Chapter 7

Final discussion

7.1: A summary of results

In this study, I showed that the expression pattern of zebrafish *lama1* is conserved during evolution. *lama1* expression in tetrapods and fish is detected in the somites, PSM, anterior CNS, neural tube, eye, otic vesicles, pro-nephric tubules and the uro-genital region. In the zebrafish, *lama1* is also observed in the notochord, hypochord, heart, jaw musculature, and the pectoral fins. Accordingly, this conserved expression pattern is associated with a conservation of mechanisms that regulate *lama1*. While loss of *Shh* in the E9.5 mouse embryo causes a loss of *Lama1* expression in the somite and neural tube, with no effect in the PSM (Anderson et al. 2009), it has no effect on somitic or neural tube *lama1* expression in the zebrafish, although there is a reduction in the PSM. However, elevated levels of Hh signalling in the zebrafish are sufficient to up-regulate *lama1* expression in the somites, PSM, anterior CNS, neural tube, vasculature/pro-nephric tubules, and the uro-genital region. Therefore, in both zebrafish and mouse embryos, Shh signalling controls *lama1* expression, although its role appears more critical within the mouse embryo. The differential effect of Shh signalling on *lama1* expression in the PSM, somites and neural tube between zebrafish and mouse indicates that distinct regulatory mechanisms have evolved in the mouse PSM, somites and neural tube.

In both species, it is unknown whether Hh regulates directly *lama1* transcription. The maintenance of some *lama1* expression in zebrafish and mouse embryos lacking Hh signalling indicates that other signalling pathways contribute to *lama1* expression. The identification of the transcription factors which control zebrafish *lama1* transcription would provide more information regarding the signalling pathways that regulate *lama1* expression. In order to identify these transcription factors, a bacterial artificial chromosome (BAC) was selected and shown using a transgenesis approach to contain all the necessary enhancer elements required to completely recapitulate *lama1* expression. Like *lama1* mRNA, *lama1* GFP expression is Hh responsive, suggesting that the direct or indirect control by Hh signalling maps to regions located in the BAC. To characterise the regions of DNA that control the expression of *lama1*, homologous recombineering approaches have been performed to delete large sequences of DNA within the *lama1* gene and its surrounding upstream and downstream DNA sequences. I have found that intron 1 in combination with 3291 bases upstream of the *lama1* transcription start site is sufficient for normal *lama1* expression. Within intron 1, bases 1 to 4415 are sufficient for the expression of *lama1* in the muscle fibres, PSM and the notochord, but cannot activate expression within the anterior CNS, neural tube, eye, otic vesicle, jaw musculature, heart, vasculature/pro-nephric tubules, uro-genital region, and the pectoral fin. Bases 4416 to 9779 of intron 1 are required, possibly in combination with bases 1 to 4415 for the expression of *lama1* in these tissues.

7.2: Laminin gene transcription

Upstream of the *lama1* transcription start site, I have identified 2 potential promoter sequences, which are characterised by the presence of TATA boxes. One is located within the 878bp minimal promoter region that is included in the Δmin.prom*lama1*GFPΔ5kb construct. This minimal promoter region drives the expression of GFP reporter gene when combined with the enhancer elements located in intron 1 of *lama1*. However, whether the TATA box is actually necessary for the transcription of zebrafish *lama1* is currently unknown. The identification of the transcription factors that bind to and activate the promoters of other Laminin genes gives an insight into the signalling pathways and mechanisms that directly regulate the transcription of Laminins. To date however, there is limited information on the transcriptional control of Laminins (Aberdam et al. 2000).

7.2.1: A conserved activation of Laminin promoters

Within *Lamc1*, a highly conserved transcriptional element, Bcn-1, has been identified in the promoter at -1077 to -20bp upstream of *Lamc1* (Kawata et al. 2002). Smad3/4 activate *Lamc1* transcription when their interacting partner TFE3 (a b-HLH transcription factor that contains a leucine zipper region for dimerisation and DNA binding) binds to the Bcn-1 element (Kawata et al. 2002). It was later shown that Insulin-like growth factor-binding protein-5 (IGFBP-5) triggers the process resulting in the transcription of *Lamc1*, as IGFBP-5 treatment to rat glomerular mesangial cells leads to dephosphorylation and cleavage of Filamin A (an actin binding protein), resulting in the recruitment of Smad3/4 to Filamin A and Smad3/4 translocation to the nucleus (Abrass and Hansen 2010). *Lamc1* expression is also responsive to TGF-β, IL-1β (Suzuki et al. 1996; Dolez et al. 2011) and glucose (Phillips et al. 1999).

