4.3: Discussion

Enhancers, in combination with promoters, are part of the molecular machinery that controls gene expression in a spatio-temporal manner (Kowalczyk et al. 2012). A single enhancer unit capable of binding one or several transcription factors can regulate gene expression. Alternatively, multiple enhancers that are large distances apart can act in combination to control gene expression (Müller et al. 1999). In both of these scenarios, enhancers can be great distances away from the transcriptional start site (Irimia et al. 2012). For this reason, BAC constructs injected into zebrafish embryos may fail to activate reporter gene expression, if the enhancers lie outside of the region of DNA that is incorporated into the BAC. Likewise, if enhancers are spread throughout a genomic locus (see *Figure 1.18*), some enhancers may be missing from the BAC construct used. In this case, some tissues may express the reporter gene, while others may not. Enhancers that regulate temporal gene expression may also be missing from the BAC construct, and so the timing of reporter gene expression may be altered compared to the reported endogenous gene expression.

In this chapter, I have assessed whether BAC zC34A17 contains enhancer sequences capable of activating a GFP reporter gene inserted into the *lama1* gene at different stages of zebrafish embryonic development.

4.3.1: BAC zC34A17 initiates GFP expression correctly

BAC zC34A17 contains the zebrafish *lama1* gene locus, and is capable of activating tissue specific GFP reporter gene expression in a pattern similar to that of *lama1*. Maternal *lama1* transcripts are detected as early as the two-cell stage, but GFP mRNA is first expressed around the 1000-cell stage (3hpf) in transiently injected embryo. Therefore, zygotic *lama1* transcription is most likely initiated around 3hpf. GFP fluorescence is first detected between the 50-75% epiboly stage. The time lag between the expression of GFP mRNA and the detection of GFP fluorescence could represent the time needed to synthesise GFP protein.

GFP fluorescence is later restricted to the forming chordamesoderm and neural plate at the 3-somite stage, and then expressed throughout the paraxial mesoderm, notochord and neural tube by the 6-somite stage of transiently injected embryos. By the 15-somite stage, GFP is also strongly detected in the PSM, throughout the somite in both fast and slow muscle domains, the anterior CNS, and the forming eye and otic vesicle, in both transiently injected and stable line embryos. From 25hpf, GFP fluorescence is also detected in the vasculature and the uro-genital region, and then in the forming jaw musculature and pectoral fins by 48hpf. Importantly, initiation of GFP fluorescence fully recapitulates the expression of *lama1* at these early stages of development, revealing that all the enhancers required for initiation of *lama1* expression throughout the embryo are present in BAC zC34A17.

It is known that GFP expression can be affected by the surrounding DNA sequences at the site of BAC integration. This can result in variable GFP expression patterns in stable transgenic zebrafish lines, that are generated from identical DNA constructs (Ertzer et al. 2007). Encouragingly, all progeny obtained from the five founder fish display the same expression pattern of GFP, indicating that GFP expression is specifically controlled by sequences in BAC zC34A17.

4.3.2: Persistent GFP expression is not caused by the loss of repressor elements

In transient transgenic zebrafish, GFP fluorescence persists longer than the expression of *lama1* in some tissues. For instance, GFP fluorescence is detected in fast and slow muscle fibres, vasculature and pro-nephric tubules, notochord, and the neural tube, at stages in which *lama1* expression has been down-regulated. Nevertheless, GFP is also maintained in structures which normally still express *lama1* at 74hpf, including the eye, jaw musculature, heart, otic vesicles, and the midbrain-hindbrain boundary (Paulus and Halloran 2006; Semina et al. 2006).

