5.2: Results

5.2.1: Characterisation of the minimal promoter region of zebrafish *lama1*

Using McPromoter software (http://tools.igsp.duke.edu/generegulation/McPromoter/), and F Prom software (http://linux1.softberry.com/berry.phtml?topic=promoter), I have identified two potential promoter sequences for the zebrafish *lama1* gene, using the programmes default threshold value setting of +0.005. Both sites are characterised by the presence of a TATA box, and are located at positions -825 and -2399 bases from the *lama1* translational start site. Although one might argue that these sites are located too far upstream from *lama1*, it is unknown how big the 5'UTR is (section 3.3.1). Therefore, these potential promoter sequences may actually be closer to the *lama1* transcription start site than they appear.

I have cloned a fragment from the *lama1* BAC transgene (FL) into pCRII TOPO vector, that consists of the GFP reporter gene and 839 bases immediately upstream of the currently annotated *lama1* start site that is exon 1. This construct should contain the minimal promoter required to drive GFP in the zebrafish. If this is the case, upon injection of the minimal promoter construct (Min.prom*lama1*GFP) into one-cell stage zebrafish embryos, only very weak GFP expression, if any, should be observed at any stage of development.

Indeed, in striking contrast to the GFP expression obtained when full-length BAC zC34A17 GFP (*lama1* BAC (FL)) is injected, no specific GFP expression is detected in the zebrafish embryo at any stage of development after injection of Min.prom*lama1*GFP (*Figure 5.1*II). At the 15-17-somite stage, a few GFP positive cells are observed in only 2/55 embryos in the PSM and in the anterior CNS (*Figure 5.1*IIB,C). At 26hpf, GFP expression again was only detected in a few cells of the anterior of the embryo, in 2/76 embryos (*Figure 5.1*IIE,F).

These results represent the basal levels of GFP expression, and confirm that all tissue specific GFP expression observed after injection of the full *lama1* BAC (FL) is due to the activity of enhancers located within BAC zC34A17. Therefore, deletion analyses of BAC zC34A17 should yield tissue specific changes in GFP expression and allow the identification of putative enhancer sequences.

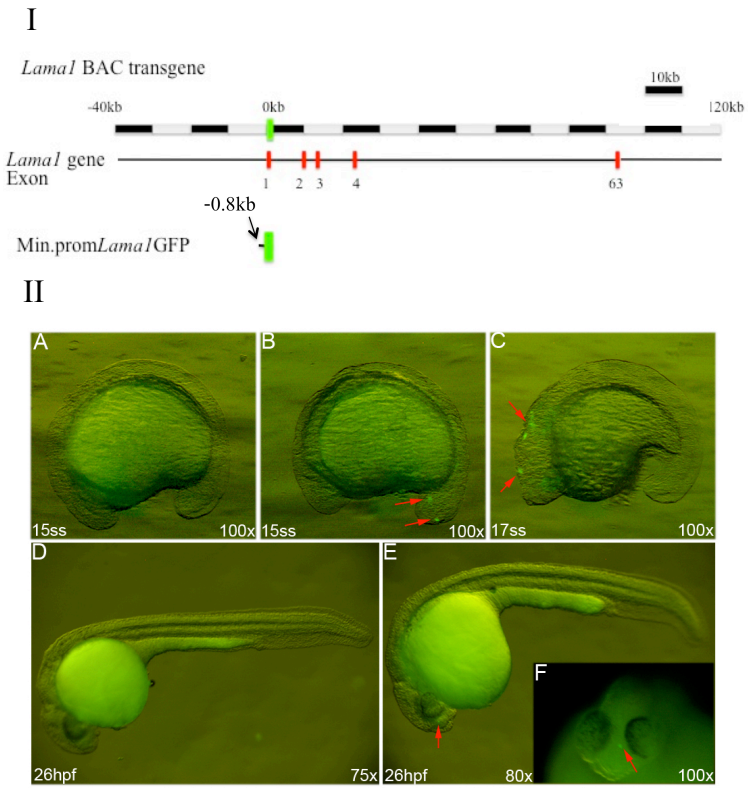

Figure 5.1: I: *A schematic representation of the lama1 gene and the minimal promoter GFP construct (Min.promlama1GFP)*. The minimal promoter construct consists of GFP and 839 bases directly upstream of the *lama1* start site. Red blocks represent exons, green block represents the GFP reporter gene. II: *Injection of 50pg Min.promlama1GFP construct into the one-cell stage embryo produces basal transcriptional levels of GFP expression*. Lateral images of embryos injected with *50pg Min.promlama1GFP* construct (A-E). Two different 15-somite stage embryos with either an absence of GFP expression (A), or just two cells expressing GFP within the PSM (red arrows) (B). A cell expressing GFP detected in the anterior CNS at the 17-somite stage (C) and at 26hpf in a lateral (E) and dorsal view (F) (marked by red arrows). The anterior of lateral views is to the left. The magnification of each image is indicated on each panel.

5.2.2: DNA alignment software helps to define conserved DNA sequences upstream of zebrafish *lama1*

Through evolution, enhancer elements have undergone purifying selection, whereby they retain the ability to bind sequence specific regulatory proteins (Thomas et al. 2003). Due to high levels of evolutionary constraint enhancer elements are often conserved. The identification of putative enhancer regions based on sequence conservation is known as phylogenetic footprinting (Tagle et al. 1988), and now often relies on in silico analysis.

In BAC zC34A17, there is 39.7kb of DNA sequence that lies upstream of the *lama1* start site, and 27.7kb of DNA sequence downstream of the last exon (exon 63). I hypothesise that enhancer regions controlling zebrafish *lama1* expression are located within these upstream and downstream sequences of DNA.

Using bioinformatics to perform multiple alignments of the *lama1* locus in different species, (ecrbrowser.dcode.org) (genome.lbl.gov/vista/customAlignment.shtml), I searched for conserved

sequences of DNA, which could harbour enhancer activity. Sequences of DNA which share 70% homology between two species over 100bp are represented with red peaks at the appropriate DNA sequence location. I identified these red peaks upstream and downstream, as well as in the introns and exons of *lama1* (*Figure 5.2a, 5.2b, 5.2c*). Analysis of the DNA sequences represented by these red peaks upstream of the *lama1* translational start site reveals that the majority of the red peaks correspond to repeat DNA sequences, and thus are unlikely to be enhancers. This includes all of the conserved regions identified in the alignments against *Xenopus*, chicken, opposum, human, cow, monkey and mouse (*Figure 5.2a*). However, two regions that are 400bp and 348bp long located at -1157 and -2408 bases upstream of the *lama1* translational start site (green arrow), respectively, contain DNA sequences conserved with the *Fugu* fish (over 70% homology over 100bp) (red arrows in *Figure 5.2a, left hand side*).

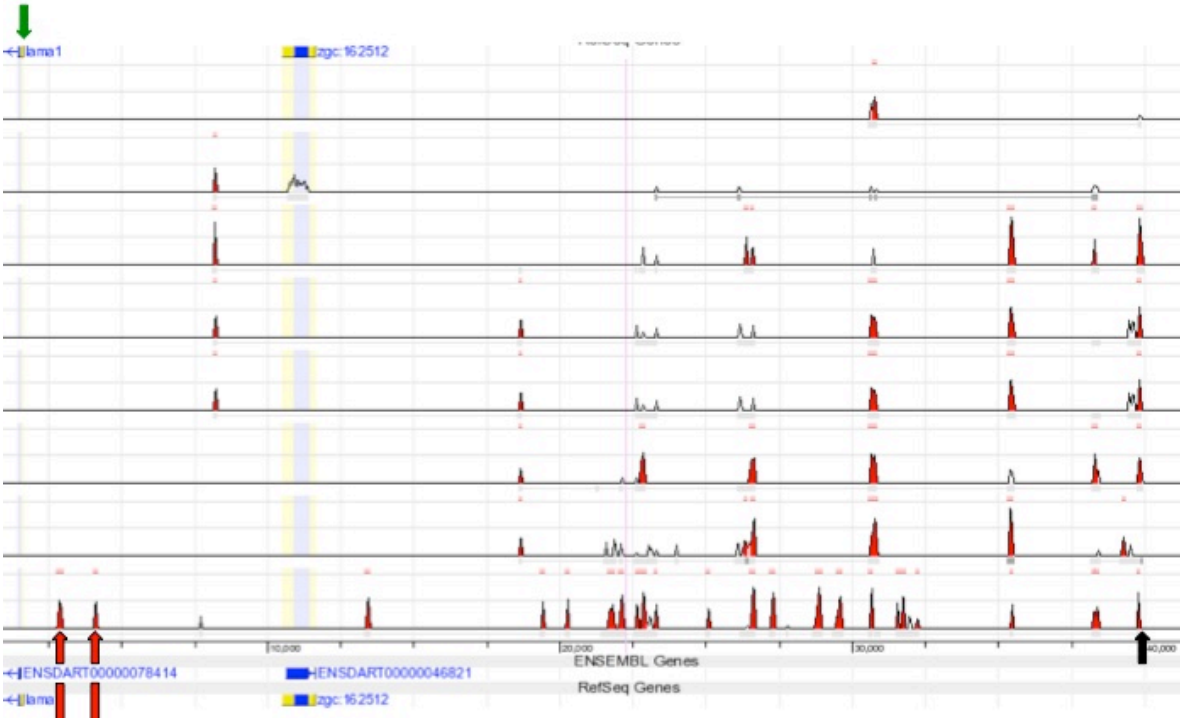

Figure 5.2a: *A screen shot from ECR Browser displaying conserved DNA sequences upstream of the lama1 start site* (green arrow). BAC zC34A17 has 39,759 bases upstream of the *lama1* translational start site, the limit of which is marked by the black arrow (bottom right of the image). Zebrafish *lama1* sequence is aligned against, from top to bottom, *Xenopus*, chicken, opossum, human, cow, monkey, mouse, and *Fugu* fish. Areas of significant conservation (100bp over 70% homology) are marked with red peaks. Sequences conserved with *Fugu* are marked by red arrows, and are located at +1157 and +2408 bases upstream of the zebrafish *lama1* start site (left of the image).

Alignment of zebrafish and *Fugu* sequences downstream of the *lama1* coding sequence reveals only one site of conservation, which is 1.4kb downstream of the last exon in zebrafish *lama1* (exon 63), and 102 bases long (*Figure 5.2b*). The rest of the downstream sequence that is present within BAC zC34A17 does not contain any DNA sequence with more than 70% homology over 100bp, when aligned against *Fugu*. This is shown by the lack of peaks in the alignment (*Figure 5.2b*). Alignment of zebrafish and *Fugu lama1* intragenic sequence using VISTA does not reveal any conservation within the introns, although 18 exons are identified as having over 70% homology (*Figure 5.2c*) (genome.lbl.gov/vista/customAlignment.shtml). The exon numbers to which each of these conserved peaks corresponds to are annotated in *Figure 5.2c*. ECR browser however identifies 2 regions of DNA which share homology between zebrafish and *Fugu*, within intron 1 of zebrafish *lama1* (section 5.2.8). Overall, these sequence alignments are not highly informative about the presence of potential enhancer elements, although the sequence that is -3.3kb upstream, and the sequence that is 1.4kb downstream of the *lama1* transcript could harbour some enhancer activity.

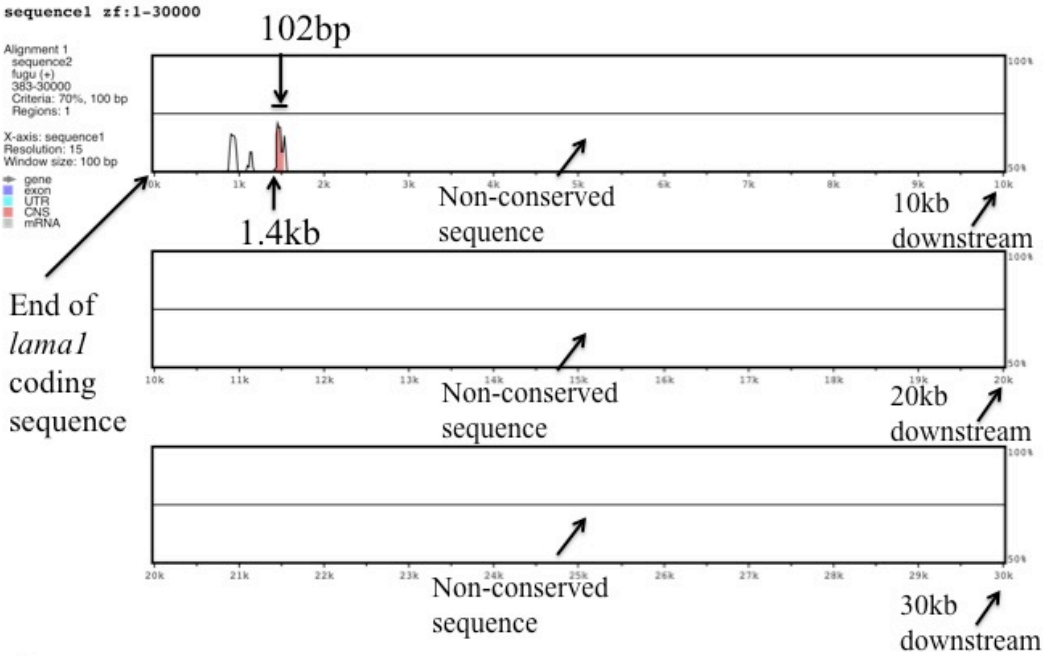

Figure 5.2b: *A screen shot from VISTA alignment displaying regions of zebrafish DNA downstream of lama1 with homology to sequences downstream of Fugu lama1.* Only one red peak is identified within 30kb of downstream sequence, corresponding to DNA sequence with over 70% homology over 100bp, and this is 102bp in length and located 1407bp downstream of exon 63 of *lama1*.

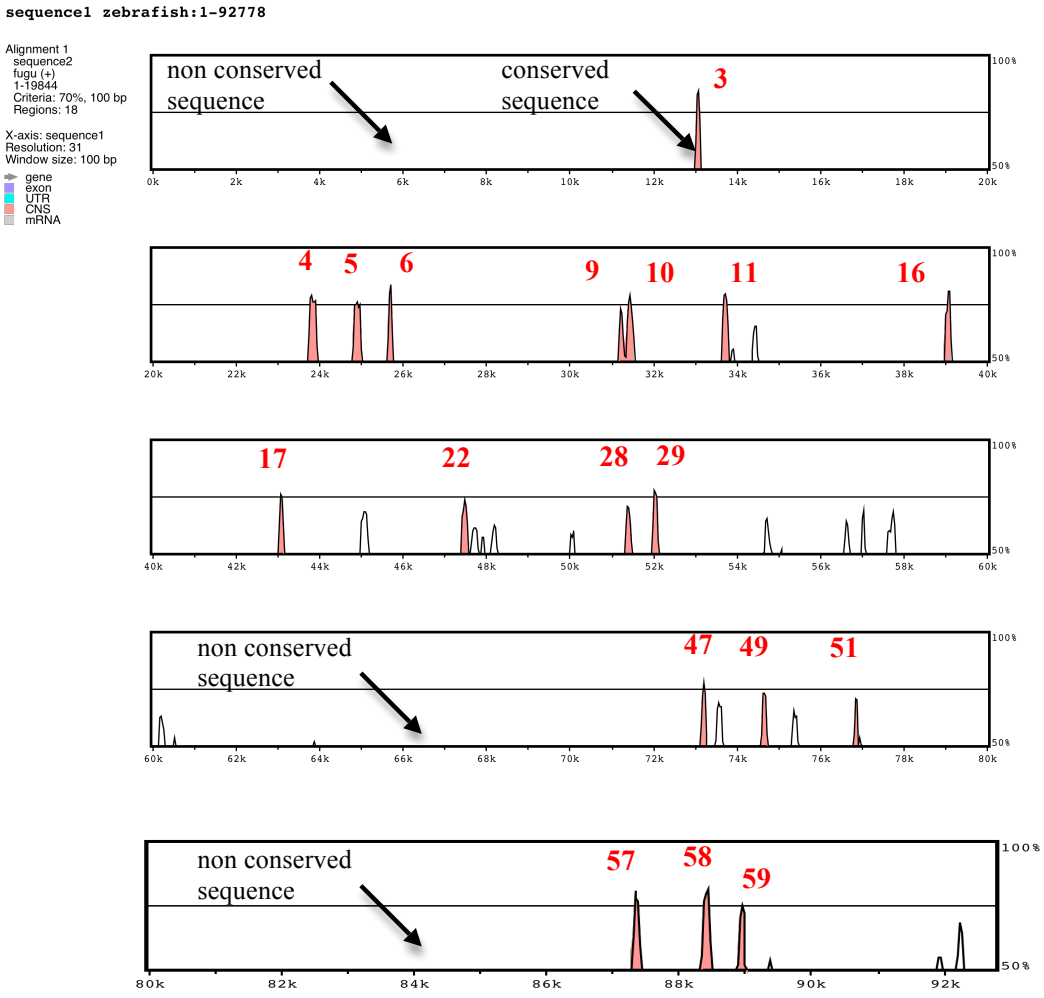

Figure 5.2c: *A screen shot from VISTA alignment displaying intragenic regions of zebrafish lama1 sequence which have homology to Fugu lama1*. Over 92kb of sequence was aligned and 18 regions were identified as having more than 70% DNA sequence homology over 100bp, with each one corresponding to an exon of *lama1*. Red numbers indicate the exon number of zebrafish *lama1* that each conserved peak has homology to.

