science 298 (5594): 827-30.

Tempe, D., Casas, M., Karaz, S., Blanchet-Tournier, M. F. and Concordet, J. P. (2006). Multisite protein kinase A and glycogen synthase kinase 3beta phosphorylation leads to Gli3 ubiquitination by SCFbetaTrCP. Mol Cell Biol 26 (11): 4316-26.

Ten Dijke, P., Yamashita, H., Sampath, T. K., Reddi, A. H., Estevez, M., Riddle, D. L., Ichijo, H., Heldin, C. H. and Miyazono, K. (1994). Identification of type I receptors for osteogenic protein-1 and bone morphogenetic protein-4. J Biol Chem 269 (25): 16985-8.

Thayer, S. P., Di Magliano, M. P., Heiser, P. W., Nielsen, C. M., Roberts, D. J., Lauwers, G. Y., Qi, Y. P., Gysin, S., Fernandez-Del Castillo, C., Yajnik, V., Antoniu, B., Mcmahon, M., Warshaw, A. L. and Hebrok, M. (2003). Hedgehog is an early and late mediator of pancreatic cancer tumorigenesis. Nature 425 (6960): 851-6.

Theil, T., Aydin, S., Koch, S., Grotewold, L. and Ruther, U. (2002). Wnt and Bmp signalling cooperatively regulate graded Emx2 expression in the dorsal telencephalon. Development 129 (13): 3045-54.

Thien, H. and Ruther, U. (1999). The mouse mutation Pdn (Polydactyly Nagoya) is caused by the integration of a retrotransposon into the Gli3 gene. Mamm Genome 10 (3): 205-9.

Thomas, J. W., Touchman, J. W., Blakesley, R. W., Bouffard, G. G., Beckstrom-Sternberg, S. M., Margulies, E. H., Blanchette, M., Siepel, A. C., Thomas, P. J., Mcdowell, J. C., Maskeri, B., Hansen, N. F., Schwartz, M. S., Weber, R. J., Kent, W. J., Karolchik, D., Bruen, T. C., Bevan, R., Cutler, D. J., Schwartz, S., Elnitski, L., Idol, J. R., Prasad, A. B., Lee-Lin, S. Q., Maduro, V. V., Summers, T. J., Portnoy, M. E., Dietrich, N. L., Akhter, N., Ayele, K., Benjamin, B., Cariaga, K., Brinkley, C. P., Brooks, S. Y., Granite, S., Guan, X., Gupta, J., Haghighi, P., Ho, S. L., Huang, M. C., Karlins, E., Laric, P. L., Legaspi, R., Lim, M. J., Maduro, Q. L., Masiello, C. A., Mastrian, S. D., Mccloskey, J. C., Pearson, R., Stantripop, S., Tiongson, E. E., Tran, J. T., Tsurgeon, C., Vogt, J. L., Walker, M. A., Wetherby, K. D., Wiggins, L. S., Young, A. C., Zhang, L. H., Osoegawa, K., Zhu, B., Zhao, B., Shu, C. L., De Jong, P. J., Lawrence, C. E., Smit, A. F., Chakravarti, A., Haussler, D., Green, P., Miller, W. and Green, E. D. (2003). Comparative analyses of multi-species sequences from targeted genomic regions. Nature 424 (6950): 788-93.

Thompson, J. D., Higgins, D. G. and Gibson, T. J. (1994). CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res 22 (22): 4673-80.

Timmer, J. R., Wang, C. and Niswander, L. (2002). BMP signaling patterns the dorsal and intermediate neural tube via regulation of homeobox and helix-loop-helix transcription factors. Development 129 (10): 2459-72.

Timmer, J., Johnson, J. and Niswander, L. (2001). The use of in ovo electroporation for the rapid analysis of neural-specific murine enhancers. Genesis 29 (3): 123-32.

Tolwinski, N. S., Wehrli, M., Rives, A., Erdeniz, N., Dinardo, S. and Wieschaus, E. (2003). Wg/Wnt signal can be transmitted through arrow/LRP5,6 and Axin independently of Zw3/Gsk3beta activity. Dev Cell 4 (3): 407-18.

