Development of a PhiC31 system for functional characterisation of *cis*-regulatory elements in reporter transgenic zebrafish

by

Jennifer Anne Roberts

A thesis submitted to the

University of Birmingham

for the degree of

DOCTOR OF PHILOSOPHY

Department of Medical and Molecular Genetics

School of Clinical and Experimental Medicine

College of Medical and Dental Sciences

The University of Birmingham

September 2013

Abstract

Current methods to study enhancers *in vivo* involved integrating DNA sequences into the genome in a random position, in zebrafish the transposase Tol2 is utilised. However, this random genome integration can result in position effects which cause variations in reporter gene expression patterns preventing conclusions from being drawn about the enhancer being tested. To address the issue of variable position effects when using random integration techniques to study enhancers, this project aimed to establish a PhiC31 integrase system of targeted transgenesis in zebrafish. Firstly, we designed and tested a multicomponent system including a recipient vector, donor vector and PhiC31integrase mRNA. These components were co-injected into one cell stage embryos to test the integration function of PhiC31 integrase in zebrafish. The integration was detected by designing the recipient and donor vectors such that a switch from GFP to mCherry lens expression marks embryos with legitimate recombination. Over 97% of fluorescent embryos showed mCherry expression in the lens after co-injection of the three components indicating that integration of the two vectors by PhiC31 integrase is highly efficient. Legitimate recombination was shown to have occurred using PCR and sequencing techniques from total RNA extracted from embryos. A recipient transgenic line was made into which the donor vector and PhiC31 integrase RNA were co-injected. This resulted in around 70% of fluorescent embryos showing mCherry expression in the lens which suggests targeted genome integration is also efficient in zebrafish embryos. To establish whether variability of positional effects could be reduced using this system, an enhancer-promoter-reporter construct was modified to

include donor vector sequences and injected into recipient line transgenic zebrafish embryos. Expression patterns observed in stable lines made with either Tol2 transposase or PhiC31 integrase suggest that variability in position effects is reduced when using the PhiC31 integrase system. Together, these results provide evidence that the PhiC31 integrase system I have tested is suitable for a number of uses in zebrafish, including enhancer screening.

Acknowledgements

I would like to thank Dr Ferenc Mueller for giving me the opportunity to study for this PhD. I would also like to thank Dr Yavor Hadzhiev for his advice throughout the project. Thanks also go to our collaborator, Dr Darius Balciunas, who has contributed throughout the project giving advice and ideas.

I also want to thank all previous and current lab members who I have thoroughly enjoyed working alongside these past few years: Sundeep Dhillon, Eva Doro, Jochen Gehrig, Harmeet Gill, Emma Kenyon, Nan Li, Irene Miguel, Padma Putta, Matt Rawlings and Andreas Zaucker,.

I am especially grateful to Holly Smith and James Bull for their support and advice. Finally, I thank my family and friends for all of their support and encouragement during my studies.

Table of Contents

List of Figures

List of abbreviations

3C	Chromatin conformation capture
AD	Activation domain
APP	Amyloid precursor protein
BAC	Bacterial artificial chromosome
BEAF	Boundary element associated factor
BFP	Blue fluorescent protein
bp	Base pair
BSA	Bovine Serum Albumin
CD	Campomelic dysplasia
CFP	Cyan fluorescent protein
ChIP-seq	Chromatin immunoprecipitation-sequencing
CNE	Conserved non-coding elements
CPE	Core promoter element
DBD	DNA binding domain
DNA	deoxyribonucleic acid
dNTP	deoxynucleotide
Dpf	Days post fertilisation
ECL+	Enhanced chemiluminescence
EDTA	Ethylenediaminetetraacetic acid
ER	Estrogen receptor
ERE	Estrogen response element
FRT	Flippase recognition target
GFP	Green fluorescent protein
GTF	General transcription factor
H3K27ac	Histone H3 lysine 27 acetylation
H3K27me3	Histone H3 lysine 27 trimethylation
H3K4me1	Histone H3 lysine 4 monomethylation
hpf	Hours post fertilisation
IFN-β	Interferon β

inr	Initiator
kb	Kilo base pairs
krt4	Keratin4 minimal promoter
MAR	Matrix attachment region
Mb	Mega base pairs
NEB	New England Biolab
PBST	Phosphate buffered saline with Tween-20
PCR	Polymerase chain reaction
PIC	Pre-initiation complex
PTU	Phenyl-2-thiourea
RMCE	Recombination-mediated cassette exchange
RPM	Rotations per minute
SB	Sleeping beauty
SDS	Sodium dodecyl sulfate
SHH	Sonic hedgehog
TAF	TBP-associated factor
TALEN	Transcription activator-like effector nucleases
TBS	Tris buffered saline
TEMED	Tetramethylethylenediamine
TF	Transcription factor
TSS	Transcription start site
UCR	Ultra-conserved regions
UCSC	University of California Santa Cruz
UTR	Untranslated region
YFP	Yellow fluorescent protein
zp3	zona pellucida 3
ZRS	zone of polarizing activity regulatory sequence

1 INTRODUCTION

1.1 Gene Regulation

Controlled differential gene expression is vital for cell survival, cell differentiation and cellular processes including apoptosis, proliferation and specific cellular function (Beuling et al. 2012). The regulation of gene expression predominantly occurs at the stage of transcription initiation and in eukaryotes RNA Polymerase II is responsible for transcribing protein-coding genes (Sims et al. 2004). The initiation of transcription can be regulated by regulatory DNA that effects gene expression but is not protein-coding sequence (cis-regulatory elements) (Bonn & Furlong 2008). These *cis*-regulatory elements can be adjacent to the core promoter, referred to as proximal promoters (Tanimoto et al. 2008); or they can be up to 1 mega base (Mb) away from the promoter, known as enhancers (Amano et al. 2009). The combination of these regulatory sequences leads to strict spatial and temporal control of tissue- and stage-specific genes (Sharpe et al. 1998). Loss of function of these regulatory sequences can cause a number of genetic diseases and so it has become clear that the study of regulatory DNA is highly important to understand these genetic disorders. For example, the lack of strict spatial and temporal control of sonic hedgehog (SHH) expression in humans can lead to the growth of extra digits, preaxial polydactily, when SHH is ectopically expressed in the anterior limb during development (Maas et al. 2011).

1.1.1 Core Promoter

The core promoter is the minimum DNA sequence required for transcription to be initiated and usually spans a region 50 base pairs (bp) up- and down-stream of the transcription start site (Butler & Kadonaga 2002). Initiation of transcription is dependent upon recruitment of the pre-initiation complex (PIC) to the core promoter (Becker et al. 2002). The PIC is comprised of RNA polymerase II and associated factors (including TFIIA, TFIIB, TFIID, TFIIE, TFIIF, and TFIIH). Different proteins in the pre-initiation complex bind to different motifs within the core promoter (Thomas & Chiang 2008). A promoter motif is a sequence of DNA that specific transcription factors bind to (Schmid 2008). There are two models of recruitment of the PIC to the core promoter; the first is a sequential model as the components of the PIC are recruited to the core promoter separately. The other model suggests recruitment of the PIC as a pre-formed holoenzyme (Ossipow et al. 1995).

Core promoter motifs

The main core promoter motifs are listed below (Thomas & Chiang 2008).

- TATA box (TATA) – Usually approximately 30 bp upstream of the transcription start site (TSS) and is bound by TBP, a subunit of TFIID.

- Initiator (Inr) – encompasses the transcription start site and is bound by two TBP-associated factors (TAF), TAF1 and TAF2 which are part of TFIID.

- Downstream core promoter element (DPE) – bound by TFIID (thought to be subunits TAF6 and TAF9) and is important for basal transcription. This functions with the Inr and the distance between these motifs is strictly regulated.

There have been several studies showing that core promoter motifs and their positions in the core promoter can influence long-range *cis*-regulation of transcription; these will be discussed below.

1.1.2 Proximal promoters

A proximal promoter is a DNA sequence directly upstream, up to a few hundred basepairs, of the core promoter (Figure 1-1A; Maston et al. 2006). It often contains regulatory DNA which activators and repressors can bind to and influence spatial and temporal control of gene expression.

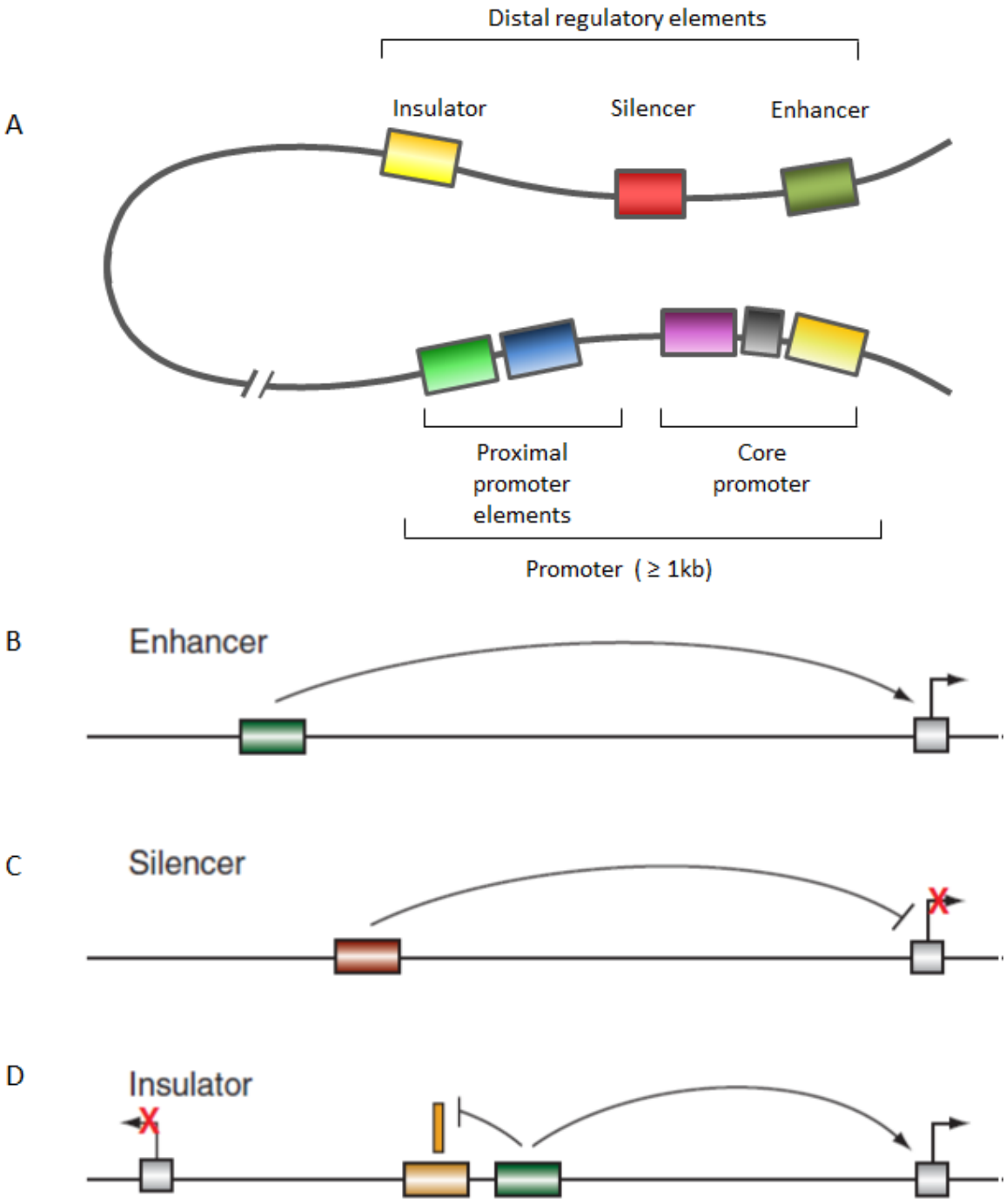

Figure 1-1 Schematic showing cis-regulatory elements influencing transcription

(A*) shows the positions of the cis-regulatory elements in relation to the core promoter. (B-D) show how each of the cis-regulatory elements exerts their control over gene expression. Figure adapted from (Maston et al. 2006).*

1.1.3 Distal-acting cis-regulatory regions

Cis-regulatory elements are sequences of DNA that can influence gene expression by either activation or repression of transcription and are located on the same chromosome as the core promoter. These *cis*-regulatory regions can act over large distances and are said to be distal-acting. The binding of the pre-initiation complex components to the core promoter motifs often requires activator proteins and mediators and it is through these proteins that long-range *cis*-regulatory elements can influence gene expression (Figure 1-1).

1.1.3.1 Enhancers

Enhancers are *cis*-regulatory elements which usually act over large distances to activate transcription of specific genes (Figure 1-1B). The activator protein(s) bind to the enhancer through their DNA binding domain (DBD) and in turn interact with either the mediator complex or directly with general transcription factors to aid the recruitment of the pre-initiation complex (Figure 1-2) (Alberts et al. 2008).

Activator proteins bound to the enhancer element are thought to interact with transcription factors bound to the core promoter and hence the enhancer and promoter are brought into close proximity (Figure 1-1A). Enhancers often contain a number of transcription factor binding sites that are bound by activators (Thomas & Chiang 2008). The balance of relevant transcription factors in the cell is dependent on signalling cascades influenced by many factors including the cell's environment, and the process

of differentiation (Zhou & Snider 2006). These transcription factors converge on the enhancer (called an enhanceosome) and the enhancer activates transcription from the target gene promoter (Pan & Nussinov 2011). The interferon-β (IFN-β) enhanceosome has been studied extensively and serves as a model for the activation of transcription by enhancers (Panne 2008). This IFN-β enhancer is located between two nucleosomes, one of which blocks the TATA box of the IFN-β promoter. A number of transcription factors including ATF-2/c-Jun dimer, IRF-3, IRF-7 and NFκB bind sequentially to the enhancer sequence. This enhancesome has been shown to recruit CBP/p300 (chromatin remodeling enzymes) which re-models the chromatin revealing the TATA box of the IFN-β promoter and hence activating transcription (Merika et al. 1998).

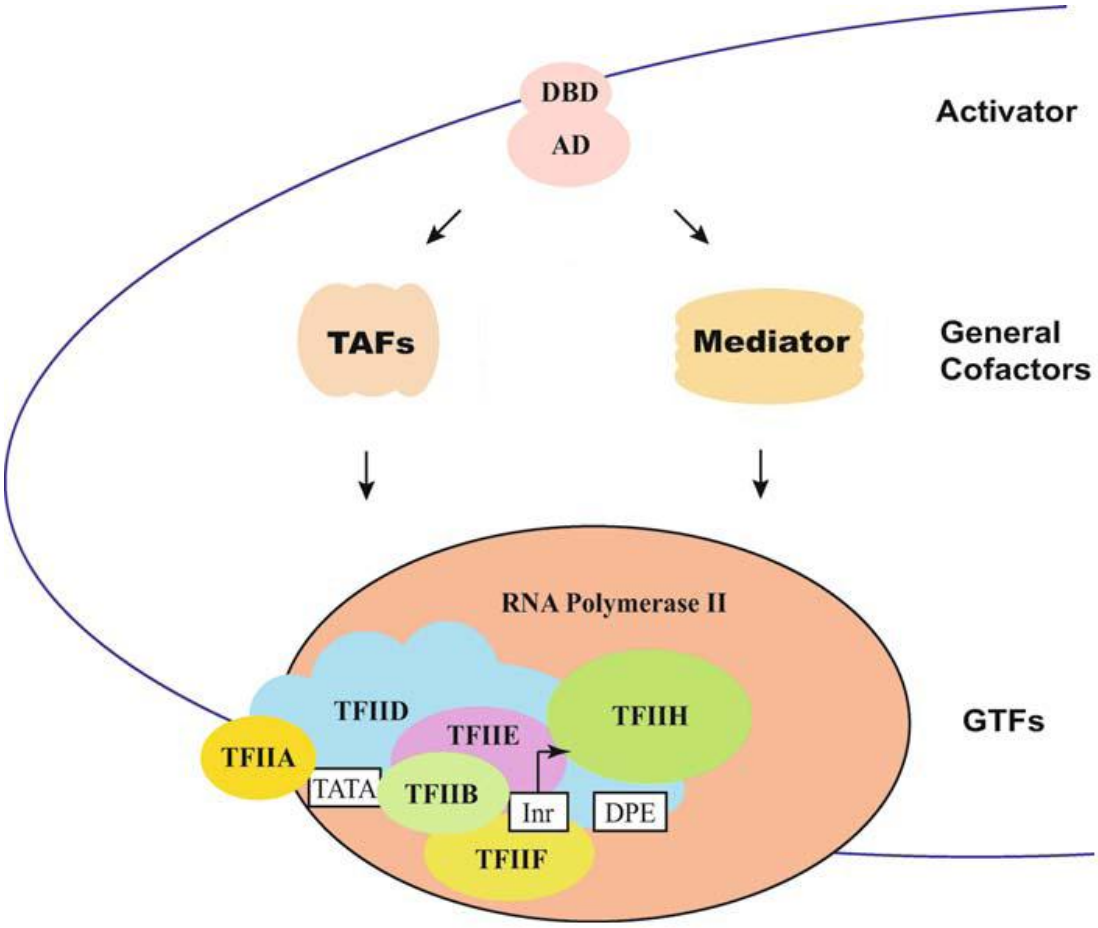

Figure 1-2 Schematic of cis-regulatory element function in regulation of gene transcription

The activator protein binds to the enhancer via its DNA binding domain (DBD) and can interact with TBP-associated factors (TAFs) or mediator proteins via the activation domain (AD) which can aid the recruitment of the general transcription factors (GTFs) to various elements of the core promoter including the TATA box (TATA), initiator (Inr) and the downstream promoter element (DPE). Adapted from (Thomas & Chiang 2008).

1.1.3.1.1 Identification of enhancers

Historically *cis*-regulatory elements have been difficult to identify as they can be located at large distances from their target genes in intergenic or intronic regions. However, recent developments in whole genome sequencing of many organisms have made this

task much easier. Some of the methods employed for the identification of *cis*-regulatory elements are discussed below.

Computational methods

Comparative genomics has aided the identification of cis-regulatory DNA (Duret & Bucher 1997; Allende et al. 2006). Functional DNA including protein-coding and *cis*-regulatory elements are known to be conserved throughout evolution which allows them to be identified by comparing the DNA sequences from the genomes of different species. Comparative genomics involves comparing genomes of either many different similar species or of two (or more) evolutionally distant species, for example human and fish (Hardison et al. 1997; Goode et al. 2003; Sandelin et al. 2004). Whole genomes can be aligned using these methods and the putative enhancers are then functionally tested *in vivo*.

As enhancers are known to contain clusters of transcription factor (TF) binding sites, these sites can be used to predict the location of *cis*-regulatory DNA (Berman et al. 2002). This method has limited success on its own as not all TF binding site clusters are enhancers (creates false positives) but it can be useful when used in conjunction with other methods of identifying enhancers (Yáñez-Cuna et al. 2013).

In vivo methods

A recent publication by Visel and colleagues used a ChIP-seq method where chromatin was immunoprecipitated with the p300 protein, which is known to be associated with enhancers (Visel et al. 2009). The chromatin fragments were then sequenced and aligned to the original genome to locate the *cis*-regulatory elements. This method was also effective in predicting when and where the putative enhancers were active depending on what tissue and stage were analysed. This has become a widely used tool to identify enhancers, however, p300 only binds to a proportion of enhancers so not all enhancers can be identified in this manner (Jin et al. 2011).

Recent technological advances have resulted in a number of chromatin signatures associated with enhancers being identified. Histone H3 lysine 4 monomethylation (H3K4me1) and histone H3 lysine 27 acetylation (H3K27ac) are marks of active enhancers while histone H3 lysine 27 trimethylation (H3K27me3) is a marker for 'poised' enhancers (Rada-Iglesias et al. 2011). Genome-wide ChIP-seq experiments pulling down enhancer-associated factors like p300 or modified histone marks such as H3K27ac can now be used as a tool to aid the identification of enhancers in fully sequenced genomes (Spicuglia & Vanhille 2012). ChIP-seq can also be carried out on specific tissues or cell types of interest to identify active enhancers in that tissue or cell type (Lo et al. 2011).

Cis-regulatory elements often fall in nucleosome free regions to allow access for transcription factors to bind the DNA; this leads to DNase hypersensitivity. These DNase hypersensitive regions (regulatory DNA) can be located using a DNase-chip method described in (Crawford et al. 2006). These nucleosome-free regions can also be identified using a method called Formaldehyde-assisted isolation of regulatory elements (FAIRE). This method involves cross-linking the DNA *in vivo* using formaldehyde and then shearing the DNA; the isolated DNA is fluorescently labelled and hybridised to a microarray (Giresi et al. 2007).

Using a combination of both multi-species sequence comparisons and *in vivo* techniques allows enhancers to be more readily identified and studied.

1.1.3.1.2 Verification of putative enhancers in vivo

There have been several methods used to study enhancers *in vivo*. A targeted approach involves cloning a CNE sequence upstream of an enhancer-susceptible promoter and a reporter gene. This has been introduced into zebrafish eggs in a vector and the embryos were analysed for transient reporter expression (Shin et al. 2005). Another approach utilised the Tol2 transposon system to integrate the CNE-reporter constructs into the zebrafish genome (Fisher et al. 2006). Many functionally tested enhancers are now recorded in databases, for example VISTA enhancer browser contains information about functionally tested human enhancers (Visel et al. 2007) and the conserved non-coding

orthologous regions (CONDOR) database contains information about conserved non-coding elements tested *in vivo* in zebrafish.

Non-targeted enhancer trap methods have also been used to identify and study enhancer function which gives the advantage of genomic context (Ellingsen et al. 2005). A novel approach was described in 2008 using enhancer-trap methods within bacterial artificial chromosomes (BACs) (Shakes et al. 2008). A basal promoter-reporter gene construct was inserted randomly into several places along the genomic DNA in the BAC using a transposon system. The various modified BACs were then introduced into zebrafish embryos and the reporter expression pattern used to map *cis*-regulatory elements in the genomic DNA of the BAC. In this study they also found that tissue-specificity directed by an enhancer can be context dependent. An enhancer of amyloid precursor protein (APP) found within intron 1 directed expression in the neurons when functioning with all other *cis*-regulatory elements, however when acting with the basal promoter, expression was restricted to the notochord (Shakes et al. 2008).

One example of a gene whose *cis*-regulatory elements have been studied at length is the sonic hedgehog (*shh*) gene. Shh is involved in many signalling pathways controlling differentiation during development and so it's spatiotemporal expression must be tightly regulated. A variety of the methods described above for identifying and testing putative enhancers were used in the discovery of the enhancers influencing Shh expression. Firstly, three intronic enhancers (ar-A, ar-B and ar-C) were identified in zebrafish that

direct Shh expression in the notochord and floor plate (Müller et al. 1999). These enhancers were identified by cloning fragments of DNA from around the *shh* locus and co-injecting them with a minimal promoter-reporter construct into early zebrafish embryos which were analysed for reporter expression during development (Muller et al. 1999). DNA fragments that drove expression of the reporter were further probed by carrying out deletions to identify the functional enhancer. Another *shh* enhancer (ZRS) was identified 1 megabase upstream of the *shh* TSS that mediates expression in the developing limb bud of the mouse (Lettice et al. 2003). This enhancer was identified when a preaxial polydactyly disease model mouse was investigated to find the mutation causing the genetic defect. In this investigation, an 8.5 kb DNA fragment was co-injected with a minimal promoter-reporter construct and was found to contain regulatory sequences. Comparative genomics was then employed to identify a conserved non-coding region within the 8.5 kb DNA fragment which was mapped to an intron of a nearby gene, limb region 1 (LMBR1). Three more enhancers were identified that directed Shh expression in the ventral forebrain in the mouse (Jeong et al. 2006). These were identified using an enhancer trap construct to screen a 1 MB region around *shh* using a BAC. Another three *shh* enhancers were identified recently that control expression in the epithelial lining of the oral cavity, lung and gut (Sagai et al. 2009); they were identified by comparing the genomes of several mammals and teleost fish. They were then functionally tested and found to recapitulate endogenous Shh expression pattern in the epithelial lining.

1.1.3.1.3 Core promoters impose specificity on enhancer interaction

In *DrosoPhila*, expression of the decapentaplegic gene (*dpp*) is controlled by an enhancer that is closer to two other genes, *oaf* and *slh*. Due to these genes showing different expression patterns, it was hypothesised that the enhancer specifically directed expression from the *dpp* promoter and not the promoters of *oaf* and *slh* (Merli et al. 1996). To test this theory, Merli et al replaced the *oaf* promoter with a *dpp*-compatible promoter and *oaf* was shown to display the same expression pattern as endogenous dpp. This shows that at least a proportion of promoters must contain a sequence or sequences that can affect their interaction with long-range enhancers. In this sort of scenario we refer to *dpp* as the target gene and *oaf* and *slh* are bystander genes. Specific motifs and how they are involved in specificity of interaction with enhancers are discussed below.

TATA-containing versus TATA-less promoters

Ohtsuki and colleagues showed that some enhancers in *DrosoPhila* preferentially interact with TATA-containing promoters (AE1 and IAB5) and some interact with TATA-containing and TATA-less promoters equally (NEE) (Ohtsuki et al. 1998). P transposon constructs were made containing an enhancer and two reporter genes under the control of two different core promoters and gene expression was analysed in transgenic *DrosoPhila* embryos by *in situ* hybridization using RNA probes. The constructs contained different combinations of core promoters; two TATA-containing promoters, two DPE/Inr-containing promoters or one promoter from each class. This study

demonstrated that some enhancers show selectivity of promoters depending on core promoter elements *in vivo* (Figure 1-3*).* This method did not eliminate position effects and so the results observed could be as a result of another *cis*-regulatory element in the area where the transgenes were integrated into the genome.

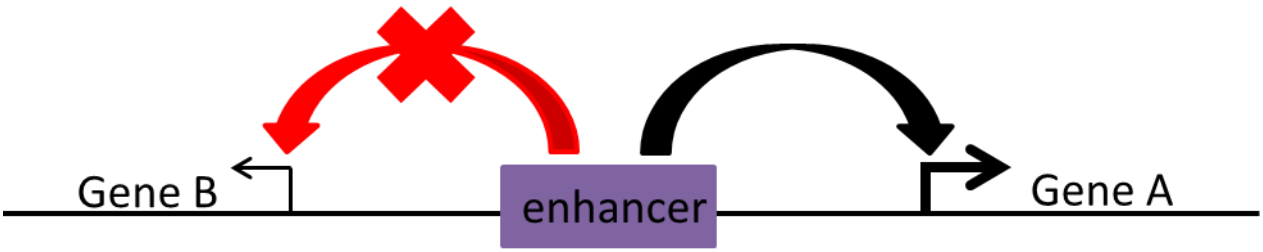

Figure 1-3 Schematic showing an enhancer preferentially binding to one gene over another

Figure adapted from (Ohtsuki et al. 1998)

TATA versus DPE-containing promoters

Butler and Kadonaga used an enhancer trap method to compare two promoters in the same genomic position (Butler & Kadonaga 2001). The enhancer trap construct was a transposon vector (see page 24) containing a pair of loxP sites flanking a TATA-GFP reporter and a pair of flippase recognition target (FRT) sites flanking a DPE-GFP reporter. The promoters were identical apart from the presence of either a TATA box or a DPE element and expression was shown to be dependent on these elements. The transposon was introduced into the *DrosoPhila* genome and stable enhancer trap lines were created. The Cre and FLP recombinases were then used to make pairs of sister *DrosoPhila* lines that contained either TATA-GFP or DPE-GFP in exactly the same position. From 18 pairs of *DrosoPhila* lines analysed they identified three DPE specific

enhancers, one TATA specific enhancer and 14 enhancers that did not show specificity for either TATA- or DPE-containing promoters. This method eliminated differential position effects on the transgenes between the pairs of lines and so the results do suggest that the TATA box and DPE element show specificity for certain enhancers.