Interestingly, both *Lama1* and *Lamc1* promoters in human and rodents contain no TATA or CAAT boxes, and they have multiple transcription initiation sites (Niimi et al. 2003; Niimi et al. 2004; Piccinni et al. 2004). Like *Lama1*, *Lamc1* contains many GC rich motifs (O'Neill et al. 1997) similar to the Sp1 transcription factor binding site (GGGCGG), and an identical Bcn-1 transcriptional element (CCCCGCCCACCTCGCGC) (Suzuki et al. 1996). Sp1 transcription factors are zinc finger transcription factors, which are important for promoter activation in the absence of a TATA box (Briggs et al. 1986; O'Neill et al. 1997). Supporting the idea that these sequences are functional, over-expression of Sp1 in rat hepatocytes increases the expression of Laminin γ1 (Lietard et al. 1997), and Sp1 can also act synergistically with Krüppel-like factor (KLF4), which binds the Bcn-1 sequence to activate *Lamc1* transcription (Higaki et al. 2002). KLF4 also plays a role in the regulation and expression of *Lama*3A in MCF10A breast epithelial cells, and the decreased Laminin α3A expression detected in breast cancer cells correlates with a decreased KLF4 activity (Miller et al. 2001). As discussed in chapter 1, the KLF proteins also

have a role in activation of the *Lama1* promoter in various cell lines derived from *Drosophila*, mouse, and human (Piccinni et al. 2004; Niimi et al. 2006).

Taken together, these results suggest that multiple transcription factors such as Sp1 proteins, KLF, and Smad proteins regulate Laminin gene transcription in a co-operative manner, and that these can be induced by a variety of signalling pathways including TGF-β. TATA boxes are not involved in the transcription of *Lama1* or *Lamc1* in both the rodent and human, raising the possibility that the potential TATA box I identified upstream of zebrafish *lama1* is not required for *lama1* expression, or that promoter sequences have diverged during evolution to accommodate additional complexity in *lama1* regulation.

7.3: Functional conserved enhancer sequences are often located in introns

It is not uncommon for enhancer sequences to be located within the introns of a gene, in particular within intron 1 (Haeussler and Joly 2011) (*Table 7.1*). Numerous enhancer elements control the expression of *shh* in zebrafish and mice, some of which are located in introns 1 and 2, and intron 2, respectively (Epstein et al. 1999; Müller et al. 1999; Ertzer et al. 2007). The enhancer elements controlling *shh* transcription in the zebrafish function synergistically to control the precise spatial, temporal, and quantitative aspects of the *shh* expression pattern (Ertzer et al. 2007). Expression of *shh* in the floor plate is driven by the enhancer ar-B, located in zebrafish *shh* intron 1. However, correct initiation of *Shh* expression in the floor plate requires enhancer ar-C (located in intron 2 of *shh*) in addition to enhancer ar-B, despite the fact that enhancer ar-C cannot efficiently drive floor plate expression alone (Ertzer et al. 2007). Similarly, my data show that enhancers which regulate *lama1* expression in the zebrafish are spread throughout intron 1, as well as in a region that is 3291 bases upstream of the transcriptional start site, and in sequences that are likely to reside downstream of exon 2 (*Figure 5.32*).

Within the ar-C enhancer, a highly conserved region of 240 bases shares sequence similarity with the mouse SFPE2 enhancer, which drives *Shh* expression in the floor plate (Müller et al. 1999; Müller et al. 2002). This 240 base sequence drives reporter gene expression in the zebrafish with a pattern similar to that of the full ar-C sequence, that is, *shh* transcription within the notochord and zona limitans intrathalamica (Ertzer et al. 2007). This indicates that highly conserved sequences are more likely to have been subjected to selective pressure during evolution in order to preserve their transcriptional function as enhancers (Thomas et al. 2003). This suggests that the two conserved sequences identified within intron 1 of *lama1* may have enhancer function.

However, despite sequence conservation, enhancers may display divergent function. For instance, the SFPE2 sequence directs *Shh* expression in the floor plate in the mouse, whilst the 61% identical ar-C sequence mediates *shh* expression in the forebrain and notochord, and only very weakly in the floor plate of the zebrafish embryo (Ertzer et al. 2007). Similarly, zebrafish *shh*

enhancer ar-A drives expression in the notochord, yet despite significant conservation with the mouse *Shh* intron 1 (Müller et al. 2002), the mouse intron 1 sequence does not drive any *Shh* expression in the mouse (Jeong et al. 2006; Ertzer et al. 2007). Overall, studies show that variations in tissue-specific expression do occur between structurally conserved enhancers. This could result from micro-changes in the DNA sequence, which may affect transcription factor binding sites (Haeussler and Joly 2011). For example, in the mouse limb bud, *Shh* is expressed in the posterior of the limb bud in the zone of polarising activity (ZPA) and is controlled by an enhancer sequence, which is located 1Mb upstream of the transcriptional start site (Lettice et al. 2003). This enhancer sequence is highly conserved, and is also identified in *Fugu*. The homologous sequence in *Fugu* can drive *shh* expression in the ZPA of the mouse limb bud, demonstrating sequence and gene expression conservation (Lettice et al. 2003). However, a single base pair mutation in this regulatory sequence modifies the enhancer activity and results in ectopic *Shh* expression in the anterior margin of the limb bud, in both mouse and human (Lettice et al. 2003). This leads to preaxial polydactyly. Variation in tissue-specific gene expression, despite the presence of conserved enhancer sequences, is also the result of combinatorial regulatory mechanisms. This involves differences in the transcription factor micro-environment, which influence the transcription of a gene and refine gene expression patterns (Spitz and Furlong 2012). Structurally similar enhancers can also be modified by the presence or absence of other enhancers which share a similar activity (Spitz and Furlong 2012). These enhancers are known as secondary or shadow enhancers, and can function to shield the gene expression driven by the primary enhancer, from environmental perturbation such as changes in the transcription factor micro-environment (Spitz and Furlong 2012).