Tsukui, T., Capdevila, J., Tamura, K., Ruiz-Lozano, P., Rodriguez-Esteban, C., Yonei-Tamura, S., Magallon, J., Chandraratna, R. A., Chien, K., Blumberg, B., Evans, R. M. and Belmonte, J. C. (1999). Multiple left-right asymmetry defects in Shh(-/-) mutant mice unveil a convergence of the shh and retinoic acid pathways in the control of Lefty-1. Proc Natl Acad Sci U S A 96 (20): 11376-81.

Tyurina, O. V., Guner, B., Popova, E., Feng, J., Schier, A. F., Kohtz, J. D. and Karlstrom, R. O. (2005). Zebrafish Gli3 functions as both an activator and a repressor in Hedgehog signaling. Dev Biol 277 (2): 537-56.

Uchikawa, M., Ishida, Y., Takemoto, T., Kamachi, Y. and Kondoh, H. (2003). Functional analysis of chicken Sox2 enhancers highlights an array of diverse regulatory elements that are conserved in mammals. Dev Cell 4 (4): 509-19.

Uchikawa, M., Takemoto, T., Kamachi, Y. and Kondoh, H. (2004). Efficient identification of regulatory sequences in the chicken genome by a powerful combination of embryo electroporation and genome comparison. Mech Dev 121 (9): 1145-58.

Umbhauer, M., Djiane, A., Goisset, C., Penzo-Mendez, A., Riou, J. F., Boucaut, J. C. and Shi, D. L. (2000). The C-terminal cytoplasmic Lys-thr-X-X-X-Trp motif in frizzled receptors mediates Wnt/beta-catenin signalling. Embo J 19 (18): 4944-54.

Van De Wetering, M., Cavallo, R., Dooijes, D., Van Beest, M., Van Es, J., Loureiro, J., Ypma, A., Hursh, D., Jones, T., Bejsovec, A., Peifer, M., Mortin, M. and Clevers, H. (1997). Armadillo coactivates transcription driven by the product of the Drosophila segment polarity gene dTCF. Cell 88 (6): 789-99.

Van Der Hoeven, F., Schimmang, T., Vortkamp, A. and Ruther, U. (1993). Molecular linkage of the morphogenetic mutation add and the zinc finger gene Gli3. Mamm Genome 4 (5): 276-7.

Van Genderen, C., Okamura, R. M., Farinas, I., Quo, R. G., Parslow, T. G., Bruhn, L. and Grosschedl, R. (1994). Development of several organs that require inductive epithelial-mesenchymal interactions is impaired in LEF-1-deficient mice. Genes Dev 8 (22): 2691-703.

Varjosalo, M., Li, S. P. and Taipale, J. (2006). Divergence of hedgehog signal transduction mechanism between Drosophila and mammals. Dev Cell 10 (2): 177-86.

Vavouri, T., Mcewen, G. K., Woolfe, A., Gilks, W. R. and Elgar, G. (2006a). Defining a genomic radius for long-range enhancer action: duplicated conserved non-coding elements hold the key. Trends Genet 22 (1): 5-10.

Verrijzer, C. P. and Van Der Vliet, P. C. (1993). POU domain transcription factors. Biochim Biophys Acta 1173 (1): 1-21.

Von Mering, C. and Basler, K. (1999). Distinct and regulated activities of human Gli proteins in Drosophila. Curr Biol 9 (22): 1319-22.

Vortkamp, A., Franz, T., Gessler, M. and Grzeschik, K. H. (1992). Deletion of GLI3 supports the homology of the human Greig cephalopolysyndactyly syndrome (GCPS) and the mouse mutant extra toes (Xt). Mamm Genome 3 (8): 461-3.

Vortkamp, A., Gessler, M. and Grzeschik, K. H. (1991). GLI3 zinc-finger gene interrupted by translocations in Greig syndrome families. Nature 352 (6335): 539-40.

Vortkamp, A., Gessler, M., Le Paslier, D., Elaswarapu, R., Smith, S. and Grzeschik, K. H. (1994). Isolation of a yeast artificial chromosome contig spanning the Greig cephalopolysyndactyly syndrome (GCPS) gene region. Genomics 22 (3): 563-8.

Vortkamp, A., Heid, C., Gessler, M. and Grzeschik, K. H. (1995). Isolation and characterization of a cosmid contig for the GCPS gene region. Hum Genet 95 (1): 82-8.