5.2.3: A homologous recombineering strategy to delete specific regions of DNA sequences within BAC zC34A17

Ultramer homology arms that target sequences of BAC zC34A17 DNA that are long distances apart from each other allow for the deletion of all BAC zC34A17 DNA sequences that lie between the ultramer homology arms (*Figure 2.2* and *5.3*). Deletion analyses to identify enhancers controlling zebrafish *lama1* expression can be performed using this approach.

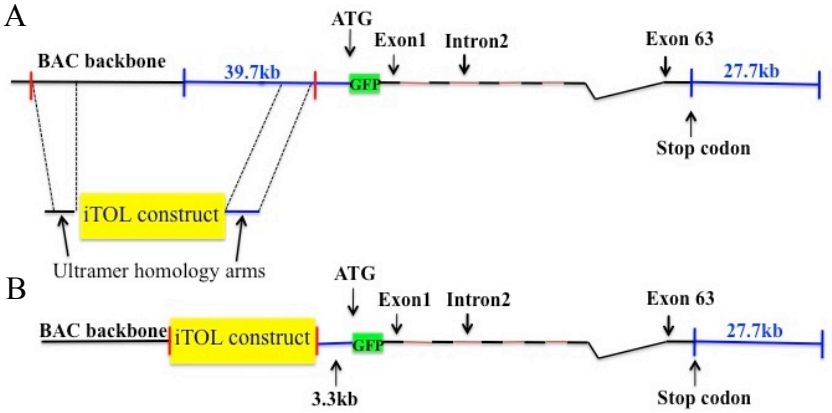

Figure 5.3: *A schematic representation of the process of deleting regions of DNA by homologous recombineering*. A: For upstream DNA sequence deletions, ultramer arms which have homology to sequences on either side of the region intended to be deleted are added to the iTOL2-AMP construct by PCR. The forward ultramer shares homology with the BAC backbone (vector) DNA sequence, whilst the reverse ultramer is homologous to the upstream DNA sequence. Reverse ultramers are designed against sequences that are -3291 bases upstream of the *lama1* start site (section 5.2.4), or sequences that are -878 bases upstream (section 5.2.10). B: After homologous recombineering, the iTOL2-AMP construct replaces all DNA sequence between the two regions that share homology with the ultramer arms. Depicted is the deletion that leaves -3291 bases upstream of the *lama1* start site (section 5.2.4). Deletions downstream of the *lama1* start site are performed in the same way, except that a Kanamycin resistance gene is inserted instead of iTOL2-AMP. This allows for the selection of colonies correctly recombined with the Kanamycin gene that already contain upstream deletions with the iTOL2-AMP insertion.

5.2.4: The region located between –39,759 to -3291 bases upstream of the *lama1* start site does not contain enhancer sequences

Phylogenetic foot printing analysis indicated the possibility that two putative enhancers (*Figure 5.2a*) are located within 3.3kb upstream of the *lama1* start site, but no other conserved sequence was found between -3.3kb and -39.7kb. Therefore, my first deletion aimed at eliminating sequences between -3.3kb and -39.7kb to verify whether any transcriptional activity was associated with this DNA region. The new construct called Δ3.3kb*lama1*GFP contained also the intragenic and downstream sequences (see *Figure 5.5*I). Following homologous recombineering of the iTOL2-AMP construct with the BAC zC34A17-GFP reporter construct, PCR testing was performed to ensure that a successful recombination and deletion of upstream DNA sequence had occurred.

PCR testing on BAC minicultures and a BAC midiprep (*Figure 5.4*) confirms successful deletion of the upstream DNA sequence, and that the GFP reporter gene and downstream DNA sequence remain intact.

	CmF	1
Primer pair	CmF, Upstream reg.2 R	Lama1 F1, Lama1 R1
Size (bases)	1638	1200

Table 5.1: *Primers used for BAC PCR analysis*: Primer pair CmF consists of a forward primer with homology to the Chloramphenicol resistance gene in the BAC backbone (vector DNA), and a reverse primer that lies just downstream of the reverse ultramer homology sequence. A 1638 base band is produced when homologous recombineering has been successful. Primer pair 1 consists of primers that lie either side of the GFP reporter gene, and produce a 1200 bases fragment.

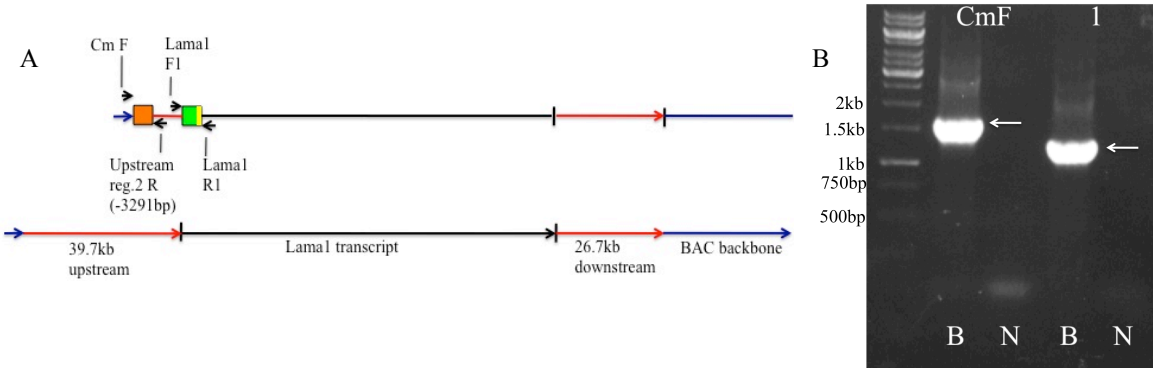

Figure 5.4: *PCR analysis of Δ3.3kblama1GFP BAC midi prep DNA confirms successful deletion of the upstream DNA sequence.* A: A schematic representation of the BAC construct and primer positions. B: Primer pair CmF is the expected size and only works if the upstream DNA sequence has been removed. Primer pair 1 confirm the presence of an intact GFP reporter gene. Primers used for each reaction, and the expected fragment size (indicated by white arrows) are listed in *Table 5.1*. Size of ladder is indicated on the left-hand side. Abbreviations: BAC DNA (B), no DNA control (N).

50pg BAC Δ3.3kb*lama1*GFP was injected at the one-cell stage. At the 15-19-somite stage, BAC construct Δ3.3kb*lama1*GFP shows no change in GFP expression compared to the full-length BAC (*lama1* BAC transgene (FL)) (*Figure 5.5*IIA,E). GFP positive cells are observed in the telencephalon and diencephalon of the anterior CNS, (*Figure 5.5*IIB,F) eye (*Figure 5.5*IIB,F), otic vesicles (*Figure 5.5*IIB), notochord (*Figure 5.5*IIB,C,F,G), muscle fibres (*Figure 5.5*IID,F,G) and the PSM (*Figure 5.5*IID,F) as in the full length BAC (*Figure 5.5*IIA,E). By 25hpf, Δ3.3kb*lama1*GFP BAC drives GFP expression in all tissues reported above (*Figure 5.6*C,Ci), as well as in the floor plate (*Figure 5.6*C,Ci), the ventral vasculature, the uro-genital region, and the epidermis (*Figure 5.6*D). This fully recapitulates the GFP expression observed in *lama1* BAC transgene (FL)-injected embryos (*Figure 5.6*A,B). As expected, GFP is also initiated in the pectoral fin and forming jaw muscles/branchial arches at 49hpf (*Figure* 5.7Ci,D). GFP expression is also clearly observed in the heart (*Figure 5.7*Ci). Finally, GFP expression is detected in the telencephalon and midbrain-hindbrain boundary of the anterior CNS (*Figure 5.7*Ci,D), the muscle fibres (*Figure 5.7*C,E), the eye, the otic vesicles (*Figure 5.7*Ci), the ventral vasculature and the uro-genital region at 49hpf (*Figure 5.7*E). Even at 97hpf, GFP is maintained in the muscle fibres, eye, otic vesicles, anterior CNS, and in the vasculature (*Figure 5.8B,C*), as in *lama1* BAC (FL)-injected embryos (*Figure 5.8*A). Therefore, the GFP expression pattern derived from transgenic zebrafish injected with Δ3.3kb*lama1*GFP BAC fully recapitulates the reporter gene expression

observed with the full length BAC. Results indicate that the DNA sequence between –39,759 and -3291 bases upstream of the *lama1* start site is not required for any spatial or temporal aspects of *lama1* expression in the zebrafish. This finding is in line with the absence of conserved DNA sequences in this upstream region and therefore suggests that enhancers are located downstream of -3.3kb.

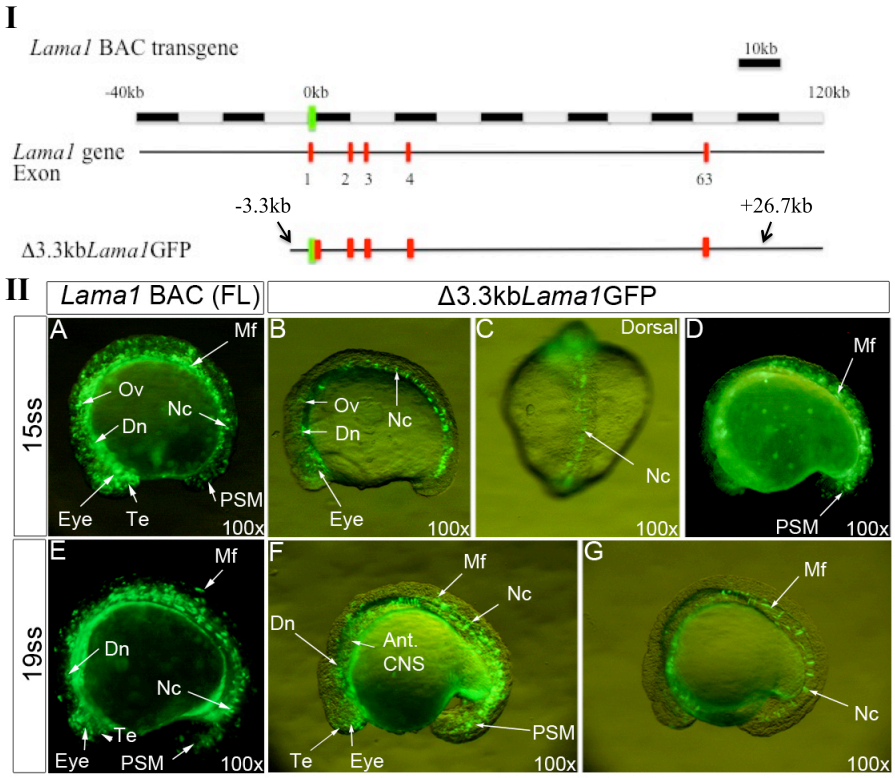

Figure 5.5: I: *A schematic representation of the lama1 gene and the Δ3.3kblama1GFP construct*. The Δ3.3kb*lama1*GFP construct consists of -3291 bases upstream of the *lama1* start site, the GFP reporter gene, and all sequences downstream of the *lama1* start site. Red blocks represent exons, green block represents the GFP reporter gene. II: *Analysis of GFP expression from 15-19-somite stages after injection of 50pg Δ3.3kblama1GFP DNA at the one-cell stage*. GFP expression was compared to *lama1* BAC full-length (FL)-injected embryos (A,E). In 15-somite stage embryos injected with *Δ3.3kblama1GFP DNA,* GFP is detected in the eye (B), diencephalon (B), otic vesicle (B), notochord (B,C), muscle fibres (D), and the PSM (D), recapitulating the expression observed in *lama1* BAC (FL)-injected embryos (A). At the 19-somite stage, *Δ3.3kblama1GFP DNA*-injected embryos express GFP in a pattern matching that observed in *lama1* BAC (FL)-injected embryos (E). GFP is expressed in the anterior CNS in the telencephalon and diencephalon (F), eye (F), muscle fibres (F,G), notochord (G), and the PSM (F,G). The anterior of lateral views is to the left. The anterior of dorsal views is at the top. The magnification of each image is indicated on each panel. Abbreviations: diencephalon (Dn), otic vesicle (Ov), notochord (Nc), muscle fibre (Mf), pre-somitic mesoderm (PSM), telencephalon (Te), anterior (Ant.)

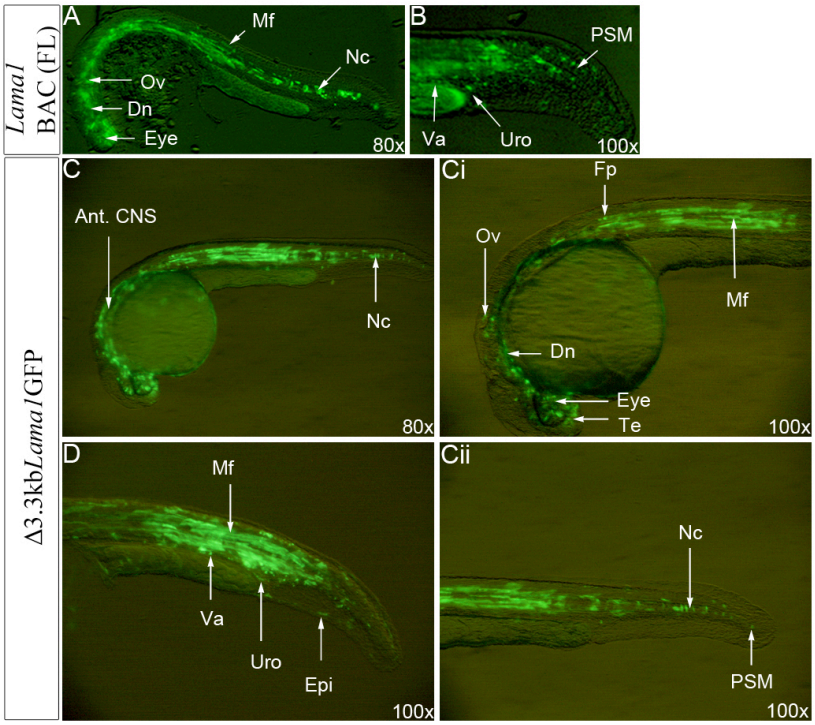

Figure 5.6: *Analysis of GFP expression in 25hpf embryos after injection of 50pg Δ3.3kblama1GFP construct DNA at the one-cell stage*. GFP expression was compared to embryos injected with *lama1* BAC (FL) (A,B). Embryos injected with Δ3.3kb*lama1*GFP DNA express GFP throughout the anterior CNS including the telencephalon, diencephalon (C,Ci), eye (C,Ci), otic vesicles (Ci), floor plate (Ci), muscle fibres (Ci,D), and the notochord (C,Cii). A separate embryo shows expression in the muscle fibres, epidermis, vasculature/pro-nephric tubule region, and the uro-genital region (D). Anterior is to the left. Magnification is indicated. Abbreviations: anterior (Ant.), notochord (Nc), telencephalon (Te), diencephalon (Dn), otic vesicles (Ov), floor plate (Fp), muscle fibres (Mf), vasculature (Va), uro-genital region (Uro), epidermis (Epi), pre-somitic mesoderm (PSM).

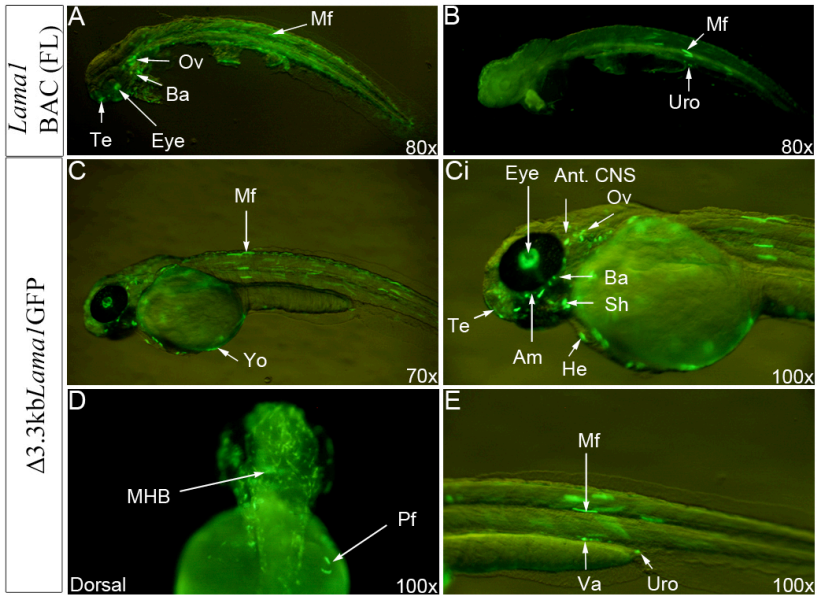

Figure 5.7: *Analysis of GFP expression in 49hpf embryos after injection of 50pg Δ3.3kblama1GFP DNA at the one-cell stage*. GFP expression was compared to embryos injected with *lama1* BAC (FL) (A,B). Embryos injected with Δ3.3kb*lama1*GFP DNA express GFP in the muscle fibres (C,E), yolk sac (C), telencephalon (Ci), anterior CNS (Ci), eye (Ci), branchial arches and forming jaw muscles (Am and Sh) (Ci), otic vesicles, (Ci) heart (Ci), vasculature (E) and the uro-genital region (E). A dorsal view also reveals GFP expression in the midbrain-hindbrain boundary and pectoral fins (D). Anterior of lateral views is to the left. Anterior of dorsal view is at the top. Abbreviations: muscle fibres (Mf), yolk (Yo), telencephalon (Te), anterior (Ant.), otic vesicle (Ov), branchial arch (Ba), heart (He), sternohyoideus (Sh), adductor mandibulae (Am), midbrain-hindbrain boundary (MHB), pectoral fin (Pf), vasculature (Va), uro-genital region (Uro).