Currently, it is unknown whether the two conserved sequences found in intron 1 of zebrafish *lama1*, which share homology with *Fugu*, have transcriptional activity in *Fugu*. Further approaches are required to determine which tissues express *lama1* under the influence of these conserved sequences. Alternatively, as zebrafish intron 1 does not share significant homology with mouse *lama1* intron 1, it is possible that these conserved DNA sequences may not be functionally active. Instead, *lama1* enhancers could lie in non-conserved regions of DNA sequence within intron 1 (section 7.3.3).

Gene	**Species**	**Reference**
Sall1	Chicken	(Izumi et al. 2007)
Hoxb4	Mouse, *Fugu*	(Aparicio et al. 1995)
Hob2	Mouse, Chicken	(Maconochie et al. 1997)
Pax6	Mouse	(Kleinjan et al. 2004)
Shh	Zebrafish, mouse	(Müller et al. 1999)

Table 7.1: Examples of genes which contain functional enhancers within their introns, and the species that this has been identified in.

7.3.1: Multiple transcription factors could activate *lama1* enhancer elements

I have uncovered that multiple enhancer elements control the expression of zebrafish *lama1*. This, combined with the fact that *lama1* is expressed in a complex and dynamic pattern during zebrafish embryonic development, suggests that different signalling pathways co-operate together to control *lama1* transcription through combinatorial regulatory mechanisms. Supporting this statement, I found that *lama1* expression in the zebrafish embryo is only partially regulated by Hh signalling, and in the mouse embryo (Anderson et al. 2009).

Previous studies have also reported the involvement of combinatorial regulatory mechanisms in the control of other Laminin subunits. For instance, retinoic acid which controls *Lama1* expression in mouse F9 cells, is also required for the direct transcription of *Lamb1* (Vasios et al. 1989; Aberdam et al. 2000). Transient co-transfection analysis of potential upstream enhancer regions along with a heterologous thymidine kinase promoter revealed activation of the *Lamb1* gene in RA-treated F9 cells (Vasios et al. 1989). Deletion and mutagenesis experiments identified a 46-bp pair RA-responsive element (RARE) between -477bp and -432bp of the *Lamb1*. In vivo studies in mice have confirmed that this RARE site acts in a combinatorial manner with other enhancers to drive tissue-specific *Lamb1* transgene expression (Sharif et al. 2001; Sharif et al. 2004). Indeed, the 0.7kb promoter construct containing the RARE site drives expression of *Lamb1*/LacZ transgene in the pro-spermatogonia (Sharif et al. 2001; Sharif et al. 2004), whereas elements located between -1.4kb and -0.7kb drive *Lamb1* transgene expression in the developing kidney and the ovary, and cis-regulatory elements located between -2.5kb and -1.4kb drive transgene expression in the adult kidneys (Sharif et al. 2001; Sharif et al. 2004). Cis-regulatory elements between -3.9kb and -2.5kb relative to the transcription start site are also required for *Lamb1* expression in the cortex, striatum, hippocampus and thalamus of the brain (Sharif et al. 2001; Sharif et al. 2004). Therefore, tissue-specific *Lamb1* expression is driven by upstream cis-regulatory elements, acting in combination with the RARE element.

Likewise, signalling pathways may act in a combinatorial manner to provide the full expression pattern of *lama1* in the zebrafish. These may include pathways already discussed in this thesis, such as Integrin β1, FGFs, and TGF-β (Neubauer et al. 1999; Aumailley et al. 2000; Li et al. 2001; Niimi et al. 2003; Futaki et al. 2004). To identify if any of these signalling pathways directly regulate *lama1* transcription, further deletion analyses and over-expression approaches need to be carried out. Bioinformatic analyses of intron 1 of zebrafish *lama1* can aid in the identification of putative binding sites which may possess the ability to drive GFP reporter gene expression.

Using MatInspector software (www.genomatix.de), I have searched for the presence of transcription factor binding sites within intron 1 of zebrafish *lama1* (*Figure 7.1*). In the first 4415 bases of intron 1, the sequence sufficient for *lama1* expression in the muscle fibres, notochord,