Position and distance between promoter motifs

By sequence comparison and bioinformatics of promoters (between +500 and -100 relative to TSS), Vardhanabhuti et al identified the importance of the position of certain motifs and the distance between motifs in the regulation of gene expression (Vardhanabhuti et al. 2007). They also suggest that the combination of these motifs correlates with promoters of genes showing tissue-specific expression.

The core promoter sequence as a whole is important for transcription regulation

Wolner and Gralla aligned the MLP (high basal transcription) and E4 (strong activation-driven transcription) promoters according to the TATA box which was the only region showing conservation in 122 TATA-containing promoters that were compared (Wolner & Gralla 2000). The promoters were then divided into blocks which were then substituted between the promoters. For example the TATA block of wild type MLP was replaced by the TATA block of E4. By comparing the differences in general and activated *in vitro* transcription levels, they made the following conclusions: TATA and Inr elements affect general transcription levels but not the ratio of basal to activated transcription levels; however, the two blocks flanking the TATA block affect both general transcription and

the degree of activation. This study confirmed that not only the core promoter elements influence transcription regulation; the rest of the promoter sequence can also have an effect.

Core promoter elements and TAFs in regulation of gene expression

Following on from Wolner and Gralla's work looking at GC-rich regions flanking the TATA box, these regions were defined as the B recognition element (BRE) core promoter element (upstream GC-rich region) and the downstream GC-rich region was referred to as the G track. Chen and Manley show that a mutation in the BRE significantly increases general expression levels; however the activation potential of this promoter was the lowest of the promoters tested (Chen & Manley 2003). A mutation in the G track caused a modest reduction in general expression levels; however this promoter showed the greatest activation potential.

1.1.3.2 Insulator sequences

Insulator sequences can act as boundaries between euchromatin (transcriptionally active chromatin) and heterochromatin (transcriptionally inactive chromatin) preventing the spread of repressive chromatin structure (Sun & Elgin 1999). Insulators also show enhancer blocking activity; when an insulator sequence lies between an enhancer and a promoter, the interaction between those elements is prevented (Figure 1-1C) (Cai & Levine 1995; Kuhn & Geyer 2003). Insulator sequences can cause tertiary DNA structures, contain protein-binding sites and are associated with matrix attachment

regions (MARs) (Michel et al. 1993). These three findings could all influence chromatin structure and regulation of gene expression. However, when two *gypsy* insulator sequences are located between an enhancer and a promoter, the insulator function can be lost (Cai & Shen 2001). This pairing of insulators could result in activation of specific promoters by enhancers due to the specific chromatin structure, eg looping out, bringing the promoter and enhancer into close proximity (Maksimenko et al. 2008). Other insulator pairs do not show this loss of blocking activity (Kuhn et al. 2003). The study of insulator sequences has mainly been carried out using *DrosoPhila* as a model and although they are known to block enhancer activity, the mechanism is yet to be discovered. Insulators are known to contain protein binding motifs for example, CCCTC-binding factor (CTCF) in all eukaryotes (Bell et al. 1999) and boundary element-associated factor (BEAF) in *DrosoPhila* (Jiang et al. 2009). An insulator sequence has been identified in the proximal promoter of the spermatid-specific mouse gene, *SP10* which tethers the proximal promoter to the nuclear matrix in somatic cells and therefore represses transcription (Abhyankar et al. 2007). This demonstrates that insulator sequences can show similar functions in vertebrates as they do in *DrosoPhila*. Cohesin and CTCF play an important role in regulation of gene expression although the mechanisms have not yet been determined (Wendt & Peters 2009).

Promoter targeting sequences have been found to act as anti-insulator sequences which allow enhancers to interact with specific promoters despite there being an insulator sequence between them (Sipos et al. 1998; Zhou & Levine 1999; Calhoun et al. 2002;

Calhoun & Levine 2003; Chen et al. 2005; Akbari et al. 2008). This mechanism, if it applies to a number of genes, has great potential to influence the specificity of enhancer-promoter interactions.

1.1.3.3 Silencers

Silencers are specific DNA sequences that act to repress transcription from a specific promoter (Figure 1-1C); they are considered the opposite of enhancers.

1.1.4 Mechanisms of Enhancer-Promoter Interaction

The three most common models for enhancer-promoter interaction over long distances are tracking, linking and looping and will be discussed in more detail in the following sections.

Tracking

Tracking involves transcription factors binding to enhancer sequences, which then scan the DNA until they reach a promoter and transcription is activated (Herendeen et al. 1992). This model provides a possible mechanism of insulator sequences known to block enhancer activity when present between the enhancer and promoter.

Linking

This model proposes that transcription factors binding to the enhancer induces the binding of proteins along the DNA between the enhancer and the promoter allowing their interaction (Bulger & Groudine 1999). A protein thought to be involved in this mechanism was identified in *DrosoPhila*, called Chip (Morcillo et al. 1996) . Chip-related proteins have since been identified in vertebrates (Morcillo et al. 1997); mouse Clim-1 and -2 (Bach et al. 1997) and xenopus Xldb1 (Agulnick et al. 1996). These proteins are known to bind LIM (Lin-11, Isl-1 and Mec-3) domain-containing proteins but the precise functions in enhancer-promoter interaction facilitation are unknown.

Looping

This model suggests that protein-protein interactions occur between proteins bound to the enhancer and promoter-bound proteins with the DNA in between looping out (Bulger & Groudine 1999; Petrascheck et al. 2005). However, it has been shown that enhancer activity is position-independent in eukaryotes; this suggests there must be other mechanisms involved in facilitating the interaction.

Interaction Specificity

It was originally thought that enhancers were indiscriminate and would interact with any and all promoters within the locus. This idea came from enhancer trap methods where one promoter was combined with many enhancers *in vivo* (Weber et al. 1984; Wilson et

al. 1989). If this were true then all genes within the same genomic locus with a given enhancer would have the same expression patterns. The first example of enhancer-promoter specificity was shown in *DrosoPhila* studying the causes of different expression patterns of *gsb* and *gsbn* found in the same locus (Li & Noll 1994). To date, many enhancer-promoter interactions have been studied and have been found to be specific. So far there have been several proposed mechanisms for enhancer-promoter specificity but it seems there is no single universal mechanism, as it is likely combinations of these models are working simultaneously (Sipos & Gyurkovics 2005).

1.2 Mutations in cis-regulatory elements are involved in human disease

Many developmental genes are known to have *cis*-regulatory elements that control their expression. There are a number of examples of genetic disorders caused by translocations of chromosomes or point mutations which affect a genes' *cis*-regulatory region and not the coding sequence. This emphasises the crucial role that *cis*-regulatory elements (for example enhancers) have during development. To understand the mechanism of promoter-enhancer interaction and the specificity of this interaction could greatly increase our knowledge of these genetic disorders.

One example of a condition caused by a mutation in an enhancer is a rare syndrome affecting skeletal development and sex determination, campomelic dysplasia (CD). Mutations in the developmental gene *SOX9* have been shown to cause cases of CD

(Foster et al. 1994) (Wagner et al. 1994). It has also been shown by several groups that certain forms of the disorder can result from chromosome translocations affecting *cis*-regulatory elements and not alterations in the coding sequence of *SOX9* itself (Wunderle et al. 1998) (Pfeifer et al. 1999).

Similar DNA alterations have also been found in preaxial polydactyly patients, where a long range enhancer of the sonic hedgehog (*SHH*) gene (zone of polarising activity regulatory sequence (ZRS) enhancer) has a point mutation causing abnormal expression of the SHH developmental regulator (Gurnett et al. 2007). A case has been described where a translocation breakpoint lies within the fifth intron of the *LMBR1* gene (the location of the ZRS enhancer) in a patient with preaxial polydactyly (Lettice et al. 2002).

Aniridia is the absence of an iris and is thought to be caused by loss of one copy of the *PAX6* gene, Fantes and colleagues identified 2 patients with translocations that did not affect the *PAX6* gene itself. They suggested that this could be due to positional effects possibly affecting long distance *cis*-regulatory elements (Fantes et al. 1995).

Little is known about the mechanisms of interaction of enhancers and promoters and even less about the specificity of these interactions therefore warranting further study.

1.3 Zebrafish as a model organism

Zebrafish has been established as a great model for a wide variety of genetics experiments, these are discussed below. There are many advantages of using this model organism, including; external development of embryos, transparency of embryos, fully sequenced genome, relatively rapid development and the production of hundreds of embryos from a single crossing of an adult pair. Between 100 and 200 fertilised eggs can be collected from each pair of adult fish every week. This allows large scale experiments to be carried out on large numbers of embryos. The external development and transparency of the embryos allows the study of the development of organ systems during early development (Glass & Dahm 2004). It also allows the use of fluorescent reporters used to test putative enhancer sequences *in vivo* (Iklé et al. 2012).

Other methods used in genetics utilising this model include: morpholino gene knock-down (Nasevicius & Ekker 2000); targeted, germline genome mutagenesis using transcription activator-like effector nucleases (TALENs) (Dahlem et al. 2012; Bedell et al. 2012); disease modelling (Berman et al. 2003; Gama Sosa et al. 2012); behavioural genetics in adult fish (Norton & Bally-Cuif 2010); and screens for novel drugs (Langheinrich et al. 2002; Langheinrich 2003).

There are also many online resources aiding researchers using the zebrafish model. These include: zfin – The Zebrafish Model Organism Database (Sprague et al. 2001); Ensemble – a source of the annotated human, mouse and zebrafish genomes (Flicek et

al. 2012); CONDOR – an online database of conserved non-coding elements (CNEs) (Woolfe et al. 2007); and UCSC (University of California Santa Cruz) Genome Browser – allows convenient visualisation of genomes of a number of species (DiBiase et al. 2006).

1.3.1 Zebrafish development

In 1995 a paper was published detailing the developmental stages of zebrafish larvae up until 5 days post fertilisation (dpf) (Kimmel et al. 1995). This provides the information on all of the nomenclature used to describe development in the zebrafish larvae. Zebrafish development is far more rapid than other vertebrates; all major organs develop within the first 3 dpf and so these can easily be visualised while the embryo is still transparent. It also means that a sexually mature fish can be grown from fertilisation within around 3-4 months (Westerfield 2000), making transgenic zebrafish work much quicker than in mouse, for example.

1.3.2 Uses of zebrafish transgenesis

Transgenesis is a widely used method in zebrafish research and it is used for a number of types of study. For example, zebrafish transgenesis is used in tissue-labelling and *in vivo* imaging (Gilmour et al. 2002), disease-modelling (Langenau et al. 2005), the study of regulatory DNA (Kenyon et al. 2011) and the study of gene function (Min et al. 2012). Due to the popularity of transgenesis methods across the zebrafish research community, it is vital to optimise these methods to allow more research to continue.

1.4 Transgenesis technologies

A transgenic organism is one in which the deoxyribonucleic acid (DNA) sequence has been permanently altered by insertion of a foreign sequence of DNA. The original method of making transgenic zebrafish was by injecting linearised DNA into early embryos which can lead to concatemerisation of these fragments and integration into a chromosome in a random location (Stuart et al. 1988). More recent transgenesis methods utilise various transposon systems to integrate a single copy of a transgene into a random position in the genome. Examples include the P element transposon in *DrosoPhila* (Rubin & Spradling 1982) and other transposons in vertebrate models (Hogan et al. 1982; Kawakami et al. 1998). This involves cloning the transgene into a vector containing transposon arms (recognition sites) that facilitate the integration of the DNA sequence between them. In vertebrate models, including zebrafish, the mRNA of the transposase (enzyme that facilitates transposition) is co-injected with the DNA containing the transgene flanked by transposon arms into early embryos. The transgene becomes integrated into the genome in a proportion of cells. If the transgene is present in the gametes of the organism it can be passed on to some of it's offspring which will contain the transgene in all cells.

1.4.1 Transposons – SB and Tol2

Sleeping Beauty (SB) is a synthetic transposon system developed from inactive transposons of the Tc1/ mariner superfamily found in teleost fish genomes (Ivics et al. 1997). SB has been shown to function in mice (Luo et al. 1998), human cells (Ivics et al.

1997) and in zebrafish (Davidson et al. 2003). This system was found to increase transposition efficiency from 5 % to 31 % compared to standard plasmid micro-injection transgenesis techniques in zebrafish (Davidson et al. 2003).

The Tol2 gene was found in the genome of the teleost fish, medaka, and is from the hAT (hobo, Activator, Tam3) family of transposases (Koga et al. 1996). The Tol2 transposon system has been known to give transgenesis efficiencies of up to 50% compared to approximately 5-10% for the previous methods of micro-injecting plasmid DNA (Urasaki et al. 2006). This involves cloning Tol2 recognition sites either side of the DNA sequence to be integrated and then co-injecting this construct with Tol2 transposase mRNA (Urasaki et al. 2006). The efficiency of inheritance of the transposon by progeny of transgenic zebrafish was also greatly increased (Kawakami, Takeda, et al. 2004). It has been tested in zebrafish (Kawakami et al. 1998), mice (Kawakami & Noda 2004) and xenopus cells (Kawakami, Imanaka, et al. 2004). The Tol2 transposon system was shown to facilitate single copy transposition (by a cut and paste mechanism) as opposed to the numerous copies (20-100) observed using the SB system (Urasaki et al. 2008). The Tol2 transposon system is now widely used in zebrafish transgenesis for such purposes as the generation of stable transgenic lines, enhancer traps and transient expression assays (Kawakami 2007).

These transposon systems allowed much more efficient transgenesis of several model organisms. However, a disadvantage of using transposon systems is that integration into

the genome is random and so positional effects can vary greatly between the same construct injected into different embryos.

1.4.2 Site-specific reversible recombinases – Cre and Flp

To overcome the problem of variable position effects due to random integration, site-specific integration methods have been studied as a way of integrating different transgenes into the same locus in the genome. Cre and flippase (Flp) are two well-known site-specific phage recombinases and have been used in transgenesis methods. Cre recombinase and Flp mediate site-specific recombination between loxP and FRT sites respectively. Cre and Flp are both part of the tyrosine recombinase family. The mechanism involves cleavage of a single strand of DNA in the middle of two palindromic sequences of the loxP site and ligation occurs between swapped strands of DNA in another cleaved loxP site resulting in the formation of the Holliday junction (Hoess & Abremski 1985). This homologous recombination can be visualised in Figure 1-4 which shows the intermediates involved in the recombination between two loxP sites.

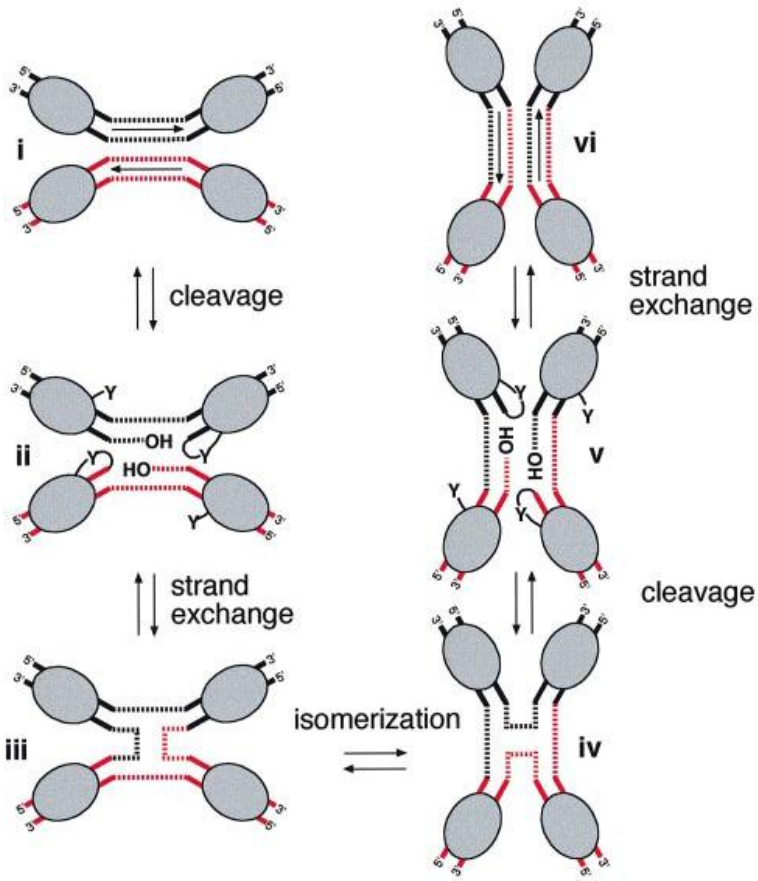

Figure 1-4 Mechanism of recombination between loxP sites catalysed by Cre

Schematic showing the steps and intermediates involved in recombination between two loxP sites during Cre-mediated reactions. Single strands of both loxP sites are cleaved and are ligated together after strand exchange. Isomerisation then takes place before the other strands are cleaved and ligated together after strand exchange. Figure adapted from (Gopaul et al. 1998)

Cre and Flp have been used to integrate two different transgenes into the same genomic locus in *DrosoPhila* (Siegal & Hartl 1996). This was done by cloning two transgenes between transposon arms; one flanked with loxP sites and the other with FRT sites. The construct was then injected into *DrosoPhila* embryos to make a transgenic line. Excision of either of the transgenes could then be induced using either Cre or Flp. This method has also been used for co-placement of two transgenes in the

mouse genome (Tanimoto et al. 2008). This method has limitations though; as only two transgenes can be compared in the same genomic position and each pair of sister lines will have a different transgene location.

1.4.3 Site-specific irreversible recombinase – ϕC31 integrase

More recently the possibility of using an integrase from the ϕC31 phage has been studied. The ϕC31 integrase facilitates the synapse of attP and attB integration sites, the natural function being to integrate the phage genome into the bacterial host genome (Kuhstoss & Rao 1991).

PhiC31 integrase is part of the serine recombinase family which have a different mechanism of recombination than the tyrosine recombinases described above. PhiC31 integrase binds to two recognition sites, attP and attB, and cleaves all four DNA strands simultaneously leaving a staggered break. Covalent bonds are formed between the 5' DNA end and the integrase; one half of the complex then rotates 180° relative to the other and re-ligation of the strands occurs (Smith et al. 2004).

The recombination between attP and attB sites in a foreign organism was first investigated in *Escherichia coli* (Thorpe & Smith 1998). PhiC31 integrase was shown to be directional due to the non-identical recognition sites. The recombination of attP and attB by PhiC31 integrase was also shown to occur *in vitro* proving that no co-factors were required (Thorpe & Smith 1998). Excision of DNA by recombination was shown to

work efficiently (>50%) in human cell lines transfected with a PhiC31 integrase expression vector containing attP and attB (Groth et al. 2000). In 2004, PhiC31 integrase was used to target integration of a donor vector containing attB into an attP-containing transgenic *DrosoPhila* line genome (Figure 1-5) (Groth et al. 2004).

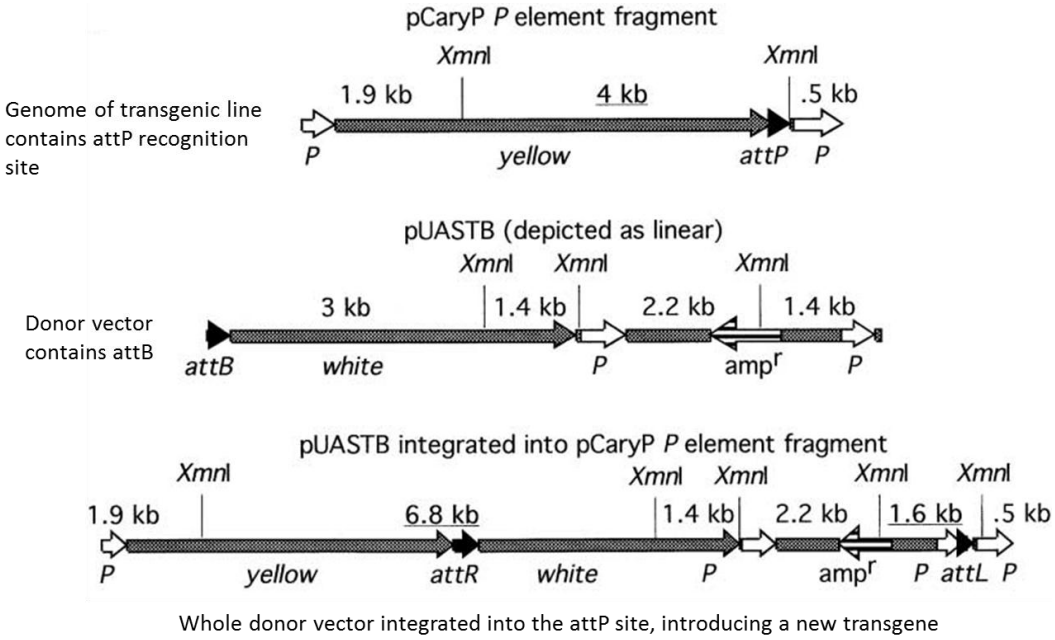

Figure 1-5 Strategy to target integration of a donor vector into the genome of a transgenic DrosoPhila embryo

A transgenic DrosoPhila recipient line was made using P element transposon to introduce an attP target site into the genome. Injection of the donor vector into recipient line embryos resulted in the integration of the whole vector into the attP site. Figure adapted from (Groth et al. 2004)

A different strategy was employed in mice which showed that recombination-mediated cassette exchange could target a transgene in a donor vector to the target site of a transgenic recipient line mouse (Belteki et al. 2003). This involved using pairs of inverted attP sites in the recipient line and pairs of inverted attB sites flanking the transgene in

the donor vector. The integration of a donor vector into a target site of a transgenic mouse was shown to be inherited in the germline. An RMCE transgenesis system has also been developed in *DrosoPhila* which utilises PhiC31 integrase (Figure 1-6) (Bateman et al. 2006).

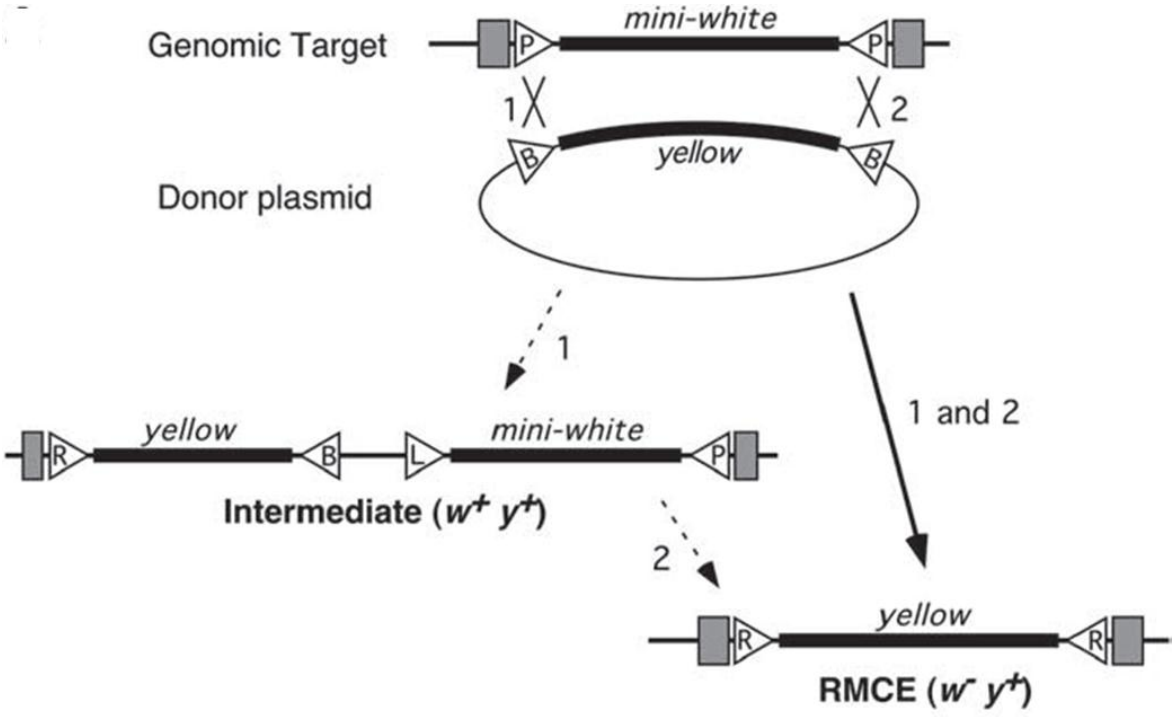

Figure 1-6 PhiC31 integrase transgenesis in DrosoPhila

Schematic showing the targeted integration of a donor vector into the genome of a transgenic fruit fly. Within the genome of the transgenic fly there is mini-white gene flanked by inverted attP sites. The donor vector contains yellow gene flanked by inverted attB sites. During recombination-mediated cassette exchange, an intermediate is formed. Figure from (Bateman et al. 2006)

Since starting this project, the study of PhiC31 integrase use in zebrafish has been published. In 2010, Lister published a paper showing that PhiC31 integrase can facilitate intramolecular excision of a DNA sequence flanked by an attP site and an attB site

(Figure 1-7). Excision was also shown to occur in the genome of a stable transgenic line when embryos were injected with PhiC31 integrase RNA. Lister later showed that PhiC31 integrase-mediated excision could be restricted to specific tissues (Lister 2011).

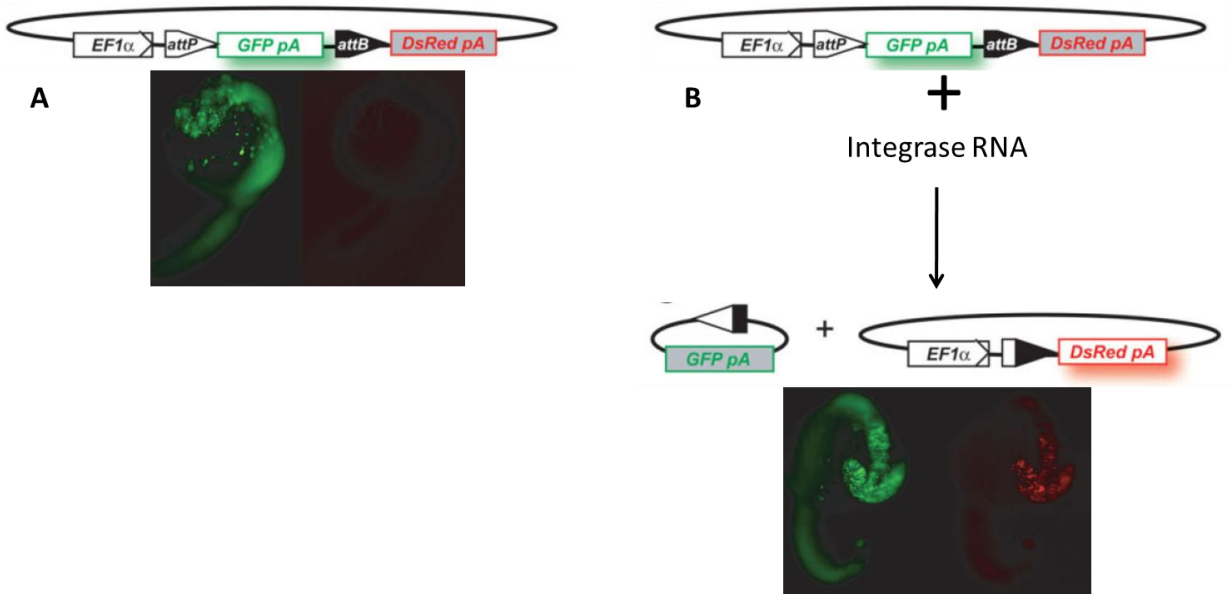

Figure 1-7 Integrase facilitates recombination between attB and attP in zebrafish

(A*) Injection of the depicted DNA construct alone results in GFP expression in the zebrafish. (B) Co-injection of the DNA construct and PhiC31 integrase RNA results in intramolecular excision of GFP resulting in mosaic expression of both GFP and DsRed. Figure adapted from (Lister 2010).*

An RMCE strategy has also been developed in zebrafish (Figure 1-8) (Hu et al. 2011). Firstly, a recipient transgenic line was made containing a cFos promoter sequence upstream of a GFP sequence that were flanked by attP sites. Upstream of the 5' attP site was a Sox10 enhancer driving expression in a number of tissues including the spinal cord. When the donor construct containing attB and mCherry was co-injected into transgenic embryos, RMCE occurred and mCherry was observed showing an

expression pattern consistent with Sox10. This showed that targeted integration into the genome of a transgenic zebrafish was possible.