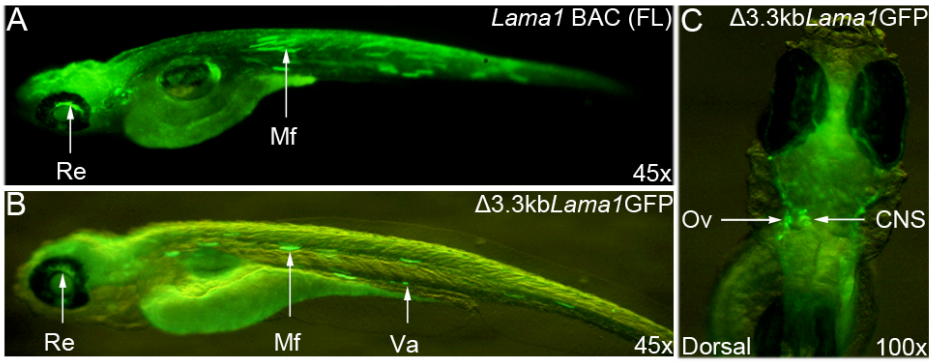

Figure 5.8: *Analysis of GFP expression at 97hpf after injection of 50pg Δ3.3kblama1GFP construct DNA into the one-cell stage embryo*. GFP expression was compared to embryos injected with *lama1* BAC (FL) (A). Lateral view embryos injected with Δ3.3kb*lama1*GFP DNA express GFP in the retina (B), muscle fibres (B), and vasculature (B), recapitulating the GFP expression pattern generated in embryos injected with *lama1* BAC (FL) (A). A dorsal view of another embryo shows GFP in the otic vesicles and CNS. The anterior of lateral views is to the left of each panel. The anterior of the dorsal view is at the top of the panel. Magnification is indicated. Abbreviations: retina (Re), muscle fibres (Mf), otic vesicles (Ov), central nervous system (CNS), vasculature (Va).

5.2.5: The DNA sequence -3291 bases upstream of the *lama1* start site has only very weak enhancer activity

To investigate whether enhancers controlling *lama1* expression are situated within the DNA sequence that lies directly -3291 bases upstream of the *lama1* start site, I cloned this region along with the GFP into the pCRII TOPO vector (Invitrogen). Injection of this construct (3.3kb*lama1*GFP-TOPO) into the one-cell stage embryo results in no detectable GFP expression at the 6-somite stage (n = 39 embryos). By the 20-somite stage, GFP can be observed in a few cells of the trunk in 10/58 embryos and in the anterior CNS in 3.4% embryos (n = 2/58) (*Figure 5.9*IIB,C) similar to the expression observed in Min.prom*lama1*GFP construct injected embryos (4% embryos express GFP in just a single cell). However, unlike Min.prom*lama1*GFP-injected embryos, specific GFP expression is detected in the muscle fibres (2/58 embryos, 3%) and in notochord cells (1/58 embryos) at the 20-somite stage (*Figure 5.9*IIC). At 25hpf, muscle fibre expression is detected in 9/36 embryos (25%) although only a few fibres express weak GFP per embryo (*Figure 5.9*IID,E). A few GFP fluorescent cells are also observed within the anterior CNS region of the embryo (4/36 (11%) embryos) and the trunk region (7/36 (19%) embryos) (*Figure 5.9*IID,E). Again, the intensity of this GFP fluorescence is extremely low and can only be seen when highly exposed (red arrows) (*Figure 5.9*IID). In contrast, embryos injected with Min.prom*lama1*GFP DNA only have a few fluorescent cells in 2/76 embryos (3%), at 26hpf (*Figure 5.1*IIE).

These results suggest that the region located -3291 bases upstream of the *lama1* start site is not sufficient for the full activation of *lama1*, although it may harbour some weak enhancer activity compared to the Min.prom*lama1*GFP construct. In particular, the construct 3.3kb*lama1*GFP-

TOPO was not able to activate reporter gene expression in early embryonic stages as does *lama1* BAC transgene (FL). However, 3.3kb*lama1*GFP-TOPO could drive some GFP expression at later stages (from the 20-somite stage), in particular in muscle and notochord cells. Nevertheless, as this expression was significantly weaker than that observed with *lama1* BAC transgene (FL), I conclude that the *lama1* enhancers are located downstream of the *lama1* or within the introns of the gene.

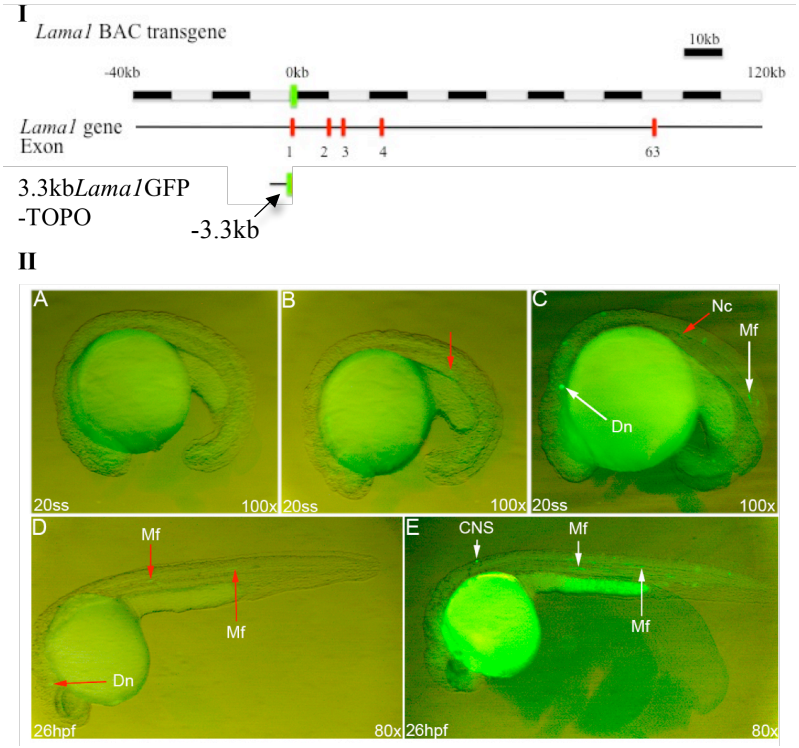

Figure 5.9: I: *A schematic representation of the lama1 gene and the Δ3.3kblama1GFP-TOPO construct.* The Δ3.3kb*lama1*GFP-TOPO construct consists of -3291 bases directly upstream of the *lama1* start site and the GFP reporter gene. Red blocks represent exons, green block represents the GFP reporter gene. II: *Analysis of GFP expression after injection of 50pg Δ3.3kblama1GFP-TOPO DNA at the one-cell stage*. A-C: Lateral views of 20-somite stage embryos at 100x magnification showing no GFP expression (A), or a single cell with weak GFP fluorescence (B) (red arrow), or a couple of cells expressing GFP in the anterior CNS with some GFP expression in a few muscle fibres (C). Weak GFP is also detected in the notochord (C) (red arrow). D and E: Lateral views of 26hpf embryos at 80x magnification show a single GFP fluorescent cell in the diencephalon region (D) (red arrow), in addition to two muscle fibres which express GFP weakly (D) (red arrows). A highly exposed image of a separate embryo reveals GFP within several muscle fibres, and in a single cell in the anterior CNS region (E). Anterior is to the left. Abbreviations: notochord (Nc), diencephalon (Dn), muscle fibres (Mf), central nervous system (CNS).

5.2.6: Enhancers controlling *lama1* are not located downstream of the *lama1* gene

Sequences upstream of zebrafish *lama1* are not required for *lama1* activation. Therefore, I next tested whether enhancers controlling *lama1* expression were located downstream of the gene. VISTA alignment identified one putative enhancer region 1.4kb downstream of *lama1* (*Figure 5.2b*), and to test whether it has enhancer function, I planned to delete downstream DNA sequences in the construct Δ3.3kb*lama1*GFP using homologous recombineering to insert a Kanamycin resistance gene, isolated from pCRII TOPO vector, in place of the DNA sequences 3' of the *lama1* gene. A loss of GFP expression in some or all tissues expressing *lama1* in the zebrafish embryo would be an indication that regulatory elements have been lost. Analysing GFP expression at different stages of development would also allow me to assess whether the initiation or maintenance of the gene is altered upon downstream DNA sequence deletion.

PCR analysis confirms the successful deletion of the DNA sequences downstream of the last *lama1* exon (exon 63) and the absence of the region upstream of -3291bp (*Figure 5.10* and *Table 5.2*). The GFP reporter gene is still present and intact, as is the sequence encoding the exons and introns.

	Ex63	**12.7**	**37/38**	**1**
Primer pair	Ex63 F, BAC backbone Rev	12.7kb up F, 12.7kb up R	Ex 37/38 F, Ex 37/38 R	Lama1 F1, Lama1 R1
Size (bases)	1369	501 (should not work)	985	1200

Table 5.2: *Primers used for BAC PCR analysis.* Primer pair Ex63 consists of a forward primer that binds to the last exon of *lama1* (exon 63), and a reverse primer that lies just downstream of the reverse ultramer homology sequence in the BAC backbone (vector DNA). A 1369 base band is produced when homologous recombineering has been successful. Primer pair 12.7 amplify a sequence located 12.7kb upstream of the *lama1* start site. This region of DNA is absent in Δ3.3kb*lama1*GFPΔex63 BAC construct DNA. Primer pair 37/38 amplify a sequence that spans exon 37, intron 37, and exon 38, and should be present in Δ3.3kb*lama1*GFPΔex63 BAC DNA. Primer pair 1 consists of primers that lie either side of the GFP reporter gene, and produce a fragment that is 1200 bases.

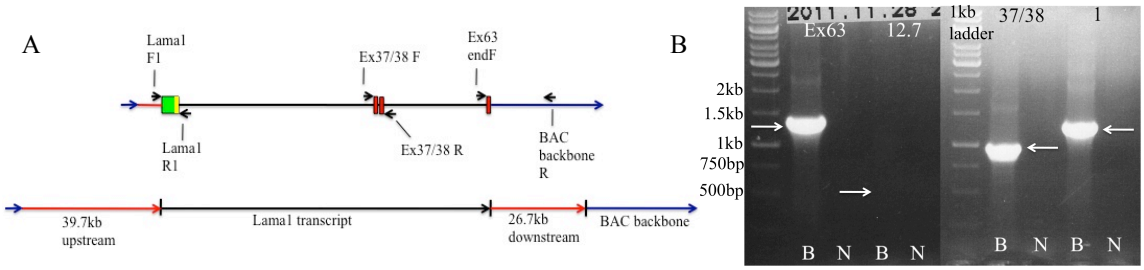

Figure 5.10: *PCR analysis of Δ3.3kblama1GFPΔex63 BAC midi prep DNA confirms successful deletion of DNA sequence downstream of exon 63.* A: A schematic representation of the BAC construct and primer positions. B: Primer pair Ex63 is the expected size and only works if the DNA downstream of exon 63 has been removed. Primer pair 12.7 confirms the absence of upstream DNA sequence. Primer pair 37/38 and primer pair 1 confirm the presence of intact exon-intron sequence and the GFP reporter gene, respectively. Primers used for each reaction, and the expected fragment size (indicated by white arrows) are listed in *Table 5.2*. Size of ladder is indicated on the left-hand side. Abbreviations: BAC DNA (B), no DNA control (N).

Deletion of the sequences upstream of the *lama1* start site, between -39,759 and -3291, and of all sequences downstream of the *lama1* coding sequence (Δ3.3kb*lama1*GFPΔex63) has no effect GFP expression. At the 15-19-somite stage, strong GFP expression continues to be observed throughout the anterior CNS, in the telencephalon and diencephalon (*Figure 5.11*IIB,E), eye (*Figure 5.11*IIB,E), otic vesicles (*Figure 5.11*IIB), notochord (*Figure 5.11*IIA,D,E), muscle fibres (*Figure 5.11*IIA,B,C,E), and the PSM (*Figure 5.11*IIB,E).

By 25hpf, expression is identical to that observed upon injection with *lama1* BAC transgene (FL) (*Figure 5.6*A,B). GFP is detected in the telencephalon and diencephalon of the anterior CNS, the eye, otic vesicles (*Figure 5.12*A,Aii), the floor plate of the anterior CNS and neural tube (*Figure*

*5.12*Aii), the somitic muscle fibres and the notochord (*Figure 5.12*A,Ai). Expression is also observed in the ventral vasculature region (*Figure 5.12*B).

At 49hpf, GFP expression is strongly maintained in the tissues mentioned above, as well as in the uro-genital region and the epidermis (*Figure 5.13*A,B,C,D,F). GFP is also initiated on time in the pectoral fins and the forming jaw muscles (*Figure 5.13*C,E). GFP expression in the heart is initiated with a similar timing to zebrafish injected with *lama1* BAC transgene (FL) (*Figure 5.13*A).

Therefore, GFP expression driven by the construct Δ3.3kb*lama1*GFPΔex63 presents a spatio-temporal pattern in all tissues identical to that of the GFP pattern generated with the *lama1* BAC transgene (FL). This demonstrates unequivocally that the enhancers controlling *lama1* expression are not located downstream of the *lama1* gene.

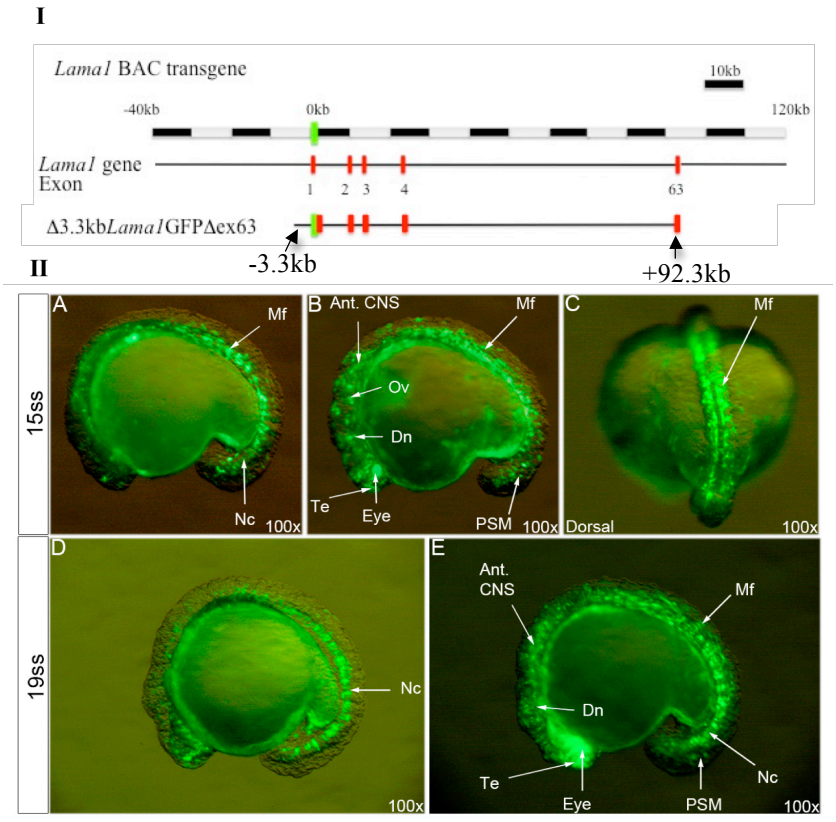

Figure 5.11: I: *A schematic representation of the lama1 gene and the Δ3.3kblama1GFPΔex63 construct*. The Δ3.3kb*lama1*GFPΔex63 construct consists of -3291 bases upstream of the *lama1* start site, the GFP reporter gene, and all the introns and exons of *lama1*. DNA sequences downstream of exon 63 have been deleted. Red blocks represent exons, green blocks represent the GFP reporter gene. II: *Analysis of GFP expression at the 15-19-somite stage after injection of 50pg Δ3.3kblama1GFPΔex63 BAC construct DNA into the one-cell stage embryo*. A-C: Lateral (A,B) and dorsal (C) views of 15-somite stage embryos at 100x magnification showing GFP expression in the muscle fibres (A,B,C), notochord (A), telencephalon and diencephalon of the anterior CNS (B), the eyes (B), otic vesicles (B), and the PSM (B). D and E: Lateral views of 19-somite stage embryos at 100x magnification showing strong expression of GFP in the notochord (D,E), telencephalon and diencephalon of the anterior CNS (E), eyes (E), muscle fibres (E) and the PSM (E). Anterior is to the left. Anterior of dorsal view image is at the top. Abbreviations: muscle fibre (Mf), notochord (Nc), telencephalon (Te), diencephalon (Dn), otic vesicle (Ov), anterior (Ant.), pre-somitic mesoderm (PSM).