Vortkamp, A., Lee, K., Lanske, B., Segre, G. V., Kronenberg, H. M. and Tabin, C. J. (1996). Regulation of rate of cartilage differentiation by Indian hedgehog and PTH-related protein. Science 273 (5275): 613-22.

Wagner, K., Mincheva, A., Korn, B., Lichter, P. and Popperl, H. (2001). Pbx4, a new Pbx family member on mouse chromosome 8, is expressed during spermatogenesis. Mech Dev 103 (1-2): 127-31.

Wakaguri, H., Yamashita, R., Suzuki, Y., Sugano, S. and Nakai, K. (2008). DBTSS: database of transcription start sites, progress report 2008. Nucleic Acids Res 36 (Database issue): D97-101.

Walterhouse, D., Ahmed, M., Slusarski, D., Kalamaras, J., Boucher, D., Holmgren, R. and Iannaccone, P. (1993). gli, a zinc finger transcription factor and oncogene, is expressed during normal mouse development. Dev Dyn 196 (2): 91-102.

Waltzer, L. and Bienz, M. (1998). Drosophila CBP represses the transcription factor TCF to antagonize Wingless signalling. Nature 395 (6701): 521-5.

Wang, B. and Li, Y. (2006). Evidence for the direct involvement of {beta}TrCP in Gli3 protein processing. Proc Natl Acad Sci U S A 103 (1): 33-8.

Wang, B., Fallon, J. F. and Beachy, P. A. (2000). Hedgehog-regulated processing of Gli3 produces an anterior/posterior repressor gradient in the developing vertebrate limb. Cell 100 (4): 423-34.

Wang, G. G., Pasillas, M. P. and Kamps, M. P. (2006). Persistent transactivation by meis1 replaces hox function in myeloid leukemogenesis models: evidence for co-occupancy of meis1-pbx and hox-pbx complexes on promoters of leukemia-associated genes. Mol Cell Biol 26 (10): 3902-16.

Wang, G., Wang, B. and Jiang, J. (1999). Protein kinase A antagonizes Hedgehog signaling by regulating both the activator and repressor forms of Cubitus interruptus. Genes Dev 13 (21): 2828-37.

Wang, Q. T. and Holmgren, R. A. (2000). Nuclear import of cubitus interruptus is regulated by hedgehog via a mechanism distinct from Ci stabilization and Ci activation. Development 127 (14): 3131-9.

Wang, Y., Yin, L. and Hillgartner, F. B. (2001). The homeodomain proteins PBX and MEIS1 are accessory factors that enhance thyroid hormone regulation of the malic enzyme gene in hepatocytes. J Biol Chem 276 (26): 23838-48.

Washington Smoak, I., Byrd, N. A., Abu-Issa, R., Goddeeris, M. M., Anderson, R., Morris, J., Yamamura, K., Klingensmith, J. and Meyers, E. N. (2005). Sonic hedgehog is required for cardiac outflow tract and neural crest cell development. Dev Biol 283 (2): 357-72.

Wasserman, W. W. and Sandelin, A. (2004). Applied bioinformatics for the identification of regulatory elements. Nat Rev Genet 5 (4): 276-87.

Wasserman, W. W., Palumbo, M., Thompson, W., Fickett, J. W. and Lawrence, C. E. (2000). Human-mouse genome comparisons to locate regulatory sites. Nat Genet 26 (2): 225-8.