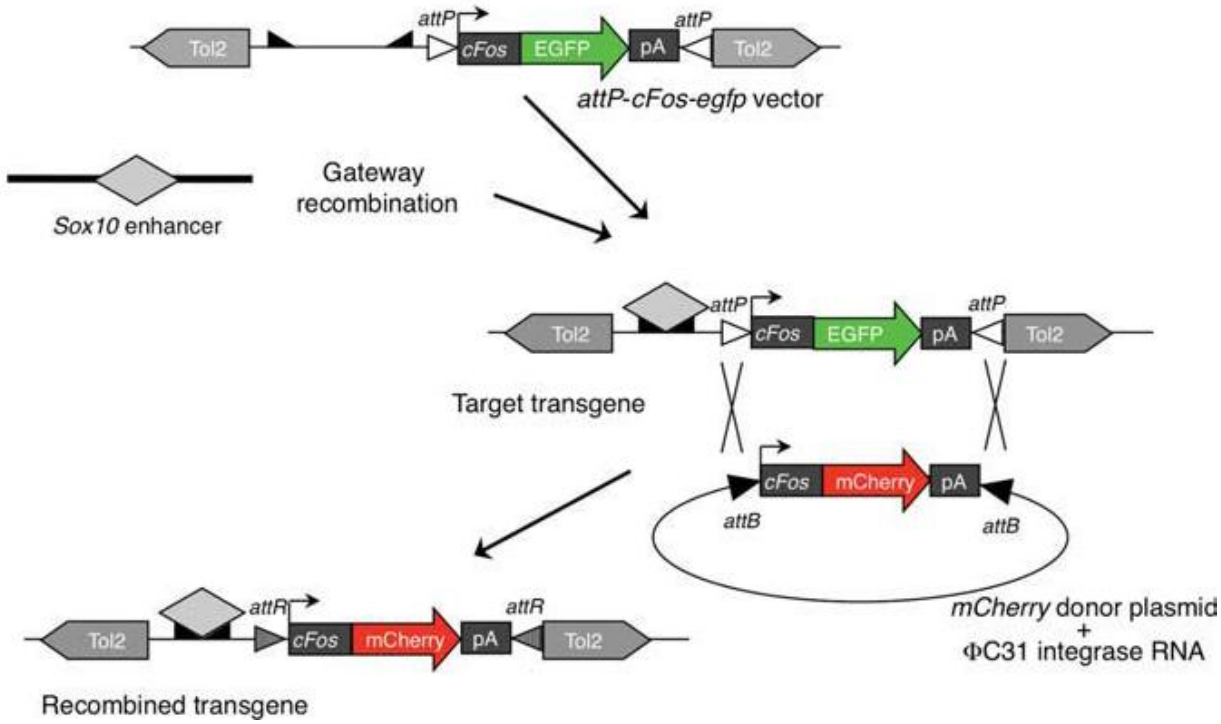

Figure 1-8 Strategy for recombination-mediated cassette exchange in zebrafish

This figure is from (Hu et al. 2011) and shows the strategy employed for targeted recombination-mediated cassette exchange facilitated by PhiC31 integrase in zebrafish.

There are natural-occurring att sites in vertebrate genomes, for example, two have been identified in the mouse genome (Thyagarajan et al. 2001) and several in the DrosoPhila genome (Groth et al. 2004). These sites, termed ‘pseudo-attP’ sites, may be recognised by integrase and cause integration at these loci.

1.4.4 Use of BACs in transgenesis

Due to their ability to hold very large sequences of DNA, bacterial artificial chromosomes (BACs) have become a very useful tool in many areas of research. They have been shown to provide rescue of knock-down phenotypes in transgenic mice due to the endogenous control of the rescue gene contained within the BAC (Antoch et al. 1997). Linearised BACs have been used to make transgenic mice (Chrast et al. 1999) and more recently, Tol2 has been used to increase efficiency of BAC transgenesis in mice and zebrafish (Suster et al. 2009). The increase in fully sequenced genomes has resulted in BAC libraries being widely available and easily obtainable (Vintersten et al. 2008) and so the BAC has become a popular tool in genetic studies. BACs are also used in disease modelling, in 2003 a Down syndrome model mouse was published using BAC transgenesis (Kazuki et al. 2003).

BACs can be used to study cis-regulatory elements as their large size allows the integration of a whole genomic locus encompassing all *cis*-regulatory elements (Jessen et al. 1999). BACs can also be modified to insert a fluorescent reporter sequence immediately upstream of the translational start site of the gene of interest (Heintz 2001; Yang et al. 2006). This allows the visualisation of the reporter gene driven by an endogenous promoter accompanied by all of the *cis*-regulatory elements which is expected to recapitulate endogenous expression patterns precisely which cannot be achieved when studying cis-regulatory elements as a solitary unit.

1.5 Aims and objectives

A study has not yet been carried out studying enhancer-promoter specificity *in vivo* within the endogenous genomic locus. It seems likely that a combination of all described mechanisms of specificity result in the differential gene regulation of the vast number of genes in the eukaryotic genomes. We know that the study of gene regulation is essential to understand many genetic disorders and the amount of research in this area reflects that.

The aim of this project was to develop a transgenesis system that would allow the study of enhancer-promoter interactions within their genomic environment. The zebrafish was selected as it is the ideal animal model for this type of experiment as the embryos are transparent and fluorescent reporters can be used to visualise expression patterns driven by *cis*-regulatory DNA.

In order to carry out such an experiment, a number of techniques needed to be developed. Firstly, in order to alleviate variability in position effects when studying a transgene integrated randomly into the zebrafish genome, a targeted integration system was developed. PhiC31 integrase is used frequently in *DrosoPhila* transgenesis and has been shown to function in mammalian cells. This led us to hypothesise that it would also work in zebrafish and allow us to target transgene integration into the same genomic locus resulting in consistent position effects between different constructs. The

development of a PhiC31 integrase system in zebrafish will be useful in any experiments where subtle changes in gene expression will affect the results.

In vertebrate genomes, long-range enhancers are known to be located up to 1 MB away from their target gene. In order to study such enhancers, both target gene and enhancer must be present in the same DNA construct as well as intervening DNA that may contain other regulatory DNA essential for endogenous expression; for example insulators. This is just not possible using standard vectors and so BAC transgenesis needs to be employed to allow long-range enhancers to be studied. The modification of the BACs to introduce fluorescent reporter genes to label the target gene expression pattern will allow a visual read out of enhancer interactions with the target promoter.

To be able to target the integration of a modified reporter-containing BAC into the same position in the genome would enable the study of enhancer-promoter interactions *in vivo*. The BAC could then be mutated in specific locations thought to be involved in enhancer-promoter specificity; for example specific motifs of the core promoters. To understand the mechanisms by which enhancers can interact with specific promoters would be extremely useful in the study and treatment of genetic disorders involving cis-regulatory DNA.

2 MATERIALS AND METHODS

2.1 Materials

All chemicals described throughout were obtained from Sigma-Aldrich (Poole, UK) unless otherwise stated.

2.1.1 Molecular Biology

Pfu polymerase	Promega, Southampton, UK
PCR extender	5prime, Nottingham,UK
Oligonucleotides	Sigma-Aldrich, Poole, UK
Agarose	Bioline Reagents, London, UK
DNA ladder 1 kb	New England Biolab, Hitchin, UK
DNA ladder 100 bp	New England Biolab, Hitchin, UK
6x gel loading dye	New England Biolab, Hitchin, UK
Go taq polymerase	Promega, Southampton, UK
dNTPs (deoxynucleotides)	Fisher Scientific, Loughborough
Restriction enzymes	New England Biolab, Hitchin, UK
Triton-X-100	Sigma-Aldrich, Poole, UK
T4 DNA ligase	Promega, Southampton, UK
T4 DNA polymerase	Promega, Southampton, UK
Phenol	Sigma-Aldrich, Poole, UK
Phennol:Chloroform	Sigma-Aldrich, Poole, UK
ECL+ (enhanced chemiluminescence)	GE Healthcare

2.1.2 Antibodies

Anti-Bacteriophage PhiC31 Integrase antibody	Abcam, Cambridge, UK

Secondary	Abcam, Cambridge, UK

2.1.3 Bacterial Culture

Luria Broth (LB)	Sigma-Aldrich, Poole, UK
LB agar	Sigma-Aldrich, Poole, UK
Ampicillin	Sigma-Aldrich, Poole, UK
Kanamycin	Sigma-Aldrich, Poole, UK
Tetracycline	Sigma-Aldrich, Poole, UK
Chloramphenicol	Sigma-Aldrich, Poole, UK
Chemical competent cells	Bioline Reagents, London, UK

2.1.4 Kits

NucleoSpin plasmid	Camlab Ltd, Cambridge, UK
NecleoSpin gel and PCR clean up kit	Camlab Ltd, Cambridge, UK
NucleoBond Xtra Midi	Camlab Ltd, Cambridge, UK
NecleoBond BAC 100	Camlab Ltd, Cambridge, UK
mMessage machine SP6	Ambion, Warrington, UK
Blunt cloning	Clontech, France

2.1.5 Zebrafish work

Phenol Red Solution	Sigma-Aldrich, Poole, UK
Microfiltration columns	Pall, Ann Harbor, USA
Proteinase K	Sigma-Aldrich, Poole, UK
Gentamycin	Fisher Scientific, Loughborough
1-phenyl 2-thiourea (PTU)	Sigma-Aldrich, Poole, UK
Protease (*Streptomyces griseus)*	Sigma-Aldrich, Poole, UK

2.1.6 Zebrafish lines

The zebrafish used in this project were AB* wild type strain.

2.1.7 Consumables

Standard laboratory consumables for example Eppendorf and Falcon tubes are not included here.

Microloader tips	Eppendorf
Nuclease-free water	Sigma-Aldrich, Poole, UK
X-ray film	Kodak

2.1.8 Equipment

PCR machine	GeneFlow Lab cycler
Agarose gel tanks	Bio-Rad
Stereomicroscope SMZ645	Nikon, Kingston, UK
Digital microscope camera DFC300 FX	Nikon, Kingston, UK
Centrifuge	Thermo Scientific, Loughborough, UK
Nanodrop	Thermo Scientific, Loughborough, UK
Microfuge	Thermo Scientific, Loughborough, UK
Bacterial incubator	Thermo Scientific, Loughborough, UK
Embryo incubator	Thermo Scientific, Loughborough, UK
Waterbath	Fisher Scientific, Loughborough
Temperature block/shaker	Eppendorf Thermomixer
Gas microinjector	Tritech Research, Los Angeles, USA
Flaming-Brown needle puller	Sutter Instruments, Novato, USA

2.1.9 Solutions

TBE buffer

90 mM Tris base, 90 mM Boric acid, 2 mM EDTA (Ethylenediaminetetraacetic acid), pH 8

E3 medium

5 mM NaCl, 0.17 mM KCl, 0.33 mM $CaCl_2$, 0.33 mM $MgSO_4$, gentamycin

DNA extraction buffer

0.1 M Tris pH9, 0.1 M NaCl, 0.05 M EDTA, 0.2 M Sucrose, 0.5 % SDS (Sodium dodecyl sulfate), 750 µg/ml Proteinase K

PTU

0.0045% PTU in 1 x E3 medium

SDS sample buffer

63 mM Tris-HCl pH 6.8, 10% glycerol, 5% ß-mercaptoethanol, 3.5% sodium dodecyl sulphate

SDS running buffer

192 mM Glycine, 250 M Tris-Base and 0.1% SDS

Transfer buffer

192 mM Glycine, 250 mM, Tris-Base and 20% methanol

2.2 Methods

2.2.1 Molecular Techniques

2.2.1.1 Molecular Cloning

Polymerase Chain Reaction (PCR)

The primers were diluted to 10 μM in a volume of 50 μl: 5 μl forward primer, 5 μl reverse primer and 40 μl nuclease-free water. The PCR reaction was set up as follows:

Reagent	Volume
Nuclease-free water	326 μl
10 x buffer	40 μl
10 mM dNTPs	8 μl
10 μM forward and reverse primer mix	20 μl
DNA template (200 ng/μl)	2 μl
Pfu I enzyme (1 unit/ μl)	4 μl
Total reaction volume	**400 μl**

The PCR machine was then programmed as follows:

Temperature	Time	
94 °C	2 min	
94 °C	30 sec	35 cycles
60 °C	1 min	
72 °C	2 min	
72 °C	5 min	
12 °C	Hold	

The annealing temperature (60 °C) and extension time are variable depending on the primers used and the length of the PCR fragment respectively. Any alterations made are described in the text below. If the PCR product was too long to use Pfu polymerase, the PCR extender system (5prime) was used instead.

Electrophoresis gel

To make a 1 % agarose gel to separate DNA of different sizes; 1 g of agarose powder was dissolved per 100 ml of 1 x Tris-borate-EDTA (TBE) buffer. This was heated until the agarose dissolved and allowed to cool to 50-60 °C. Five µl of ethidium bromide was added and the agarose poured into a gel tray containing a comb and allowed to solidify. This was then transferred to an electrophoresis tank containing 1 x TBE buffer and the comb was removed. As a marker, 5 µl of New England Biolab 1 kb or 100 bp DNA ladder were loaded onto the gel. Loading dye was added to the samples (1 µl per 5 µl of sample) which were then loaded onto the gel. The tanks were connected to a power pack set at 80-90 V and ran for 30-60 minutes depending on the amount of separation required.

Gel extraction

To isolate the PCR fragment, the Machery-Nagel NucleoSpin gel and PCR clean-up kit was used. Firstly, DNA loading dye (Promega) was added to the PCR reaction and it was run on a 1 % agarose gel. The appropriate DNA band was then cut out using a Ultra Violet (UV) illuminator and excess agarose was removed. The agarose containing

the DNA was then weighed and dissolved in NTI buffer at a ratio of 200 µl to 100 mg of agarose. This was heated to 60 °C on a shaker to allow the agarose to dissolve and then cooled on ice. The solution was transferred to a column and incubated at room temperature for 1 minute. The column was centrifuged at 15,000xg for 1 minute. The flow through was discarded and 700 µl of buffer NT3 was added to the column and the centrifugation was repeated. Again the flow through was discarded and the wash step repeated with 500 µl of buffer NT3. The flow through was discarded and the column was centrifuged for 2 minutes to dry the membrane. The column was transferred to a sterile 1.5 ml Eppendorf tube and 30 µl of nuclease-free water was added to the column and incubated at room temperature for 1 minute. The tubes were centrifuged at 15000xg for 1 minute and the column was discarded. The eluate was stored at -20 °C.

Enzyme digests and gel extraction

Enzyme digest reactions were assembled as follows:

Reagent	**Amount**
Nuclease-free water	Up to 50 µl
Appropriate restriction enzyme buffer	5 µl
DNA (vector or PCR product)	10 µg or 5µg
Bovine Serum Albumin (BSA)	0.5 µl
Restriction Enzyme	10 units
Total reaction volume	**50 µl**

The digests were incubated for 2 hours at 37 °C. The DNA was then separated on a 1% agarose gel and the appropriate fragments purified as described above.

Ligation reaction

The ligation reactions were assembled as follows:

Reagent	Amount
Nuclease-free water	Up to 20 µl
Ligase buffer	2 µl
Vector DNA	100 ng
Insert DNA	Molar ratio 3 insert:1 vector
T4 DNA ligase	1 µl
Total reaction volume	**20 µl**

The ligation reaction was incubated at 16 °C overnight.

Transformation

The ligation reaction was then used to transform chemical competent *E. coli* cells by adding 10 µl of the ligation reaction mixture to 50 µl cells (incubated on ice for 15 mins) and heat shocking at 42 °C for 45 secs. Subsequently, 400 µl of LB media was added to the cells and they were incubated on a shaker at 37 °C for 30 – 60 minutes. The 400 µl was then spread on an LB agar plate with an antibiotic selection appropriate for the vector. The plates were then incubated at 37 °C overnight and checked for colonies the

next day. Colonies from each plate were picked and used to inoculate 5 ml LB medium. These cultures were grown overnight at 37 °C and 220 rpm (revolutions per minute).

Minipreps

To isolate the plasmid DNA from the bacterial cultures, the Machery-Nagel Nucleospin plasmid kit was used. The bacteria were pelleted by centrifugation at 10,000xg for 1 minute. The supernatant was discarded and the bacterial pellet was re-suspended in 250 μl of Buffer A1. Next, 250 μl of Buffer A2 was added and the tube was inverted gently 4-6 times. This was then left for approximately 3 or 4 minutes for the lysate to clear. After this time, 300 μl of Buffer A3 was added and the tube inverted gently 4-6 times. The bacterial lysate was the centrifuged at maximum speed (15,000xg) in a microcentrfuge for 10 minutes. The supernatant was transferred to a Spin Column set up in a 2 ml Collection Tube and this was centrifuged at 15,000xg for 1 minute. The flow through was discarded and the Spin Column was replaced in the Collection Tube; 600 μl of Buffer A4 was added to the Column and it was centrifuged again at 15,000xg for 1 minute. The flow through was discarded and the Column was centrifuged for another minute at maximum speed to dry the membrane. The Column was then transferred to a 1.5 ml Eppendorf tube and 50 μl of nuclease-free water was added to the column. This was incubated at room temperature for 1 minute and then centrifuged at 15,000xg for 1 minute. The column was then discarded and the eluted DNA was stored at -20 °C.

Test enzyme digests

A sample of the miniprep DNA was digested using two or three restriction enzymes to ensure the PCR product had ligated into the vector as expected. They were set up as follows:

Reagent	**Amount**
Nuclease-free water	Up to 20 µl
Appropriate buffer	2 µl
BSA	0.2 µl
Appropriate enzyme(s)	2 – 3 units
Total reaction volume	**20 µl**

The digests were incubated at 37 °C for 2 hours, and the reaction was run on a 1 % agarose gel to check the correct band sizes were present.

Sequencing

Vector sequencing was carried out externally by Beckman Coulter Genomics. The plasmid samples were prepared by diluting the sample to 100 ng/µl and oligonucleotides were diluted to 5 µM.

The results were analysed by aligning the sequencing data with the known sequence of the insert sequence and checking for any differences. This analysis was carried out

using the software SerialCloner and when required, Cap3 sequence assembly program available online (http://pbil.univ-lyon1.fr/cap3.php).

Midipreps

Firstly, a 100 ml culture was set up and incubated overnight at 37 °C and 220 rpm. The next day the cells were harvested by centrifugation at 6000xg and 4 °C for 20 minutes. The supernatant was discarded and the pellet resuspended in 8 ml of Resuspension Buffer + RNase A (ribonuclease). Then, 8 ml of Lysis Buffer was added and the tube inverted gently 4-6 times. This was left at room temperature for 3-4 minutes before adding 8 ml of Neutralisation Buffer. The lysate was cleared by centrifugation at 6000xg and 4 °C for 10 minutes and then the supernatant was decanted into the column filter. Once all of the solution had passed through the filter and column, 5 ml of Equilibration Buffer was added to wash out the lysate in the filter. The filter was then removed and 8 ml of Wash Buffer was added to the column. After flow through of the Wash Buffer, the column was placed in a 50 ml Falcon tube and 5 ml of Elution Buffer was added. The column was then discarded and 3.5 ml of isopropanol was added to the eluate. This was mixed by inverting the tube and then centrifuged at 15,000xg and 4 °C for 30 minutes. The supernatant was removed and 5 ml of 70 % ethanol was added to the pellet. Centrifugation was repeated and the ethanol removed. The pellet was allowed to air dry completely before resuspension in 500 µl of nuclease-free water.

Determination of nucleic acid concentration

The concentration of nucleic acid solutions were measured using the Nanodrop spectrophotometer. The absorption at a wavelength of 260 nm was measured to determine nucleic acid concentration. Absorption ratios of 260/280 nm and 260/230 nm were checked to ensure there were no contaminants (proteins or ethanol) that would affect downstream protocols.

2.2.1.2 *In vitro* Transcription

Five micrograms of the appropriate pCS2+ vector was linearised downstream of the polyA tail with NotI restriction enzyme. A 50 µl digest was set up containing 5 µl of NEB buffer 4, 0.5 µl BSA, 5 µg plasmid DNA, 5 units NotI and made up to a total of 50 µl with nuclease-free water. The reaction was incubated for 2 hours at 37 °C and the DNA was separated on a 1 % agarose gel. The DNA was then isolated using the Machery-Nagel Gel and PCR clean up kit and eluted in 50 µl nuclease-free water. The *in vitro* transcription reaction was set up as follows:

Reagent	Amount
Nuclease-free water	Up to 60 µl
2x NTP/CAP	30 µl
10x reaction buffer	6 µl
Linear DNA template	3 µg
Enzyme	6 µl
Total reaction volume	**60 µl**

This was incubated at 37 °C for 2 hours. Next, 3 µl of TURBO DNase was added and the reaction was incubated for a further 15 minutes at 37 °C. In order to stop the reaction, 345 µl of nuclease-free water and 45 µl of Ammonium Acetate Stop Solution were added and the solution was mixed gently. The Ribonucleic Acid (RNA) was extracted by adding 450 µl of phenol:chloroform and mixing. The phases were then separated by centrifugation at 15,000 xg for 20 minutes and the aqueous phase was removed and transferred to a fresh tube. An equal volume of chloroform was added and the solution mixed. The centrifugation was repeated and the top phase was transferred to a fresh tube. The RNA was then precipitated by adding 0.2 volumes of 5 M ammonium acetate and 2 volumes of 100 % ethanol. The RNA was pelleted by centrifugation at 15,000 xg for 30 minutes at 4 °C. The supernatant was removed and discarded, 500 µl of 75 % ethanol was then added and the centrifugation was repeated. The supernatant was discarded and the pellet was allowed to air-dry for 5-10 minutes. The RNA was then resuspended in 50 µl nuclease-free water and stored at -80 °C.

2.2.2 Zebrafish Embryo Manipulation

2.2.2.1 Microinjections

The day prior to injections, pairs of zebrafish were set up in cages, separated by a divider. The morning of injections the divider was removed and the embryos were collected after 10 minutes. The embryos were injected with the injection solution (50 ng/µl DNA construct, 15 ng/µl Tol2 RNA (where required), 1 x phenol red solution and 1% rhodamine (where required)) and incubated at 28 °C in water from the fish system. The afternoon after the injections the water was replaced with E3 medium supplemented with Gentamycin. The embryos were incubated at 28 °C for up to five days. Fish that were to be grown up were transferred to the zebrafish facility on day 5 and feeding was started.

2.2.2.2 Genomic DNA Isolation

Ten – twenty 5 dpf embryos from the same pair crossing were pooled and transferred to a 1.5 ml Eppendorf tube, as much liquid was removed as possible and 770 µl of DNA extraction buffer was added depending on the number of embryos. The samples were then vortexed briefly and incubated for 2 – 3 hours at 55 °C until digestion was completed. The samples were then heated to 95 °C for 10 minutes to inactivate the Proteinase K. Samples were centrifuged at 8000 rpm for ≥10 minutes to remove any debris, the supernatant was transferred to a fresh tube and stored at -20 °C. When required, 2 µl of the DNA was added to a 25 µl PCR reaction.

2.2.2.3 Imaging

The zebrafish embryos were imaged using a Nikon SMZ1500 fluorescence stereomicroscope for both bright field and fluorescence images. The imaging software used was NIS elements.

2.2.2.4 Western Blotting

Protein samples were extracted from zebrafish embryos and Sodium Dodecyl Sulphate PolyAcrylamide Gel Electrophoresis (SDS-PAGE) was employed to separate proteins by size. This protein separation was then transferred onto a nitrocellulose membrane which was subsequently blotted with primary and secondary antibodies to label specific proteins of interest.

Sample Preparation

Embryos between 2 and 4 cell stage were dechorionated using a pronase solution. As the first couple of embryos are dechorionated, the whole dish is rinsed 3 times each time in a litre of fish water (2 parts tap water and 1 part deionised water). After dechorionation, the embryos are kept in an agarose lined petri dish. When the embryos reached high stage the yolk was removed from 20 embryos using a hypodermic needle and the cells were transferred to a 1.5 ml Eppendorf tube. As much of the liquid was removed as possible and 150 µl of SDS sample buffer was added. The cells were homogenised thoroughly using a pestle until the solution was uniform. This solution was

then incubated in a waterbath at 100 °C for 5 minutes and centrifuged at 15,000xg for 2 minutes. The supernatant was transferred to a fresh Eppendorf tube.

SDS-PAGE

The 10 % SDS polyacrylamide resolving gel was made as follows:

Reagent	**Amount**
Acrylamide	3.4 ml
Water	3.8 ml
1.5 M Tris-HCl pH 8.8	2.5 ml
10 % Ammonium persulfate	200 µl
10 % SDS	100 µl
TEMED (Tetramethylethylenediamine)	10 µl
Total volume	**~10 ml**

This was then immediately pipetted into 1 mm glass plates. Isopropanol was added slowly to the top of the gel and it was allowed to set. Once the resolving gel had set, it was rinsed thoroughly to remove the isopropanol. The stacking gel was made as follows:

Reagent	Amount
Acrylamide	1 ml
Water	3.4 ml
1.5 M Tris-HCl pH 8.8	1.5 ml
10 % Ammonium persulfate	90 µl
10 % SDS	60 µl
TEMED	5 µl
Total volume	**~6 ml**

This was pipetted on top of the resolving gel, a comb was immediately added and it was allowed to set.

Sample loading buffer was added to the protein samples and they were incubated at 100 °C for 5 minutes. A protein molecular weight ladder was loaded into the first well followed by the samples in subsequent wells. The gel was run in SDS running buffer for 1-2 hours at 100 V.

Transfer

The transfer apparatus was prepared as depicted in Figure 2-1. The sponge, membrane and filter paper were all soaked in transfer buffer prior to preparation.

Figure 2-1 Layout of materials required for protein transfer

Figure from Abcam Western Blot Protocol

(http://www.abcam.com/index.html?pageconfig=resource&rid=13045)

The transfer is run for an hour and to prevent over-heating, an ice pack is placed in the transfer buffer.

Immunoblotting

The membrane was blocked in 10 % dried milk powder (Marvel) dissolved in Tris buffered saline (1xTBS) at 4 °C overnight on a roller. The primary and secondary antibodies were diluted in 10 % dried milk powder. After blocking overnight, the diluted antibody was added and the membrane was incubated at room temperature for 1 hour on a roller. The membrane was then washed for 5 minutes in PBST and this step was repeated 3 times. The membrane was blocked again in 10 % dried milk powder for 5 minutes at room temperature on a roller. The diluted secondary antibody was then added and incubated for 1 hour at room temperature after which the wash steps were repeated as above. The membrane was then incubated with enhanced chemo-

luminescence plus (ECL+) following manufacturer's guidance (GE Healthcare) for 5 minutes. The bands were visualised using X-ray film which was developed.

2.2.3 Making a Transgenic Line

The following describes the steps taken to make a stable transgenic zebrafish line.