and the PSM, there are potential transcription factor binding sites for factors such as MyoD, Mef2, Hand family members, Pax3, ZIC, TCF and Six3, in addition to the Gli proteins. These factors are often associated with and expressed in the somites and PSM of zebrafish and mouse embryos (Pownall et al. 2002). Although the significance of these transcription factor binding sites remains to be determined, one may speculate that they may be involved in the expression of *lama1* in these tissues, and drive expression of GFP in these tissues in Δmin.prom*lama1*GFPΔ5kb-injected embryos. There are also 2 Smad (the effectors of BMP signalling) binding sites detected within this region of intron 1 (*Figure 7.1*). BMP-mediated signalling is commonly associated with an antagonism of Hh signalling, such as in the neural tube (Patten and Placzek 2002) and during the patterning of the myotome (Maurya et al. 2011). Therefore, as Hh signalling plays a role in the regulation of *lama1* expression in the somites of the zebrafish and mouse, it may be worth examining the effect that BMP signalling has on *lama1* expression, and whether it affects *lama1* expression in response to Hh signalling. pSmads are known to associate with Gli repressor proteins and repress Hh signalling activity in the myotome, restricting the domain of Engrailed-expressing muscle pioneer and medial fast fibres (Dolez et al. 2011; Maurya et al. 2011). However, *lama1* expression in the somite is unaffected by a loss of Hh signalling or an increase in Gli repressor protein, suggesting that pSmad does not directly repress *lama1* transcription with Gli repressor. Nevertheless, inhibition of BMP signalling causes an expansion of the muscle pioneer/adaxial cells and medial fast fibre population (Maurya et al. 2011). At late somitogenesis stages, *lama1* expression in the somite is restricted to the adaxial cells, and as a result, inhibition of BMP could indirectly cause up-regulation of *lama1* expression in the somite. This intronic sequence also contains three potential retinoic acid receptor (RXR) recognition sites (*Figure 7.1*). As in the case of *Lama1* and *Lamb1* in the mouse (Sharif et al. 2001; Futaki et al. 2004; Sharif et al. 2004), retinoic acid signalling may therefore have a transcriptional role in the control of zebrafish *lama1*.

Enhancers driving *lama1* expression in the anterior CNS, neural tube, eye, otic vesicle, jaw musculature, heart, vasculature/pro-nephric tubules, uro-genital region, and the pectoral fin are located between bases 4416 to 9779 within intron 1. This region of intron 1 also contains the transcription factor binding site sequences mentioned above. The fact that both intronic sequences have a similar set of transcription factor binding sites suggests that the differential expression driven by these intronic sequences could be the result of a loss of combinatorial regulatory mechanisms, where essential enhancers in the 3' half of intron 1 are required in combination with enhancer elements in the 5' of intron 1. It is also possible that there are transcription factor binding sites not detected by my MatInspector analysis, which contribute to the control of *lama1* expression in the anterior CNS, neural tube, eye, otic vesicle, jaw musculature, heart,

vasculature/pro-nephric tubules, uro-genital region, and the pectoral fin, that are not present in the 5' half of intron 1.