Waterston, R. H., Lindblad-Toh, K., Birney, E., Rogers, J., Abril, J. F., Agarwal, P., Agarwala, R., Ainscough, R., Alexandersson, M., An, P., Antonarakis, S. E., Attwood, J., Baertsch, R., Bailey, J., Barlow, K., Beck, S., Berry, E., Birren, B., Bloom, T., Bork, P., Botcherby, M., Bray, N., Brent, M. R., Brown, D. G., Brown, S. D., Bult, C., Burton, J., Butler, J., Campbell, R. D., Carninci, P., Cawley, S., Chiaromonte, F., Chinwalla, A. T., Church, D. M., Clamp, M., Clee, C., Collins, F. S., Cook, L. L., Copley, R. R., Coulson, A., Couronne, O., Cuff, J., Curwen, V., Cutts, T., Daly, M., David, R., Davies, J., Delehaunty, K. D., Deri, J., Dermitzakis, E. T., Dewey, C., Dickens, N. J., Diekhans, M., Dodge, S., Dubchak, I., Dunn, D. M., Eddy, S. R., Elnitski, L., Emes, R. D., Eswara, P., Eyras, E., Felsenfeld, A., Fewell, G. A., Flicek, P., Foley, K., Frankel, W. N., Fulton, L. A., Fulton, R. S., Furey, T. S., Gage, D., Gibbs, R. A., Glusman, G., Gnerre, S., Goldman, N., Goodstadt, L., Grafham, D., Graves, T. A., Green, E. D., Gregory, S., Guigo, R., Guyer, M., Hardison, R. C., Haussler, D., Hayashizaki, Y., Hillier, L. W., Hinrichs, A., Hlavina, W., Holzer, T., Hsu, F., Hua, A., Hubbard, T., Hunt, A., Jackson, I., Jaffe, D. B., Johnson, L. S., Jones, M., Jones, T. A., Joy, A., Kamal, M., Karlsson, E. K., Karolchik, D., Kasprzyk, A., Kawai, J., Keibler, E., Kells, C., Kent, W. J., Kirby, A., Kolbe, D. L., Korf, I., Kucherlapati, R. S., Kulbokas, E. J., Kulp, D., Landers, T., Leger, J. P., Leonard, S., Letunic, I., Levine, R., Li, J., Li, M., Lloyd, C., Lucas, S., Ma, B., Maglott, D. R., Mardis, E. R., Matthews, L., Mauceli, E., Mayer, J. H., Mccarthy, M., Mccombie, W. R., Mclaren, S., Mclay, K., Mcpherson, J. D., Meldrim, J., Meredith, B., Mesirov, J. P., Miller, W., Miner, T. L., Mongin, E., Montgomery, K. T., Morgan, M., Mott, R., Mullikin, J. C., Muzny, D. M., Nash, W. E., Nelson, J. O., Nhan, M. N., Nicol, R., Ning, Z., Nusbaum, C., O'connor, M. J., Okazaki, Y., Oliver, K., Overton-Larty, E., Pachter, L., Parra, G., Pepin, K. H., Peterson, J., Pevzner, P., Plumb, R., Pohl, C. S., Poliakov, A., Ponce, T. C., Ponting, C. P., Potter, S., Quail, M., Reymond, A., Roe, B. A., Roskin, K. M., Rubin, E. M., Rust, A. G., Santos, R., Sapojnikov, V., Schultz, B., Schultz, J., Schwartz, M. S., Schwartz, S., Scott, C., Seaman, S., Searle, S., Sharpe, T., Sheridan, A., Shownkeen, R., Sims, S., Singer, J. B., Slater, G., Smit, A., Smith, D. R., Spencer, B., Stabenau, A., Stange-Thomann, N., Sugnet, C., Suyama, M., Tesler, G., Thompson, J., Torrents, D., Trevaskis, E., Tromp, J., Ucla, C., Ureta-Vidal, A., Vinson, J. P., Von Niederhausern, A. C., Wade, C. M., Wall, M., Weber, R. J., Weiss, R. B., Wendl, M. C., West, A. P., Wetterstrand, K., Wheeler, R., Whelan, S., Wierzbowski, J., Willey, D., Williams, S., Wilson, R. K., Winter, E., Worley, K. C., Wyman, D., Yang, S., Yang, S. P., Zdobnov, E. M., Zody, M. C. and Lander, E. S. (2002). Initial sequencing and comparative analysis of the mouse genome. Nature 420 (6915): 520-62.

Weinmann, A. S., Yan, P. S., Oberley, M. J., Huang, T. H. and Farnham, P. J. (2002). Isolating human transcription factor targets by coupling chromatin immunoprecipitation and CpG island microarray analysis. Genes Dev 16 (2): 235-44.

Westerfield, M., Wegner, J., Jegalian, B. G., Derobertis, E. M. and Puschel, A. W. (1992). Specific activation of mammalian Hox promoters in mosaic transgenic zebrafish. Genes Dev 6 (4): 591-8.

Whiting, J., Marshall, H., Cook, M., Krumlauf, R., Rigby, P. W., Stott, D. and Allemann, R. K. (1991). Multiple spatially specific enhancers are required to reconstruct the pattern of Hox-2.6 gene expression. Genes Dev 5 (11): 2048-59.