Firstly, approximately 200 embryos were co-injected with the DNA construct containing the desired transgene flanked by Tol2 arms and Tol2 transposase mRNA. These were incubated in E3 medium supplemented with gentamycin at 28.5 °C for 5 days in an incubator. Each day, the embryos were observed under a fluorescence stereomicroscope. The healthy looking embryos were selected and taken to the zebrafish facility and a feeding regime was started. After 3-4 months, once the fish are deemed big enough, they were crossed pairwise with a wild type fish. The F1 embryos were then screened for fluorescence and any embryos showing specific expression were once again transferred to the zebrafish facility to be grown up. If required, a number of lines were established from different founder fish. Once the F1 generation were sexually mature, they were crossed together to produce homozygotes, heterozygotes and non-transgenics. Embryos showing fluorescence were selected before transferring them back to the zebrafish facility. Once the F2 generation were sexually mature, they were outcrossed to wild type fish and homozygotes and heterozygotes were identified by the proportion of their offspring showing reporter gene expression: 100 % and 50 % respectively.

3 INVESTIGATION OF THE FUNCTION OF PhiC31 INTEGRASE IN ZEBRAFISH

3.1 Introduction

3.1.1 Variation in position effects during random integration transgenesis

Currently, the most popular method for the study of enhancers in zebrafish involves utilising Tol2 transposase to facilitate the integration of an enhancer-promoter-reporter constructs into the zebrafish genome. Throughout early development the tissue-specific expression pattern of the enhancer being tested may be established by monitoring fluorescence in the developing zebrafish embryo.

A major disadvantage of this method is that integration of the construct into the genome by Tol2 transposase is random. Therefore, the transgene in each injected embryo will be incorporated into a different locus within the genome. The position of the transgene may influence expression patterns of the reporter gene in a number of ways; for example, if a transgene integrates close to a genomic enhancer, it could interact with the promoter of the construct, driving the reporter gene expression. An example is depicted in Figure 3-1, where 4 different zebrafish transgenic lines resulting from injections of the same construct under the same conditions are presented. The embryos were injected with a construct containing a putative enhancer of the estrogen-related receptor ɣ (errɣ) gene. The keratin four minimal promoter was selected as it has been shown to interact with a number of different known enhancers (Gehrig et al. 2009). For more details on this

construct, refer to Chapter 5. Each of the four lines exhibit different expression patterns and variances in the expression intensity. This data supports the commonly held belief that due to the random nature of construct integration and thus differences in flanking DNA composition, inconsistency in the interference experienced by the enhancer-promoter-reporter transgene results in differences in reporter gene expression patterns.

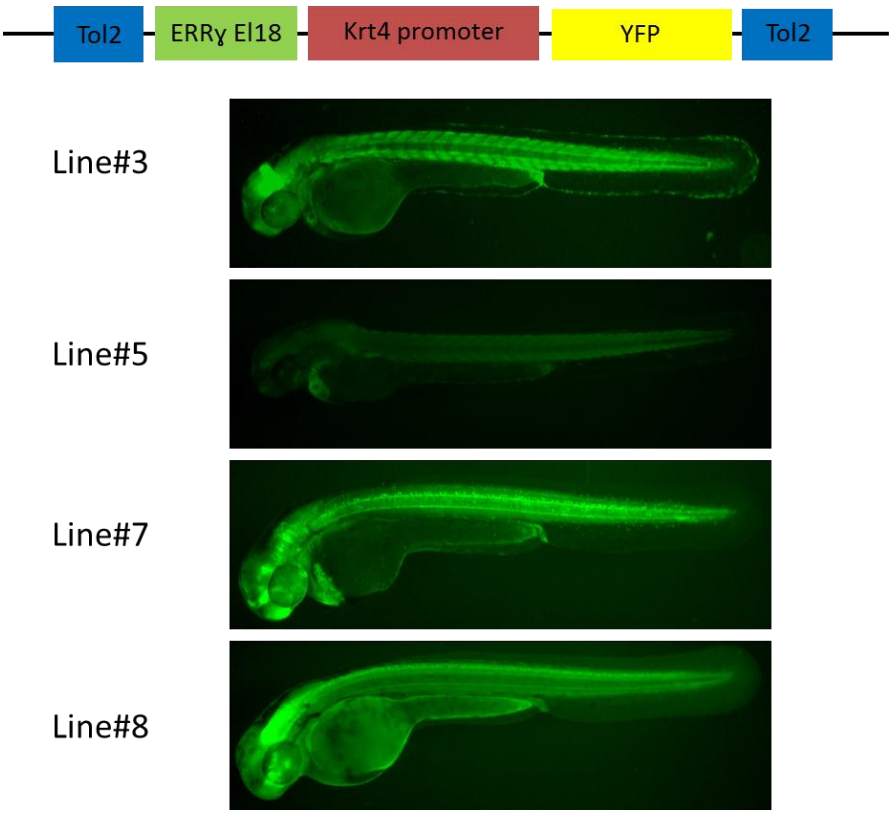

Figure 3-1 Tol2 transgenesis can result in large variability in positional effects

Images of F1 embryos resulting from different founder zebrafish injected with the same construct and Tol2 RNA. Individual lines show variability in intensity and expression pattern thought to be caused by random integration into different genomic loci. Figure adapted from Yavor Hadzhiev. The gateway vector contains Tol2 arms to facilitate integration (blue rectangles), ERRγ EI18 – a putative enhancer sequence (yellow rectangle), keratin 4 minimal promoter sequence (pink rectangle) and reporter gene venus GFP (green rectangle).

If construct integration could be targeted to the same genomic locus, thus overcoming the issues with random integration it would make zebrafish an even more attractive model for the study of cis-regulatory DNA.

PhiC31 integrase is a protein already utilised to develop transgenic *DrosoPhila* lines and has been very successful. PhiC31 integrase facilitates the recombination between two recognition sites: attP and attB for more detailed information refer to section 1.4.3.

3.1.2 Aims

Current methods of studying enhancers *in vivo* using the zebrafish model involve transgenesis protocols that result in random genome integration. This is widely thought to lead to positional effects specific to the genomic context of the transgene, which can make it difficult to draw conclusions about expression patterns driven by specific enhancers. In order to overcome the inconsistencies associated with random integration, the aim of this work is to design a method to target integration of constructs to a specific locus within the zebrafish genome.

Firstly, we need to establish whether PhiC31 integrase RNA can be translated by the zebrafish machinery when injected into the early embryo. Micro-injection of RNA is the most common method of providing early embryos with functional protein. The next objective is to find out if PhiC31 integrase can facilitate integration between a single attP site and a single attB site in the zebrafish model. We also aim to facilitate targeted

integration of a transgene into a specific locus of the zebrafish genome which is expected to reduce variability in position effects when injecting enhancer constructs. This integration should be detectable *in vivo* using live imaging of the zebrafish embryos without the need for laborious insertion site analysis. Therefore, we aim to identify embryos with targeted integration by utilising fluorescent reporter genes which mark targeted integration events. We aim to investigate the efficiency of PhiC31 integrase-mediated genome integration and establish a transgenesis system suitable for use in many applications in the zebrafish model; such as enhancer testing or cell labelling.

3.2 Results

3.2.1 Expression of PhiC31 in the zebrafish model

To determine whether it would be possible to utilise PhiC31 integrase in the zebrafish model, the translation of *in vitro* synthesised mRNA and function of the protein was assessed. Firstly, mRNA coding for PhiC31 was produced by cloning the mouse codon optimised PhiC31 coding sequence (Raymond & Soriano 2007) into the pCS2+ vector. This was then linearised and PhiC31 RNA was transcribed *in vitro* and was injected into early stage zebrafish embryos. Next the embryos were grown in E3 medium at 28.5°C until they reached high stage. When the embryos reached high stage (approximately 3.5hpf and over 1000 cells) the yolk was removed and the protein extracted from the cells. Western blot analysis (Figure 3-2) showed a band corresponding to PhiC31 integrase in RNA-injected samples that was not present in protein extracted from non-injected embryos. Non-specific bands are visible in both RNA-injected and non-injected samples. The non-specific bands appear stronger in the non-injected sample suggesting that the absence of the 67 kDa band cannot be due to less protein being loaded onto the SDS gel. This result suggests that PhiC31 integrase protein is produced by high stage from mRNA injected into a 1 cell stage zebrafish embryo.

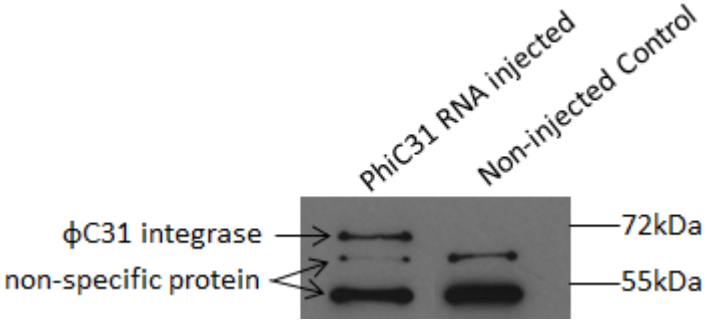

Figure 3-2 PhiC31 integrase RNA is translated when injected into early stage zebrafish embryos

The early stage embryos were injected with 20ng/µl PhiC31 RNA and 0.1% phenol red. At high stage (approximately 3.5hpf) the protein was extracted from a pool of 20 embryos for both the injected and non-injected embryos. The western blot shows PhiC31 integrase protein (67 kDa) is present in injected embryos but absent in non-injected embryos. Primary antibody: Anti-Bacteriophage PhiC31 Integrase antibody and secondary antibody: anti-rabbit IgG. Non-specific binding of antibody to a protein is used as a loading control.

3.2.2 Design and construction of vectors for investigating PhiC31 integrase function in zebrafish

To test whether PhiC31 integrase can facilitate integration between the two recognition sites attP and attB, two constructs were designed: a recipient vector which contains a docking site for integration and an integration donor vector.

Recipient construct

The recipient construct was received from a collaborator, Darius Balciunas, Temple University, USA. It contains a 390 bp promoter sequence for the lens-specific *γ-crystallin* gene; this promoter was used to restrict expression of the reporter gene contained within the vector to the lens (Figure 3-3). The lens was selected as the marker as it is a relatively small tissue that doesn't obscure any other anatomy that may be labelled by

the reporter linked to the enhancer-promoter sequences. Downstream of the lens-specific promoter is an 83 bp PhiC31 integrase recognition site attP and a GFP coding sequence with a polyA tail to stabilise the RNA being transcribed (Figure 3-3). The attP site was positioned between the promoter and reporter gene sequence to allow the targeted integration of the donor vector to be identified visually *in vivo* by the disruption of GFP expression (Figure 3-6). The γcry-attP-GFP sequence is flanked by Tol2 arms to facilitate the integration of the transgene to make transgenic recipient lines (Figure 3-3).

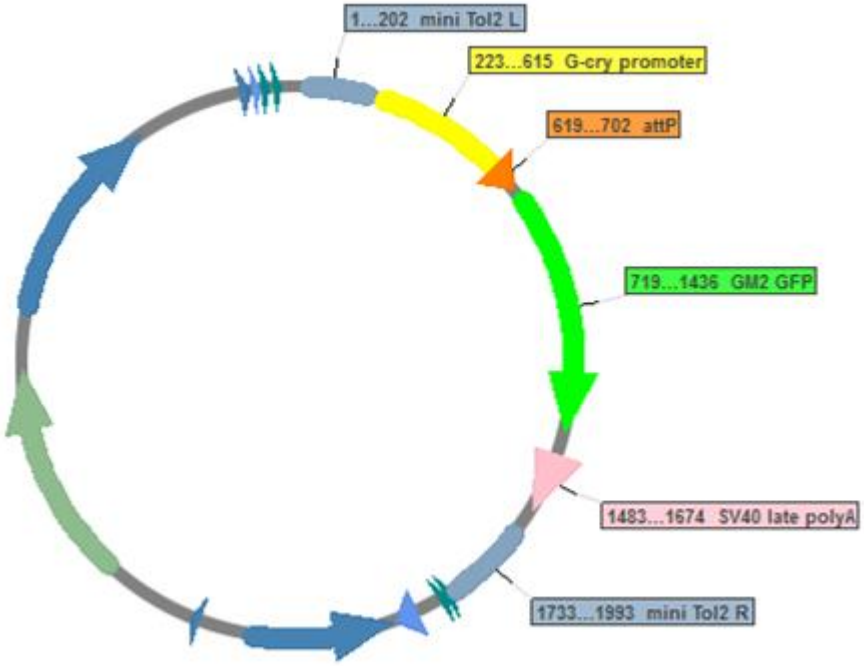

Figure 3-3 Schematic of the recipient vector

The recipient vector contains Tol2 arms, ~390 bp of γ-crystallin *promoter, 83 bp attP integrase recognition site, GFP coding sequence and a SV40 pA tail. Vector received from collaborator Darius Balcuinas at Temple University, USA.*

To test the tissue-specificity of the *γ-crystallin* promoter of the recipient vector, wild type embryos were co-injected with the recipient vector and Tol2 mRNA. The embryos were injected at 1-cell stage and incubated at 28.5 °C in E3 medium supplemented with gentamycin. They were then screened at 3-4 dpf for transient GFP expression (Figure 3-4). Of 136 embryos co-injected with recipient vector and Tol2 RNA, 98 showed GFP expression specific to the lens (72.1 %). The rest of the embryos showed no GFP expression at all. This result suggests that the *γ-crystallin* promoter gives lens-specific GFP expression which is supported by subsequent results in this chapter (Figure 3-11). This shows that the *γ-crystallin* promoter is suitable for use in the recipient vector.

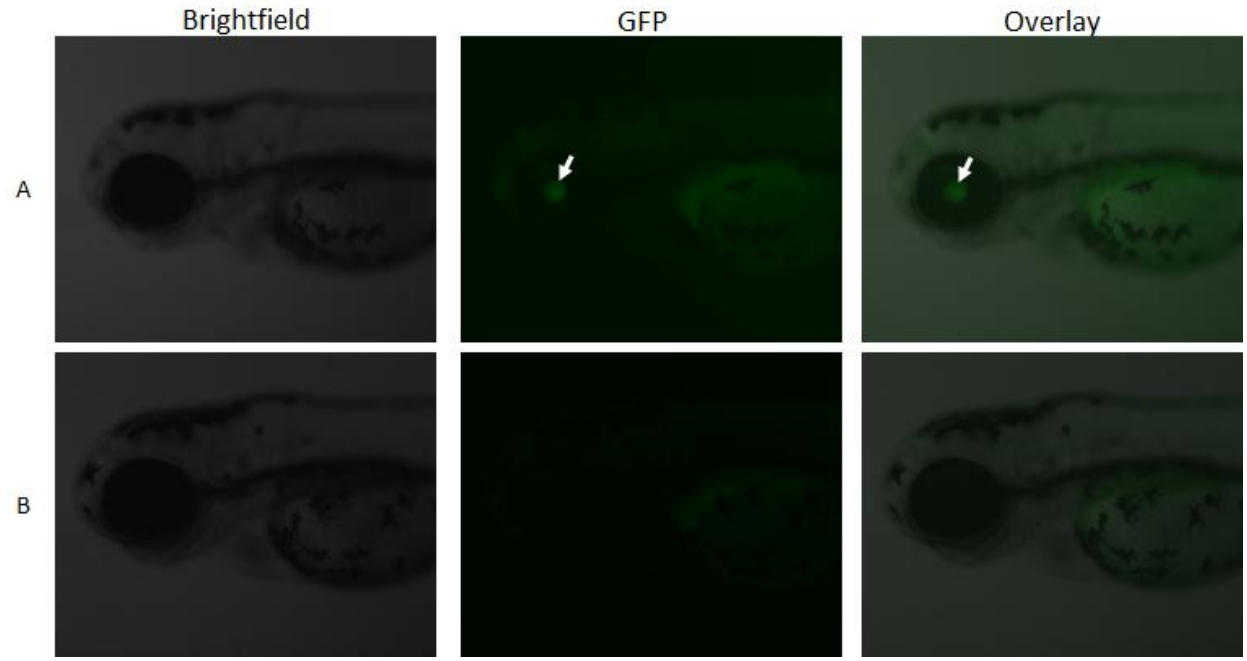

Figure 3-4 The recipient vector gives lens-specific GFP expression

(A*) Images of a 4 dpf stage embryo injected with 50ng/µl recipient vector DNA, 20ng/µl Tol2 RNA and 0.1% phenol red. (B) Images of a 4 dpf control non-injected embryo. All images are a lateral view and anterior to the left.*

Donor construct

The construct to be integrated (donor construct) will contain an attB site and a different fluorescent reporter protein (mCherry). This will enable us to identify embryos in which targeted integration has occurred as mCherry expression in the lens replaces GFP (Figure 3-6). There was no previous work published in zebrafish to evaluate the efficiency of PhiC31 integrase therefore we assumed that the number of embryos showing targeted integration of the two constructs may be limited and a screening step using fluorescent reporters may be required to detect the correct integration-containing embryos.

The donor vector was constructed by joining the attB sequence (from the fly vector pUASTattB (Bischof et al. 2007)) to the mCherry coding sequence with polyA tail by PCR of the two components. The reverse primer for the attB sequence contained a linker sequence that was complimentary to a linker sequence in the forward primer used to amplify the mCherry-pA sequence (Figure 3-5A). These two PCR products were combined and with the use of the attB forward primer and pA reverse primer another PCR was carried out to combine the two components due to the complementarity of the linker incorporated into the first sets of primers (Figure 3-5A). This PCR product was then cloned into pJET2.1 using the Clontech blunt-ended cloning kit. The resulting vector is depicted in Figure 3-5B.

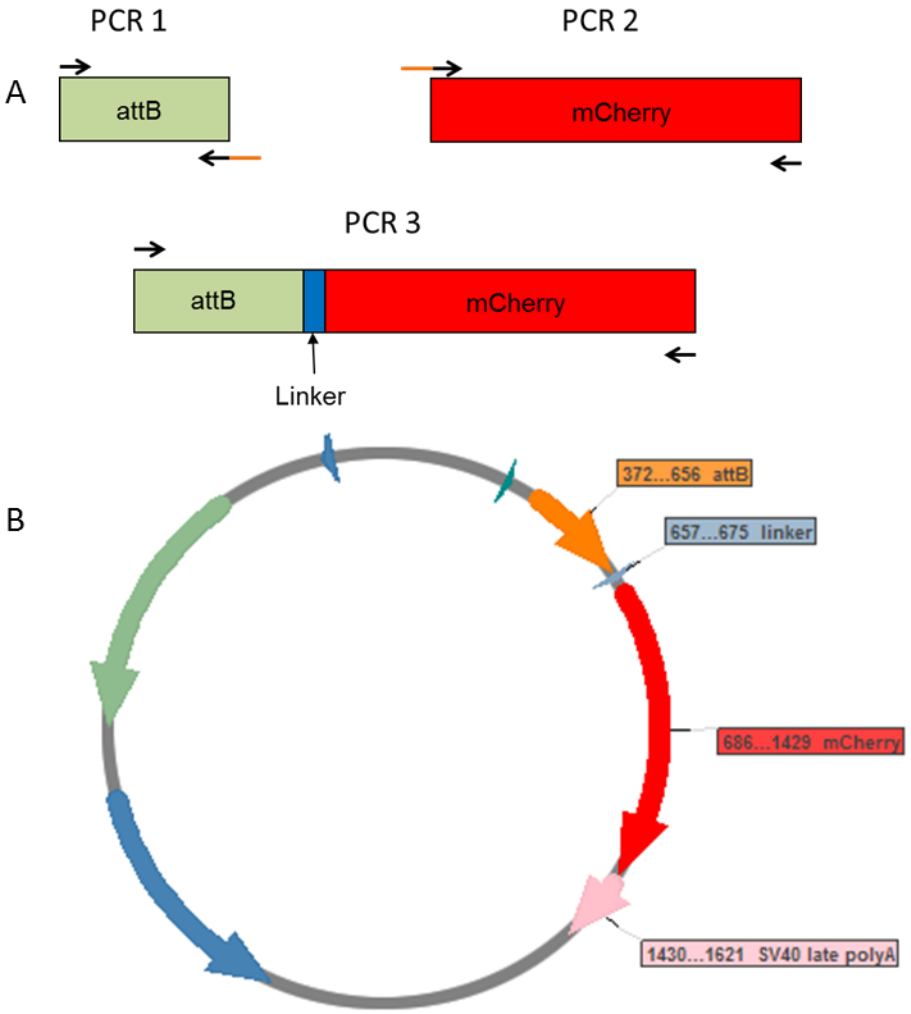

Figure 3-5 Schematic of donor vector and its construction

(A*) A schematic of how the vector was constructed. Black arrows show PCR primers and the orange lines represent linker sequences included in the primers (B) A vector map of the donor construct (backbone is pJET1.2).*

In order to assess whether PhiC31 integrase can facilitate integration between the recipient and donor vectors, the experiment depicted in Figure 3-6 was designed. The co-injection of the recipient and donor vectors with PhiC31 integrase mRNA into wild type embryos was expected to result in integration between the vectors in an unknown proportion of the embryos.

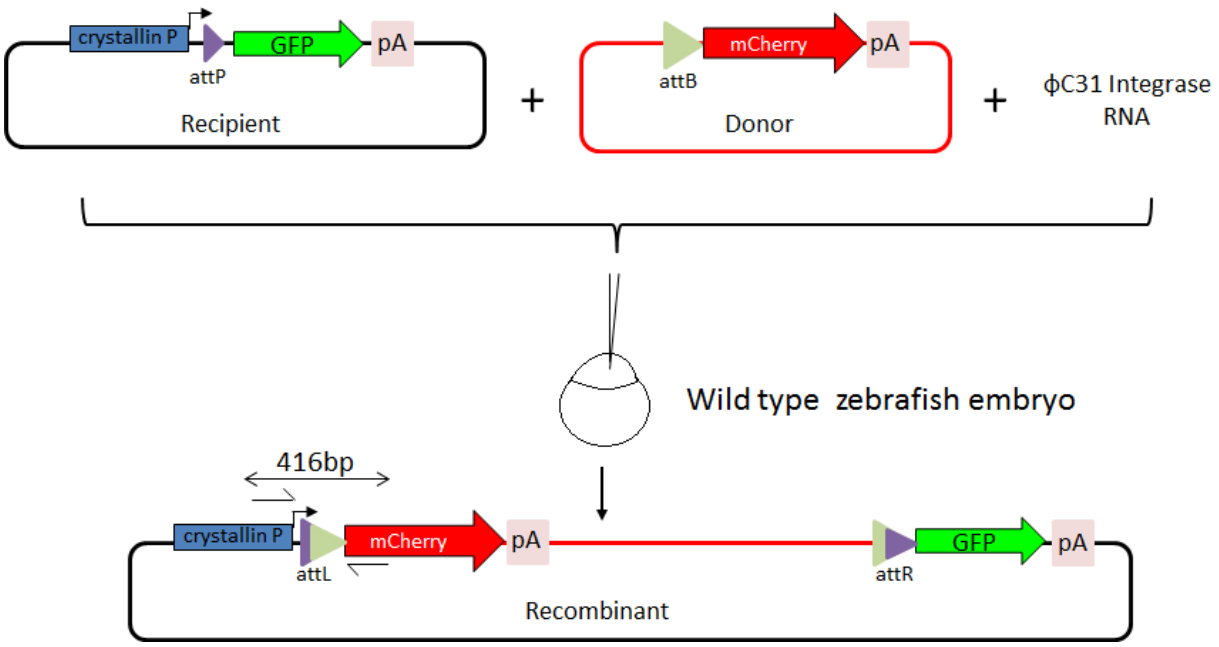

Figure 3-6 Experiment design to test PhiC31 integrase function in zebrafish

Co-injection of the recipient vector, donor vector and PhiC31 Integrase RNA is expected to result in the recombination of the two vectors, facilitated by PhiC31 integrase.

3.2.3 Targeted integration of the donor vector into the recipient vector facilitated by PhiC31 integrase

Next I needed to test whether PhiC31 integrase can facilitate integration between attP and attB sites in zebrafish. To do this, the recipient vector, donor vector and PhiC31 Integrase RNA were co-injected into early stage wild type zebrafish embryos with the expectation that PhiC31 integrase would facilitate the integration of the two vectors (Figure 3-6).

The embryos were grown as described previously and screened at 4 dpf for GFP and mCherry expression in the lens. Over 300 embryos were screened and none of them

showed mCherry expression in the lens which was expected to mark targeted integration.

3.2.3.1 Evidence for targeted integration events by plasmid sequence analysis

A lack of mCherry reporter activity could be due to a number of reasons: the efficiency of PhiC31 integrase could be too low for the mCherry to be detected visually, PhiC31 integrase may not facilitate integration in zebrafish or there could be a problem with the transcription or translation of mCherry. To check if integration does occur but the reporter expression is insufficient to detect, I have chosen to study the sequence of the integration site by PCR and sequencing. Firstly, total DNA was extracted from three pools of 10 embryos co-injected with recipient vector, donor vector and PhiC31 integrase mRNA and three pools of 10 embryos injected with recipient and donor vector only. A PCR was performed using a forward primer specific to the *γ-crystallin* promoter and a reverse primer specific to the mCherry coding sequence. If integration occurs then a 416 bp PCR fragment should be produced (Figure 3-6). When the recipient and donor vectors were injected with PhiC31 integrase mRNA, a band at approximately 416 bp was visible on an agarose gel (Figure 3-7A). The band was not present when the recipient and donor vectors were injected without PhiC31 integrase mRNA. This suggests that the targeted integration does not occur randomly, it only occurs in the presence of PhiC31 integrase. This was confirmed by a second PCR using a *γ-crystallin* forward primer and GFP reverse primer. A 1.1 kb (kilo base pairs) PCR product was expected to be produced in both samples as it would detect the presence of the non-

integrated recipient vector. This confirmed that PCR-quality DNA was present in the control sample. The 416 bp PCR product was sequenced (carried out by Beckman Coulter Genomics) and a comparison to the known attR sequence (Bateman et al. 2006) confirmed that the recombination was legitimate (Figure 3-7B-C). Together, these data show that PhiC31 integrase can facilitate integration between attP and attB sites in zebrafish. However, these experiments do not provide information about the frequency of targeted integration in embryos.