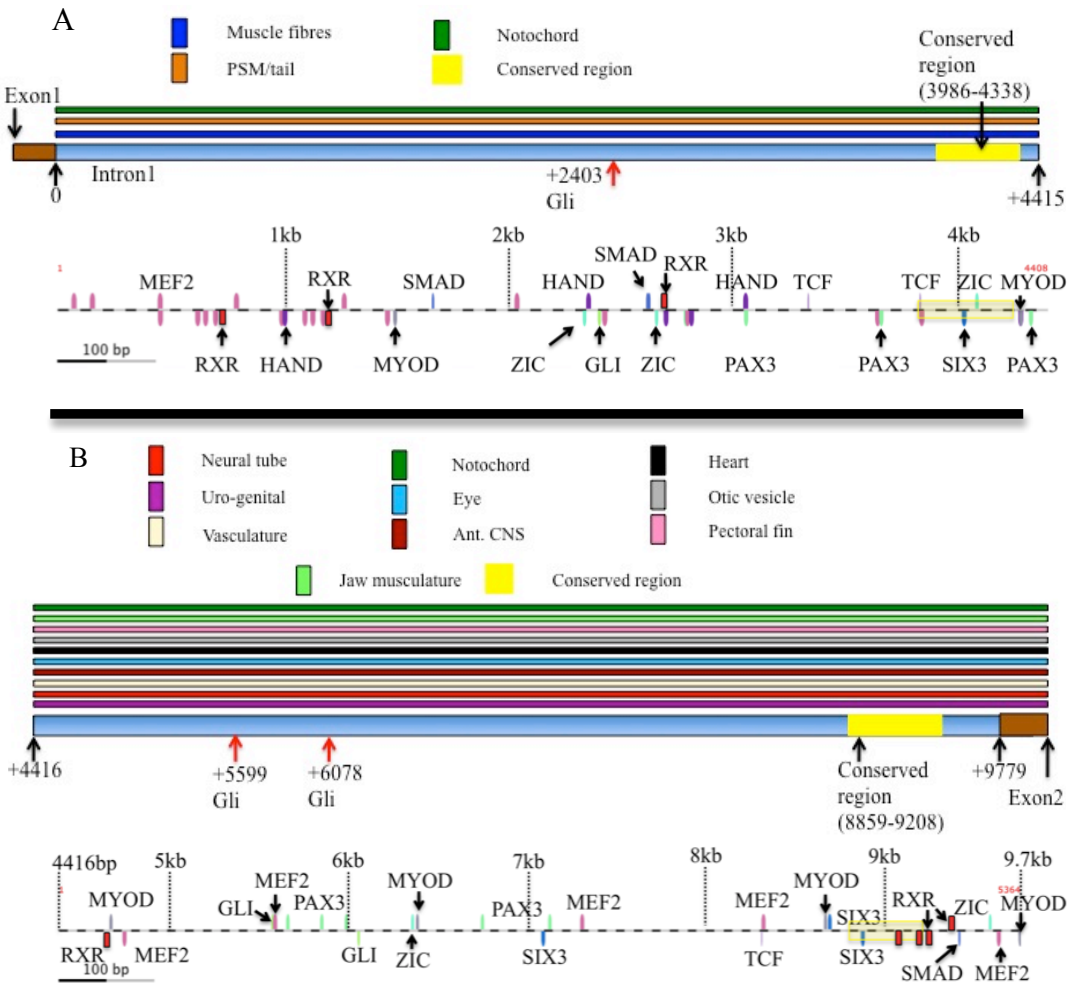

Figure 7.1: *Schematic representation of the putative transcription factor binding sites located between bases 1-4415 (A) and 4416-9779 (B) within intron 1 of zebrafish lama1*. Transcription factor binding sites are indicated with a colour-coded bar. This data was generated using MatInspector software. Above each sequence of transcription factor binding sites is a schematic showing the predicted positions of tissue-specific enhancers which control *lama1* expression in the zebrafish. A colour-coded key is shown above these schematics.

Analysis of the transcription factor binding sites present in potential enhancer elements of a gene does not always reveal which transcription factors and signalling pathways actually function to regulate the expression of the gene. Highly occupied transcription (HOT) enhancers have recently been reported in a variety of organisms, including *Drosophila* and human (Farley and Levine 2012). HOT enhancers are bound by many transcription factors, but generally lack transcription factor motifs, suggesting that the transcription factors are recruited non-specifically or via protein-protein interactions (Farley and Levine 2012; Kvon et al. 2012). In line with the early expression of *lama1* in the zebrafish embryo, most HOT enhancers are also associated with

increased transcriptional activity during early development (Farley and Levine 2012). Therefore, transcription factors not identified by MatInspector could function to regulate *lama1* transcription. Similarly, through the process of protein-protein interaction, multiple Gli proteins could be associated with the DNA sequence in intron 1 of zebrafish *lama1*. In support of this, Gli proteins are known to dimerise with other transcription factors in the zebrafish embryo, such as phosphorylated Smad1 (Maurya et al. 2011), and the Zic proteins (Koyabu et al. 2001), both of which have binding sites throughout intron 1 of *lama1*. In addition, Zic proteins can also bind to Gli binding sites, and Gli and Zic can regulate each other's cellular localisation and transcriptional activity (Koyabu et al. 2001).

7.3.2: The identification of *Lama1* enhancer sequences in the mouse

My data raise the possibility that the intronic sequences within mouse *Lama1* may also regulate the expression of *Lama1* in the mouse. In support of this statement, *lama1* in both the zebrafish and the mouse contains 63 exons, and intron 1 and intron 3 are the largest introns in both species, suggesting a degree of overall structural conservation of the *lama1* gene. Like in the zebrafish, intron 1 of mouse *Lama1* contains three putative Gli binding sites (unpublished data, Kalin Narov thesis). Hh signalling is critical for *Lama1* expression in the mouse somites and neural tube (Anderson et al. 2009). Therefore, these Gli binding sites could potentially play a role in the regulation of *Lama1* expression.

Thus, despite the lack of sequence conservation between mouse and zebrafish, intron 1 could contain enhancer sequences for both zebrafish and mouse *lama1* expression. These sequences could act in combination with the previously characterised enhancer and promoter sites that are located upstream of the *lama1* transcriptional start site (Niimi et al. 2003; Niimi et al. 2004). In particular, it is possible that Gli or other transcription factors with binding sites within intron 1 may co-operate with NF-Y, Sp1/3, Sox7/17, and YY1 transcription factors that have been shown to bind approximately -3165 to -3373 bases upstream of the *Lama1* transcriptional start site (*Figure 7.2*). However, transcription factor binding in this region is associated with differentiating mouse F9 carcinoma cells, and the differentiation of parietal endoderm tissue (Niimi et al. 2003; Niimi et al. 2004). Whether this upstream enhancer functions to express *Lama1* in vivo at the post-implantation stages of development is unknown. The functionality of the mouse intronic sequences and this upstream enhancer sequence could be tested in vivo using a mouse BAC transgenic approach. Mouse BACs could also be injected into the zebrafish embryo to test whether, despite sequence divergences, mouse cis-regulatory elements can drive reporter gene expression in the zebrafish. In this respect, it would be interesting to establish whether the pattern of reporter gene expression driven by the mouse BAC in the zebrafish resembles that of mouse *Lama1* or that of zebrafish *lama1*. BAC deletion analyses could be performed to identify the

sequences within the mouse BAC which are required for *lama1* expression in the zebrafish. Sequences deemed as important for *Lama1* expression could be validated with cell-culture techniques, or injection into the mouse embryo.