I believe that the persistent GFP fluorescence is not due to the loss of repressor regulatory elements from BAC zC34A17, despite data suggesting that GFP is still transcriptionally active at 74hpf (*Figure 4.17*). From 25hpf, a stable line of *lama1:*GFP zebrafish down-regulate GFP expression within the somites, neural tube, notochord, vasculature/pro-nephric tubules, and the PSM. This is observed more clearly when analysing the expression of GFP mRNA, which shows an anterior to posterior down-regulation of somitic expression at 25hpf, recapitulating the down-regulation and posteriorised gradient of *lama1*. By 49hpf, very little GFP mRNA expression remains in the trunk, and expression is fully down-regulated in the somites, uro-genital region, notochord, and the PSM by 72hpf. The down-regulation of GFP mRNA is slightly delayed compared to *lama1*, which could be due to the stability of GFP transcripts (unpublished data, Stone Elworthy). The increased stability of GFP mRNA in combination with GFP protein stability is likely to contribute to the weak GFP fluorescence detected in the somites, vasculature, and notochord at 25hpf.

Therefore, I propose that the persistent GFP expression observed in the BAC-injected embryos is pseudo-expression (Bonifer et al. 1996; Cranston et al. 2001). 5' and 3' chromosomal sequences adjacent to the transgene integration site can enhance the expression of a transgene, or even affect tissue specificity (Clark et al. 1994; Cranston et al. 2001). However, the fact that as much as 50% of embryos strongly express GFP in the muscle fibres at 49hpf (*Table 4.6*D) means I can rule out the possibility that the ectopic expression is due to positional integration effects. Epigenetic modifications including methylation can also alter the expression of a transgene (Yin et al. 2012). It is possible that methylation of *lama1* is normally responsible for its down-regulation in the somites, neural tube, notochord, vasculature/pro-nephric tubules and the PSM. Other Laminin α chains have also been associated with DNA methylation (Ii et al. 2011; Lee et al. 2012). If this is

the case, an absence of these methylation and chromatin remodeling mechanisms around the integration site in the transiently injected embryos could account for persistent GFP expression. In support of this, inhibition of DNA methylation was found to increase expression of GFP in fibroblast cell lines obtained from transgenic pigs (Yin et al. 2012). Although histone modifications are inherited, they can give rise to alterations in subsequent generations, resulting in a different epigenetic landscape and differing gene expression (Gabory et al. 2011). Therefore, it is conceivable that sequences in BAC zC34A17 only recruit the necessary methylation machinery to down-regulate GFP expression in the F1 generation. This may happen because stably integrated transgenes are likely to be lower in transgene copy number per cell, compared to the transiently injected transgene (Suster et al. 2009; Bussmann and Schulte-Merker 2011). When a cell contains multiple copies of the transient transgene, it is possible that so much DNA cannot be silenced, resulting in ectopic expression. In support of this, others have shown that transgene persistence is associated with large concatemer structures of DNA (Chen et al. 2001).

4.3.3: BAC zC34A17 is Hh responsive

In the zebrafish CNS, *Shh* is expressed in the ventral midbrain, hypothalamus, telencephalon, zona limitans intrathalamica (zli), and the floor plate of the neural tube (Krauss et al. 1993; Ertzer et al. 2007). Through loss-of-function and gain-of-function approaches, I have shown that Hh signalling has a role in regulating *lama1* expression in the zebrafish embryo. Elevated levels of Hh signalling in *ptc1/2* and dnPKA mRNA-injected embryos result in an increased expression of *lama1* in the anterior CNS at the 15-somite stage. An increased expression of *lama1* is also observed in the anterior CNS of 25hpf dnPKA mRNA-injected embryos (section 3.2.5.2), and a slight reduction of *lama1* expression in the anterior CNS may also be observed in cyclopamine-treated embryos at 25hpf (section 3.2.5.1). In support of the possibility that Hh signalling regulates *lama1* expression in the anterior CNS, I have shown in this chapter that GFP expression in transiently injected embryos is mainly observed in ventrally-located cells of the anterior CNS, a domain known to be Hh responsive. At the 20-somite stage, 47% of embryos express GFP in the ventral CNS, compared to 30% of embryos expressing GFP in the dorsal CNS (*Figure 4.20*B and *Table 4.5*B). Similarly, at 25hpf, 67% of embryos express GFP in the ventral CNS (15 GFP-expressing cells on average), compared to 47% of embryos expressing GFP in the dorsal CNS (8 GFP-expressing cells on average) (*Figures 4.16, 4.20*C and *Table 4.6*C).