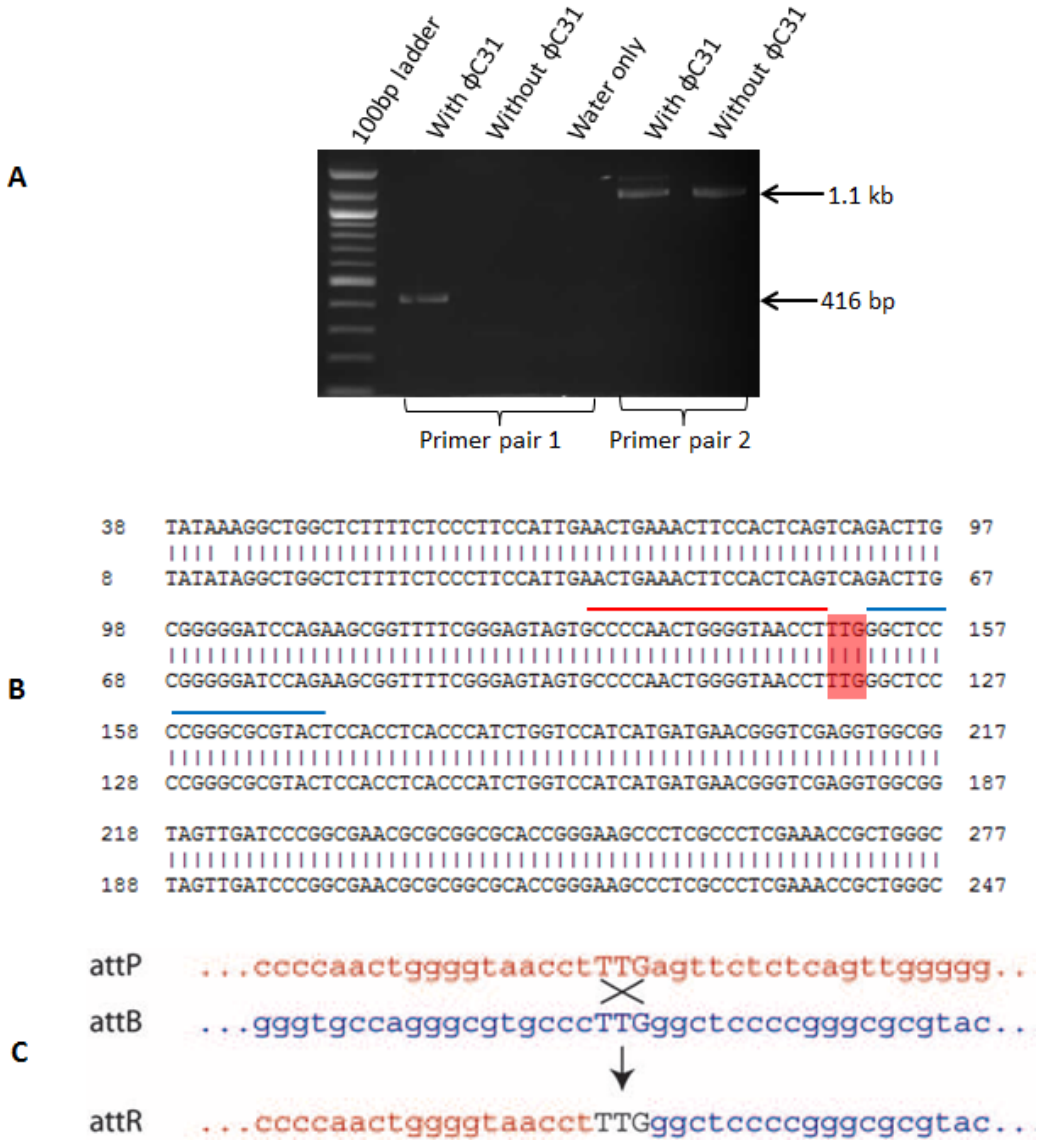

Figure 3-7 Legitimate recombination occurs between the recipient and donor vectors

(A) Image of an agarose gel showing the results of a series of PCRs using two pairs of primers on DNA extracted from embryos injected with recipient vector, donor vector and PhiC31 integrase mRNA or embryos injected with recipient and donor vectors without PhiC31 integrase mRNA. (B) Sequencing data from the 416 bp PCR product aligned to the expected sequence after recombination (according to published data (Bateman et al. 2006)). The red rectangle highlights the 3 bases at which the recombination occurs. The red and blue lines show sequence aligning to attP and attB (respectively) shown in (C). (C) Adapted from (Bateman et al. 2006) showing the partial sequences of attP, attB and attR.

3.2.4 Targeted integration of a donor construct to a predetermined genomic locus in zebrafish

My next aim was to test if PhiC31 integrase can also target integration of a donor vector into a predetermined genomic locus. In order to target integration in the zebrafish, a PhiC31 target site first has to be introduced into the genome. This will be carried out using Tol2 transgenesis to introduce the recipient sequence into the genome of wild type embryos. Due to the mosaic nature of founders, the recipient sequence could integrate into the genome of gamete cells but not lens cells. This meant that all embryos (approximately 100) co-injected with the recipient vector and Tol2 RNA were taken to the zebrafish facility to be grown to sexual maturity.

Founder recipient line fish were identified by outcrossing the adult fish to wild type fish. The F1 embryos were screened at 4 dpf and the fish which gave rise to a clutch containing embryos with GFP expression in the lens were separated. Of 36 adult fish screened, 6 transgenic recipient line founders were identified (16.7 %); none of these showed any ectopic expression which would indicate positional effects and GFP expression was specific to the lens. The germline transmission rates varied from 5-60% between the 6 lines (Table 5-1). The F1 embryos showing lens GFP expression were grown up and 6 stable transgenic lines were established. Once the F1 generation reached sexual maturity they were outcrossed again to wild type fish and the embryos screened for GFP expression. Mendelian genetics appeared to be observed suggesting that the transgene had integrated in a single genomic locus. The GFP expressing

heterozygous F2 generation were grown to adulthood and the F2 adults were used in further experiments.

To establish whether PhiC31 integrase can facilitate integration of a donor vector into the target site of a recipient line embryo, the following experiment was designed. An adult recipient line fish will be crossed with a wild type fish and the fertilised eggs will be injected with donor vector plus PhiC31 integrase mRNA or donor vector alone as a control. The schematic in Figure 3-8 shows the outline of the experiment and the expected outcome. When co-injected with PhiC31 integrase mRNA, the donor vector is expected to recombine with the attP target site in the recipient line and the whole vector be integrated into that locus. This is expected to result in the expression of mCherry in the lens replacing GFP. When the donor vector is injected alone into the recipient line embryos, it is expected that the recombination event will not occur and therefore the embryos will show GFP expression in the lens not mCherry. As a heterozygous recipient line fish will be used, approximately 50% of the offspring is expected to carry the target site for integration.

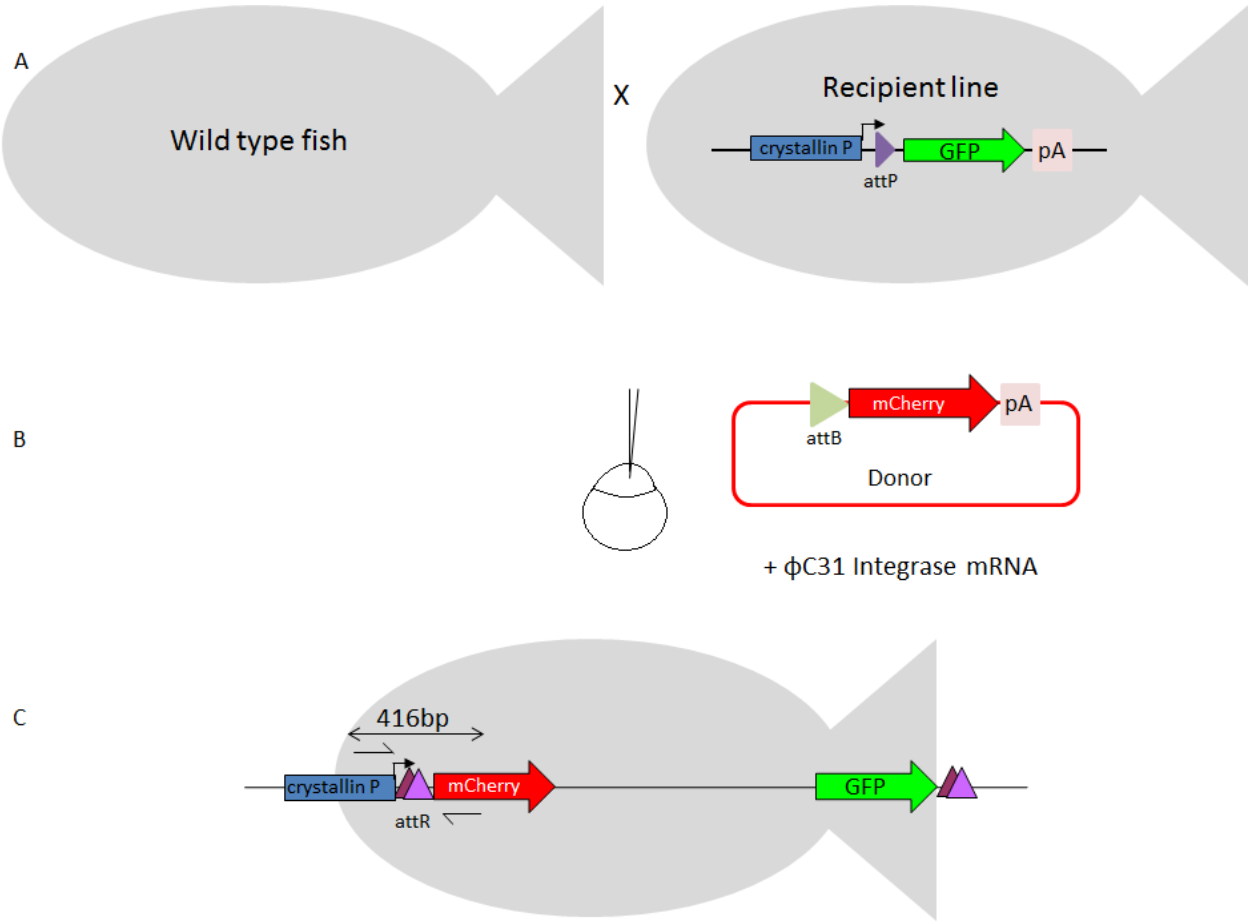

Figure 3-8 Schematic of the experiment to test the function of PhiC31 integrase to target integration into a transgenic zebrafish genome

(A) An adult recipient line fish will be crossed with a wild type fish. (B) The embryos will be co-injected with 100ng/µl donor vector, 20ng/µl PhiC31 RNA and 0.1% phenol red. (C) The expected result is that the donor vector integrates into the target site and the γ-crystallin *promoter now drives expression of mCherry, not GFP.*

In order to screen for targeted genomic integration by switching of reporter gene activity, the embryos were screened at 4 dpf for GFP and mCherry expression in the lens. Over 500 embryos were screened and none of these showed mCherry expression in the lens. However, as expected approximately 50% of embryos expressed GFP in the lens

indicating those embryos that contained the target site. This result suggests that either the donor vector does not integrate into the target site or that mCherry expression is too low to detect/ absent despite the occurrence of targeted integration.

To check whether targeted genomic integration does occur and reporter gene activity is too low to detect, the PCR test for integration was carried out on total DNA extracted from the recipient line embryos injected with donor construct and PhiC31 integrase mRNA and showed the 416 bp band was visible (Figure 3-9A). The PCR product was isolated and purified and then digested with a restriction enzyme which cuts between the γ-crystallin promoter and the attP/R (BamHI) to ensure the PCR product resulted from a legitimate integration. As a control, PCR product that had been sequenced in the previous experiment (Figure 3-7) was also digested with BamHI. These test digests confirmed that the PCR product was specific and the integration had occurred legitimately at the attP site (Figure 3-9B). This data shows that PhiC31 integrase can facilitate the integration of a donor vector into a predetermined site in the zebrafish genome. The lack of reporter gene activity switching means that the frequency of targeted integration in the embryos is still unknown.

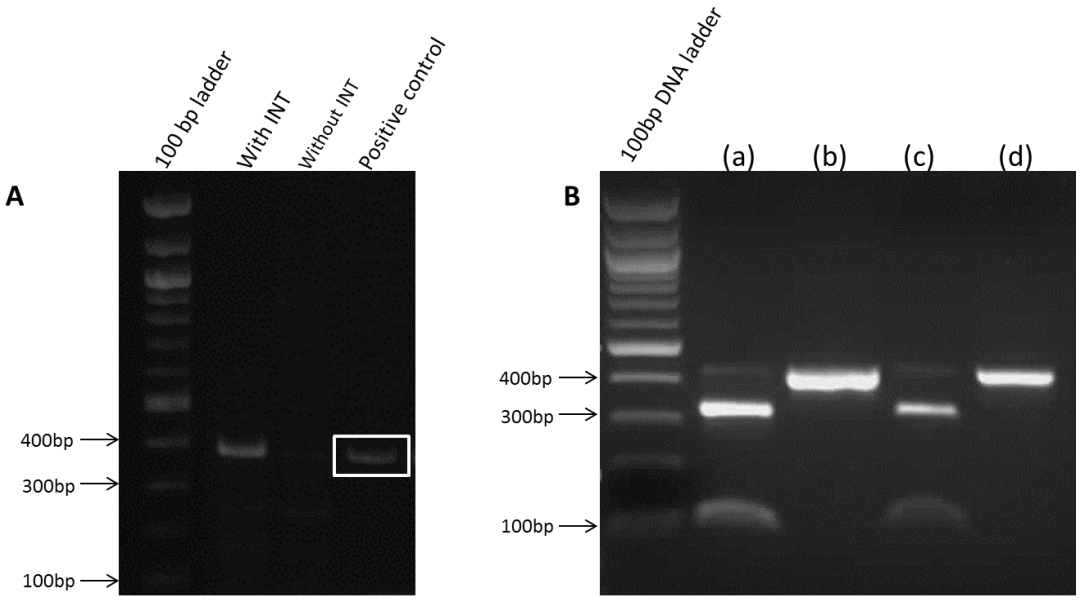

Figure 3-9 The donor vector is legitimately targeted to integration site in recipient line 1

(A) Gel image shows the junction fragment detected by PCR product using a ɣ-crystallin *forward primer and a mCherry reverse primer indicating that PhiC31 integrase has facilitated the correct integration of the donor vector. (B) Gel picture showing PCR products digested with BamHI to show the recombination occurred at the attP site downstream of the ɣ-crystallin promoter. Lane a - DNA was harvested from embryos injected with donor vector and PhiC31 integrase RNA, a PCR was carried out to amplify the integration site and the product was digested between the ɣ-crystallin promoter and the attP site with BamHI. Lane C shows sequenced PCR product from Figure 3-7 digested with BamHI as a positive control. Lanes B and D show undigested PCR products for comparison.*

3.2.5 Optimisation of the visual readout for targeted integration

Targeted integration appeared to be fairly efficient because at least 1 in 10 embryos had targeted integration indicated by the results from PCR on pools of 10 embryos, however no mCherry was detected. A hypothesis was made that translation of mCherry could be affected by the attR sequence between the *ɣ-crystallin* promoter and the start codon. Therefore, the sequence of attB was checked for alternative start codons that could be influencing translation of mCherry. The 285 bp attB sequence from pUASTattB contains

4 ATG codons in the correct orientation (Figure 3-10). Of these four ATGs, three would also be present in the attR sequence after integration. None of these start codons were in frame with mCherry and so all of them could potentially cause a frameshift which could explain the lack of mCherry expression despite evidence of targeted integration. It has previously been published that shorter attP and attB sequences also function and so an 84 bp attB sequence was checked for any ATGs (Figure 3-10). There were no alternative ATGs predicted and so this sequence was used to build a new donor vector in the same manner as described in Figure 3-5.

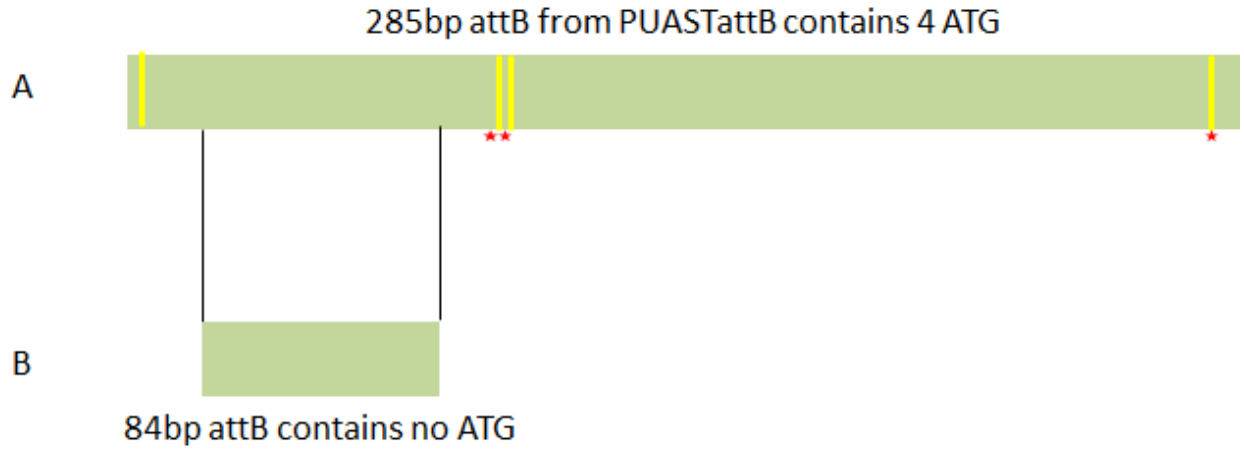

Figure 3-10 Full length attB contains start codons causing a frame shift in mCherry

(A) *Schematic showing the full length attB from pUASTattB used to make transgenic DrosoPhila. The green rectangle represents the attB sequence, the yellow lines mark the locations of the ATG codons that are in the correct orientation and the red stars represent the ATGs that would be present in the attR sequence after integration. (B) Schematic showing the 84bp attB, attachment sites of this length have been shown to function previously (Thorpe & Smith 1998).*

3.2.6 Analysis of the efficiency of PhiC31 integrase-mediated integration into the zebrafish genome

My next aim was to test whether targeted integration of a new donor vector containing a shorter attB site would result in the expression of mCherry in the lens. This new donor vector was tested by co-injection with the recipient vector and PhiC31 integrase into wild type embryos. When these embryos were screened at 4 dpf, 59.3% showed mCherry expression in one or both lenses (Figure 3-11A-B). The percentage of fluorescing embryos showing mCherry expression in the lens was over 98% (Figure 3-11C). These results suggest that PhiC31 integrase facilitates efficient targeted integration of a donor vector into a recipient vector in zebrafish. Throughout the rest of this thesis, donor vector refers to this vector with a shorter attB sequence.

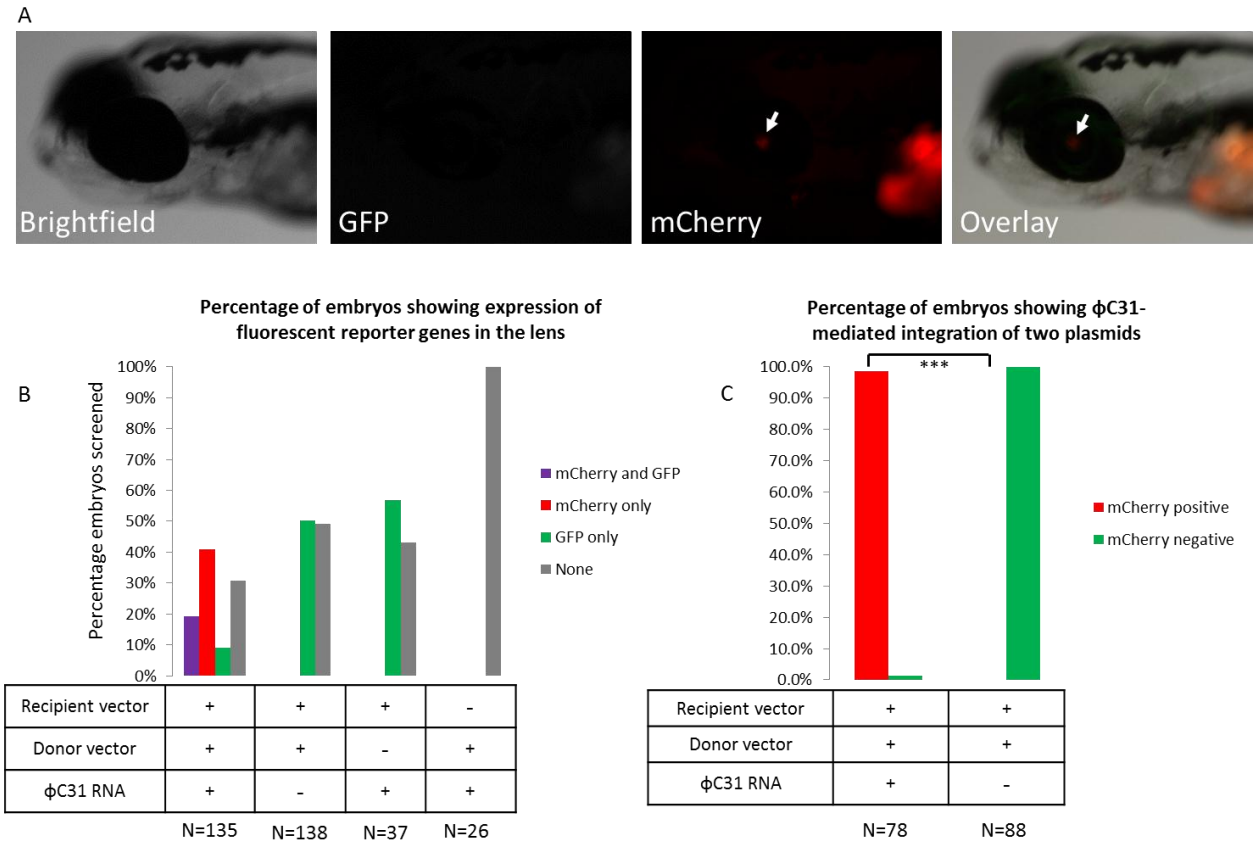

Figure 3-11 The recombination event between co-injected vectors is highly efficient

*(A) Panel A shows an image of a 4 dpf embryo expressing mCherry in the lens. (B) The bar chart shows the percentage of total embryos screened showing mCherry and/or GFP expression in the lens after a number of control solutions were injected. (C) The bar chart displays the percentage of fluorescent embryos that were positive or negative for mCherry expression in the lens. Using the chi-squared test, these results were statistically significant (** p<0.001).*

To test whether the targeted integration of the new donor vector would result in mCherry expression in the lens, the donor vector was injected alone or co-injected with PhiC31 integrase mRNA into recipient line 1 embryos. The embryos were screened at 4 dpf and the results show that more than 70% of the transgenic embryos injected with donor vector and PhiC31 integrase mRNA had mCherry expression in the lens (Figure 3-12A and B). None of the embryos injected with donor vector alone showed mCherry

expression in the lens (Figure 3-12A bottom panel). This supports the evidence that recombination between attP and attB cannot occur randomly in the absence of PhiC31 integrase collected in transient transgenic embryos (Chapter 3.2.3). Another recipient line was also tested for targeted genomic integration. Recipient line 2 embryos were co-injected with donor vector and PhiC31 integrase mRNA and showed very similar rates of targeted integration (around 70%) as recipient line 1 (Figure 3-12C). This provides evidence that PhiC31 integrase can facilitate efficient targeted integration of the donor vector into a predetermined site in the zebrafish genome.

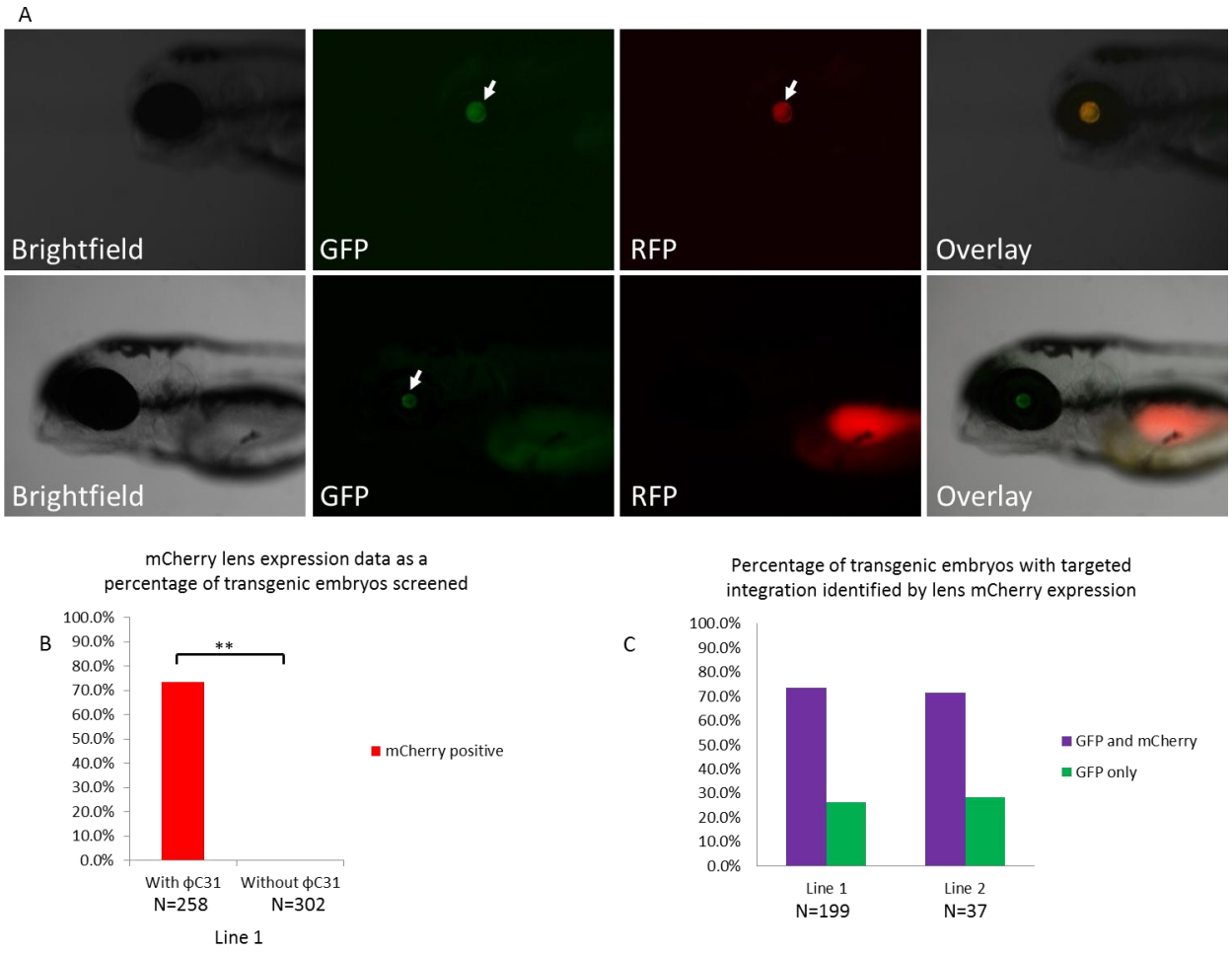

Figure 3-12 Co-injection of donor vector and PhiC31 integrase RNA results in mCherry expression in the lens of over 70% of transgenic embryos

(A) The top panel shows a recipient line 1 embryo which was injected with the donor vector and PhiC31 integrase RNA. The bottom panel shows a recipient line 1 embryo injected with donor vector only. Images are lateral view and anterior to the left.(B) The bar chart shows the percentage of transgenic recipient line 1 embryos that are mCherry positive in the lens when donor vector is injected alone vs co-injected with PhiC31 integrase RNA. (C) A bar chart showing the percentage of transgenic recipient line 1 and 2 embryos injected with the donor vector and PhiC31 integrase RNA that had mCherry expression in the lens. Recipient lines 1 and 2 are lines made from two different founders ie the integration sites are in a different locus in each line.

When screening the injected recipient line embryos for mCherry and GFP expression in the lens, both eyes of the embryo had to be screened as the integration of the donor

vector was mosaic. The majority of mCherry positive embryos only expressed mCherry in one lens (Figure 3-13). This mosaicism also explains why embryos co-expressing mCherry and GFP in the same lens were observed.

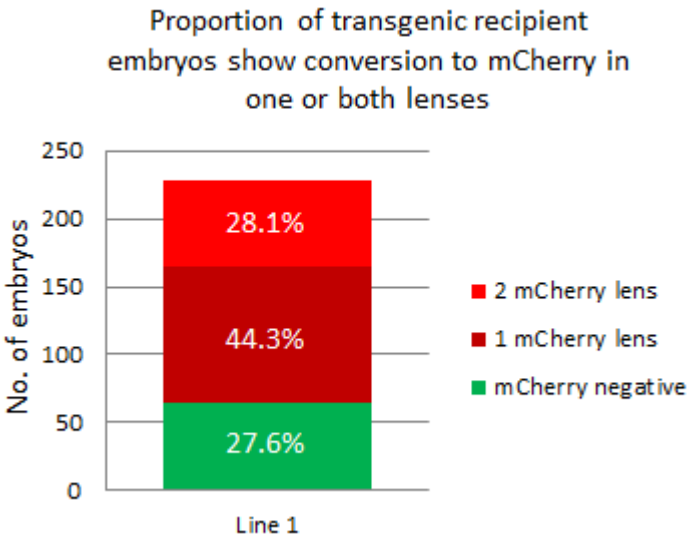

Figure 3-13 Mosaic integration of the donor construct into the recipient line 1 genome

The bar chart shows the proportion of 228 recipient line 1 embryos that show mCherry expression in one, both or neither lens.