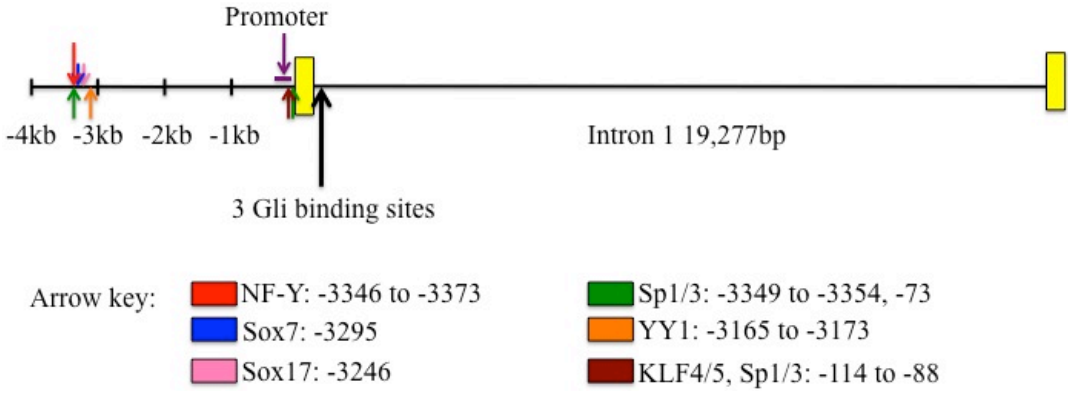

Figure 7.2: *A schematic representation of the mouse Lama1 gene, and the identified promoter and enhancer sites upstream of the Lama1 gene that control transcription.* A minimal promoter fragment -103 to -178bp upstream of the transcription start site (TSS) (horizontal purple line) drives *Lama1* expression in F9 cells, whilst KLF4/5 and Sp1 transcription factors (-114 to -73bp) bind to the *Lama1* promoter in Caco2-TC7 cells (brown arrow). The enhancers responsible for *Lama1* expression during parietal endoderm differentiation in F9 cells (-3516 to -3082) bind to Sox7 (blue arrow), Sox17 (pink arrow), NF-Y (red arrow), YY1 (orange arrow) and Sp1/3 transcription factors (green arrow). Three Gli binding site sequences are located within 153 bases of intron1 immediately after exon 1. Yellow boxes indicate positioning of exons. Intron 1 is not drawn to scale.

7.3.3: Enhancer function can be conserved despite a lack of sequence conservation

It is now well known that gene expression can be conserved despite a lack of gene sequence conservation (Fisher et al. 2006; Tamplin et al. 2011). Fisher et al. (2006) identified 13 non-coding sequences in the human *Ret* (receptor tyrosine kinase) gene that are conserved with at least three non-primate mammals. 11 out of the 13 sequences were capable of driving expression of *ret* in the zebrafish in a pattern that recapitulates endogenous zebrafish *ret* expression, despite a lack of sequence conservation between zebrafish and mammalian *Ret* (Fisher et al. 2006). Interestingly, the mammalian sequences drove reporter gene expression in cells normally not found in mammals, such as afferent neurons of the lateral line ganglia. Expression was also observed in cells within the excretory system, although these are developmentally and anatomically different between fish and mammals (Fisher et al. 2006). Therefore, the pattern of reporter gene expression derived from transgenesis studies in the zebrafish using mouse cis-regulatory elements is reminiscent of zebrafish *ret* rather than mouse *Ret*. These results indicate that either non-orthologous enhancers can function analogously, or that orthologous enhancers control *ret* expression despite evolving beyond recognition. In both instances, it is likely that the same transcription factors regulate *Ret* expression, and their binding to the teleost and mammalian enhancer elements is conserved (Fisher et al. 2006).

The fact that mammalian enhancer sequences can activate specific gene expression in the zebrafish suggests that the converse may be true, and that zebrafish sequences could activate specific gene expression in mammals, although this has rarely been tested (Müller et al. 1999; Lettice et al. 2003). I hypothesise that sequences contained within intron 1 of zebrafish *lama1* are sufficient to activate *Lama1* expression within the mouse embryo. This is the case with the zebrafish *shh* enhancer sequences located in introns 1 and 2 of *shh*, which are capable of driving *shh* expression in the notochord and floor plate of both zebrafish and mouse embryos (Müller et al. 1999).

It would be interesting to test whether sequences contained within bases 1 to 4415 of intron 1 of zebrafish *lama1* are capable of activating reporter gene expression in the muscle fibres and PSM of the mouse embryo. It would be worth examining also whether intron 1 sequence could drive expression in the mouse notochord, despite the lack of endogenous expression of *Lama1* in the notochord in E9.5 mouse embryos (Anderson et al. 2009). Similarly, the sequence 4416 to 9779 bases within intron 1 of zebrafish *lama1* could direct expression of *Lama1* within the anterior CNS, eye, otic vesicles, vasculature/pro-nephric tubules, jaw musculature, heart, and the limb of the mouse.

The isolation of DNA sequences that can drive *lama1* expression in specific tissues of the zebrafish embryo should ultimately allow the identification of the transcription factors which can bind to this given DNA sequence, through bioinformatic analyses, followed by techniques such as electromobility shift assay (EMSA) and chromatin immunoprecipitation (CHIP). Candidate regulatory proteins that may bind to the *lama1* sequence can be tested in both cases through the addition of competitive or inhibitive substrates. This will provide essential information regarding the signalling pathways and mechanisms that regulate the transcription of zebrafish *lama1*. Manipulation of these signalling pathways or transcription factors should therefore affect the expression levels of *lama1* in the zebrafish, in a tissue-specific manner.