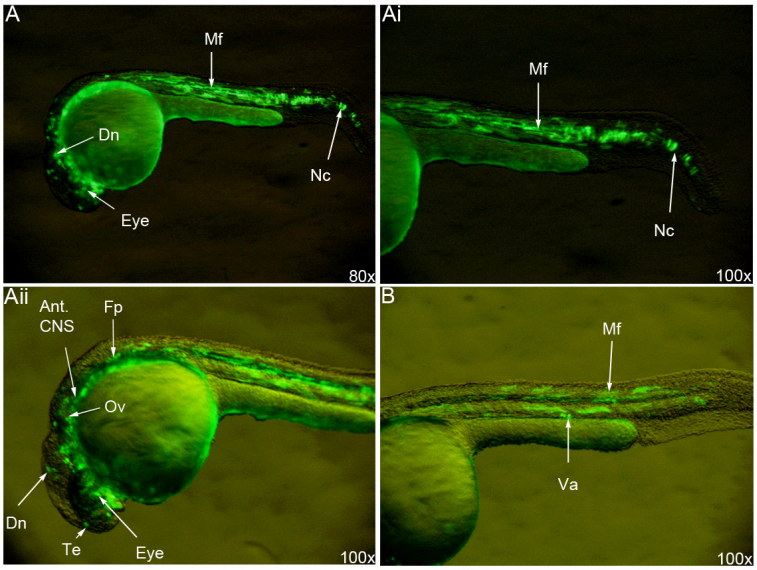

Figure 5.12: *Analysis of GFP expression at 25hpf after injection of 50pg Δ3.3kblama1GFPΔex63 BAC construct DNA at the one-cell stage*. A and Ai: Lateral views of the same embryo at 80x (A) and 100x (Ai) magnification, respectively, showing GFP expression in the eye (A), diencephalon and anterior CNS (A), muscle fibres (A,Ai) and the notochord (A,Ai). Aii: Lateral view of the embryo in panel A at 100x magnification, showing GFP expression within the telencephalon and diencephalon, floor plate of the anterior neural tube, eyes, otic vesicles, and the muscle fibres. B: Lateral view of an embryo at 100x magnification, revealing GFP expression in muscle fibres and the vasculature. Anterior is to the left. Abbreviations: muscle fibre (Mf), diencephalon (Dn), notochord (Nc), telencephalon (Te), anterior (Ant.), otic vesicle (Ov), floor plate (Fp), vasculature (Va).

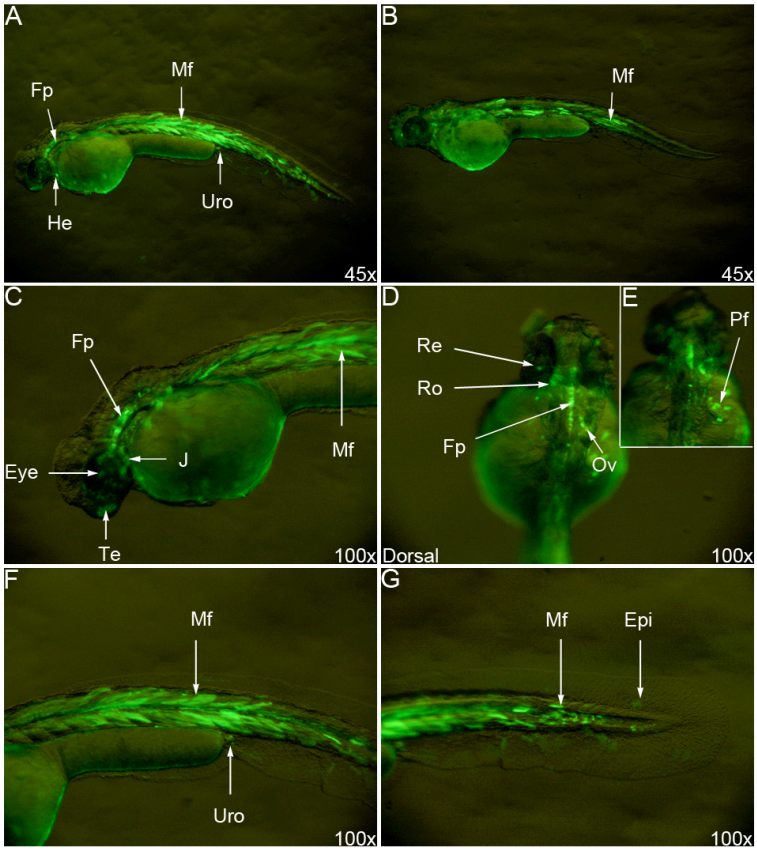

Figure 5.13: *Analysis of GFP expression at 49hpf after injection of 50pg Δ3.3kblama1GFPΔex63 BAC construct DNA at the one-cell stage*. A and B: Lateral views of two different embryos at 45x magnification showing GFP expression in the muscle fibres (A,B), heart (A), floor plate of the anterior CNS (A), uro-genital region (A), and the epidermis (A). A large variation in GFP expression in muscle fibres is observed (A,B). C and D: Lateral and dorsal views, respectively, of the embryo in panel A at 100x magnification. GFP is expressed in the floor plate (C,D), telencephalon (C), retina (C,D), forming jaw musculature (C), muscle fibres (C), the rhombencephalon region (D), and the otic vesicles (D). The inset image (E) shows a dorsal view of another 49hpf embryo with GFP expression in the pectoral fins. F and G: Lateral views of the embryo in panel A, at 100x magnification. Strong GFP expression is observed in many muscle fibres (F,G), the uro-genital region (F), and the epidermis (G). Anterior of lateral view images is to the left. Anterior of dorsal view images is at the top. Abbreviations: floor plate (Fp), heart (He), muscle fibre (Mf), uro-genital region (Uro), telencephalon (Te), jaw musculature (J), retina (Re), rhombencephalon (Ro), otic vesicle (Ov), pectoral fin (Pf), epidermis (Epi).

5.2.7: Enhancers controlling *lama1* expression are not located between introns 4 and 63

The enhancer elements controlling *lama1* expression are not located upstream or downstream of the *lama1* coding sequence (sections 5.2.4,5.2.5,5.2.6). These results suggest that the enhancers controlling all aspects of zebrafish *lama1* expression are situated within the gene itself, most likely within the intronic regions. To test this, I used homologous recombineering to delete intronic sequences in the construct Δ3.3kb*lama1*GFP, and insert a Kanamycin resistance gene, isolated from pCRII TOPO vector, in place of the *lama1* genomic DNA sequence (*Figures 2.2* and *5.14*). First, I deleted the *lama1* genomic sequence from exon 4 onwards (Δ3.3kb*lama1*GFPΔex4) (*Figure 5.15*I). Introns 1 to 3 were retained in the BAC construct because introns 1 and 3 share some sequence conservation with uncharacterised sequences identified from *Xenopus* and *Gasterosteus aculeatus*, which currently have an unknown chromosomal positioning of *lama1*. In addition, it is often the first two or three introns of a gene which contain enhancer sequences (Müller et al. 1999). The prediction was that loss of GFP expression from some or all of the tissues in the zebrafish embryo would reveal the presence of a putative enhancer. Analysis of GFP expression at different stages of development was performed to assess whether the initiation or maintenance of *lama1* expression is also altered following *lama1* genomic DNA sequence deletion. PCR analysis on BAC minicultures and a BAC midiprep confirms successful deletion of DNA sequence from exon 4 onwards within the Δ3.3kb*lama1*GFPΔex4 construct. The GFP reporter gene is still present and intact (*Figure 5.14* and *Table 5.3*).

	End	**1**	**37/38**
Primer pair	Ex4 end F, BAC backbone end R	Lama1 F1, Lama1 R1	Ex 37/38 F, Ex 37/38 R
Size (bases)	1180	1200	985 (should not work)

Table 5.3: *Primers used for BAC PCR analysis.* Primer pair End consists of a forward primer that binds the end of the forward ultramer homology sequence in exon 4, and a reverse primer that binds the end of the reverse ultramer homology sequence in the BAC backbone (vector DNA). A 1180 base band is produced when homologous recombineering has been successful, with the insertion of the Kanamycin gene in place of the exon-intron DNA sequence. Primer pair 1 consists of primers that lie either side of the GFP reporter gene, and produce a 1200 base fragment. Primer pair 37/38 amplifies a sequence that spans exon 37, intron 37, and exon 38 of the *lama1* gene. This is absent in Δ3.3kb*lama1*GFPΔex4 BAC DNA, but present in control BAC DNA.

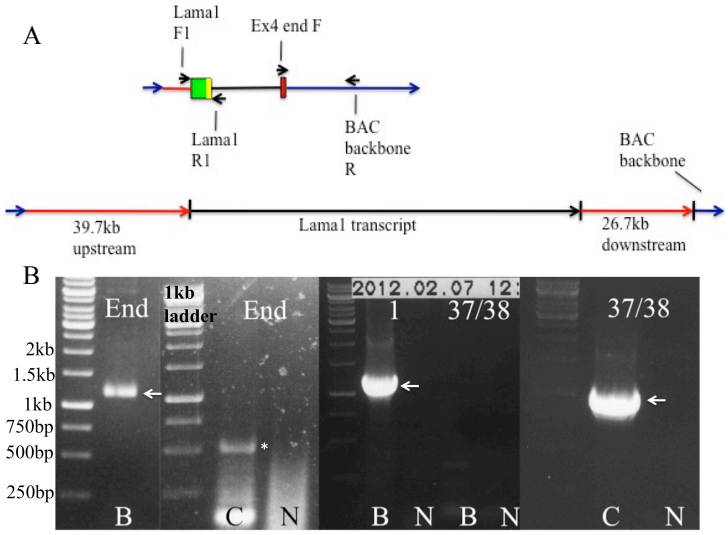

Figure 5.14: *PCR analysis of Δ3.3kblama1GFPΔex4 BAC DNA confirms successful deletion of DNA sequence downstream of exon 4*. A: A schematic representation of the BAC construct and primer positions. B: Primer pair End is the expected size and only works if the DNA sequence downstream of exon 4 has been removed. For this reason, no band is produced when using *lama1* BAC transgene (FL) construct DNA as a control. Primer pair 1 confirm the presence of the GFP reporter gene. Primer pair 37/38 confirms the absence of downstream DNA sequence in the Δ3.3kb*lama1*GFPΔex4 construct. Primers used for each reaction and the expected fragment sizes (white arrows) are listed in *Table 5.3*. Size of ladder is indicated on the left-hand side. Asterisk indicates non-specific band. Abbreviations: BAC DNA (B), *lama1* BAC transgene (FL) construct control DNA (C), no DNA control (N).

At the 19-somite stage, Δ3.3kb*lama1*GFPΔex4 injected embryos display strong GFP expression in the telencephalon and diencephalon (*Figure 5.15*IIA,C), neural tube (*Figure 5.15II*A,Ai), eye (*Figure 5.15*IIA,C,D), otic vesicle (*Figure 5.15*II*D*), muscle fibres (*Figure 5.15II*A,B,C), notochord (*Figure 5.15*IID), and the PSM (*Figure 5.15*IIA,C,D). By 25hpf, these tissues continue to express GFP (*Figure 5.16*A-C), in addition to the GFP observed in the uro-genital region (*Figure 5.16*D) and the floor plate (*Figure 5.16*C). At 49hpf, strong GFP expression remains in the telencephalon and the midbrain-hindbrain boundary (*Figure 4.17*C). GFP is also detected in the lens and retina, the otic vesicles, the forming jaw musculature, the heart (*Figure 4.17*C) and yolk sac (*Figure 4.17*A), muscle fibres (*Figure 4.17*E,F), the epidermis (*Figure 4.17*B), the ventral vasculature/pro-nephric tubule region, the uro-genital region (*Figure 4.17*F), and the pectoral fins (*Figure 5.17*D). Taken together, GFP expression driven by construct Δ3.3kb*lama1*GFPΔex4 fully recapitulates the spatial and temporal aspects of GFP expression observed in embryos injected with *lama1* BAC transgene (FL) (*Figure 5.7*A,B). This suggests that sequences upstream and downstream of the *lama1* gene, and DNA regions downstream of exon 4 are dispensible for the activation and maintenance of *lama1* expression.

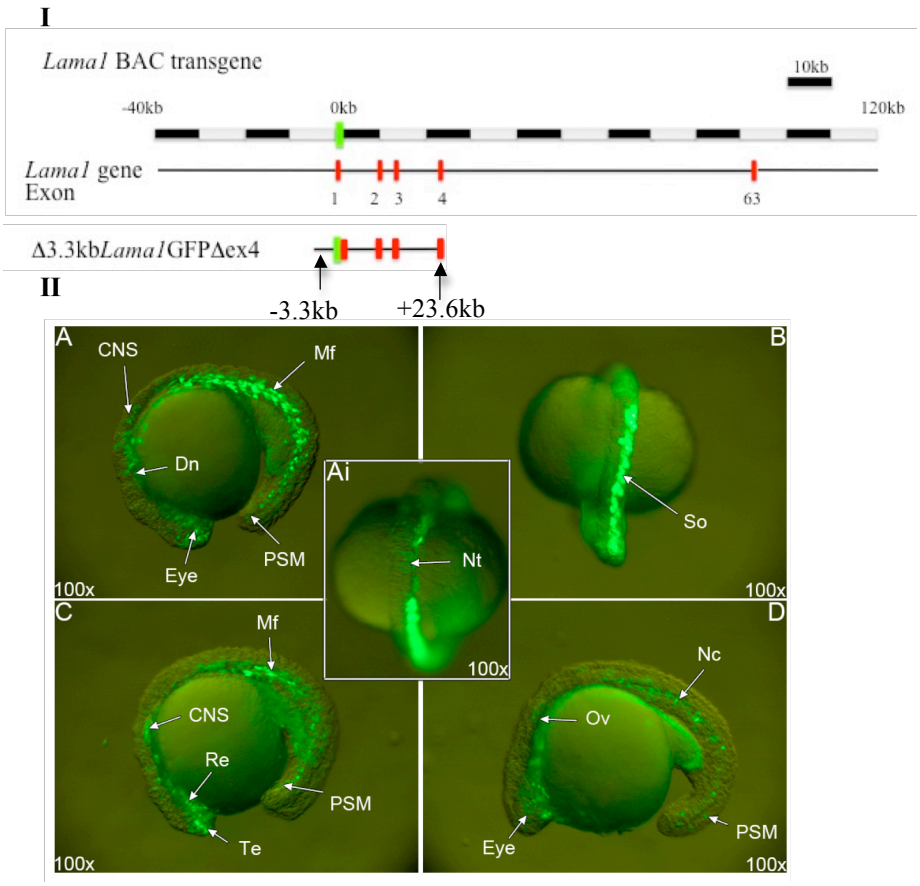

Figure 5.15: I: *A schematic representation of the lama1 gene and the 3.3kblama1GFPΔex4 construct*. The Δ3.3kb*lama1*GFPΔex4 construct consists of -3291 bases upstream of the *lama1* start site, the GFP reporter gene, exons 1 to 4, and introns 1 to 3. Red blocks represent exons, green blocks represent the GFP reporter gene. II: *Analysis of GFP expression at the 19-somite stage after injection of 50pg Δ3.3kblama1GFPΔex4 BAC DNA into the one-cell stage embryo*. A: lateral view showing GFP expression in the diencephalon and anterior CNS, eye, muscle fibres and the PSM. Ai: Dorsal view of the embryo in panel A, showing neural tube expression in the anterior CNS. B: Dorsal view with GFP expression in the somites. C: Lateral view revealing GFP expression within the telencephalon, in addition to the expression detected in the eye, anterior CNS, muscle fibres and the PSM. D: Lateral view shows GFP expression in the notochord, PSM, eye, and the otic vesicle. Anterior of lateral view images is to the left. Anterior of dorsal view images is at the top. The magnification of all images is 100x. Abbreviations: diencephalon (Dn), central nervous system (CNS), muscle fibres (Mf), pre-somitic mesoderm (PSM), neural tube (Nt), somites (So), telencephalon (Te), retina (Re), otic vesicle (Ov), notochord (Nc).

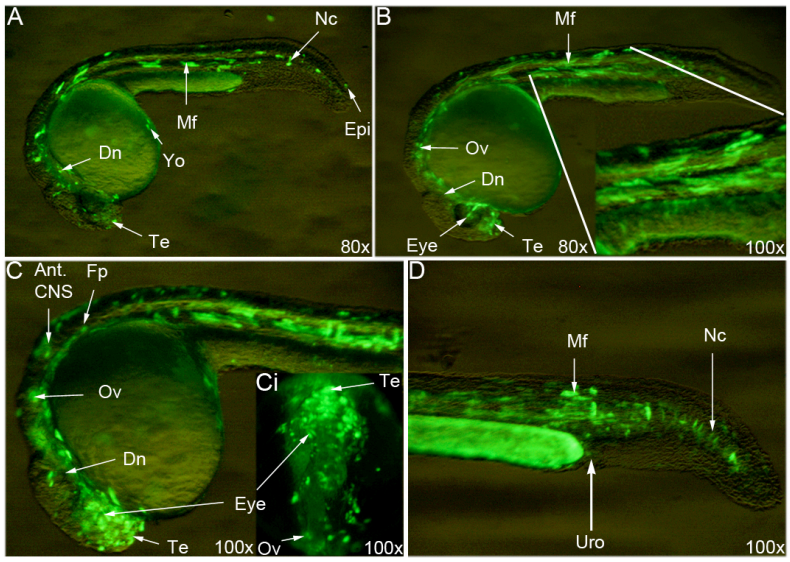

Figure 5.16: *Analysis of GFP expression at 25hpf after injection of 50pg Δ3.3kblama1GFPΔex4 BAC construct DNA into the one-cell stage embryo*. A and B: Lateral views of embryos showing GFP expression in the telencephalon, diencephalon, eye (B), otic vesicles (B), muscle fibres, notochord (A), epidermis (A), and the yolk sac (A). The magnified image in panel B shows GFP expression in the muscle fibres at 100x magnification. C: Lateral view of an embryo. GFP is detected throughout the anterior CNS, in the telencephalon, diencephalon and the floor plate. Expression is also observed in the otic vesicle, eye and muscle fibres. Ci: Dorsal view of the embryo in panel C at 100x magnification, showing GFP expression in the telencephalon, eye and the otic vesicle. D: Lateral view of the tail. GFP is detected in the muscle fibres, the uro-genital region, and the notochord. Anterior of lateral views is to the left. Anterior of dorsal view is at the top. Magnification is indicated in each image. Abbreviations: telencephalon (Te), diencephalon (Dn), yolk (Yo), muscle fibres (Mf), notochord (Nc), epidermis (Epi), otic vesicle (Ov), anterior (Ant.), floor plate (Fp), uro-genital region (Uro).