3.2.7 Survival of zebrafish injected with PhiC31 integrase mRNA

A high proportion of embryos injected with PhiC31 integrase mRNA were abnormal or did not survive to 24 hpf. To test whether the low survival of mRNA-injected embryos was specific to PhiC31 integrase mRNA, the survival of embryos injected with Tol2 mRNA was used as a comparison. The low survival rate of PhiC31 integrase mRNA-injected embryos was observed consistently throughout the experiments described in this Chapter; on average less than 35% of PhiC31 integrase RNA-injected embryos would

survive to 24 hpf (Figure 3-14). Survival of PhiC31 integrase mRNA-injected embryos was significantly reduced (p=0.0054) when compared with embryos injected with Tol2 mRNA. This result suggests that PhiC31 integrase mRNA is causing a high proportion of embryos to develop abnormally and die early in development.

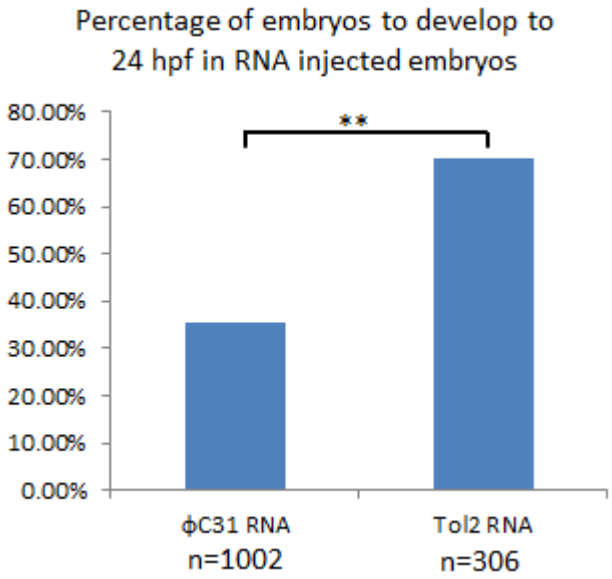

Figure 3-14 Survival of PhiC31 integrase RNA-injected embryos is significantly reduced

*The bar chart shows the percentages of PhiC31 integrase and Tol2 mRNA-injected embryos to survive to 24 hpf. Using a binomial model, the p value= 0.00554 (** represents p<0.01).*

3.3 Discussion

3.3.1 Translation of Phi31 integrase RNA by the zebrafish embryo

At the start of this project, there was no data published on using PhiC31 integrase in zebrafish and so firstly, the delivery of PhiC31 integrase protein had to be confirmed. The standard delivery in other transgenesis methods (eg Tol2 and SB) is injection of mRNA coding for the transposase (Kawakami et al. 1998). Therefore, I needed to establish whether PhiC31 integrase mRNA could be translated using the zebrafish translation machinery. I found that PhiC31 integrase protein was translated by the zebrafish embryo after micro-injection of mRNA. PhiC31 integrase mRNA was injected into early stage wild type zebrafish embryos and at 3.5 hpf, protein was extracted from the embryos and a western blot was performed. A band corresponding to PhiC31 integrase was detected in the PhiC31 integrase mRNA-injected sample but not the non-injected control. This data suggests that PhiC31 integrase expression is induced in the zebrafish by injection of mRNA. However, the western blot does not give us information about efficiency of translation or biochemical activity of the protein.

3.3.2 PhiC31 integrase transgenesis system

The development of a targeted integration system in zebrafish would have a wide range of uses. These had to be considered while designing the PhiC31 integrase system and the constructs involved so that it would be suitable in a wide range of applications.

With the input of our collaborator, Darius Balciunas, we designed an integration system which involves the integration of the whole of the donor construct. This allows the use of a single attP site in the recipient construct or line and a single attB site in the donor vector. Although integration using single recognition sites had never been tested, it had previously been published that PhiC31 integrase can induce excision of a DNA sequence flanked by an attP and attB site in mouse (Sangiorgi et al. 2008) proving that recombination reactions can occur between single attP and attB sites in a vertebrate model. This has now also been shown to occur in zebrafish (Figure 1-7). The use of single recognition sites may improve the efficiency of targeted integration by PhiC31 integrase compared with recombination-mediated cassette exchange (RMCE) utilised in *DrosoPhila* (Bateman et al. 2006) and more recently in zebrafish (Hu et al. 2011). In RMCE, two recombination reactions have to occur in order to integrate the donor sequence flanked by attB sites (Figure 1-8). This double recombination reaction may reduce the efficiency of PhiC31 integrase-mediated integration (Hu et al. 2011).

The integration of the whole construct is expected to allow the targeted integration of larger DNA sequences which may be restricted when using RMCE. This again widens the number of techniques that could be employed using the PhiC31 integrase system; for example, BAC transgenesis. We do, however, need to be aware that the integration of the whole construct leads to the integration of bacterial DNA sequences (in the vector backbone) that could influence transgene activity. Currently, there is no evidence to

suggest that the vector backbone does have a detrimental effect on the PhiC31 integrase system or activity of inserted reporter constructs.

We positioned the promoter upstream of the attP integration site in the recipient construct despite the possibility that this could lead to problems during transcription and translation of the fluorescent reporter genes. This was to ensure that the fluorescent reporter colour change would specifically identify those embryos in which targeted integration has occurred. This design meant that the donor vector did not have to contain any regulatory DNA at all (ie. a promoter). If the donor vector contained a promoter sequence upstream of the mCherry sequence, expression of mCherry could be driven if the vector integrated randomly in the genome non-related to Phi31 integrase (Figure 1-8).

The use of mouse codon optimised Phi31 integrase (Sangiorgi et al. 2008) has been shown to work significantly more efficiently in zebrafish than the phage PhiC31 integrase (Lister 2010). This supports our decision to use the mouse codon optimised PhiC31 integrase sequence to make the RNA for injections.

3.3.3 PhiC31 integrase-mediated integration between attP and attB in zebrafish

Once the system had been designed and the constructs made, I needed to ascertain whether PhiC31 integrase can function in zebrafish to integrate two constructs, each containing a recognition site (attP in the recipient and attB in the donor). I have provided data that shows that PhiC31 integrase not only can facilitate the integration between these two constructs but that it does so with an efficiency of approximately 98% (Figure 3-11). This demonstrates that PhiC31 integrase can facilitate whole vector integration between single attP and attB sites. This was a good start but is not particularly useful as this is in transient transgenics but we want to target integration into the genome. This experiment was carried out to show the system has potential and meant that I needed to make the recipient lines to target genomic integration.

3.3.4 Targeted integration into the genome of a transgenic zebrafish

In order to study *cis*-regulatory elements in zebrafish, transgenesis methods are employed to integrate a transgene into the genome. In order to use PhiC31 integrase as a transgenesis system, a recipient transgenic line needs to be made with a target integration site, attP. Using a recipient line containing a target site, I showed that PhiC31 integrase can facilitate the integration of a whole donor construct into the site (Figure 3-9). This targeted integration occurs in over 70% of embryos screened at 3-4 dpf (Figure 3-12). Targeted integration into the zebrafish genome by RMCE using PhiC31 integrase has been published previously (Hu et al. 2011); however, it has never been

shown that PhiC31 integrase can facilitate integration between single integration sites rather than RMCE. The efficiency of targeted integration in F0 embryos using RMCE has not been specified in previous publications and so we still don't know if the efficiency is improved by the use of single recognition sites.

3.3.5 Visualisation of targeted integration *in vivo*

The identification of embryos with targeted integration is essential whilst the embryo is alive in order to be able to study the effects as the embryo develops. This means that DNA extraction and PCR are not suitable for screening targeted integration. A fluorescent reporter system is the ideal way to identify those embryos in which integration of the construct has been targeted to the correct locus. After initial issues with the translation of mCherry after targeted integration were solved (Figure 3-10), our system provided a clear marker for embryos in which targeted integration had occurred.

We selected a lens-specific promoter as our collaborators have previously studied the promoter (Davidson et al. 2003) and the lens is a small tissue that doesn't interfere with potential tissues of interest. In recent publications of the use of PhiC31 integrase in zebrafish, regulatory DNA for the *sox10* (Hu et al. 2011) and *ef1α* (Lister 2010) genes have been used to drive fluorescent reporter expression. The expression of the reporter gene is therefore throughout the whole embryo meaning it is not suitable for studying novel enhancers using another fluorescent reporter gene.

3.3.6 Mosaicism of targeted genome integration

A common issue with injection of zebrafish embryos, including when studying fluorescent reporter expression patterns driven by novel enhancers, is mosaicism of the injected embryos (Fisher et al. 2006). The mosaic pattern may mean that very few cells express the fluorescent protein and therefore either cannot be detected or the tissue cannot be identified. As I have shown in Figure 3-13, targeted genomic integration in zebrafish by PhiC31 integrase is mosaic. This is supported by the observations that GFP expression is not lost even when the lens also shows mCherry expression and that the majority of embryos showing mCherry expression only had it in one eye. Therefore the integration must have occurred after the progenitor lens cells had been split into left and right. This mosaicism also occurs when using RMCE in zebrafish shown by the co-expression of both GFP and dsRed reporters (Hu et al. 2011).

To combat this problem, stable transgenic lines often have to be made using Tol2 transposase which is both time consuming (4 months to grow the fish) and costly (to maintain large numbers of founder fish in the facility). To reduce mosaicism in injected embryos would enable us to better screen F0 fish saving a lot of time and money compared to making stable lines for each enhancer construct. Reducing mosaicism would also increase the chances of germline transmission of the transgene, improving the efficiency of making stable transgenic zebrafish lines.

3.3.7 Toxicity of PhiC31 integrase

The data in Figure 3-14 clearly shows that survival of PhiC31 integrase mRNA-injected embryos is significantly reduced compared with Tol2 mRNA-injected embryos. This has also been observed by others using PhiC31 integrase in zebrafish (Hu et al. 2011). They found that when PhiC31 integrase mRNA was injected at more than 20 pg per embryo they observed developmental abnormalities and when it was injected at 40 pg per embryo the frequency of embryos showing abnormalities was high. They chose to reduce the dosage to 20 pg and have normally developed embryos. I decided to inject at higher concentrations of PhiC31 integrase mRNA, 20 ng/µl (approximately equivalent to 40 pg per embryo), because I found that the efficiency of PhiC31 integrase dropped when using lower PhiC31 integrase mRNA concentrations. A higher concentration of PhiC31 integrase mRNA, although reduces survival, allows a higher rate of integration and has the advantage of less embryos to screen and less adult fish required to find a founder. This means that fewer lines have to be grown up, saving space and cost maintaining the fish.

The cause of the toxicity of PhiC31 integrase RNA is unknown. Double-stranded RNA (dsRNA) injected at high concentrations is known to cause general developmental defects in zebrafish embryos (Zhao et al. 2001). However, the toxicity is specific to PhiC31 integrase as when Tol2 transposase RNA is injected at the same concentration,

survival is significantly higher and so dsRNA cannot be causing the developmental abnormalities.

It has previously been found that chromosomal translocations occur in human fibroblasts in the presence of PhiC31 integrase (Liu et al. 2009). One logical explanation for the toxicity in zebrafish is that PhiC31 integrase causes recombination effects between pseudo attP sites in the zebrafish genome, causing lethal translocations of chromosomes. However, using a BLASTn search of the zebrafish genome on NCBI (Altschul et al. 1990) for the 84 bp attP sequence gives 3 'somewhat similar' genomic sequences. None of these contain both the essential TTG and surrounding region required for PhiC31 integrase-facilitated recombination. This suggests that there are no pseudo attP sites in the zebrafish genome that could result in chromosome translocations in the presence of PhiC31 integrase. The cause of the toxicity specific to PhiC31 integrase mRNA still remains unknown.

3.3.8 Conclusions

PhiC31 integrase can facilitate targeted integration into the genome of a transgenic zebrafish embryo. I have designed (along with collaborators) and verified an efficient targeted genomic integration system in zebrafish, utilising fluorescent reporter genes to identify embryos in which targeted integration has occurred. This system will provide another tool for the zebrafish research community to study cis-regulatory DNA. This

system could also be used for any experiment requiring integration into a consistent genomic context.

4 COMPARING THE EFFICIENCY OF PhiC31 INTEGRASE-MEDIATED GENOME INTEGRATION USING DIFFERENT DELIVERY METHODS OF PhiC31 INTEGRASE: RNA VS MATERNAL PROTEIN

4.1 Introduction

Mosaicism of transient transgenic embryos often causes problems when studying integrated enhancer-promoter-reporter sequences (Fisher et al. 2006). This mosaicism is caused by a delay in integration of the transgene due to the production of functional protein by translation of the mRNA. This means that the PhiC31 integrase protein may only be present in a subset of the embryo's cells which results in a mosaic F0 fish (Figure 3-13). Providing functional protein earlier may reduce mosaicism and reveal enhancer pattern in founder fish. This could be a much quicker method of studying enhancers rather than growing fish to adulthood for 4 months to screen the non-mosaic F1 generation embryos. A reduction in mosaicism would also increase the efficiency of germline transmission when making stable transgenic zebrafish lines.

An oocyte-specific *cre* transgenic zebrafish line was previously made to provide Cre in the early embryo (Liu et al. 2008). This was made using the zebrafish *zona pellucida 3* (*zp3*) promoter previously characterised by the same group (Liu et al. 2006). Zona pellucida is a glycoprotein expressed specifically in the oocytes of female vertebrates (Liu et al. 2006). In this paper, they make a *zp3:gfp* transgenic line and show GFP

expression in zebrafish oocytes and early embryos until it gradually disappears by 3 dpf. The presence of GFP at the one-cell stage suggests that a *zp3:PhiC31* line will provide a functional protein at the time of micro-injection of a DNA construct. As a result, integration is expected to occur earlier and therefore mosaicism would be reduced and enhancer tissue-specificity patterns could be revealed in injected embryos.

Providing maternal PhiC31 integrase protein rather than RNA may also improve the survival of embryos.

4.1.1 Aims

To develop a transgenic line to provide PhiC31 integrase protein in the early zebrafish embryo with the expectation that it will allow earlier integration of the donor vector and therefore result in less mosaic F0 embryos. I then plan to test the function of this maternal PhiC31 integrase transgenic line by crossing the fish with a recipient line fish and injecting the embryos with the donor construct.

4.2 Results

4.2.1 Design and construction of a vector to make a maternal PhiC31 integrase transgenic line in zebrafish

With the aim of integrating an enhancer-promoter-reporter construct at an earlier developmental stage, a maternal PhiC31 integrase line female will be crossed with a recipient line male and the embryos injected with the donor construct (Figure 4-1). Due to the presence of functional PhiC31 integrase protein at the time of injection of the DNA construct is expected to result in earlier integration of the donor construct which is expected to manifest in both lenses showing mCherry expression and increase efficiency of transmission of the transgene to the next generation through the germline.

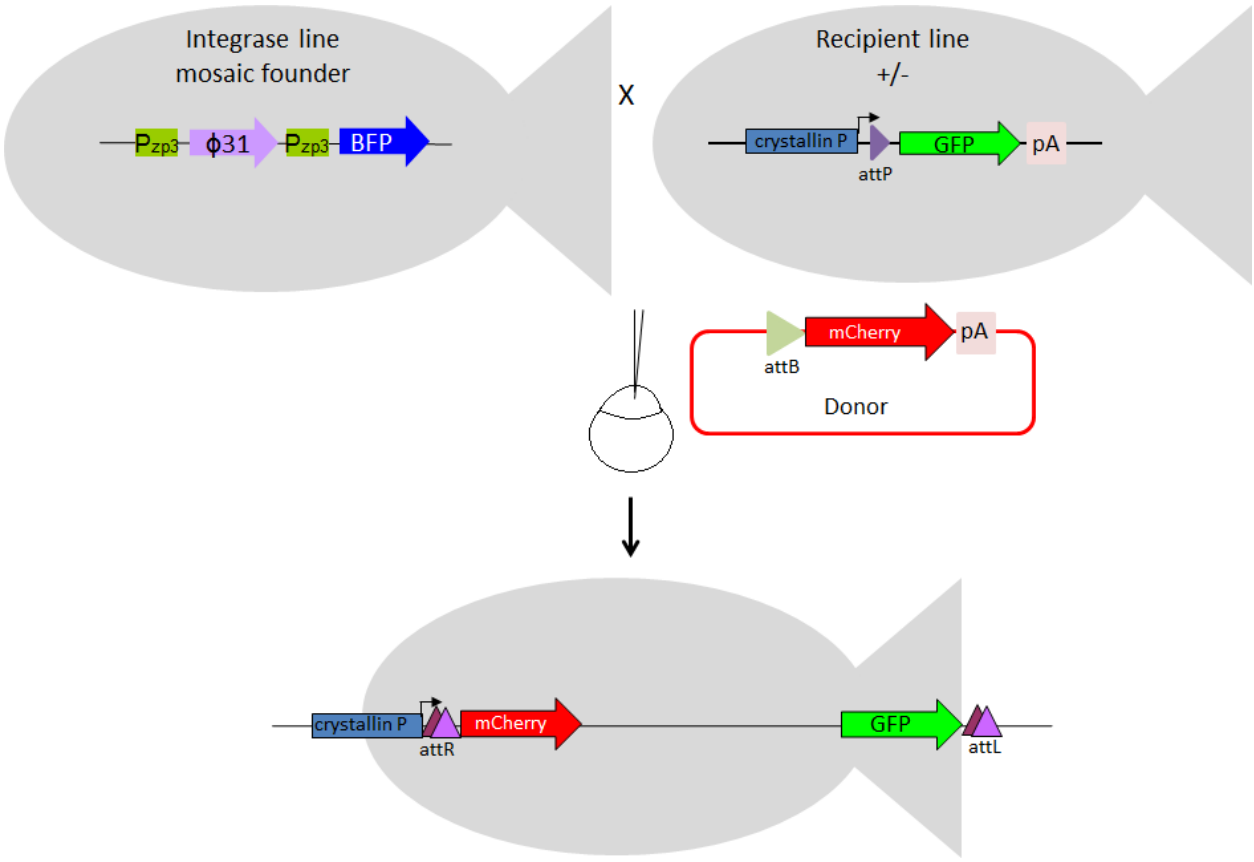

Figure 4-1 Schematic showing the use of a maternal PhiC31 line to target genomic integration

(A) represents a maternal PhiC31 integrase line female fish being crossed to a male recipient line fish. (B) Embryos will then be injected with the donor construct alone. (C) The mCherry positive embryos are expected to show a less mosaic pattern ie. they would show mCherry expression in both lenses not just one.

In order to create a maternal PhiC31 integrase transgenic line, a construct was made containing two *zp3* promoter sequences driving expression of both PhiC31 integrase and blue fluorescent protein (BFP) (Figure 4-2). BFP was included as a screening marker for identification of transgenic embryos, allowing them to be selected for analysis.

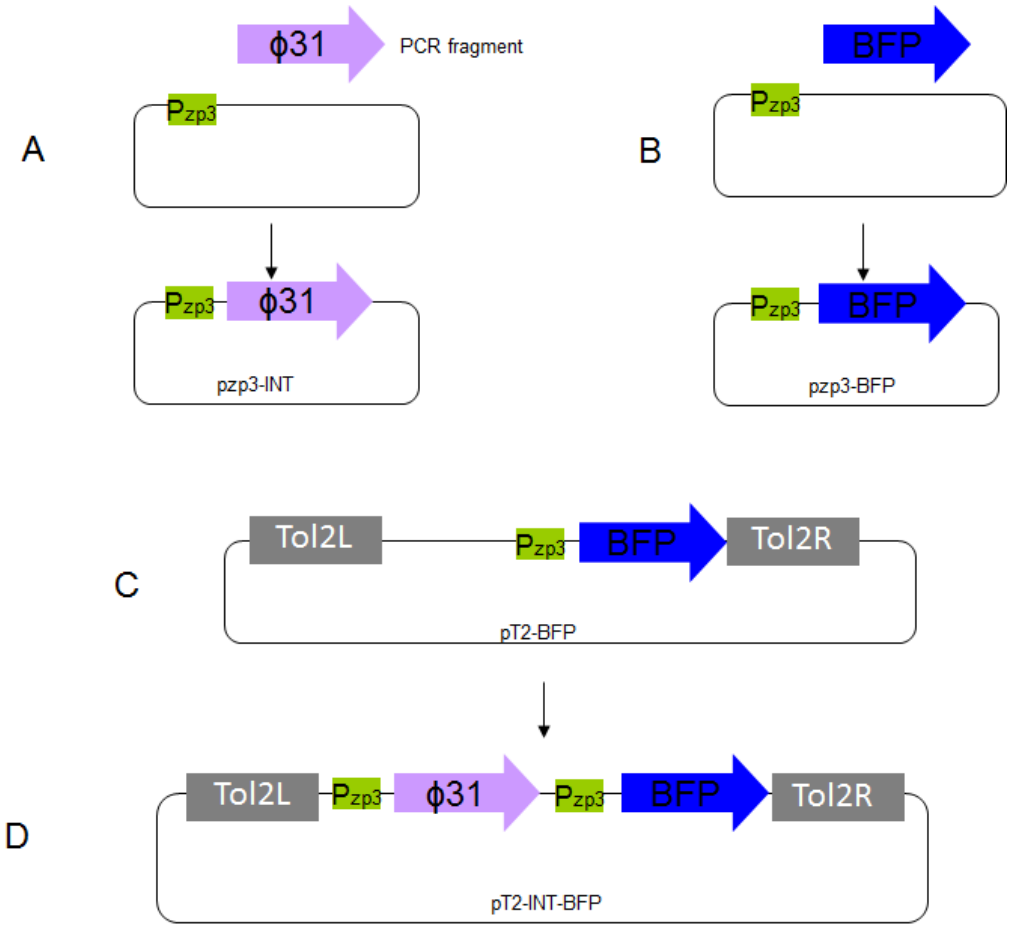

Figure 4-2 Schematic of the cloning steps to make a maternal PhiC31 integrase line

Firstly, PhiC31 integrase was cloned into a vector containing the zp3 promoter sequence (Pzp3) (A). Simultaneously, BFP was cloned into the same vector (B).The Pzp3-BFP fragment was then cloned into pT2KXIGΔin which contains Tol2 arms for DNA integration (C). Lastly the Pzp3-PhiC31 sequence was cloned upstream of the Pzp3-BFP to make pT2-INT-BFP (D).

4.2.2 Generation and identification of maternal PhiC31 integrase founder fish

To generate a stable maternal PhiC31 integrase line, the construct shown in Figure 4-2 was co-injected with Tol2 mRNA and approximately 60 embryos were grown to adult fish. For this transgenic line the injected fish cannot be screened, only the F1

generation. Once injected fish were 4-5 months old the females were outcrossed and the embryos were dechorionated at 1-cell stage and screened on a fluorescence microscope as soon as possible (Figure 4-3). Once a founder female fish was identified by the detection of BFP in her offspring, all her eggs were grown up to generate a stable line. In the meantime, the founder females were used in subsequent experiments.

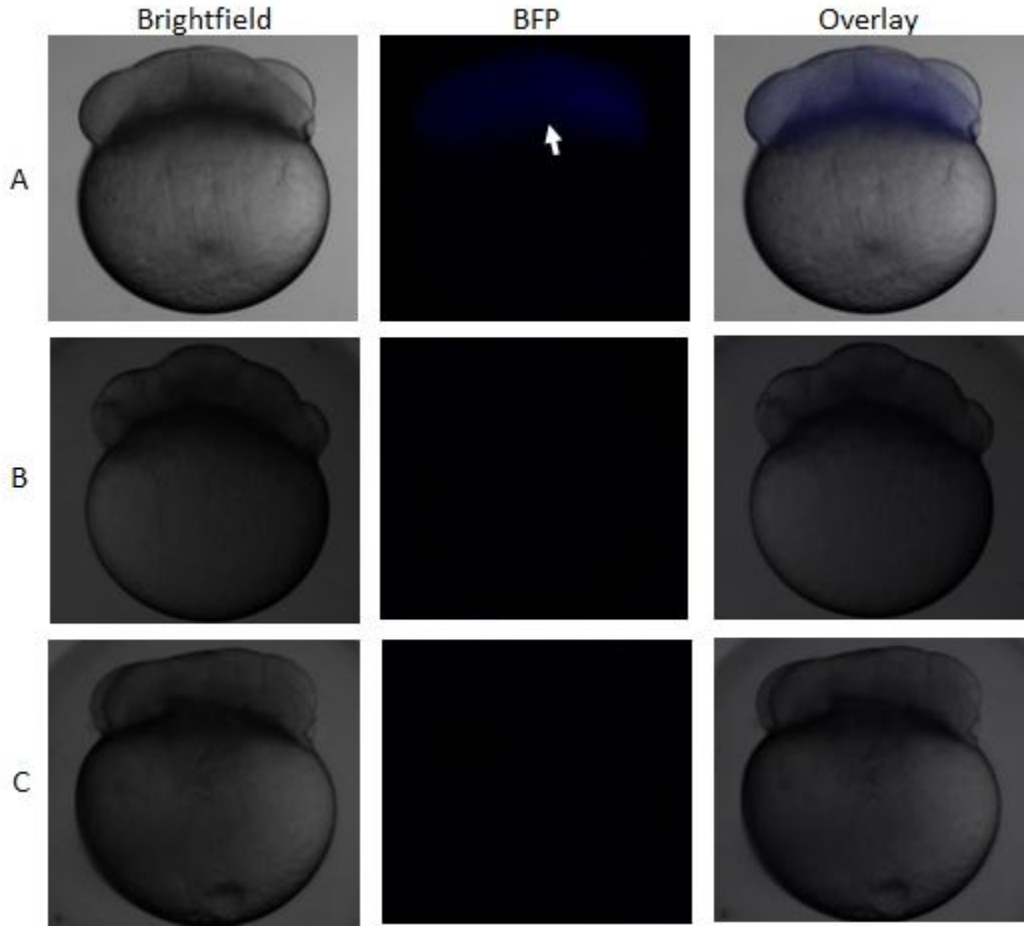

Figure 4-3 Images of a maternal PhiC31 line showing BFP expression in the early embryo

(A*) shows images of an 8 cell stage F1 BFP positive embryo from the maternal Phi31 integrase line 1. (B) shows images of an 8-16 cell stage F1 BFP negative embryo from the founder of maternal PhiC31 integrase line 1. (C) shows images of an 8 cell stage control non-injected wild type embryo.*

4.2.3 Targeted integration function of the maternal PhiC31 integrase protein

To assess whether providing PhiC31 integrase protein maternally could facilitate targeted integration into the target site of a recipient line embryo, two founder females were tested by crossing with recipient line males and injecting the embryos with the donor construct. The embryos were screened before epiboly began and the BFP positive embryos were separated from the BFP negative embryos in order to identify at later developmental stages which embryos had had maternal PhiC31 integrase. The embryos were incubated at 28.5 °C and the medium was changed every day. At 4 dpf the embryos were screened for mCherry and GFP expression in the lens.