7.4: The therapeutic potential of *lama1*

Bashful zebrafish lack Laminin α1 due to a mutation in the *lama1* gene (Stemple et al. 1996; Vihtelic et al. 2001; Pollard et al. 2006; Semina et al. 2006). This leads to the failure of notochord differentiation and a shortened body axis, due to the loss of the notochordal BM (Pollard et al. 2006). A variety of other defects are also observed in *bashful* zebrafish, including ocular abnormalities caused by degeneration of retinal BMs, and abnormal migration of axonal tracts in the eye and brain (Paulus and Halloran 2006; Semina et al. 2006). Recently, a role for *lama1* has also been established in the maintenance of the MTJ in the zebrafish, and the absence of Laminin α1 leads to detachment of muscle fibres from the MTJ and a dystrophic phenotype (Sztal et al. 2012). Laminin α1 within the MTJ is likely to be secreted from the muscle fibres, which I have

shown express *lama1*. By over-expressing *lama1* in the notochord, anterior CNS, eye, and the muscle fibres, I expect that the defects in the notochord, anterior CNS, eye, and the MTJ observed in *bashful* zebrafish would be rescued to some extent. For instance, expression of Laminin α1 could contribute to the retinal and notochordal BMs, the MTJ, and could act as an adhesive substrate for axonal tract migration.

7.4.1: Manipulating the expression of *lama1*

One of the possible applications of the research carried out in this thesis, is to apply the knowledge generated to manipulate the expression of *lama1*. However, further work is needed to delineate the boundaries of tissue-specific enhancers through further deletion analyses and transgenic approaches to test candidate enhancer sequences in the zebrafish embryo before one may consider expressing *lama1* in a tissue-specific manner. Currently, injection of the DNA sequence that includes bases 1 to 4415 of intron 1 of zebrafish *lama1* into the zebrafish is capable of activating *lama1* expression in the muscle fibres, PSM, and the notochord. Therefore, this region of DNA is a good candidate for future approaches aiming at driving the expression of *lama1* cDNA specifically within these tissues. However, because *lama1* is not detected in the muscle fibres, PSM, and the notochord after the early embryonic stages of development (Sztal et al. 2011), suggesting that the necessary regulatory proteins may not be present after this stage, such a strategy may require to provide additionally the relevant transcription factors and regulatory proteins. This may not be necessary for the enhancer sequences driving *lama1* expression in the eye, because, unlike in the muscles, *lama1* expression in the eye is not down-regulated in older embryos (Semina et al. 2006; Sztal et al. 2011).

Alternatively, once the transcription factors controlling tissue-specific *lama1* expression are identified, activation of the appropriate signalling pathways through drug treatment could provide an alternative method to activate or re-activate *lama1* expression in a tissue-specific manner. Therefore, my work opens up an exciting possibility whereby the amount of *lama1* expression and Laminin α1 protein synthesis in both the zebrafish and the mouse embryo and adult, could be modulated through manipulation of the signalling pathways which control *lama1* transcription. Re-expression of *lama1* could only be of therapeutic benefit if the gene was intact and not mutated, and the reason for the absence of Laminin α1 is due to a lack of signalling input which would normally activate *lama1* expression. Alternatively, a construct containing wild-type *lama1* cDNA could be specifically expressed within tissues of the zebrafish embryo to alleviate some of the defects associated with the *bashful* zebrafish, if it was combined with the enhancers which control tissue-specific expression.

7.4.2: Laminin α1 could be crucial to the development of muscular dystrophy treatment

The re-expression of *lama1* may be particularly important for the therapy of congenital muscular dystrophy. Congenital muscular dystrophies are characterised by muscle weakness, joint contractures and impeded motion, often a result of mutations in the skeletal muscle BM and adhesion genes (Gawlik and Durbeej 2011). In adult skeletal muscles, the BM is composed mainly of Laminin-211, Collagen Type IV, Nidogen-1 and HSPGs (Sanes et al. 1990; Miner and Yurchenco 2004). Mutation in human or mouse Laminin α2 results in congenital muscular dystrophy (CMD), a disease also modeled in *candyfloss* zebrafish, which carry a mutation in *lama2* (Hall et al. 2007). *dy/dy* mice, which carry a mutation in *Lama2* display down-regulation of Integrin α7 expression, associated with the dystrophic phenotype (Hayashi et al. 1993; Xu et al. 1994). As Integrin α7 is the main receptor for Laminin α2 (von der Mark et al. 2002) it suggests that the dystrophic phenotype is in part associated with a failure of Integrin α7β1 to signal to muscle cells in the absence of its ligand, Laminin α2. In support of this idea, humans and mice lacking Integrin α7 subunit (*Itgα7*) also develop congenital muscular dystrophy (Mayer et al. 1997; Hayashi et al. 1998). Laminin α1 can also bind to Integrin α7β1 although with a lower affinity than Laminin α2 (Talts et al. 1999). Laminin α1 is expressed at the extremities of myotubes and at the MTJ where tendon attachments are established, suggesting that Laminin α1 can regulate myoblast adhesion and myotendinous attachments (Patton et al. 1999; Sztal et al. 2012). Myoblasts have also been found to fuse with regenerating muscle fibres which re-express Laminin-111 (Bischoff 1990; Patton et al. 1999). This is in line with zebrafish data where Laminin α1 is believed to strengthen muscle fibre-MTJ attachments in the absence of *lama2* (Sztal et al. 2012). In the absence of both *lama1* and *lama2*, the muscle detachment phenotype is worsened, highlighting the compensatory effect that *lama1* has in the MTJ (Sztal et al. 2012).