Wijgerde, M., Mcmahon, J. A., Rule, M. and Mcmahon, A. P. (2002). A direct requirement for Hedgehog signaling for normal specification of all ventral progenitor domains in the presumptive mammalian spinal cord. Genes Dev 16 (22): 2849-64.

Wild, A., Kalff-Suske, M., Vortkamp, A., Bornholdt, D., Konig, R. and Grzeschik, K. H. (1997). Point mutations in human GLI3 cause Greig syndrome. Hum Mol Genet 6 (11): 1979-84.

Williams, S. K. and Tyler, J. K. (2007). Transcriptional regulation by chromatin disassembly and reassembly. Curr Opin Genet Dev 17 (2): 88-93.

Wilson, D., Sheng, G., Lecuit, T., Dostatni, N. and Desplan, C. (1993). Cooperative dimerization of paired class homeo domains on DNA. Genes Dev. 7 (11): 2120-2134.

Wilson, L., Gale, E. and Maden, M. (2003). The role of retinoic acid in the morphogenesis of the neural tube. J Anat 203 (4): 357-68.

Wilusz, J., Pettine, S. M. and Shenk, T. (1989). Functional analysis of point mutations in the AAUAAA motif of the SV40 late polyadenylation signal. Nucleic Acids Res 17 (10): 3899-908.

Wingender, E., Dietze, P., Karas, H. and Knuppel, R. (1996). TRANSFAC: a database on transcription factors and their DNA binding sites. Nucleic Acids Res 24 (1): 238-41.

Winnier, G., Blessing, M., Labosky, P. A. and Hogan, B. L. (1995). Bone morphogenetic protein-4 is required for mesoderm formation and patterning in the mouse. Genes Dev 9 (17): 2105-16.

Winter, R. M. and Huson, S. M. (1988). Greig cephalopolysyndactyly syndrome: a possible mouse homologue (Xt-extra toes). Am J Med Genet 31 (4): 793-8.

Wong, C., Rougier-Chapman, E. M., Frederick, J. P., Datto, M. B., Liberati, N. T., Li, J. M. and Wang, X. F. (1999). Smad3-Smad4 and AP-1 complexes synergize in transcriptional activation of the c-Jun promoter by transforming growth factor beta. Mol Cell Biol 19 (3): 1821-30.

Wong, G. T., Gavin, B. J. and Mcmahon, A. P. (1994). Differential transformation of mammary epithelial cells by Wnt genes. Mol Cell Biol 14 (9): 6278-86.

Woolfe, A., Goodson, M., Goode, D. K., Snell, P., Mcewen, G. K., Vavouri, T., Smith, S. F., North, P., Callaway, H., Kelly, K., Walter, K., Abnizova, I., Gilks, W., Edwards, Y. J., Cooke, J. E. and Elgar, G. (2005). Highly conserved non-coding sequences are associated with vertebrate development. PLoS Biol 3 (1): e7.

Wotton, D., Lo, R. S., Lee, S. and Massague, J. (1999a). A Smad transcriptional corepressor. Cell 97 (1): 29-39.

Wotton, D., Lo, R. S., Swaby, L. A. and Massague, J. (1999b). Multiple modes of repression by the Smad transcriptional corepressor TGIF. J Biol Chem 274 (52): 37105-10.

Wrana, J. L., Attisano, L., Wieser, R., Ventura, F. and Massague, J. (1994). Mechanism of activation of the TGF-beta receptor. Nature 370 (6488): 341-7.

Wu, C. (1980). The 5' ends of Drosophila heat shock genes in chromatin are hypersensitive to DNase I. Nature 286 (5776): 854-60.

Wu, J. W., Fairman, R., Penry, J. and Shi, Y. (2001). Formation of a stable heterodimer between Smad2 and Smad4. J Biol Chem 276 (23): 20688-94.

Y. Chen, D. H. G., C.A. Haipek, B.J. Martinsen, M. Bronner-Fraser, C.E. Krull, (2004). Characterization of chicken <I>Nf2</I>/merlin indicates regulatory roles in cell proliferation and migration. Developmental Dynamics 229 (3): 541-554.