Maternal integrase line 1 offspring showed a reduced level of survival among the BFP positive embryos compared to their sibling BFP negative embryos (Figure 4-4A). When line 1 BFP positive embryos were screened for mCherry expression in the lens, over 50% of embryos containing the target site showed targeted integration (Figure 4-4B). Unexpectedly, there were also around 11% of 'BFP negative' embryos showing mCherry expression. All of my previous data suggests targeted integration does not occur without the presence of PhiC31 integrase. Therefore, I propose that these mCherry positive embryos are embryos that were BFP positive but were not selected during the early screening due to the difficulties encountered in detecting low levels of BFP fluorescence.

Maternal PhiC31 integrase line 2 showed no reduction in survival of BFP positive embryos compared to their BFP negative siblings. When they were screened at 4 dpf for

mCherry expression in the lens, no positives were found to indicate targeted integration in either BFP positive or BFP negative embryos (Figure 4-4B).

Together, these data support the data in Chapter 3 to suggest that the concentration of PhiC31 integrase in the early zebrafish embryos influences survival of those embryos. These data also provide evidence for the hypothesis that lower PhiC31 integrase concentrations, result in a reduced frequency of targeted integration.

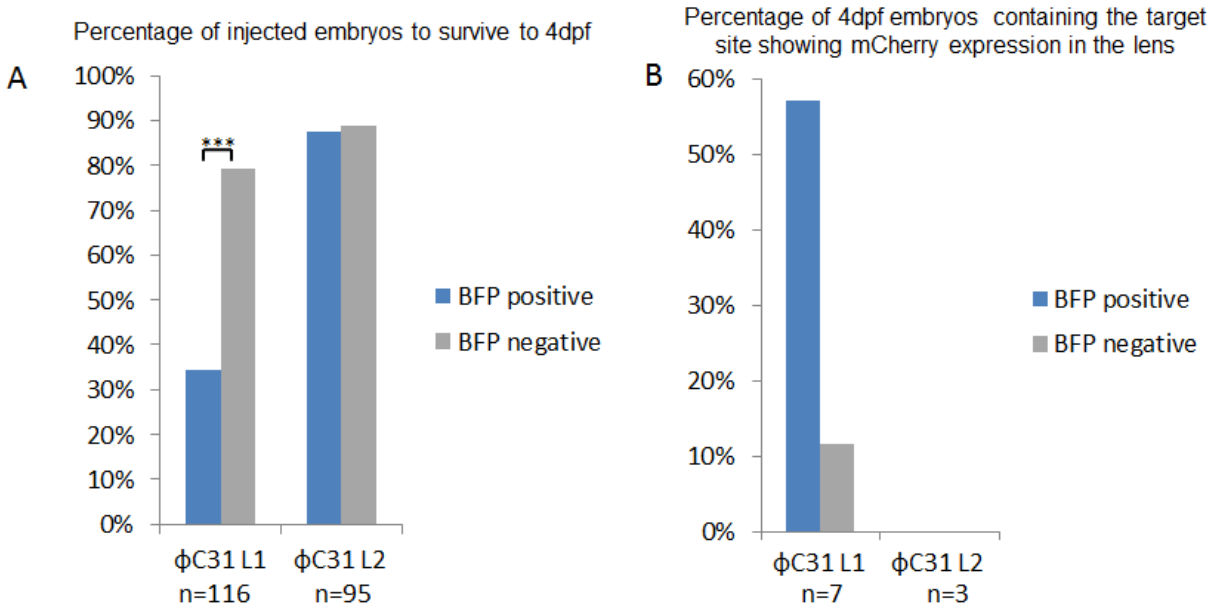

Figure 4-4 Efficiency of targeted integration by PhiC31 integrase is linked to survival of embryos

(A) *Bar chart showing the percentage of embryos from a cross between the maternal Phi31 integrase line and the recipient line injected with the donor construct to survive to 24 hpf. (*** p<0.001). (B) A bar chart displaying the percentage of embryos containing the attP target site that show mCherry expression in the lens.*

4.2.4 The effects of PhiC31 integrase delivery method on mosaicism

In order to assess whether the maternal PhiC31 integrase allowed early integration and reduced mosaicism, the percentage of mCherry positive embryos that showed expression in both lenses was analysed. The data analysed was from the experiment described in Chapter 3.2.6 when recipient line embryos were injected with donor vector and PhiC31 integrase mRNA and the proportion of embryos showing mCherry expression in one, both or neither lens was shown in Figure 3-13. The mosaicism of the embryos resulting from crosses of the recipient line fish and maternal PhiC31 integrase line fish (Chapter 4.2.3) were also assessed in the same manner, by recording whether mCherry was expressed in one, both or neither lens. When using maternal PhiC31 integrase line 1, 44.4% of mCherry positive embryos showed expression in both lenses. This was compared to the percentage of mCherry positive embryos displaying mCherry expression in both lenses when using Phi31 integrase mRNA, which was 38.3%. Binomial tests were carried out to confirm that there is no significant difference between these data (p=1.0) which suggests that providing maternal PhiC31 integrase does not result in reduced mosaicism compared to providing PhiC31 integrase mRNA.

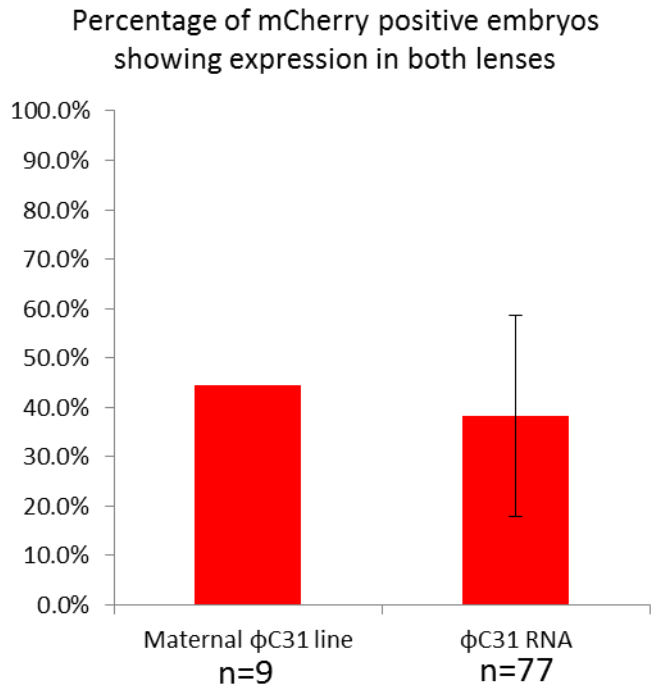

Figure 4-5 There is no evidence to suggest that the maternal PhiC31 integrase line reduces mosaicism

The bar chart compares the percentage of mCherry positive embryos that had mCherry expression in both lenses when using the maternal PhiC31 integrase line or PhiC31 integrase RNA. Error bars show 1 standard deviation from the mean of four different experiments carried out on different days.

In order to generate stable maternal PhiC31 integrase lines, F1 adult fish were screened for transgenics. To more rapidly screen for transgenic F1 fish, fin clips were taken and the genomic DNA extracted. When this genomic DNA was tested by PCR for presence of the PhiC31 integrase coding sequence, no transgenic F1 fish were identified out of 48 fish tested from 3 maternal PhiC31 integrase lines. This lack of transgenic F1 fish may be due to the reduced survival of embryos containing PhiC31 integrase.

4.3 Discussion

The aim of this work was to reduce mosaicism in embryos injected with transgenes that were integrated into the genome. A reduction in mosaicism is expected to improve the study of putative enhancer sequences in vivo in enhancer-promoter-reporter construct-injected embryos. It would also improve the efficiency of transmission of the transgene through the germline, which would be useful in many research labs who generate stable transgenic zebrafish. Previously published data show that a maternal *cre* line can facilitate excision of loxP flanked DNA sequences in the zebrafish embryo (Liu et al. 2008). We hypothesised that PhiC31 integrase could do the same and that providing functional protein at the time of injection of the donor vector would mean earlier transgene integration and therefore, less mosaicism.

4.3.1 Maternal PhiC31 integrase construct design

The oocyte-specific *zp3* promoter was selected as it has already been characterised in zebrafish and a *zp3:gfp* transgenic line showed that protein was present at the 1-cell stage. It has already been used to make a functional maternal cre transgenic zebrafish line (Liu et al. 2008) and this makes it ideal for making a maternal PhiC31 integrase line.

The fluorescent reporter used to screen the maternal PhiC31 integrase line had to be designed carefully so as not to interfere with expression patterns of other fluorescent proteins utilised in the system. As described previously, protein expressed from the *zp3* promoter doesn't degrade until around 3 dpf (Liu et al. 2006). This immediately ruled out

both GFP and mCherry/RFP as screening tools for the maternal PhiC31 integrase line. We also plan to use cyan fluorescent protein (CFP) and yellow fluorescent protein (YFP) to label enhancer-promoter expression in future experiments. This left BFP as the only fluorescent protein that we could detect without interference and bleed through from other fluorescent protein signals detected by our fluorescent microscopes. BFP is not a commonly used fluorescent protein in zebrafish and so it was tested for activity prior to making this construct. The BFP coding sequence was cloned into the pCS2+ vector and mRNA was synthesised *in vitro*. This mRNA was then injected into wild type embryos and they were screened at 24 hpf for BFP fluorescence which was easily detectable in these embryos (data not shown). This showed that BFP could be used in zebrafish and we had the means to detect it *in vivo*. Despite this data, BFP did not prove to be an effective screening tool for this maternal transgenic line. The PhiC31 integrase lines were not very easily screened as the BFP was very weak and difficult to detect. This made the screening process lengthy and not 100% accurate (Figure 4-4).

4.3.2 Targeted genomic integration using a PhiC31 maternal integrase

Despite the issues with screening, the maternal PhiC31 integrase line can facilitate targeted genome integration into a recipient line (Figure 4-4). This shows that the zp3 promoter is driving expression of the PhiC31 integrase as well as the BFP. However, the second line tested did not show any embryos with targeted integration; this could be due to variability in the concentration of PhiC31 integrase present caused by position effects as the transgene integrated randomly.

4.3.3 The effects of maternal PhiC31 integrase on mosaicism

By screening both lenses for the expression of mCherry, I provided a way of assessing the mosaicism of the injected embryos. In embryos expressing mCherry in only one lens, the integration must have occurred after the lens lineage cells split into right and left. However, embryos expressing mCherry in both lenses must have had an earlier integration resulting in a less mosaic pattern. Based on observation of integrations using both PhiC31 integrase RNA and the maternal PhiC31 integrase line, there is no evidence to suggest that a maternal PhiC31 integrase line reduces mosaicism in F0 embryos.

4.3.4 Technical problems with the use of maternal Phi31 integrase lines

The maternal PhiC31 integrase founder fish were unreliable at laying eggs which made it very difficult to obtain enough embryos to carry out the experiments. This led to lower numbers of embryos screened reducing the chances of getting statistically significant results. The cause of the lack of numbers of eggs laid is unknown; it could be related to the presence of toxic levels of PhiC31 integrase in the oocytes of the transgenic females.

When founder females were identified by clutches containing BFP expressing embryos, the clutches were grown up to make a stable transgenic line. However, when the F1 fish were screened, no transgenic fish were found out of 48 screened from 3 different maternal lines. It is possible that the transgene was selected against as the PhiC31 integrase positive ones seem to be less likely to survive.

Others have addressed the same issue by using PhiC31 integrase mRNA with an untranslated region (UTR) of the *nanos1* gene which targets the mRNA to the germline precursor cells and therefore is expected to result in more efficient transmission of the transgene to the next generation (Hu et al. 2011). Injection of PhiC31 integrase-nanosUTR mRNA is not reported to reduce survival of embryos that occurs with the PhiC31 integrase mRNA that I used.

4.3.5 Conclusions

From my experiments there is no evidence to suggest that providing PhiC31 integrase maternally reduces mosaicism and the survival of embryos does not appear to be improved. This suggests that using a maternal PhiC31 integrase line may not provide an advantage over injections carried out with PhiC31 integrase mRNA. Taking this into consideration and the problems encountered when trying to make a stable line, the maternal PhiC31 integrase line does not look to be a suitable alternative to PhiC31 integrase RNA. The use of PhiC31 integrase-nanosUTR mRNA by other researchers may improve the efficiency of transgene transmission in the germline but this will not aid the study of enhancers in F0 embryos.

5 NOVEL ENHANCER FUNCTION IDENTIFIED BY USING THE PhiC31 INTEGRASE SYSTEM

5.1 Introduction

To test if enhancer function could be reliably detected without position effects using the PhiC31 integrase transgenesis system, an enhancer-promoter-reporter construct known to show strong positional effects using Tol2 transgenesis will be investigated using the PhiC31 integrase system.

5.1.1 Tol2 transgenesis is prone to position effects masking enhancer function

During a previous study by Yavor Hadzhiev, a putative enhancer was found located in the 5th intron of the *erry* gene (Figure 5-1). The 794 bp putative enhancer sequence was cloned into a gateway vector containing the *keratin4* minimal promoter (*krt4*) and Venus YFP (Figure 3-1). This construct was co-injected with Tol2 mRNA into zebrafish embryos and the YFP expression pattern studied at approximately 36-48 hpf. Analysis of F0 embryos was unsuccessful in ascertaining whether the putative enhancer was driving expression of *erry* as the expression patterns observed were broad, varied and did not reflect the pattern expected for the putative target gene (Bardet et al. 2004; Thisse et al. 2004). In an effort to reduce the background YFP expression, stable lines were made by growing up these injected embryos. When these F0 fish were outcrossed to wild type fish and the F1 embryos were screened for YFP expression, different transgenic lines showed variability in YFP expression intensity and pattern (Figure 3-1).

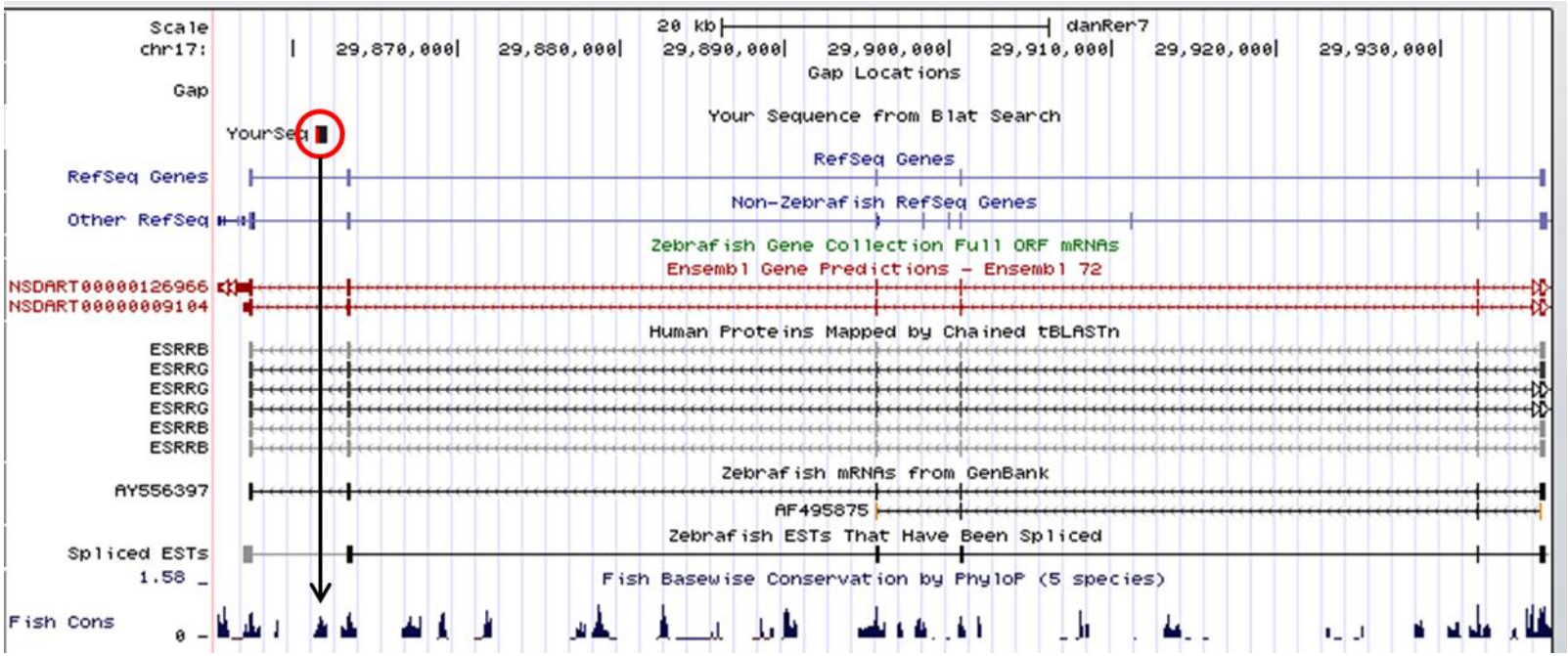

Figure 5-1 Genome browser view of the putative errγ enhancer

This UCSC genome browser view shows the putative errγ enhancer (red circle) and it's sequence conservation with 8 other species including pufferfish, human and mouse.(black arrow to conservation data in purple).

5.1.2 Estrogen-related receptor ɣ

Errɣ belongs to a group of estrogen-related receptors including Errα and Errβ. As their name suggests, they are closely related to the estrogen receptor (ER) family of receptors. However, they have no known ligands and are referred to as orphan receptors. Errɣ binds to the estrogen response element (ERE) to regulate transcription of a number of genes. They have been implicated in breast cancer and other cancers and are now the subject of research for treatments of cancer (Ariazi & Jordan 2006).

As shown in Figure 5-2, Errɣ is expressed in a restricted manner in sections of the midbrain and hindbrain as well as the tail bud, somites and pronephric tubes (Bardet et al. 2004; Thisse et al. 2004). This was not the pattern observed when studying Tol2 stable transgenic lines (Figure 3-1) and enhancer function could not be unambiguously identified.

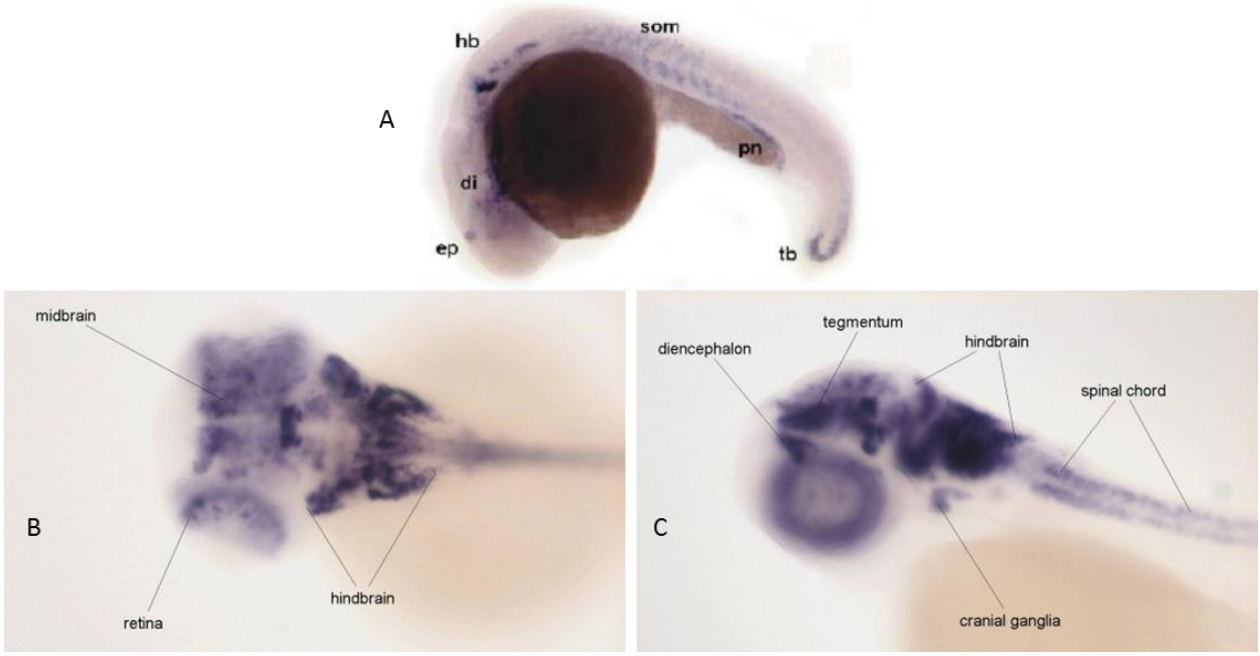

Figure 5-2 Expression pattern of Errɣ in zebrafish

(A) In situ shows the expression pattern in a 36hpf embryo – lateral view anterior to the left. Expression is visible in the epiphysis (ep), diencephalon (di), hindbrain (hb), somites (som), pronephric tubes (pn) and the tail bud (tb). Adapted from (Bardet et al. 2004). (B,C) show the expression pattern in a 48-60hpf embryo. Expression can be seen in the retina, midbrain, hindbrain and spinal cord. Adapted from (Thisse et al. 2004). (B) is a dorsal view and (C) is a lateral view, both oriented anterior to the left.

5.1.3 Aims

To test whether targeted integration of an enhancer-promoter-reporter construct can allow the detection of the enhancer's function, it will be integrated into the target site of recipient line 1. Firstly, the gateway construct will be modified to insert the attB-mCherry sequence from the donor vector. The construct will then be co-injected with either Tol2 mRNA into wild type embryos or with PhiC31 integrase mRNA into recipient line embryos. These embryos will then be screened at 48 hpf to analyse the YFP expression pattern. The expectation is that PhiC31 integrase will reproducibly restrict expression to

those tissues where the enhancer is active. Tol2 is expected to result in the varied, broad expression patterns observed previously.

5.2 Results

5.2.1 Modification and testing of the enhancer-promoter-reporter vector

Firstly, the EL18-*krt4-yfp* construct obtained from Dr Yavor Hadzhiev, was modified by cloning in the attB-*mCherry* sequence that will allow the construct to be integrated into the PhiC31 integrase target site of the recipient line. A schematic of the new vector (EL18-*krt4-yfp*-attB-*mCherry*) is shown in Figure 5-3; this was then used in the enhancer test experiments described below.

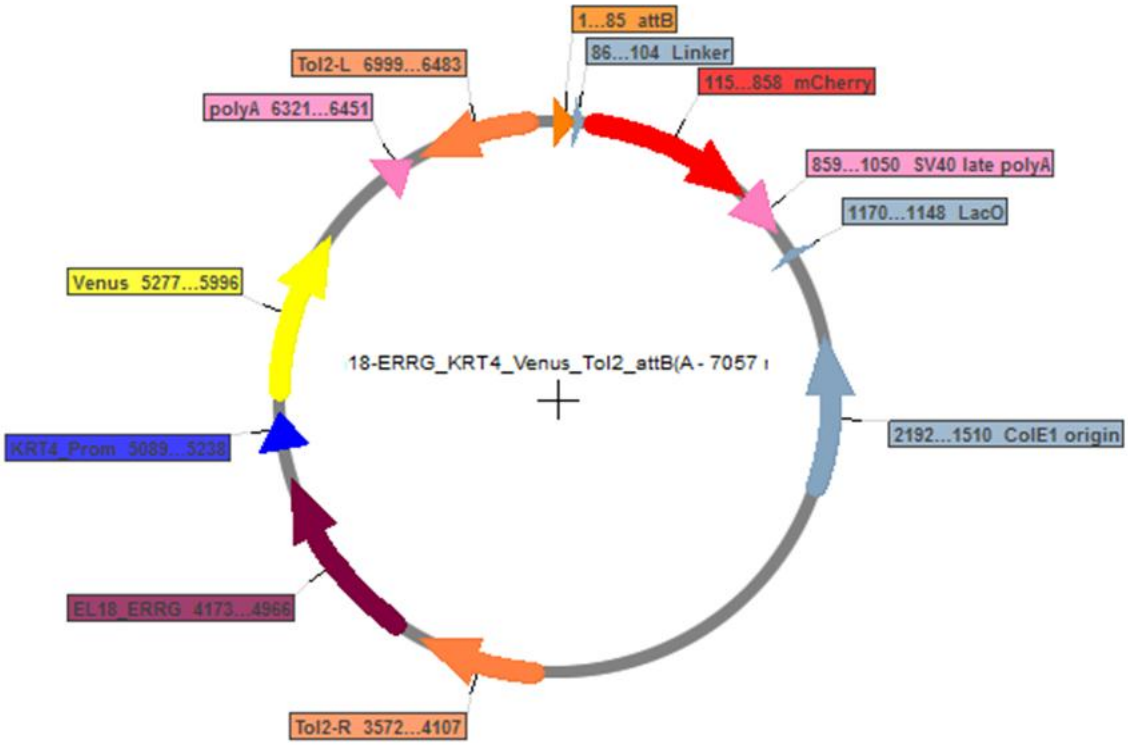

Figure 5-3 Plasmid map showing the EL18-krt4-YFP-attB-mCherry construct

*Schematic of the vector injected into both wild type and recipient line embryos. It contains Tol2 arms for integration into wild type embryos facilitated by Tol2 and the attB-*mCherry *sequence to allow integration into the recipient line by PhiC31 integrase. It also contains the EL18 putative enhancer (purple),* keratin4 *promoter (*blue*) and* venus yfp *reporter gene (yellow).*

A wild type fish is to be crossed with a heterozygous recipient line fish and the embryos will be co-injected with PhiC31 integrase mRNA and the EL18-*krt4*-*yfp*-attB-*mCherry* construct (Figure 5-4). The construct was also co-injected with Tol2 mRNA into wild type embryos to allow comparisons of the two transgenesis methods (Figure 5-5).