Together, these results raise the possibility that Laminin α1 could compensate for the loss of Laminin α2 and to some extent maintain skeletal muscle integrity.

Laminin α1 is the chain most structurally similar to Laminin α2, with which it shares 45.9% sequence homology (Saito et al. 2003). Like Laminin α2, Laminin α1 binds Dystroglycan and Integrin α7β1 in the adult muscle through interaction with LG4 and LG1-3, respectively (Talts et al. 1999; Gawlik et al. 2010). Consistent with the hypothesis of compensatory function, *dy/dy* mice show significant improvement in muscle function and life longevity when crossed with mice over-expressing Laminin α1 in the skeletal muscles and peripheral nerves (Gawlik and Durbeej 2010). Laminin α1 can also compensate for the loss of Laminin α2 in the peripheral nerve and testis (Hager et al. 2005; Gawlik et al. 2006). Truncated forms of Laminin α1 that lack LG4-5 are also capable of partially compensating for loss of Laminin α2, indicating that binding to Integrin α7 is partly responsible for recovery of the muscular dystrophy phenotype (Gawlik et al. 2010).

In agreement with this, over-expression of truncated or full-length Laminin α1 in the *dy/dy* strain up-regulates and restores Integrin α7 levels (Gawlik et al. 2010).

Overall, results show that over-expression of Laminin α1 is capable of effectively compensating for the loss of Laminin α2, and could prove crucial for the development of muscular dystrophy therapeutics, perhaps in combination with other therapeutic treatments. However, *lama1* and Laminin α1 are not expressed in the adult muscle of zebrafish, mouse or human, and so treatment would rely on exogenous application of the protein, or gene therapy. Gene therapy could depend on the identification of enhancer elements which control muscle-specific expression of *lama1*. These could drive muscle-specific expression of a construct containing wild-type *lama1* cDNA. My data indicate that enhancers controlling muscle-specific *lama1* expression in the zebrafish are located between bases 1 to 4415 of intron 1.

Alternatively, the identification of signalling pathways and transcription factors controlling *lama1* expression in the muscle fibres, based on the isolation of muscle-specific enhancer sequences in intron 1, could provide a framework for testing drugs capable of re-expressing endogenous *lama1* in adult skeletal muscles. Chemical or drug based modulation of the signalling pathways or transcription factors controlling *lama1* expression could therefore provide a potential therapeutic strategy in the treatment of congenital muscular dystrophy.

I have already shown that Hh signalling is capable of up-regulating *lama1* expression in the muscle fibres of the zebrafish embryo, although it does so only up to the 15-somite stage. As a result, agonists of the Shh pathway are unlikely to have any effect on *lama1* expression in the adult zebrafish. However, Shh signalling appears to have a more substantial role in *Lama1* expression in muscle fibres in the mouse embryo (Anderson et al. 2009), compared to the zebrafish embryo. Perhaps up-regulation of the Hh pathway in the adult mouse could cause expression of *Lama1* in skeletal muscles. If this were the case, treatment of congenital muscular dystrophy in the mouse could involve agonists of the Shh signalling pathway, that function specifically within the muscles. Currently, it is unknown if elevated levels of Hh signalling in the mouse embryo cause up-regulation of *Lama1* expression.

Taken together, my work narrows down the region of DNA that contains the muscle-specific enhancers of *lama1* expression in the zebrafish. This could potentially be of great importance in the therapy of congenital muscular dystrophy. In this respect, I have also shown that removal of DNA sequences downstream of exon 2 may remove regulatory elements which would normally repress the expression of *lama1* in the muscle fibres. Blocking the activity of this regulatory element using either a gene therapy approach or drug-based treatments could boost the expression of *lama1* in muscle fibres.

7.5: Transgenic *lama1:GFP* zebrafish can be used to study embryonic development

The generation of a stable zebrafish transgenic line expressing *lama1:GFP* is useful for a number of reasons. It allows accurate analysis of the GFP expression pattern following the manipulation of signalling pathways thought to be involved in *lama1* regulation. An alteration of GFP expression or changes in the amount of GFP-expressing cells may not be detectable in transiently injected embryos, due to mosaicism. However, when GFP is expressed throughout the embryo such as in embryos derived from the *lama1:GFP* stable line, changes to GFP expression can be more easily observed. The stable line of *lama1:GFP* zebrafish therefore offers a useful tool to further study the regulation of *lama1*. Secondly, the stable line of *lama1:GFP* zebrafish also allows the study of specific tissue and organ development or morphogenesis. For example, studies investigating fast muscle fibre development could make use of the fact that GFP is expressed throughout these fibres, allowing for the observation of muscle fibre elongation or morphogenesis. Thirdly, crossing the stable *lama1:GFP* line with *candyfloss* zebrafish could produce a zebrafish of great benefit to the field of congenital muscular dystrophy. GFP expression within the muscle fibres could allow visualisation of the dystrophic fibres that have detached from the myotendinous junctions. Using techniques to re-express *lama1* in the muscle fibres after 24hpf should correlate with the re-expression of *lama1* GFP. If this co-incides with partial recovery of the dystrophic phenotype, then this confirms that expressing *lama1* is indeed an important therapeutic tool for congenital muscular dystrophy.