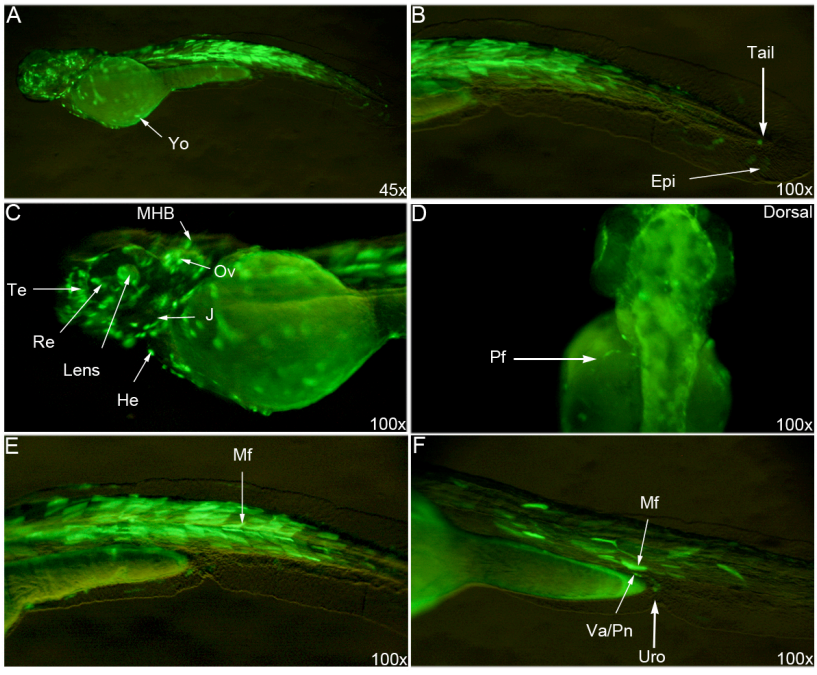

Figure 5.17: *Analysis of GFP expression at 49hpf after injection of 50pg Δ3.3kblama1GFPΔex4 BAC construct DNA into the one-cell stage embryo*. A and B: Lateral views of the same embryo at 45x (A) and 100x (B) magnification showing GFP expression in the muscle fibres, yolk sac (A) tail (B), and the epidermis (B). C and D: Lateral (C) and dorsal (D) views showing the head of different embryos at 100x magnification, showing GFP expression in the telencephalon, midbrain-hindbrain boundary, heart, lens and retina, forming jaw musculature, otic vesicles (C), and the pectoral fins (D). E: Lateral view of the embryo in panel A at 100x magnification. GFP is strongly expressed in the muscle fibres. F: Lateral view at 100x magnification shows GFP expression in the vasculature/pro-nephric tubule region, uro-genital region, and the muscle fibres. Anterior of lateral view images is to the left. Anterior of dorsal view image is at the top of the panel. Abbreviations: telencephalon (Te), yolk (Yo), muscle fibres (Mf), epidermis (Epi), otic vesicle (Ov), heart (He), retina (Re), jaw musculature (J), midbrain-hindbrain boundary (MHB), pectoral fin (Pf), vasculature/pro-nephric tubules (Va/Pn), uro-genital region (Uro).

5.2.8: Sequences downstream of exon 2 are not required for *lama1* activation

My observations (section 5.2.7) strongly indicate that the enhancers controlling *lama1* expression are located within a region encompassing the first three introns and exons of the gene, a region which equates to 23,692bp. My previous data also demonstrated a role of Hh signalling in the expression of *lama1* (sections 3.2.5.1, 3.2.5.2, and 4.2.8). However, it is unknown whether this is a direct or indirect regulation at the transcriptional level. The presence of Gli binding sites in the potential enhancer regions of *lama1* would support the possibility that Hh signalling and the Gli transcription factors are directly controlling *lama1* transcription.

To predict which introns may contain enhancers for *lama1* expression and to determine whether Hh may directly control *lama1* expression, I searched for potential Gli binding sites using Gli sequences previously identified by chromatin immunoprecipitation (ChIP) using an epitope-tagged Gli-activator protein (Vokes et al. 2007). I identified two potential Gli binding sites within intron 2, and one within intron 3 (*Table 5.4*). Using MatInspector software (http://www.genomatix.de/), I also identified three potential Gli binding sites within intron 1 of *lama1* (*Table 5.4* and *Figure 5.18*). This suggests that any of introns 1-3 could be capable of responding to Hh signalling, and possibly contain enhancers for *lama1* transcription.

	Intron1	**Intron 2**	**Intron 3**
Size (bp)	9779	2901	9353
Gli site positions	2403, 5599, 6078	922, 1290	1665

Table 5.4: *Location of putative Gli binding sites within the first 3 introns of lama1.* Intron 1 is 9779 bases, and contains three potential Gli sites identified using MatInspector software. Introns 2 and 3 are 2901 bases and 9353 bases, respectively. Two potential Gli sites are located in intron 2, and 1 potential Gli site is located in intron 3.

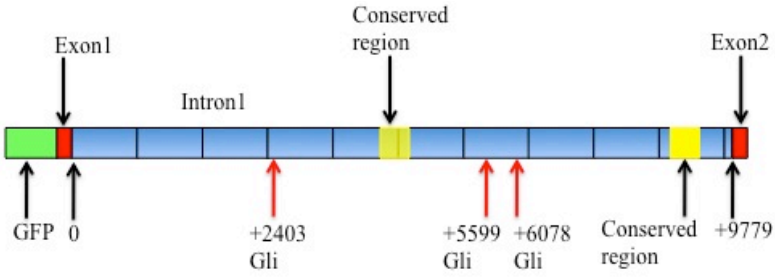

Figure 5.18: *A schematic representation of zebrafish lama1 intron 1.* Intron 1 contains three potential Gli binding sites, and the two regions of DNA that are conserved with *Fugu*.

To help determine the next deletion site, I performed new bioinformatic analyses. The ECR browser identified two regions that are 353bp and 350bp long at positions +3986 and +8859, respectively, within intron 1 that are conserved (>70%) with *lama1* in the *Fugu* genome (*Figure 5.18*), but not with other species. A 'megablast' search (http://blast.ncbi.nlm.nih.gov/) that identifies highly similar sequences does not find any significant conservation of DNA with zebrafish intron 1 of *lama1*. Thus, bioinformatic analyses do not provide significant clues about the location of potential enhancer elements. However, the conservation of 2 domains within intron 1 of *lama1* makes it a good candidate for further analyses.

Using homologous recombineering, I further deleted into the *lama1* gene by removing the genomic sequence downstream of exon 2 (Δ3.3kb*lama1*GFPΔex2) (*Figure 5.20*I) in the construct Δ3.3kb*lama1*GFP, and inserting a Kanamycin resistance gene, isolated from pCRII TOPO vector, in place of the *lama1* genomic DNA sequence.

Loss of GFP expression from some or all of the tissues in the zebrafish embryo would indicate the loss of an enhancer, and any remaining GFP expression would suggest that an enhancer lies in intron 1. Analyses of GFP expression at different stages of development would allow me to assess whether the initiation or maintenance of the gene is also altered following *lama1* genomic DNA sequence deletion.

PCR analysis of BAC minicultures and BAC midiprep DNA confirms the successful deletion of DNA sequence from exon 2 onwards within the Δ3.3kb*lama1*GFPΔex2 construct (*Figure 5.19* and *Table 5.5*). Sequences upstream of -3291bp and downstream of the *lama1* gene are absent in this construct, as demonstrated by the absence of a band when performing PCR with primers that amplify a sequence 12.7kb upstream, and primers that amplify a sequence from exon 30 to exon 31. The GFP reporter gene is still present and intact as indicated by the band generated with primer pair 1 (*Figure 5.19* and *Table 5.5*).

	End	**Test**	**1**	**12.7**	**30/31**
Primer pair	Ex2 end F, BAC backbone end R	Ex2 F, BAC backbone Rev	Lama1 F1, Lama1 R1	12.7kb up F, 12.7kb up R	Ex30/31 F, Ex30/31 R
Size (bases)	1180	1673	1200	501 (should not work)	705 (should not work)

Table 5.5: *Primers used for BAC PCR analysis.* Primer pair End consists of a forward primer that binds the end of the forward ultramer homology sequence in exon 2, and a reverse primer that binds the end of the reverse ultramer homology sequence in the BAC backbone (vector DNA). A 1180 base band is produced when homologous recombineering has been successful, with the insertion of the Kanamycin gene in place of the exon-intron DNA sequence. Primer pair Test consists of a forward primer that binds to DNA sequence just upstream of the forward ultramer homology sequence in exon 2, and a reverse primer that binds just downstream of the reverse ultramer homology sequence in the BAC backbone (vector DNA). A 1673 base band is produced when homologous recombineering has been successful. Primer pair 1 consists of primers that lie either side of the GFP reporter gene, and produce a 1200 base fragment. Primer pair 12.7 amplifies a sequence that is 12.7kb upstream of the *lama1* start site, and is absent in Δ3.3kb*lama1*GFPΔex2 BAC DNA, but present in *lama1* BAC transgene (FL) construct control DNA. Primer pair 30/31 amplifies a sequence that spans exon 30, intron 30, and exon 31 of the *lama1* gene. This is also absent in Δ3.3kb*lama1*GFPΔex2 BAC DNA, but present in *lama1* BAC transgene (FL) construct control DNA.

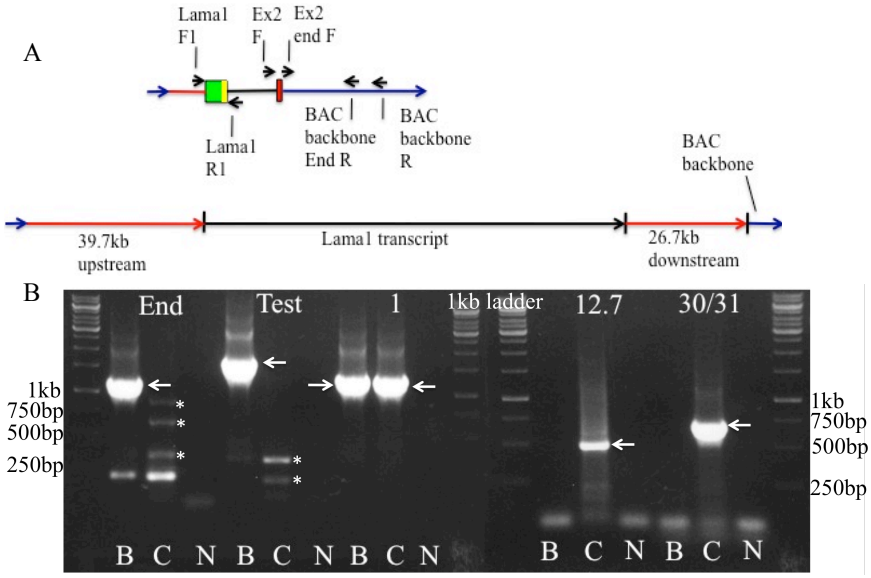

Figure 5.19: *PCR analysis of Δ3.3kblama1GFPΔex2 BAC midi prep DNA confirms successful deletion of the DNA sequence downstream of exon 2*. A: A schematic representation of the BAC construct and primer positions. B: Primer pair End is the expected size and only works if the DNA sequence downstream of exon 2 has been removed, with the associated insertion of the Kanamycin resistance gene. For this reason, a band at the expected size is not produced when using *lama1* BAC transgene (FL) construct DNA as a control. Non-specific DNA fragments are amplified. Primer pair Test is the expected size and only works if the DNA downstream of exon 2 has been removed. For this reason, a band at the expected size is not produced when using *lama1* BAC transgene (FL) construct DNA as a control. Non-specific DNA fragments are amplified. Primer pair 1 confirm the presence of the GFP reporter gene. Primer pair 12.7 confirms the absence of upstream DNA sequence in the Δ3.3kb*lama1*GFPΔex2 BAC DNA construct. Primer pair 30/31 confirms the absence of downstream DNA sequence in the Δ3.3kb*lama1*GFPΔex2 construct. Primers used for each reaction, and the expected fragment sizes (indicated by white arrows) are listed in *Table 5.5*. Asterisks indicate non-specific bands. Size of ladder is indicated on the left-hand side. Abbreviations: BAC DNA (B), *lama1* BAC transgene (FL) construct control DNA (C), no DNA control (N).

At the 19-somite stage, embryos injected with the Δ3.3kb*lama1*GFPΔex2 construct express GFP in the telencephalon and diencephalon (*Figure 5.20*IIA,B,C), otic vesicles (*Figure 5.20*IIB), eye (*Figure 5.20*IIB,D), floor plate, notochord (*Figure 5.20*IIB), muscle fibres (*Figure 5.20*IIA,D), and the PSM (*Figure 5.20*IID). These are all the observed sites of expression in *lama1* BAC transgene (FL) injected embryos (*Figure 5.5*IIA,E), suggesting there is no difference in *lama1* initiation in embryos injected with Δ3.3kb*lama1*GFPΔex2 (n = 96/102, 94%) compared with *lama1* BAC transgene (FL)-injected embryos.

By 25hpf, GFP is expressed in the telencephalon and diencephalon, lens and retina, otic vesicles, floor plate (*Figure 5.21*A,D), muscle fibres (*Figure 5.21*A,C), notochord (*Figure 5.21*A,B), and the ventral vasculature/pro-nephric tubules (*Figure 5.21*E). Again, this expression pattern (n = 101/114, 89% embryos express GFP) matches that observed in *lama1* BAC transgene (FL) injected embryos (n = 229/289, 79% embryos express GFP) (*Figure 5.6*A,B).

At 50hpf, strong GFP is detected in the telencephalon and midbrain-hindbrain of the anterior CNS (n = 16/35 (46%), compared to 47% *lama1* BAC (FL) embryos), the lens and retina (n = 17/35 (49%), compared to 21% *lama1* BAC (FL) embryos) (*Figure 5.22*A,Ai,C), the forming jaw

musculature (n = 5/35 (14%), compared to 11% *lama1* BAC (FL) embryos) (*Figure 5.22*C), the otic vesicles (n = 12/35 (34%) embryos), floor plate (n = 5/35 (14%), compared to 7% *lama1* BAC (FL) embryos) (*Figure 5.22*B), the heart (n = 4/35 (11%), compared to 12% *lama1* BAC (FL) embryos) (*Figure 5.22*D), muscle fibres (n = 22/35 (63%), compared to 50% *lama1* BAC (FL) embryos) (*Figure 5.22*C,F,G), the vasculature/pro-nephric tubules (n = 5/35 (14%), compared to 1% *lama1* BAC (FL) embryos) (*Figure 5.22*G), the uro-genital region (*Figure 5.22*H), the notochord (n = 1/35 (3%), compared to 2% *lama1* BAC (FL) embryos), and the developing pectoral fins (n = 2/35 (6%), compared to 5% *lama1* BAC (FL) embryos) (*Figure 5.22*E). Overall, 57/73 (78%) Δ3.3kb*lama1*GFPΔex2 DNA-injected embryos express GFP, compared to 58/108 (54%) *lama1* BAC (FL)-injected embryos at the same developmental age (*Figure 4.14b*).

It is clear therefore, that GFP expression in each of the tissues described above is consistent or greater than the amount of embryos expressing GFP that have been injected with *lama1* BAC transgene (FL) DNA, at 49-50hpf (*Figures 5.7*A,B and *5.23*). This suggests that the DNA sequence present in the construct Δ3.3kb*lama1*GFPΔex2 is sufficient to maintain all GFP expression. It is possible that some regulatory regions have been lost upon removal of all DNA sequence downstream of exon 2, which could normally function to suppress the expression of *lama1*, namely in the eye, muscle fibres, neural tube, and the vasculature/pro-nephric tubule region. However, micro-injection of BAC DNA into the one-cell stage embryo and the integration of the BAC DNA into the genome could also be responsible for the varying levels of GFP expression.

Importantly, at each stage analysed, the GFP expression pattern completely recapitulates the expression pattern observed in embryos injected with *lama1* BAC transgene (FL). GFP is initiated at the correct time in each tissue, and continues to be maintained as in *lama1* BAC transgene (FL) injected embryos. Therefore, the enhancers controlling GFP expression in both Δ3.3kb*lama1*GFPΔex2 and *lama1* BAC transgene (FL) injected embryos are likely to be located within intron 1.