Yagi, K., Furuhashi, M., Aoki, H., Goto, D., Kuwano, H., Sugamura, K., Miyazono, K. and Kato, M. (2002). c-myc is a downstream target of the Smad pathway. J Biol Chem 277 (1): 854-61.

Yagi, K., Goto, D., Hamamoto, T., Takenoshita, S., Kato, M. and Miyazono, K. (1999). Alternatively Spliced Variant of Smad2 Lacking Exon 3. COMPARISON WITH WILD-TYPE Smad2 AND Smad3. J. Biol. Chem. 274 (2): 703-709.

Yamada, T., Pfaff, S. L., Edlund, T. and Jessell, T. M. (1993). Control of cell pattern in the neural tube: motor neuron induction by diffusible factors from notochord and floor plate. Cell 73 (4): 673-86.

Yee, S. P. and Rigby, P. W. (1993). The regulation of myogenin gene expression during the embryonic development of the mouse. Genes Dev 7 (7A): 1277-89.

Yuan, G. C., Liu, Y. J., Dion, M. F., Slack, M. D., Wu, L. F., Altschuler, S. J. and Rando, O. J. (2005). Genome-scale identification of nucleosome positions in S. cerevisiae. Science 309 (5734): 626-30.

Zawel, L., Dai, J. L., Buckhaults, P., Zhou, S., Kinzler, K. W., Vogelstein, B. and Kern, S. E. (1998). Human Smad3 and Smad4 are sequence-specific transcription activators. Mol Cell 1 (4): 611-7.

Zechner, D., Fujita, Y., Hulsken, J., Muller, T., Walther, I., Taketo, M. M., Crenshaw, E. B., 3rd, Birchmeier, W. and Birchmeier, C. (2003). beta-Catenin signals regulate cell growth and the balance between progenitor cell expansion and differentiation in the nervous system. Dev Biol 258 (2): 406-18.

Zechner, D., Muller, T., Wende, H., Walther, I., Taketo, M. M., Crenshaw, E. B., 3rd, Treier, M., Birchmeier, W. and Birchmeier, C. (2007). Bmp and Wnt/beta-catenin signals control expression of the transcription factor Olig3 and the specification of spinal cord neurons. Dev Biol 303 (1): 181-90.

Zhang, M. Q. (1998). Identification of human gene core promoters in silico. Genome Res 8 (3): 319-26.

Zhang, M. Q. (2007). Computational analyses of eukaryotic promoters. BMC Bioinformatics 8 Suppl 6: S3.

Zhang, W., Zhao, Y., Tong, C., Wang, G., Wang, B., Jia, J. and Jiang, J. (2005). Hedgehog-regulated Costal2-kinase complexes control phosphorylation and proteolytic processing of Cubitus interruptus. Dev Cell 8 (2): 267-78.

Zhang, X. M., Ramalho-Santos, M. and Mcmahon, A. P. (2001a). Smoothened mutants reveal redundant roles for Shh and Ihh signaling including regulation of L/R symmetry by the mouse node. Cell 106 (2): 781-92.

Zhang, X., Friedman, A., Heaney, S., Purcell, P. and Maas, R. L. (2002). Meis homeoproteins directly regulate Pax6 during vertebrate lens morphogenesis. Genes Dev 16 (16): 2097-107.

Zhang, X., Rowan, S., Yue, Y., Heaney, S., Pan, Y., Brendolan, A., Selleri, L. and Maas, R. L. (2006). Pax6 is regulated by Meis and Pbx homeoproteins during pancreatic development. Dev Biol 300 (2): 748-57.

Zhang, Y., Feng, X. H. and Derynck, R. (1998). Smad3 and Smad4 cooperate with c-Jun/c-Fos to mediate TGF-beta-induced transcription. Nature 394 (6696): 909-13.

Zhao, X., Xuan, Z. and Zhang, M. Q. (2007). Boosting with stumps for predicting transcription start sites. Genome Biol 8 (2): R17.

Zhou, S., Zawel, L., Lengauer, C., Kinzler, K. W. and Vogelstein, B. (1998). Characterization of human FAST-1, a TGF beta and activin signal transducer. Mol Cell 2 (1): 121-7.

Zhuang, B. and Sockanathan, S. (2006). Dorsal-ventral patterning: a view from the top. Curr Opin Neurobiol 16 (1): 20-4.