Chapter 5

A deletion analysis to characterise the cis-regulatory sequences controlling *lama1* expression in the zebrafish

5.1: Introduction

5.1.1: Gene transcription

Gene transcription occurs in two main ways, focused initiation and/or dispersed initation (Juven-Gershon and Kadonaga 2010). Focused initation transcription begins at a single nucleotide or a cluster of nucleotides and happens in all organisms, in particular the simpler organisms. Vertebrate genes predominately have dispersed promoters, where transcription is initated from multiple weak sites over a broad region, typically 50 to 100 nucleotides (Juven-Gershon and Kadonaga 2010). These promoters are generally located within CpG islands, and their chromatin environment is marked with high levels of H3K4me3 and low levels of H3K4me1 (Heintzman et al. 2007; Heintzman et al. 2009). Many different sequence motifs have been defined at promoters, such as the TATA box. The TATA box is the best known focused core promoter motif, although it is only present in about 10% of mammalian promoters, and usually absent in dispersed promoters (Carninci et al. 2006).

Like promoters, enhancers are also cis-regulatory elements that control gene expression, although they have different characteristics. Whereas promoters initiate gene transcription and are closely associated with the transcriptional start site, enhancers are often distally located and regulate the transcription initiated from a promoter in a temporal and spatial manner (Kowalczyk et al. 2012). They can also lie downstream of the gene, or intragenically (Müller et al. 1999; Kolovos et al. 2012). Enhancers typically have about 200bp of open chromatin making them accessible to binding proteins (He et al. 2010). They can be over 100kb upstream of a transcription start site. For example, expression of *Shh* in the mouse limb is controlled by an enhancer that is about 800kb upstream of the *Shh* transcription start site, and lies within another gene (Irimia et al. 2012). Distal enhancers and their associated promoters interact when the chromatin strand forms a new, or rearranges a pre-existing, loop. This brings the enhancers, promoter, and associated transcriptional machinery into close proximity (reviewed in Kolovos et al. 2012).

The spatio-temporal expression pattern of a gene is often regulated by a mechanism known as combinatorial regulation, whereby several transcription factors act together in a transcriptional regulatory network to control gene transcription (Ravasi et al. 2010). These networks rely on the interaction of transcription factors with other transcription factors, chromatin modifiers, and co-factor proteins, which assemble together at enhancers and regulatory regions to control gene transcription (Ravasi et al. 2010; Kolovos et al. 2012). Gene expression in different tissues may be driven by different enhancers and thus controlled by different transcriptional regulatory networks. Alternatively, an enhancer may be common to several tissues, but the different molecular environment within these tissues may result in distinct spatio-temporal patterns. This can produce a fine-tuning effect on gene expression within a given tissue.

The complex and dynamic expression pattern of *lama1* during zebrafish embryonic development, coupled with the fact that Hh signalling in addition to other signalling pathways regulate *lama1*

activation, suggests that several enhancers operate to control *lama1* expression. This possibility is consistent with the fact that *Lama1* expression is unaffected in the PSM of *Shh* null mice, whereas *Lama1* expression in somites is lost (Anderson et al. 2009), suggesting that distinct enhancers operate in the PSM and in somites.

5.1.2: A recombineering approach to identify the location of enhancer sequences

I have previously shown that BAC zC34A17 contains all the enhancer sequences necessary for *lama1* expression in the embryo from its initiation at 3hpf until late embryonic stage development. In this chapter, I aim to identify the regions within this BAC that are required for *lama1* expression. The approach I used was homologous recombineering to delete large sections of DNA sequence from the BAC, followed by observation of the mosaic GFP expression pattern generated from BAC injection into the one-cell stage zebrafish embryo.

Others have previously validated the approach I took. A series of deletions within a BAC construct known to control expression of Engrailed in the midbrain-hindbrain boundary, muscle pioneer cells, and medial fast fibres of the zebrafish, resulted in the uncoupling of *eng2a* expression in the midbrain-hindbrain boundary and somite (Maurya et al. 2011). A muscle specific enhancer was identified 6kb upstream of the *eng2a* translation start site, whilst an 800bp element that controls *eng2a* expression in the midbrain-hindbrain boundary was identified 4.5kb upstream (Maurya et al. 2011).

In another case, enhancer sequences that activate *Bicc1-1* and *Pkd1L1-1* expression in the notochord of the mouse are composed of several Foxa2 transcription factor binding motifs (Tamplin et al. 2011). Expressing these motifs in the zebrafish drives expression of GFP specifically within the notochord. However, upon progressive mutation of the Foxa2 motifs, a quantifiable reduction in the number of notochord-positive embryos is observed (in addition to a weaker GFP signal intensity) (Tamplin et al. 2011). Mutation of all motifs together abolishes all notochordal GFP expression (Tamplin et al. 2011). These results indicate that deletion of DNA sequences in transient zebrafish assays can give a quantifiable effect and reveal sequences and motifs required for gene expression.

In general, the identification of enhancer elements is largely based on bioinformatic analyses, in which conserved non-coding elements (CNEs) are hypothesised to play a role in gene regulation (Haeussler and Joly 2011). However, CNEs are often found to have no role in gene activation (Ertzer et al. 2007; Haeussler and Joly 2011), and conversely, non-conserved elements can also regulate the activation state of a gene (Fisher et al. 2006). For example, although *Bicc1-1* and *Pkd1L1-1* mouse enhancer sequences have some conservation with humans, no conservation has been retained with the zebrafish. Despite this, these enhancers function in the same way in both mouse and zebrafish by specifically activating gene expression within the notochord. Therefore, mammalian enhancers that lack sequence conservation can still function (Fisher et al. 2006;

Tamplin et al. 2011), meaning that sequence conservation alone is not a sufficient indication of a cis-regulatory module or enhancer.