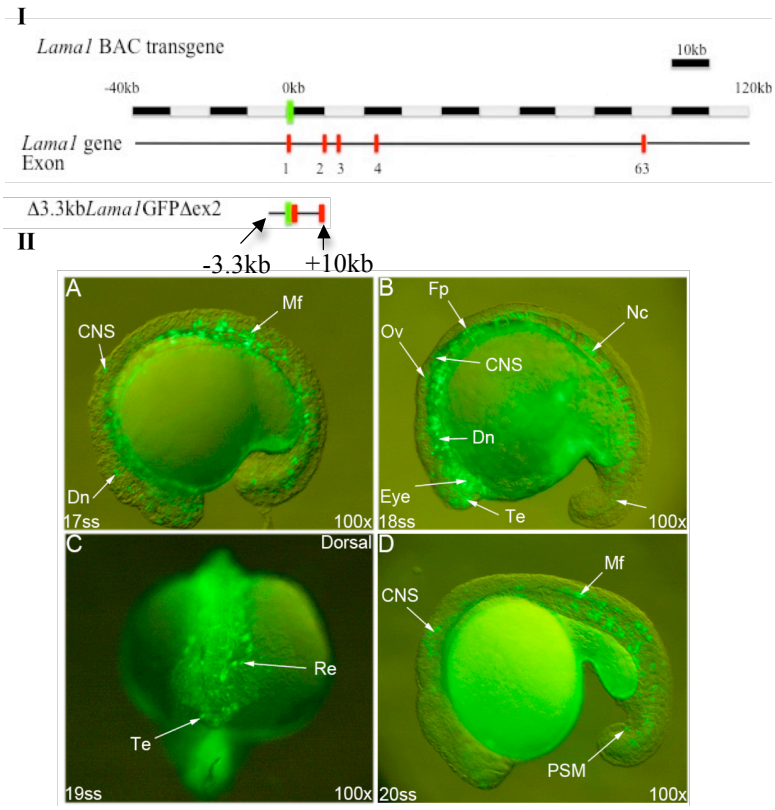

Figure 5.20: I: *A schematic representation of the lama1 gene and the Δ3.3kblama1GFPΔex2 construct*. Δ3.3kb*lama1*GFPΔex2 construct consists of -3291 bases upstream of the *lama1* start site, the GFP reporter gene, exons 1 and 2, and intron 1. Red blocks represent exons, green blocks represent the GFP reporter gene. II: *Analysis of GFP expression at the 17-20-somite stage after injection of 50pg Δ3.3kblama1GFPΔex2 BAC construct DNA into the one-cell stage embryo*. A-D: Lateral views of 17 (A), 18 (B), and 20 (D)-somite stage embryos and a dorsal view of a 19-somite stage embryo (C), at 100x magnification. GFP is expressed in the telencephalon (C), diencephalon (A,B) anterior CNS (A,B,D), floor plate (B), forming eye (B) and retina (C), otic vesicles (B), muscle fibres (A,D), notochord (B) and the PSM (D). Anterior of lateral views is to the left. Anterior of dorsal view image is at the bottom of the panel. Abbreviations: diencephalon (Dn), central nervous system (CNS), muscle fibres (Mf), telencephalon (Te), otic vesicle (Ov), floor plate (Fp), notochord (Nc), retina (Re), pre-somitic mesoderm (PSM).

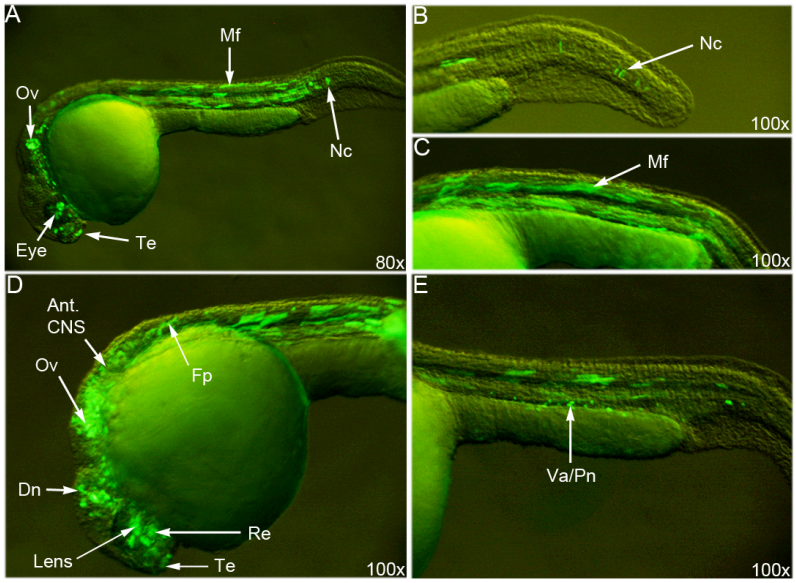

Figure 5.21: *Analysis of GFP expression at 26hpf after injection of 50pg Δ3.3kblama1GFPΔex2 BAC construct DNA into the one-cell stage embryo*. A lateral view of an embryo at 80x magnification reveals GFP in the telencephalon and diencephalon, the eye, the otic vesicles, muscle fibres, and the notochord (A). Other embryos also express GFP in the notochord (B) and the muscle fibres (C) at 100x magnification. D: Lateral view of the head of an embryo at 100x magnification, with GFP detected in the telencephalon and diencephalon, the lens and retina, the otic vesicle, the floor plate, and the muscle fibres (D). GFP is also observed in the vasculature/pro-nephric tubule region (E). Anterior is to the left of the panels. Abbreviations: diencephalon (Dn), anterior (Ant.), central nervous system (CNS), muscle fibres (Mf), telencephalon (Te), otic vesicle (Ov), floor plate (Fp), notochord (Nc), retina (Re), pre-somitic mesoderm (PSM), vasculature/pro-nephric tubules (Va/Pn).

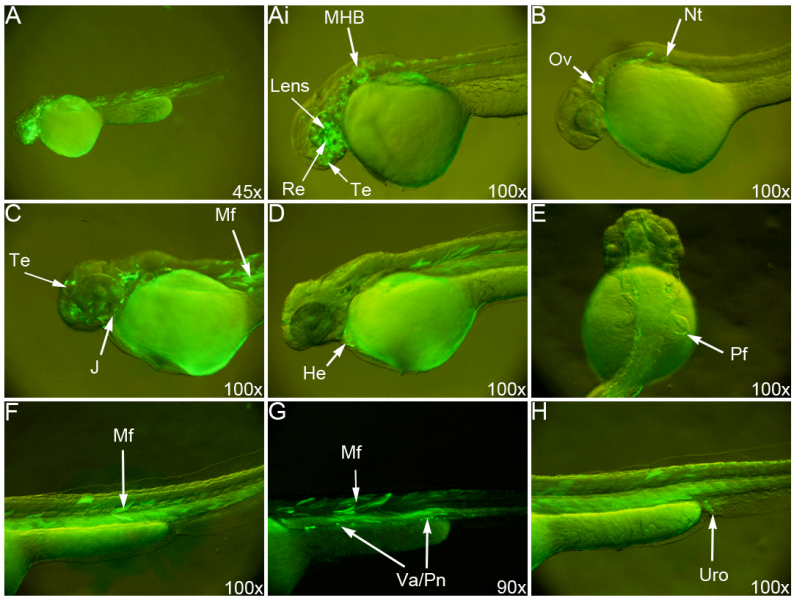

Figure 5.22: *Analysis of GFP expression at 50hpf after injection of 50pg Δ3.3kblama1GFPΔex2 BAC construct DNA into the one-cell stage embryo*. A-D and F-H are lateral views, panel E is a dorsal view. GFP is detected in the muscle fibres and anterior CNS region at 45x magnification (A). Ai: the embryo in panel A at 100x magnification shows GFP expression in the telencephalon and midbrain-hindbrain boundary, and the lens and retina (Ai), whilst another embryo expresses GFP in the otic vesicles and anterior neural tube, at 100x magnification (B). C: Another embryo at 100x magnification shows GFP in the forming jaw musculature, in addition to the described expression in the telencephalon and muscle fibres. Expression is also detected in the heart (D) and the forming pectoral fins (E), at 100x magnification. GFP is detected in the muscle fibres at 100x or 90x magnification, respectively, (F,G), along with vasculature and pro-nephric tubule expression of GFP (G). GFP continues to be expressed in the uro-genital region (H). Anterior of lateral views is to the left. Anterior of dorsal view image is at the top. Abbreviations: retina (Re), telencephalon (Te), midbrain-hindbrain boundary (MHB), otic vesicle (Ov), neural tube (Nt), jaw musculature (J), muscle fibre (Mf), heart (He), pectoral fin (Pf), vasculature/pro-nephric tubules (Va/Pn), uro-genital region (Uro).

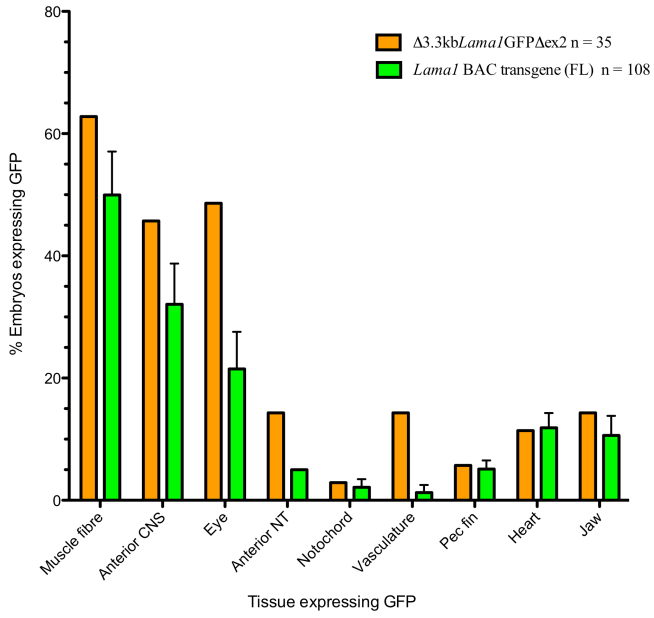

Figure 5.23: *Comparison of GFP expression in 49-50hpf embryos injected with Δ3.3kblama1GFPΔex2 DNA or lama1 BAC (FL) DNA*. The number of embryos expressing GFP in Δ3.3kb*lama1*GFPΔex2 injected embryos is similar to, or greater than the number of embryos expressing GFP in *lama1* BAC transgene (FL) injected embryos. Only one batch of 35 embryos were analysed in this experiment (out of a total of 73 embryos analysed), and so error bars are not shown.

5.2.9: Intron 1 is required for *lama1* transcription

The construct Δ3.3kb*lama1*GFPΔex2 is sufficient to allow complete GFP reporter gene expression in my transgenesis analyses (section 5.2.8). Previously, I have shown that a construct consisting of the GFP reporter gene and -3291 bases upstream of the *lama1* start site is insufficient to allow proper reporter gene activation. This suggests that the GFP expression detected in Δ3.3kb*lama1*GFPΔex2 injected embryos is the result of enhancer activities located in intron 1. To confirm this finding, I deleted all sequences downstream of exon 1 in the construct Δ3.3kb*lama1*GFP, by inserting a Kanamycin resistance gene, isolated from pCRII TOPO vector, in place of the *lama1* genomic DNA sequence. The insertion occurred after the first three bases of intron 1, therefore leaving only -3291 bases upstream of the *lama1* start site, the GFP reporter gene, exon 1, and just the first three bases of intron 1 (Δ3.3kb*lama1*GFPΔex1) (*Figure 5.25*I). Enhancers can also be associated with coding exon sequences (Tumpel et al. 2008). Δ3.3kb*lama1*GFPΔex1 construct therefore allows me to test the possibility that exon 1 contains enhancer activity capable of regulating *lama1* expression.

I hypothesised that exon 1 of *lama1* would not contain any enhancer sequences and that the construct Δ3.3kb*lama1*GFPΔex1 will be insufficient to allow proper GFP reporter expression. I predicted that only a few embryos between the 19-somite stage to 25hpf would contain just a few GFP fluorescent cells, and a small number of embryos would express GFP in a few muscle fibres or the notochord, as I previously observed when using the 3.3kb*lama1*GFP-TOPO DNA (*Figure 5.9*II).

PCR analysis on BAC minicultures and a BAC midiprep confirms successful deletion of DNA sequence from exon 1 onwards within the Δ3.3kb*lama1*GFPΔex1 construct (*Figure 5.24* and *Table 5.6*). This is demonstrated by the absence of a band when performing PCR with primers that amplify a sequence from exon 37 and exon 38, and primers that amplify a sequence from exon 30 to exon 31. The failure of primer pair 1 to produce a band confirms the loss of intron 1 from Δ3.3kb*lama1*GFPΔex1 DNA. The GFP reporter gene is still present and intact (primer pair A) (*Figure 5.24* and *Table 5.6*).

	37/38	30/31	Ex1	A	1
Primer pair	Ex 37/38 F, Ex 37/38 R	Ex30/31 F, Ex30/31 R	Ex1 end F, BAC backbone end R	Lama1 F1, GFP R	Lama1 F1, Lama1 R1
Size (bases)	985 (should not work)	705 (should not work)	1369	648	1200 (should not work)

Table 5.6: *Primers used for BAC PCR analysis.* Primer pairs 37/38 and 30/31 amplify sequences that span exon 37, intron 37, and exon 38, and exon 30, intron 30, and exon 31 of the *lama1* gene, respectively. These exon and intron DNA sequences are absent in the Δ3.3kb*lama1*GFPΔex1 BAC construct, but present in *lama1* BAC transgene (FL) construct control DNA. Primer pair Ex1 consists of a forward primer that binds the end of the forward ultramer homology sequence in exon 1, and a reverse primer that binds the end of the reverse ultramer homology sequence in the BAC backbone (vector DNA). A 1369 base band is produced when homologous recombineering has been successful, with the insertion of the Kanamycin gene in place of the exon-intron DNA sequence. Primer pair A consists of a forward primer that binds to DNA sequence just upstream of the GFP reporter gene, and a reverse primer that binds within the GFP reporter gene. A 648 base band is produced if the GFP reporter gene is present in the correct location in BAC zC34A17. Primer pair 1 consists of primers that lie either side of the GFP reporter gene. However, the reverse primer is located in intron 1 of *lama1*, and so is absent in the Δ3.3kb*lama1*GFPΔex1 BAC construct. However, intron 1 is present in *lama1* BAC transgene (FL) construct control DNA.

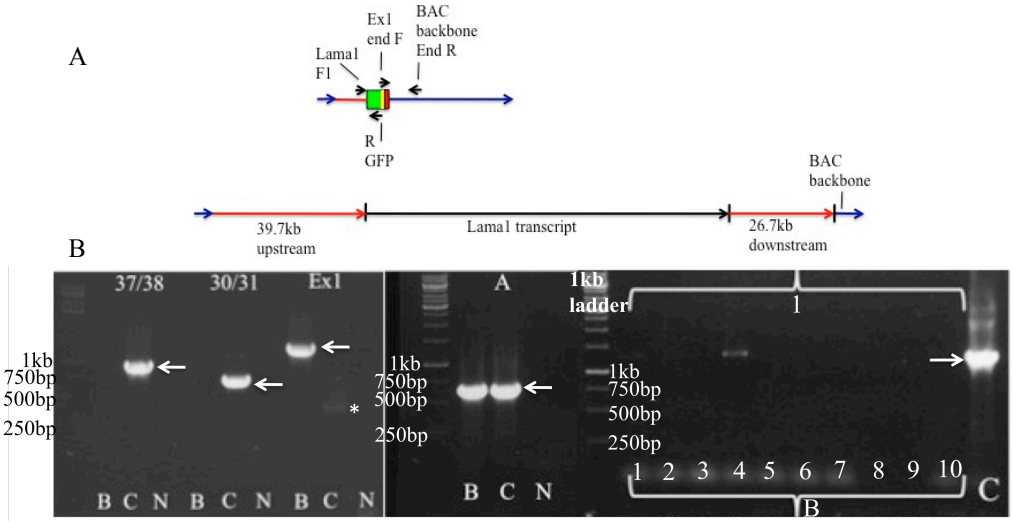

Figure 5.24: *PCR analysis of Δ3.3kblama1GFPΔex1 BAC midi prep DNA confirms successful deletion of the sequences downstream of exon 1.* A: A schematic representation of the BAC construct and primer positions. B: Primer pairs 37/38 and 30/31 confirm the absence of upstream DNA sequence in the construct Δ3.3kb*lama1*GFPΔex1. Primer pair Ex1 is the expected size and only works if the sequence downstream of exon 1 has been removed, with the associated insertion of the Kanamycin resistance gene. For this reason, a band at the expected size is not produced when using *lama1* BAC transgene (FL) construct DNA as a control. Primer pair A confirms the presence of the GFP reporter gene. Primer pair 1 confirms the absence of intron 1 in 9 out of 10 Δ3.3kb*lama1*GFPΔex1 BAC mini prep samples. Primers used for each reaction, and the expected fragment sizes (indicated by white arrows) are listed in *Table 5.6*. Asterisk indicates non-specific band. Size of ladder is indicated on the left-hand side. Abbreviations: BAC DNA (B), *lama1* BAC transgene (FL) construct control DNA (C), no DNA control (N).

At the 6-somite stage, GFP expression is not detected in embryos injected with Δ3.3kb*lama1*GFPΔex1 DNA (n = 64) (*Figure 5.25*IIA). At the 20-somite stage, there is very little detectable GFP expression. GFP flourescent cells are observed the anterior CNS (n = 6/87 (7%), compared to 3.4% of 3.3kb*lama1*GFP-TOPO embryos (n = 2/58)), and in the trunk and somite region (n = 7/87 (8%), compared to 17.2% of 3.3kb*lama1*GFP-TOPO embryos (n = 10/58)). Specific expression within the muscle fibres is detected in 3/87 (3%) embryos, although only very weakly (*Figure 5.25*IIB). Embryos injected with 3.3kb*lama1*GFP-TOPO construct also expressed GFP in 3% embryos (n = 2/58) (*Figure 5.9*C). The consistent results between embryos injected with these two constructs (Δ3.3kb*lama1*GFPΔex1 and 3.3kb*lama1*GFP-TOPO) suggest that no enhancer element is situated in exon 1 of *lama1* and that all elements required for *lama1* expression in the zebrafish are located within intron 1. At 24-26hpf, 34/195 (17%) embryos show weak GFP-expressing cells in the anterior CNS and head region, and 21/195 (11%) embryos display GFP expression in the trunk and somite (*Figure 5.25*IIC,D). The GFP expression is again comparable to the expression detected in embryos injected with 3.3kb*lama1*GFP-TOPO DNA (11% and 19% of embryos expressing GFP fluorescent cells in the anterior CNS and the trunk, respectively, n = 36) (*Figure 5.9*IID,E), but more than the amount of embryos that express GFP fluorescent cells in Min.prom*lama1*GFP DNA injected embryos (3% embryos, n = 2/76) (*Figure 5.1*IIE,F). 5/195 Δ3.3kb*lama1*GFPΔex1 DNA-injected embryos show specific GFP expression in the muscle fibres (3% embryos) (*Figure 5.25*IID,E), and 3/195 (2%) embryos show specific GFP expression in the retina. By 49hpf, no specific GFP expression is detected, although 2/78 (3%) embryos show some weak GFP fluorescence (*Figure 5.25*IIF,G).