In addition to a role of Hh signalling in regulating *lama1* expression in the anterior CNS, I have demonstrated that elevated levels of Hh signalling cause increased *lama1* expression in the neural tube, vasculature, uro-genital region, hypochord, and the pre-somitic mesoderm at 25hpf (section 3.2.5.2). This is confirmed by my finding that a significant increase in the number of GFP-expressing cells are observed along the length of the neural tube and within the uro-genital region in embryos with increased Hh pathway activation caused by dnPKA mRNA injection (*Figure*

4.22). Increased numbers of GFP-expressing cells are also observed in the muscles, vasculature, and the PSM of dnPKA mRNA-injected embryos, although not to a significant extent. The lack of significance in the GFP up-regulation in the muscles, vasculature and the PSM of dnPKA mRNA-injected embryos could be attributed to the integration effect during BAC DNA trangenesis. Although dnPKA mRNA injection increases the ratio of Gli activator to Gli repressor (Hammerschmidt et al. 1996), it does not alter the amount of BAC DNA that integrates into the genome. Therefore, if the BAC DNA has not incorporated into the genomic DNA of cells of the muscles, PSM or the vasculature or if it is integrated in transcriptionally repressed areas (Clark et al. 1994; Bonifer et al. 1996), Gli transcription factors would not be able to activate GFP expression. The smaller size of the PSM and ventral vasculature tissues reduces the chance of BAC integration into their cells. Finally, I have previously shown that the muscle fibres resist up-regulation of *lama1* in response to Hh signalling at 25hpf, presumably due to the presence of other repressive signalling mechanisms. These repressive signalling mechanisms could be operating to prevent significant up-regulation of GFP expression in dnPKA mRNA-injected embryos. Manipulation of Hh signalling in embryos derived from the stable *lama1* GFP zebrafish transgenic line will address whether Hh signalling influences reporter gene expression in the muscles, vasculature and PSM.

4.3.4: BAC zC34A17 contains all necessary regulatory elements for zebrafish *lama1* expression

To conclude, BAC zC34A17 is likely to contain all the enhancer sequences necessary for the correct spatio-temporal initiation of *lama1* expression in the zebrafish embryo. This means that the enhancers lie within a region including 39,759 bases upstream of the *lama1* transcript, 26,741 bases downstream of the *lama1* transcript, and the intragenic sequence.

GFP is detected in the PSM at 25hpf, a tissue which requires Hh signalling for the expression of *lama1*. Both GFP and *lama1* expression is altered upon manipulation of the Hh signalling pathway. Therefore, if *lama1* expression in the PSM is directly controlled by Hh signalling, we would expect that the enhancer elements responsible for PSM expression contain functional Gli binding sites. Indeed, within zebrafish *lama1*, I have identified 3 potential Gli binding sites within intron 1, 2 within intron 2, and 1 within intron 3 (section 5.2.8).

lama1 expression is unaffected in the majority of zebrafish tissues in the absence of Hh signalling or when levels of Hh signalling are elevated. This indicates that other signalling pathways must operate to control *lama1* expression.

Altogether, these observations raise the possibility that several different enhancer elements are present in BAC zC34A17, and a combination of different transcription factors bind to these sequences to control the activation of *lama1*. As the PSM and uro-genital region down-regulate expression of *lama1* in embryos lacking Hh signalling at 25hpf, and other tissues do not, this

suggests that tissue-specific enhancers control *lama1* expression in the PSM and uro-genital region. Enhancer elements controlling *lama1* expression in other tissues of the embryo could be clustered together, or they could be spread throughout the *lama1* genomic locus.