These results are in agreement with my observations made from injection of the 3.3kb*lama1*GFP-TOPO construct into one-cell stage embryos. They confirm that -3291 bases upstream of the *lama1* start site are not capable for complete *lama1* activation, although some weak enhancer and promoter activity is likely to be present. The lack of GFP expression at the 6-somite stage also confirms that the lack of GFP expression observed at later stages is not just due to a defect in maintenance of GFP expression. There is a failure to initiate GFP expression. This means that the enhancers controlling *lama1* expression do in fact lie in intron 1, and their presence is necessary for *lama1* activation. However, it is unknown whether the enhancers in intron 1 are sufficient for *lama1* activation.

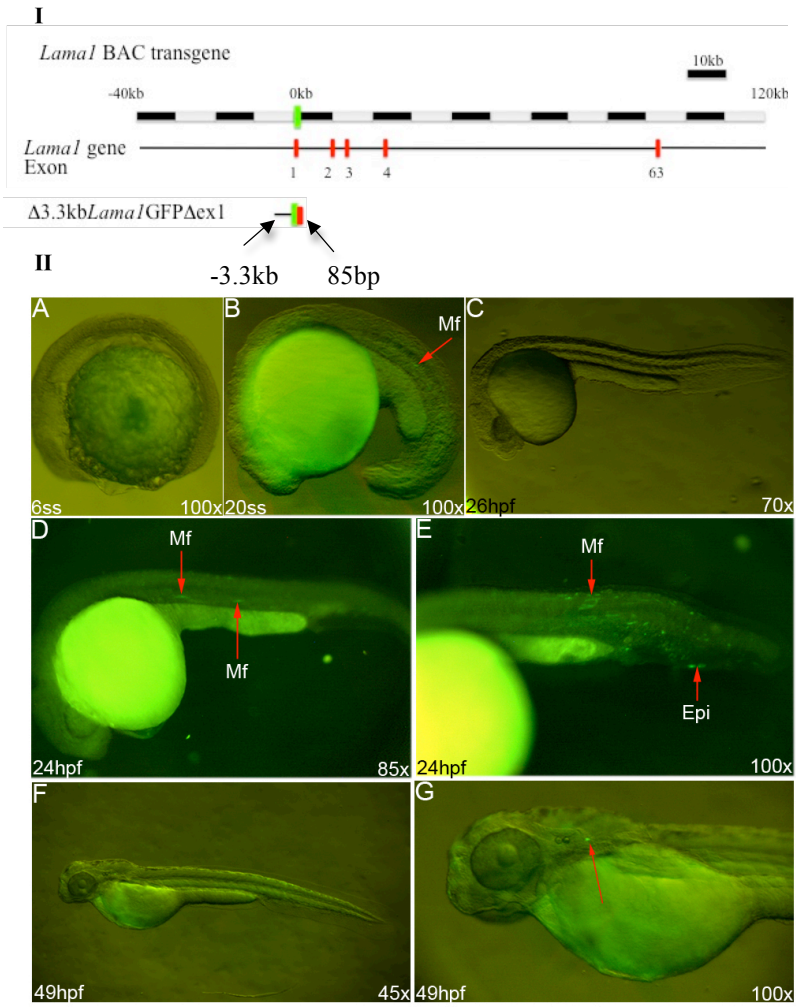

Figure 5.25: I: *A schematic representation of the lama1 gene and the Δ3.3kblama1GFPΔex1 construct*. The Δ3.3kb*lama1*GFPΔex1 construct consists of -3291 bases upstream of the *lama1* start site, the GFP reporter gene, and exon 1. There are no intron or downstream DNA sequences present. Red blocks represent exons, green blocks represent the GFP reporter gene. II: *Analysis of GFP expression after injection of 50pg Δ3.3kblama1GFPΔex1 BAC DNA into the one-cell stage embryo*. A-C: Lateral views of a 6 (A), and 20 (B)-somite stage embryo at 100x magnification, and a 26hpf embryo at 70x magnification (C). No GFP is observed in these 6-somite (A) and 26hpf stage (C) embryos, whilst weak GFP expression is detected in a single muscle fibre at the 20-somite stage (B) (red arrow). D and E: Highly exposed lateral views of 24hpf embryos at 85x (D) and 100x (E) magnification. Weak expression of GFP is detected in a few muscle fibres (D,E) and the epidermis (E) (red arrows). F and G: Lateral views of a 49hpf embryo at 45x (F) and 100x (G) magnification. GFP expression is absent (F), except for a single GFP fluorescent just posterior to the otic vesicle (G) (red arrow). Anterior is to the left. Abbreviations: muscle fibre (Mf), epidermis (Epi).

5.2.10: 4415 bases of intron 1 are sufficient for *lama1* GFP expression in the muscle fibres

To establish whether intron 1 is sufficient on its own for *lama1* activation, I removed the remaining upstream sequence (-3291 bases) left in the Δ3.3kb*lama1*GFPΔex2 construct using homologous recombineering to insert an iTOL2-AMP fragment. The new construct generated kept -878 bases directly upstream of the *lama1* start site, thereby keeping the *lama1* minimal promoter element. Part of intron 1 was also removed using homologous recombineering to insert a Kanamycin resistance gene, in place of the *lama1* genomic DNA sequence, causing a deletion of sequences downstream of base +4415 of intron 1 (*Figure 5.28*I). This deletion removed one of the two conserved regions identified in intron 1, and two of the three potential Gli binding sites (*Figure 5.26*). Analysis of GFP expression generated from this construct would establish whether this region contains enhancer elements. A total loss of GFP expression in these embryos injected with Δmin.prom*lama1*GFPΔ5kb could either indicate that the region -3291bp to -878bp or/and the region +4416 to +9779 bases within intron 1 are essential.

Based on previous findings that the -3291bp upstream sequence has a very weak enhancer activity in comparison to the enhancers within intron 1 (*Figure 5.9*II), I predicted that removal of the -3291bp upstream sequence would have no significant effect upon GFP expression. However, I expected that removal of the region +4416 to +9779 bases of intron 1 would cause a loss of GFP expression from some, or all of the tissues within the zebrafish embryo. Any remaining GFP expression would indicate the presence of enhancer elements located in the remaining conserved region of intron 1, including the Gli binding site located 1568 bases upstream of the conserved region, at position +2403 within intron 1. Analysis of GFP expression at different stages of development would allow me to assess whether the initiation or maintenance of the gene is also altered with *lama1* genomic DNA sequence deletion.

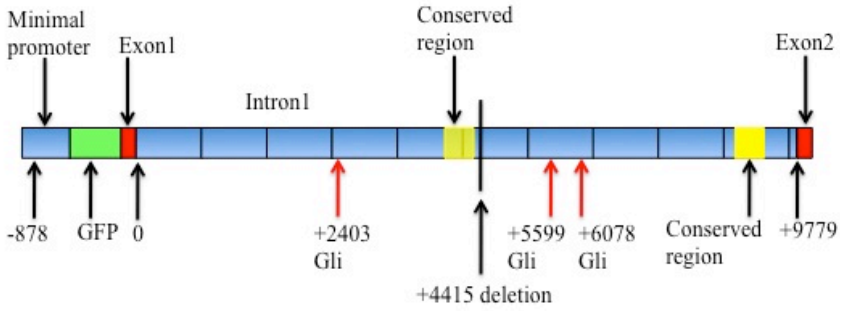

Figure 5.26: *A schematic representation of intron 1 of zebrafish lama1 in BAC zC34A17, including the minimal promoter sequence upstream of the GFP cassette.* Intron 1 is 9779 base long, and contains two regions of DNA that are conserved with intron 1 of *lama1* in *Fugu* fish (yellow boxes at positions 3986-4338 and 8859-9208 of intron 1). Three potential Gli binding sites are located in intron 1, at positions +2408, +5599, and +6078, marked with red arrows. The site of deletion within intron 1 is marked with a black arrow at position +4415. The sequence to the left of position +4415 is contained within the Δmin.prom*lama1*GFPΔ5kb BAC construct.

	M	**I**	**12.7**	**1**	**30/31**
Primer pair	iTol end F, Min prom. R	Intron1 F, BAC backbone Rev	12.7kb up F, 12.7kb up R	Lama1 F1, Lama1 R1	Ex30/31 F, Ex30/31 R
Size (bases)	2030	1627	501 (should not work)	1200	705 (should not work)

Table 5.7: *Primer pairs used for BAC PCR analysis.* Primer pairs M consists of a forward primer with homology to the end of the forward iTOL ultramer sequence, and a reverse primer with homology to the sequence just after the minimal promoter reverse ultramer. A 2030 base band is produced when homologous recombineering has been successful, with the insertion of the iTOL2-AMP construct in place of the upstream DNA sequence. Primer pair I consists of a forward primer with homology to the sequence just upstream of the forward ultramer in intron 1, and a reverse primer with homology to the sequence just downstream of the reverse BAC backbone (vector DNA) ultramer. A 1627 base band is produced when homologous recombineering has been successful, with the insertion of the Kanamycin gene in place of the exon-intron, and downstream DNA sequence. Primer pair 1 consists of primers with homology to sequence that is either side of the GFP reporter gene, and produce a 1200 base fragment. Primer pair 30/31 amplifies a sequence that spans exon 30, intron 30, and exon 31 of the *lama1* gene. This is absent in Δmin.prom*lama1*GFPΔ5kb, but present in *lama1* BAC transgene (FL) construct control DNA.

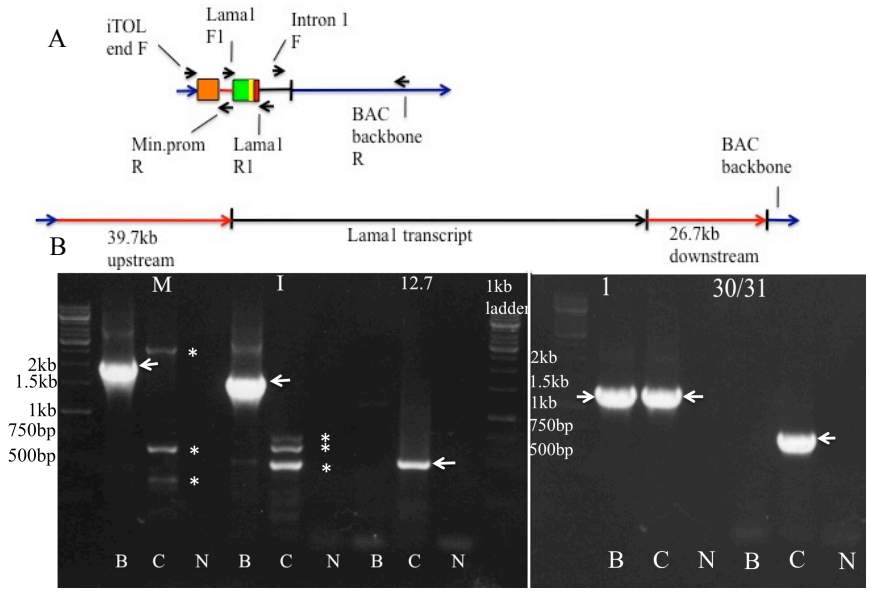

Figure 5.27: *PCR analysis of Δmin.promlama1GFPΔ5kb BAC midi prep DNA confirms successful deletion of the upstream and downstream DNA sequences*. A: A schematic representation of the BAC construct and primer positions. B: Primer pair M is the expected size and is only present if the upstream DNA sequence has been removed, with the associated insertion of the iTOL2-AMP construct. For this reason, a band at the expected size is not produced when using *lama1* BAC transgene (FL) construct DNA as a control. Non-specific DNA fragments are amplified in this case. Primer pair I is the expected size and is only present if the downstream DNA sequence has been removed, with the associated insertion of the Kanamycin resistance gene. For this reason, a band at the expected size is not produced when using *lama1* BAC transgene (FL) construct DNA as a control. Non-specific DNA fragments are amplified in this case (asterisks). Primer pairs 12.7 and 30/31 confirm the absence of upstream and downstream DNA sequences, respectively. Primer pair 1 confirms the presence of the GFP reporter gene. Primers used for each reaction, and the expected fragment sizes (indicated by white arrows) are listed in *Table 5.7*. Size of ladder is indicated on the left-hand side. Abbreviations: BAC DNA (B), *lama1* BAC transgene (FL) construct control DNA (C), no DNA control (N).

At the 19-20-somite stage, embryos injected with Δmin.prom*lama1*GFPΔ5kb show tissue specific GFP expression in the muscle fibres (n = 98/115 (85%), compared to 78% 15-somite stage *lama1* BAC (FL) embryos (n = 58), notochord (n = 27/115, (23%), compared to 37% *lama1* BAC (FL) embryos (n = 30), and the PSM (n = 16/115 (14%), compared to 30% *lama1* BAC (FL) embryos (n = 30) (*Figure 5.28II*A,B,D). Although the number of embryos expressing GFP in the notochord and PSM is reduced in comparison to *lama1* BAC (FL)-injected embryos, expression in these tissues is strong and much more widespread in comparison to the few expressing cells observed in embryos injected with Δ3.3kb*lama1*GFPΔex1 BAC construct (*Figure 5.25*IIB). Despite widespread GFP expression in these tissues, there is a noticeable absence of expression in the anterior CNS and eye. Only 13% of embryos (n = 15/115) express GFP in the anterior CNS (*Figure 5.28*IIC) (compared to 47% *lama1* BAC (FL) embryos (n = 30), and this expression consists of just a few GFP fluorescent cells (similar to Δ3.3kb*lama1*GFPΔex1 embryos (7%)). GFP expression in the eye occurs in just 1/115 embryos at the 19-20-somite stage, and this is in just a single cell. GFP expression in the anterior CNS and eye driven by the construct

Δmin.prom*lama1*GFPΔ5kb therefore correlates with the GFP expression pattern obtained after injection of the Δ3.3kb*lama1*GFPΔex1 BAC construct DNA.

At 25-26hpf, GFP is again strongly expressed in the muscle fibres (n = 106/123 (86%) embryos, comparable to 76% of *lama1* BAC (FL)-injected embryos (n = 289) in embryos injected with Δmin.prom*lama1*GFPΔ5kb) (*Figure 5.29*A-E). Expression is also detected in the ventral vasculature region (n = 4/123, 3%) (*Figure 5.29*E), the uro-genital region (1%, compared to 11% *lama1* BAC (FL) embryos), and the notochord (n = 2/123 (2%), compared to 30% of *lama1* BAC (FL) embryos) (*Figure 5.29*F), but not in the PSM (*Figure 5.29*A,B,F). As in 19-20-somite stage embryos, the number of embryos expressing GFP in the notochord and PSM is reduced in comparison to *lama1* BAC (FL)-injected embryos. The number of embryos expressing GFP in the notochord of Δmin.prom*lama1*GFPΔ5kb-injected embryos at 25hpf is also significantly reduced compared to the number of embryos expressing GFP in the notochord at the 19-20-somite stage (23% embryos express GFP) (*** p-value = 0.0002) (*Figure 5.31*). This could suggest that, although GFP expression is initiated on time in the notochord, maintenance of GFP expression in the notochord is affected in embryos injected with Δmin.prom*lama1*GFPΔ5kb construct DNA. The enhancers that maintain notochordal *lama1* expression may therefore be lost following deletion of DNA sequences downstream of base +4415 in intron 1. However, there is also a reduction in the number of embryos expressing GFP in the notochord between the 15-somite stage and 25hpf in embryos injected with *lama1* BAC (FL) DNA (p-value = 0.1686, not significant) (*Figure 4.14a*), although the difference is small compared to Δmin.prom*lama1*GFPΔ5kb-injected embryos (37% 20-somite stage (n = 30), 30% 25hpf (n = 289) in *lama1* BAC transgene (FL)-injected embryos compared to 23% 19-somite stage (n = 115), 2% 25hpf (n = 123) Δmin.prom*lama1*GFPΔ5kb-injected embryos).

Similarly, there is a reduction in GFP expression in the PSM between the 19-somite stage and 25hpf in Δmin.prom*lama1*GFPΔ5kb-injected embryos. 14% of 19-somite stage embryos (n = 115) express GFP in the PSM, whilst no 25hpf embryos (n = 123) express GFP in the PSM (*Figure 5.31*). In comparison, 30% of 20-somite stage (n = 30) *lama1* BAC (FL)-injected embryos express GFP in the PSM, and 21% of 25hpf embryos (n = 289) (*Figure 4.14b*). Enhancers that maintain the expression of GFP in the tail may also be lost in the min.prom*lama1*GFPΔ5kb BAC construct following removal of all DNA sequences downstream of +4416 in intron 1. In agreement, no GFP is detected in the notochord or PSM at 49hpf in min.prom*lama1*GFPΔ5kb-injected embryos, but is observed in the notochord in 2% of *lama1* BAC (FL) (n = 108) and 3% of Δ3.3kb*lama1*GFPΔex2-injected embryos (n = 35), and in the tail in 20% of *lama1* BAC (FL)-injected embryos, at 49hpf. The -3291bp sequence upstream of the *lama1* gene may also be required for the maintenance of *lama1* expression in the notochord and PSM.

In line with the observations made at the 19-20-somite stage, widespread GFP expression is

lacking in the anterior CNS at 25hpf (*Figure 5.29*A). 34% (n = 123) of embryos do express some GFP in the anterior CNS (compared to 67% *lama1* BAC (FL) embryos (n = 289)), although only in one or two cells (*Figure 5.29*A). A few embryos show GFP expression in several cells of the anterior CNS (*Figure 5.29*B,D), and 2% of embryos express GFP within the retina (n = 123) (*Figure 5.29*C). In contrast, 44% of *lama1* BAC (FL)-injected embryos display widespread GFP expression in the retina (n = 289).

By 49hpf, the number of embryos expressing GFP in the anterior CNS and eye is still reduced in comparison to embryos injected with Δ3.3kb*lama1*GFPΔex2 DNA (*Figure 5.22*Ai) or *lama1* BAC transgene (FL) DNA (*Figure 5.7*A). Only 11/105 (10%) embryos show GFP expression in the anterior CNS, and this expression mainly consists of just a few GFP fluorescent cells (*Figure 5.30*A). By comparison, 46% of Δ3.3kb*lama1*GFPΔex2-injected embryos (n = 35) strongly express GFP throughout the anterior CNS. Expression of GFP in the retina (*Figure 5.30*Ai), heart (*Figure 5.30*C), uro-genital region (*Figure 5.30*E), and the otic vesicle (*Figure 5.30*C) is only observed in 1/105 (1%) embryos. Expression in the heart and otic vesicle consists of only one GFP-expressing cell. No embryo expressing GFP in the pectoral fin was recorded. In contrast, GFP is expressed in the heart of 11% and 12% of Δ3.3kb*lama1*GFPΔex2 and *lama1* BAC transgene (FL)-injected embryos, respectively, in the otic vesicles of 34% of Δ3.3kb*lama1*GFPΔex2-injected embryos, in the uro-genital region of 3% of *lama1* BAC transgene (FL)-injected embryos, and in the pectoral fins in 6% and 5% of Δ3.3kb*lama1*GFPΔex2 (n = 35) and *lama1* BAC transgene (FL)-injected embryos (n = 108), respectively.

However, expression of GFP within the muscle fibres remains as strong as the expression observed in embryos injected with *lama1* BAC (FL) (50% embryos) or Δ3.3kb*lama1*GFPΔex2 (63% embryos) DNA (compare *Figures 5.30*A,B,D,E, *5.7*A, and *5.22*C,F,G). Widespread expression is detected in 66/105 (63%) embryos. Despite this persistent and strong GFP expression in the muscle fibres at 49hpf, the number of embryos that express GFP in this tissue has significantly decreased compared to 25hpf embryos (* p-value = 0.0351) (*Figure 5.31*). This is in agreement with the data shown in *Figures 4.14a* and *4.15*. There is also a significant increase in the number of embryos that do no express GFP in any tissue at 49hpf (** p-value = 0.0036) (*Figure 5.31*), also in agreement with the data shown in *Figure 4.14b* and *4.15*, which shows that GFP expression, in addition to the number of GFP-expressing embryos, is down-regulated in most tissues of *lama1* BAC (FL)-injected embryos with increasing developmental age.

Results suggest that the minimal promoter of *lama1* likely lies in the first 878 bases upstream of the gene, and enhancers controlling GFP expression within the muscle fibres, pre-somitic mesoderm, and the notochord are located within the first 4415 bases of intron 1. GFP reporter gene expression is correctly initiated in these tissues, and widespread and intense GFP signal is detected at comparable levels to embryos injected with *lama1* BAC transgene (FL) DNA or

Δ3.3kb*lama1*GFPΔex2 DNA at the 19-20-somite stage. However, the maintenance of GFP expression in the notochord and PSM may have been affected upon removal of either bases +4416 to +9779 within intron 1 or the region -3291 to -878bp upstream of the start site. Interestingly, GFP reporter gene expression in the anterior CNS, eye, otic vesicle, pectoral fin, and the heart is lacking at all developmental stages analysed. This suggests that enhancers controlling the correct initiation of *lama1* in these tissues are absent. As GFP expression was consistently absent or very weak up to 49hpf, it indicates that GFP expression is not simply delayed in these tissues. It is likely therefore that enhancers located between bases +4416 and +9779 in intron 1 control the initiation of *lama1* in the anterior CNS, eye, otic vesicle, pectoral fin and the heart. Although GFP expression is observed in the vasculature and uro-genital region, it is expressed in fewer embryos compared to *lama1* BAC (FL)-injected embryos, raising the possibility that for efficient expression of *lama1* in these tissues, enhancers located between bases +4416 and +9779 in intron 1 may be required in addition to enhancers located between bases +1 and +4415.

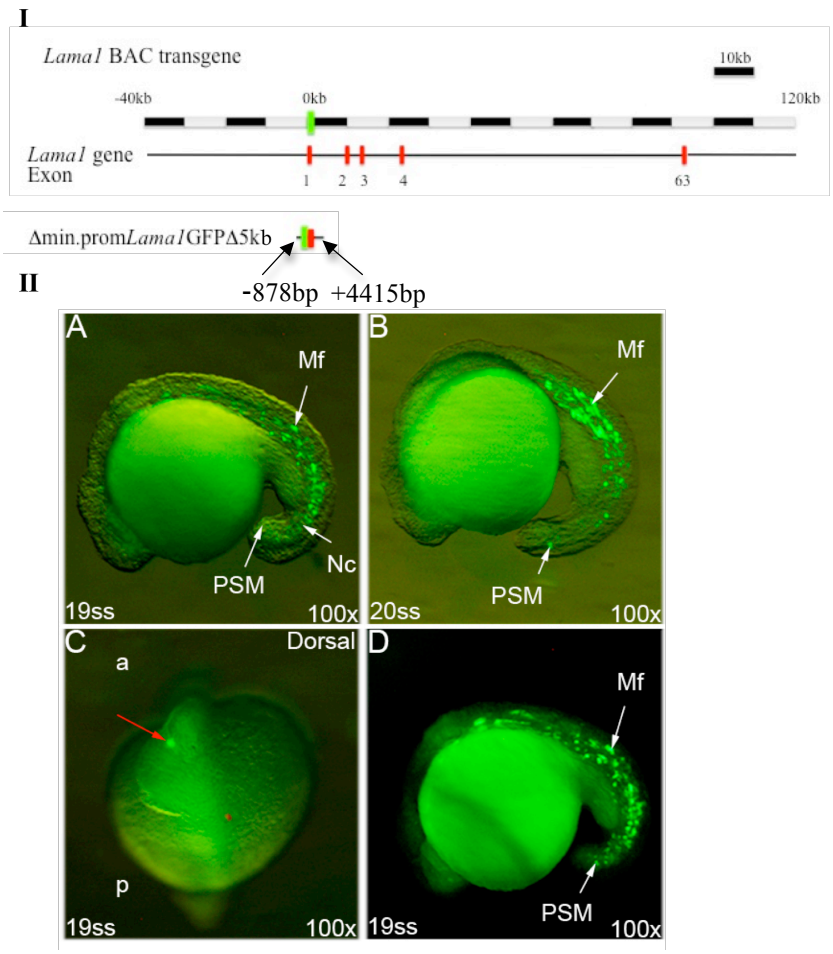

Figure 5.28: I: *A schematic representation of the lama1 gene, and the Δmin.promlama1GFPΔ5kb construct.* The Δmin.prom*lama1*GFPΔ5kb construct consists of -878 bases directly upstream of the *lama1* start site, the GFP reporter gene, exon 1, and the first 4415 bases of intron 1. All DNA sequence downstream of this point is deleted. Red blocks represent exons, green blocks represent the GFP reporter gene. II: *Analysis of GFP expression at the 19-20-somite stage after injection of 50pg Δmin.promlama1GFPΔ5kb BAC construct DNA into the one-cell stage embryo.* A-D: Lateral views of 19-somite stage (A,C,D) or 20-somite stage (B) embryos at 100x magnification. GFP expression is detected in the muscle fibres (A,B,D) notochord (A) and PSM (A,B,D). Note the absence of GFP expression in the anterior CNS of the embryos in panels A, B and D. A dorsal view of another embryo shows a single GFP fluorescent cell in the telencephalon region (C) (red arrow). Anterior of lateral views is to the left. Anterior of dorsal view is at the top of the panel. Abbreviations: muscle fibre (Mf), pre-somitic mesoderm (PSM), notochord (Nc), anterior (a), posterior (p).

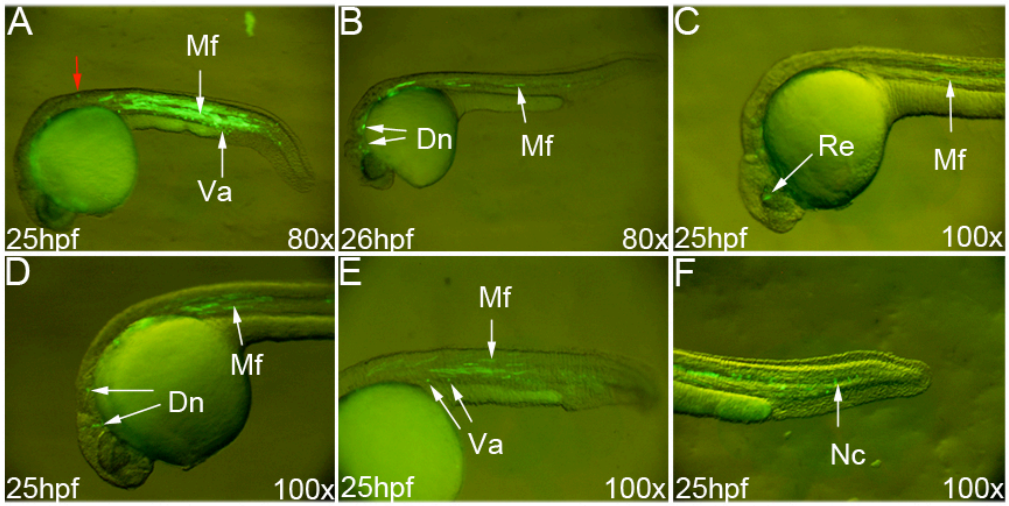

Figure 5.29: *Analysis of GFP expression at 25hpf after injection of 50pg Δmin.promlama1GFPΔ5kb BAC construct DNA into the one-cell stage embryo*. A-F: Lateral views of 25hpf embryos at 80x (A,B) or 100x (C-F) magnification. Strong GFP expression is observed in the muscle fibres (A-E) and the vasculature (A,E), and is also noticeable in the notochord (F). Note that every embryo shown continues to expresses GFP in the muscle fibres. A few embryos express GFP in the diencephalon (B,D) and the retina (C), whilst some embryos display a single GFP-expressing cell in the anterior CNS (A) (red arrow). Anterior is to the left. Abbreviations: muscle fibre (Mf), vasculature (Va), diencephalon (Dn), retina (Re), notochord (Nc).

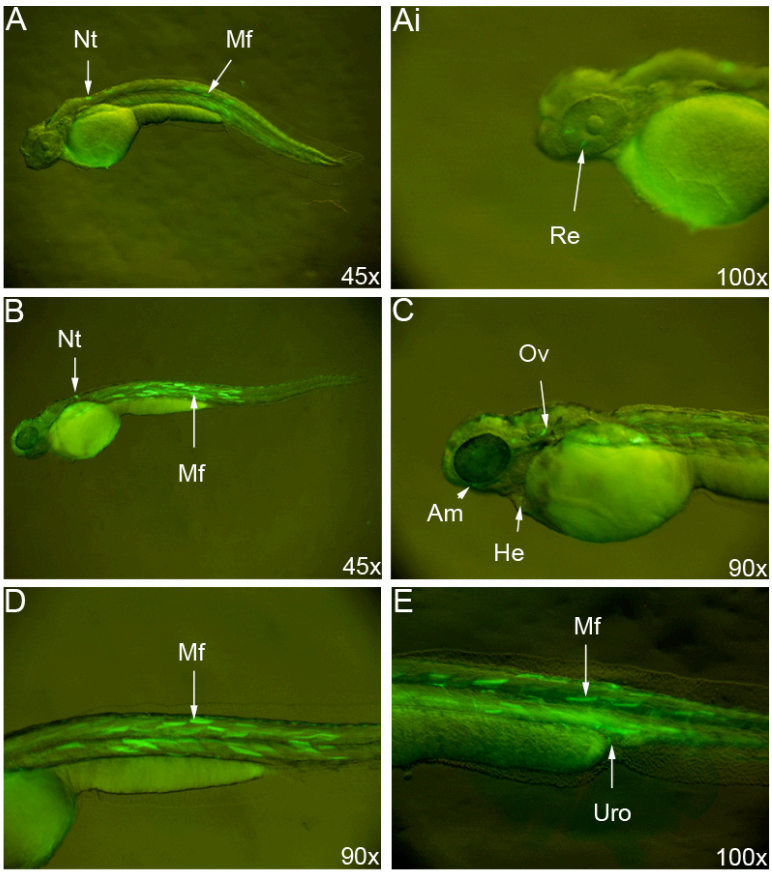

Figure 5.30: *Analysis of GFP expression at 49hpf after injection of 50pg Δmin.promlama1GFPΔ5kb BAC construct DNA into the one-cell stage embryo*. A and Ai: Lateral views of the same embryo at 45x (A) and 100x magnification (Ai) showing GFP expression in the neural tube (A), muscle fibres (A), and retina (Ai). B and D: B and D: Lateral views at 45x (B) and 90x (D) magnification with an increased amount of GFP expression in the muscle fibres and neural tube in comparison to panel A. C: Lateral view at 90x magnification, showing GFP in the otic vesicle, heart, and jaw muscle (Am). E: Lateral image at 100x magnification showing GFP in the muscle fibres, and a small region of the uro-genitial region. Anterior is to the left. Abbreviations: muscle fibre (Mf), neural tube (Nt), retina (Re), heart (He), otic vesicle (Ov), adductor mandibulae (Am), uro-genital region (Uro).

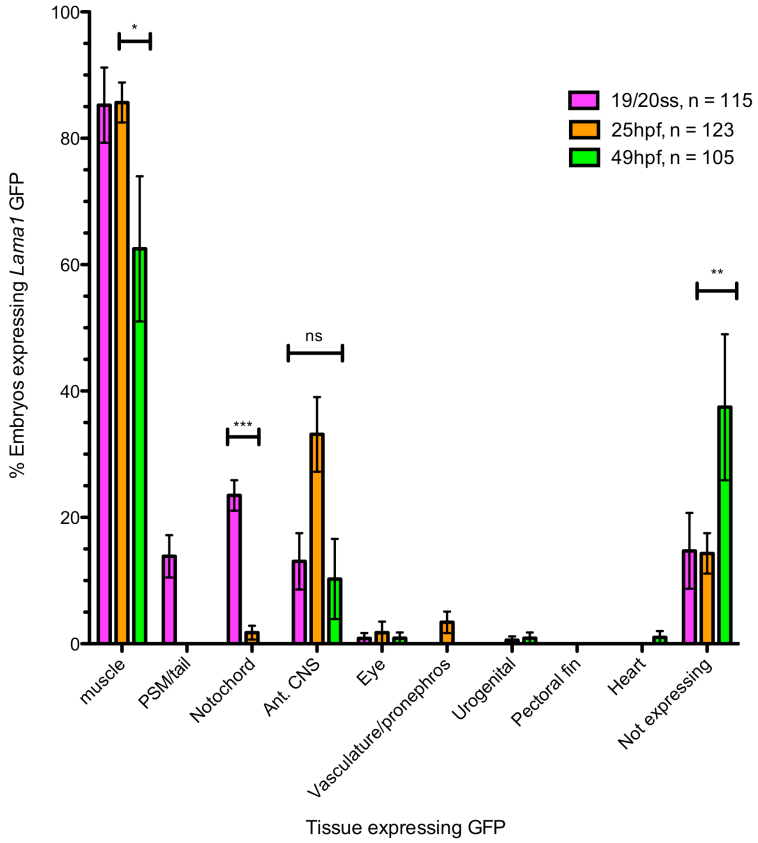

Figure 5.31: *A comparison of GFP expression in embryos injected with min.promlama1GFPΔ5kb BAC construct at different stages of development*. In agreement with the data shown in *Figures 4.14a* and *4.15* which show that GFP is down-regulated in embryos with increasing developmental age, there is a significant down-regulation of GFP expression after 25hpf (p-value = 0.0351), although GFP expression still remains at 49hpf in the muscle fibres. There is also a significant reduction in the amount of embryos expressing GFP in the notochord and PSM between the 19-20-somite stage and 25hpf (p-value = 0.0002), and in the number of embryos expressing any GFP between 25hpf and 49hpf (p-value = 0.0036). PSM expression is only detected at the 19-20-somite stage. Statistical test used was the un-paired, two-tailed t-test. n = the number of zebrafish embryos analysed. ns = not significant. Standard error of the means is shown on the graph. Abbreviations: anterior CNS (ant. CNS), pre-somitic mesoderm (PSM).