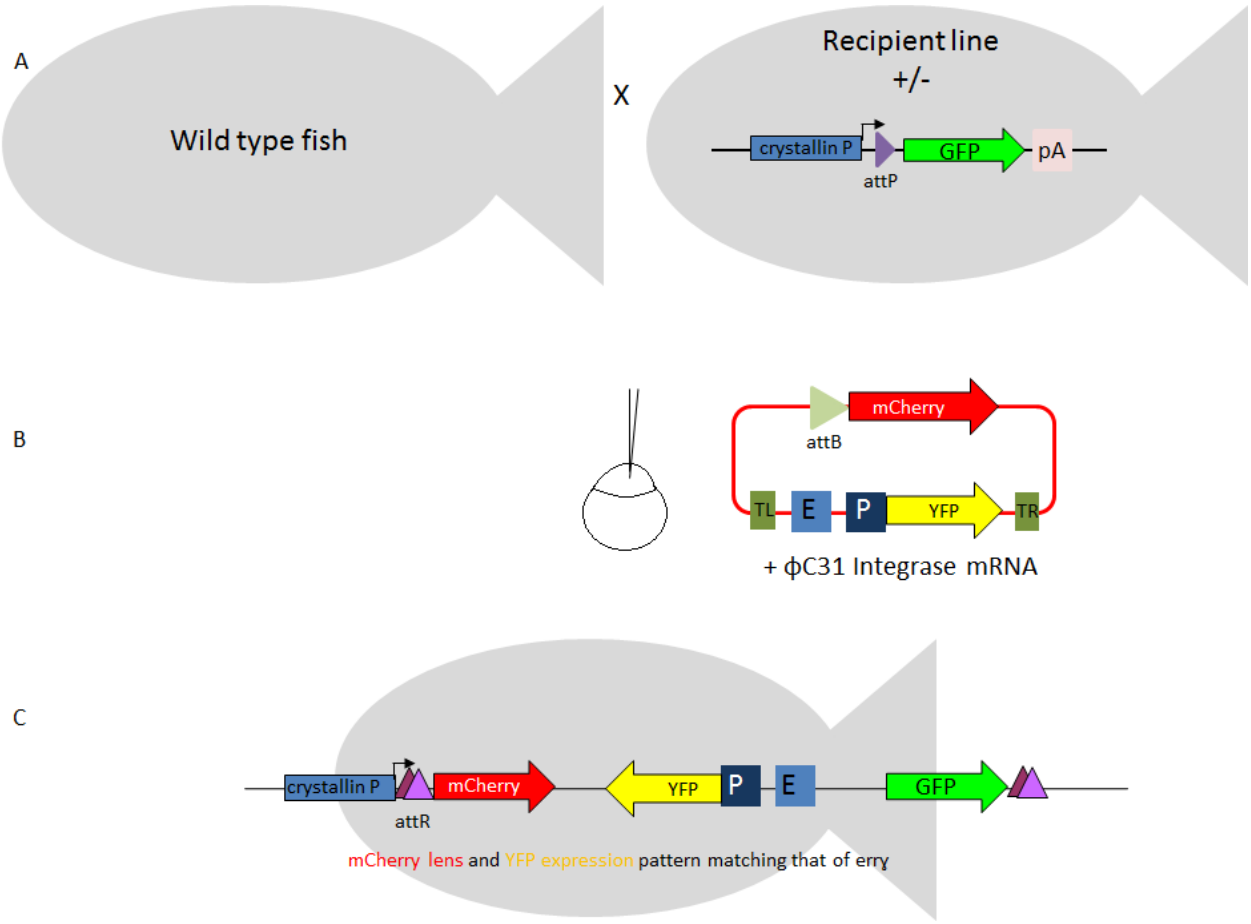

Figure 5-4 Schematic showing targeted integration of an enhancer-promoter-reporter construct

*A heterozygous recipient line fish will be outcrossed to a wild type fish (A). The early embryos are to be co-injected with an enhancer(E)-promoter(P)-reporter construct containing the donor sequence (attB-*mCherry*) and PhiC31 integrase mRNA. This vector will still contain the Tol2 arms (TL and TR) (B). Approximately 50% of these embryos will be transgenic and contain the PhiC31 integrase target site. Of these, a proportion are expected to show mCherry expression in the lens due to targeted integration into the correct locus and YFP expression labelling expression directed by the enhancer (C). The rest will show GFP in the lens and no YFP expression as it has not been integrated.*

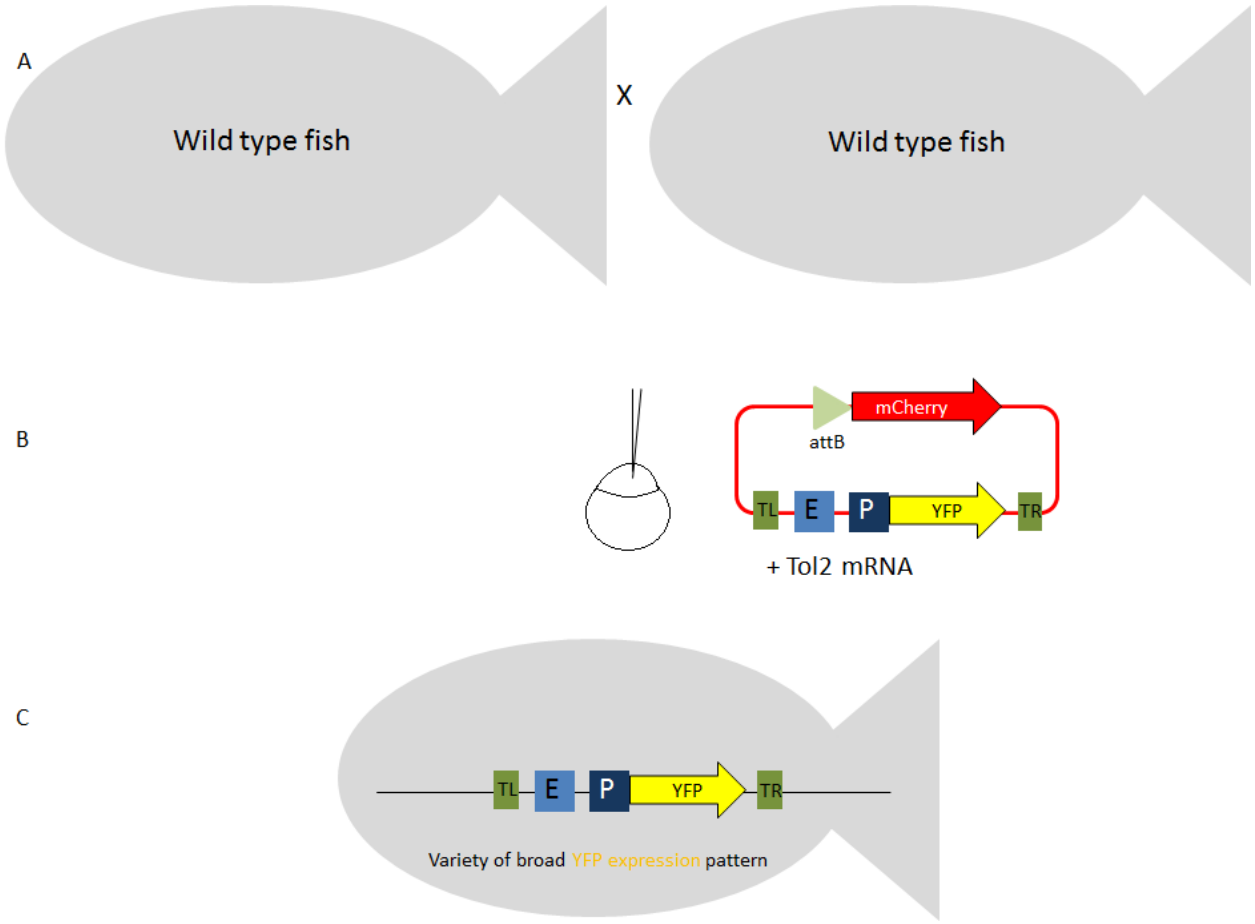

Figure 5-5 Schematic to show screening of enhancer using Tol2 transgenesis

Two wild type fish will be crossed (A) and the early embryos will be co-injected with an enhancer-promoter-reporter construct containing the donor sequence (attB-mCherry) and Tol2 mRNA (B). A proportion of these embryos will be transgenic and have YFP expression labelling expression directed by the enhancer but may also show position effects (C).

5.2.2 YFP expression in the brain region is more restricted and more reproducible using PhiC31 integrase as compared with Tol2 transgenesis

The EL18-*krt4-yfp*-attB-*mCherry* was then co-injected into wild type embryos with Tol2 mRNA. The wild type embryos injected with Tol2 mRNA and the enhancer-containing donor construct were screened at 36-48 hpf when Errγ is known to be expressed in zebrafish. Embryos showed mosaic expression throughout many tissues including

muscle fibres and skin cells as well as various tissues in the head. There was no discernible common expression domain that fit with existing expression data for Errγ.

To test whether the YFP pattern would be more restricted and reproducible when integration is targeted, the EL18-*krt4*-*yfp*-attB-*mCherry* construct was also injected into the recipient line fish along with PhiC31 integrase mRNA. These embryos showed very low levels of YFP expression, often only one or two points of YFP fluorescence were detected but these were consistently in the hindbrain region. When the expression of YFP in certain key tissues was compared between the Tol2 and PhiC31 integrase transgenesis systems, it became clear that YFP expression using PhiC31 integrase was more restricted and more frequently observed in the hindbrain region (Figure 5-6). This data suggests that PhiC31 integrase may also not be suitable for screening enhancers in F0 embryos as we could not reveal the pattern of YFP expression matching that of endogenous Errγ.

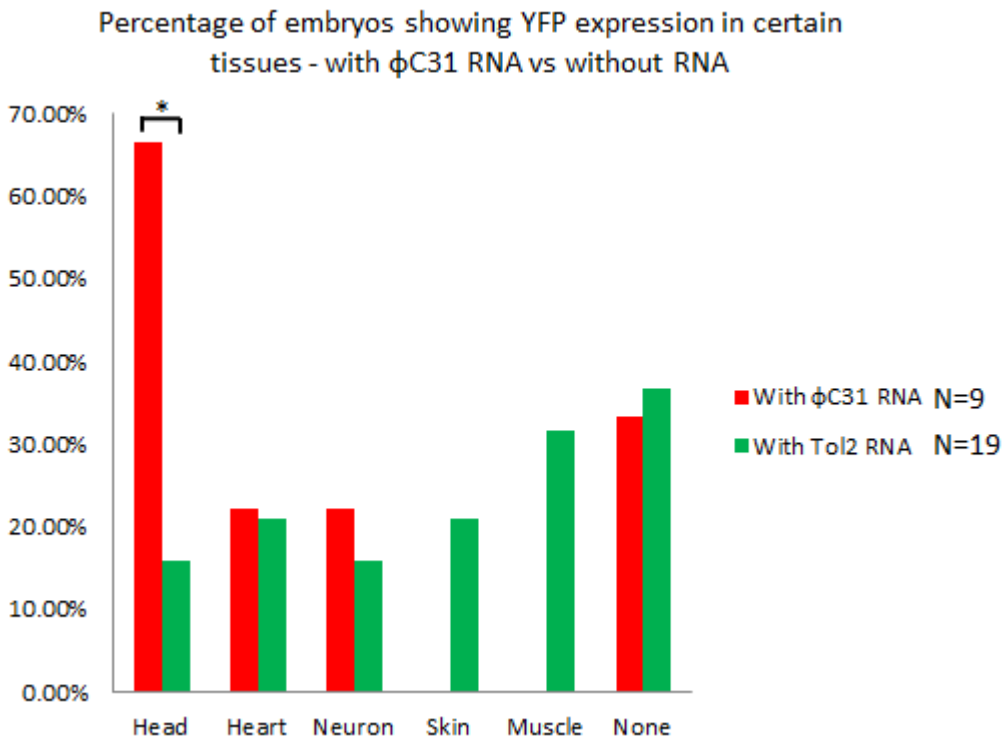

Figure 5-6 Graph showing percentage of embryos expressing transient YFP in a variety of tissues

The bar chart shows the percentage of 48 hpf embryos showing YFP expression in the most common tissues. (p=0.023)*

5.2.3 Targeted integration gives a more restricted and reproducible YFP expression pattern

In order to overcome issues with mosaicism, the injected embryos described in the section above were grown to adulthood in order to make stable transgenic lines. For the PhiC31 integrase founders, only the embryos showing mCherry expression in the lens were grown up. The Tol2 founders were also selected based on YFP expression; only those that showed YFP expression were grown up. Once these fish reached sexual maturity at 4 months, they were outcrossed to wild type fish and their embryos were screened for YFP expression in order to study enhancer activity in the F1 embryos. The

YFP expression pattern was documented at 36 hpf and 60 hpf. Embryos showing YFP expression were separated and both pools of embryos were screened for mCherry expression in the lens. It was found that all embryos showing YFP expression also showed mCherry expression in the lens. Also, all embryos showing GFP expression in the lens did not show any YFP expression. This suggests that the only integration was into the target site and argues against random integrations independent of PhiC31 integrase.

The YFP expression pattern of the recipient line embryos that were co-injected with EL18-*krt4-yfp*-attB-*mCherry* showed similarities (Figure 5-7) to the endogenous Errγ expression pattern obtained by in situ hybridisation (Bardet et al. 2004; Thisse et al. 2004). The YFP expression is more restricted than the pattern seen in *in situ* hybridisations; this is to be expected as an enhancer usually drives expression in a subset of the full expression domain.

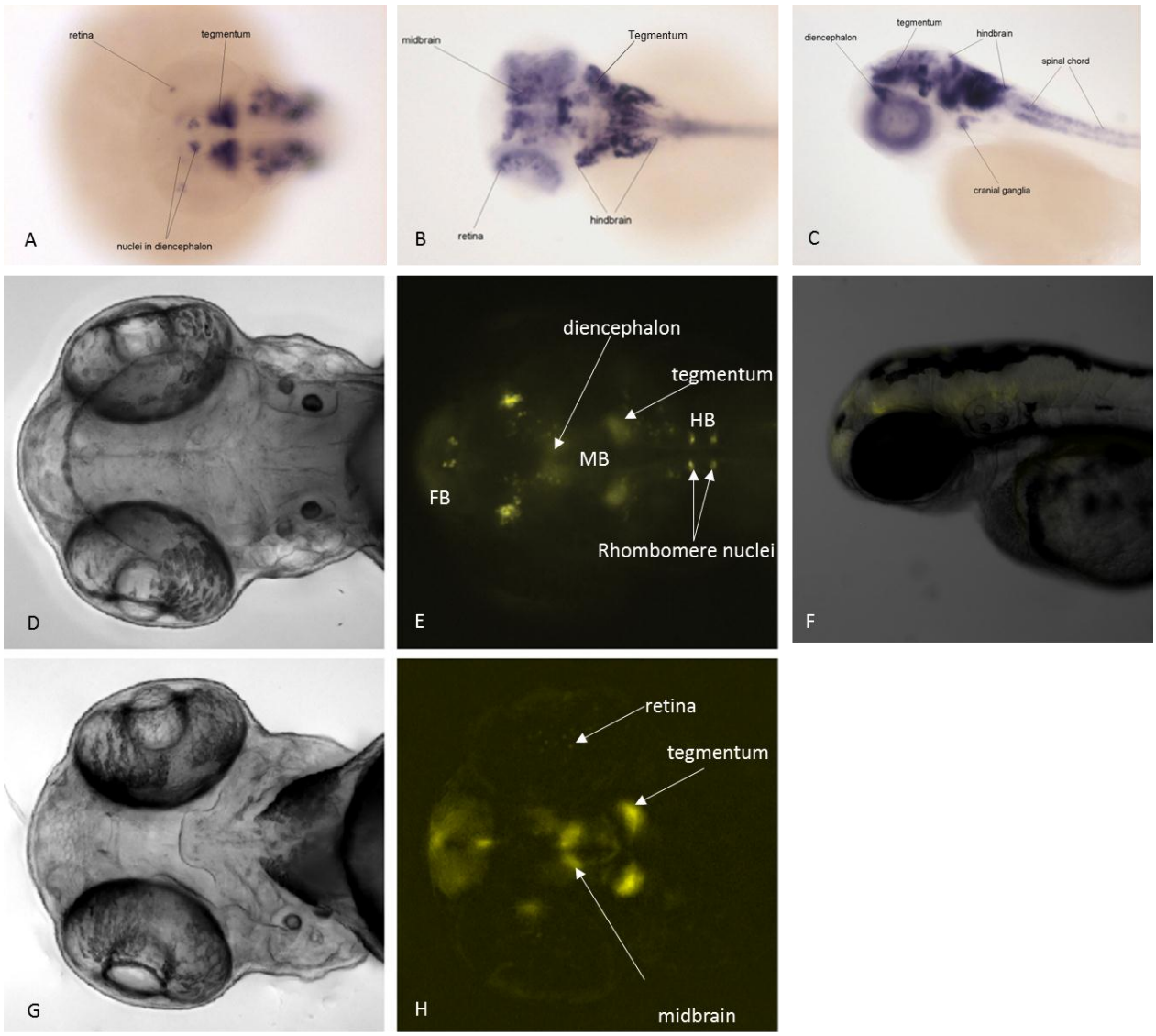

Figure 5-7 EL18-krt4-yfp shows similarities to the endogenous Errγ expression pattern when integrated into the target site

(A) *In situ showing expression pattern of Errγ at prim-15 to prim-25 stage (30-42 hpf). Image was adapted from (Bardet et al. 2004) - a dorsal view with anterior to the left. (B and C) In situ showing expression of errγ at high-pec to long-pec stage (42-60 hpf). Image was adapted from (Thisse et al. 2004) - a dorsal view (B) and lateral view (C) with anterior to the left. Brightfield (D and G) and YFP fluorescent channel (E and H) images of a 60 hpf F1 embryo with targeted integration of the EL18-krt4-YFP construct. (D and E) images are dorsal with anterior to the left (G and H) images are ventral with the anterior to the left. (F) shows a lateral view of a 48 hpf embryos with targeted integration of EL18-krt4-YFP-attB-mCherry. FB= forebrain, MB= midbrain, HB= hindbrain*

5.2.4 Comparison of PhiC31 and Tol2 transgenesis in stable lines

To test whether the F1 generation of lines made using the PhiC31 integrase system showed a more restricted pattern than the F1 of the lines made using Tol2 transgenesis, adult founders were outcrossed and the F1 embryos screened for YFP and mCherry expression. Currently, 2 transgenic F1 fish made using PhiC31 integrase have been found from 13 adult fish outcrossed (Table 5-1). Both of these lines show an almost identical restricted YFP expression pattern specific to small tissues within the brain consistent with data for endogenous Errγ expression (Figure 5-8).

In total, 4 transgenic F1 fish have been identified from 20 adult fish made using the Tol2 transgenesis method (Table 5-1). Three of these lines are displayed in Figure 5-8 and show different YFP expression patterns that are far broader than the PhiC31 integrase embryos shown in panels 4 and 5. This variability in YFP expression pattern in the Tol2 lines means that no conclusions can be drawn about the similarities between expression driven by the putative enhancer and that of endogenous Errγ.

This data suggests that using the PhiC31 integrase system to study enhancers in F1 embryos solves the issue of mosaicism and reduces the background YFP expression resulting from position effects that frequently affect enhancer studies using Tol2 transgenesis.

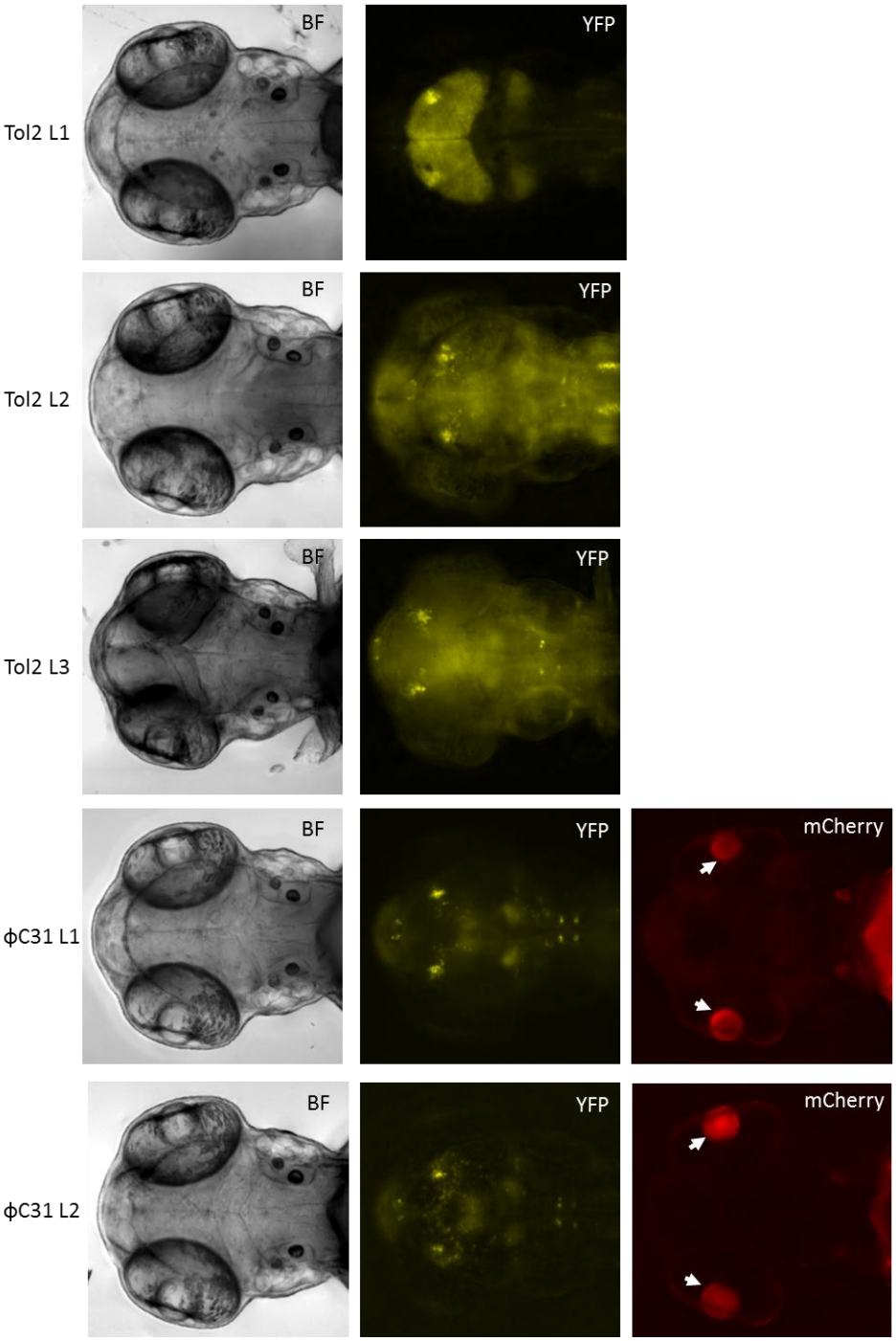

Figure 5-8 Stable lines – Tol2 compared to PhiC31 integrase transgenesis

Images of 60 hpf F1 embryos from a stable line created using Tol2 (Tol2 L1, 2 and 3) or PhiC31 integrase (PhiC31 L1 and 2). Brightfield (BF), YFP channel (YFP) and RFP channel (mCherry) images of dorsal view with anterior to the left.

Line name	Founders/total fish screened	Transgenesis rate	Germline transmission rates
γcry-attP-GFP	6/36	16.67%	5-60%
Tg(zp3:PhiC31;zp3:BFP)	5/49	10.20%	9-33%
Errγ Tol2	4/20	20.00%	14-80%
Errγ PhiC31	2/13	15.38%	8-14%

Table 5-1 Summary table of rates of transgenesis and germline transmission

For each transgenic line made, data on the number of founders identified from the number of fish screened. The percentage of F1 embryos that were transgenic is also included as the rate of germline transmission.

5.3 Discussion

The aim of this work was to establish whether the PhiC31 integrase system could be used to study putative enhancers that show too much variability in position effects when using Tol2 transgenesis techniques.

5.3.1 Suitable putative enhancer to test for position effects

To test the PhiC31 integrase system thoroughly, we needed a construct known to be susceptible to position effects and so EL18-*krt4-yfp* was chosen as the expression pattern in four different stable lines was very variable when using Tol2 (Figure 3-1).

5.3.2 The use of the PhiC31 integrase system for screening F0 embryos

When embryos injected with EL18-*krt4-yfp*-attB-*mCherry* and PhiC31 integrase were analysed for YFP expression, it was found that the signal was much weaker and more restricted than embryos injected with the same construct along with Tol2 mRNA. The expression patterns in PhiC31 integrase injected embryos were more frequently observed in the hindbrain region which is consistent with the endogenous Errγ pattern (Figure 5-6). This was a good indication that PhiC31 integrase reduces position effects, however the F0 embryos did not reveal whether EL18 is an enhancer of the *errγ* gene or not because the embryos were mosaic. The mosaicism meant that not enough YFP expression was detected to make conclusions about the putative enhancer.

5.3.3 Novel *in vivo* data to support the *in silico* evidence that the putative enhancer found is responsible for driving expression of *errγ*

Stable transgenic lines were made using either Tol2 or PhiC31 integrase in order to obtain non-mosaic embryos to see if the enhancer could be studied in the F1 generation. Three PhiC31 integrase lines have been identified so far and these show almost identical expression patterns restricted to specific tissues in the mid- and hindbrain (Figure 5-7) (data for third not shown). These expression domains appear to be consistent with that of Errγ *in situ* data in zebrafish (Bardet et al. 2004; Thisse et al. 2004). Four Tol2 lines have so far been identified and each of them shows YFP expression in the same tissues as the PhiC31 lines but they also show ectopic expression that varies between the different Tol2 lines (Figure 5-7). Not only is there stronger YFP expression in the Tol2 lines, there are differences in the expression domains which cannot be explained by the expression being stronger.

The expression pattern in the two PhiC31 integrase lines is not identical to the expression pattern shown in the *in situ* image of an embryo at the same stage. This is likely due to the tissue-specific nature of enhancers; a gene may have several *cis*-regulatory DNA sequences for the different expression domains. Silencers and other negative *cis*-regulatory DNA sequences are not present in the EL18-*krt4*-*yfp*-attB-*mCherry* construct which may also explain why the YFP expression pattern does not exactly recapitulate the endogenous Errγ pattern. Many spatially restricted genes including developmental regulators like the hox genes (Sharpe et al. 1998) and sonic

hedgehog (Müller et al. 1999; Woolfe et al. 2004) have a number of *cis*-regulatory elements controlling their expression.

In situ hybridisation labels RNA which is much less stable than fluorescent proteins. Using in situs provides a snapshot in time when RNA is expressed. Using reporter genes means that even after expression has stopped, the fluorescent protein is still visible allowing visualisation of expression from earlier developmental stages. This may explain why the YFP expression shows a more similar pattern to the prim-15 to prim-25 stage embryo in Figure 5-7A. It also suggests that the enhancer is active at the prim-15 to prim-25 as the delay in detection of the YFP may be due to the time taken for translation of the reporter gene. As it is difficult to compare expression data from *in situ*s and *in vivo* imaging, therefore *in situ* hybridisation experiments using a YFP probe will be carried out to be directly compared with in situ data using an Errγ probe (see section 6.4).

5.3.4 Indicator for targeted integration

We selected the lens as the marker for the recipient line as it is a small tissue which doesn't interfere with any tissues in the head and trunk that may be labelled by the fluorescent reporter labelling enhancer expression. However, this tissue does not develop until 48hpf and sufficient fluorescent protein has to accumulate in order to visualise expression. This means that embryos cannot be screened until 3 dpf for targeted integration which is after Errγ has begun to be expressed. Although this is not

ideal, the YFP-expressing embryos were separated and the two batches of embryos screened separately for mCherry expression in the lens. This still seems to be an effective way of studying putative enhancers *in vivo*.

5.3.5 Conclusions

Using the PhiC31 integrase system has allowed me to show *in vivo* data of a novel enhancer that has not previously been published. The data in this chapter shows that the El18 enhancer drives expression in tissues that appear to be consistent with the endogenous *erry* pattern. Previous experiments using this enhancer sequence could not tell us the expression pattern due to the broad expression and variability in patterns between different Tol2 lines. This shows the PhiC31 integrase system could be used to study novel enhancers *in vivo*; however, it may not be suitable for all applications of zebrafish transgenesis as the reduced survival of injected embryos and the mosaic F0 mean that it takes about six months in total to study a putative enhancer.

6 DISCUSSION

6.1 Development of a PhiC31 integrase transgenesis system

The use of the zebrafish model to study *cis*-regulatory DNA is widespread due to the great advantages of this model. However, these studies are sometimes hindered by position effects when using popular random integration transgenesis methods. This was shown in Figure 3-1 which shows variable, broad expression patterns in F1 embryos when the founders were injected with the same enhancer construct. This enhancer construct contained a putative enhancer thought to interact with a target gene (*erry*) whose expression is restricted to the brain, spinal cord, retina, pronephric tubes and the tail bud. The variable, broad expression patterns observed using Tol2 transgenesis are thought to be as a result of differences in the flanking genomic DNA surrounding the transgene as it integrates randomly.

In order to reduce variability in position effects compared with random integration, the aim of this project is to target each transgene integration to the same genomic locus meaning that the flanking genomic DNA will be constant. PhiC31 integrase was selected as a targeted integration method as it is already utilised in *DrosoPhila* transgenesis (Bateman et al. 2006).

Using the PhiC31 integrase system designed with the help of a collaborator I was able to show that a donor construct can be integrated into a predetermined genomic locus of a transgenic zebrafish embryo. Our system allows the detection of embryos in which

legitimate targeted integration has occurred by the switching of fluorescent reporter gene expression in the marker tissue (the lens). This rapid *in vivo* detection of targeted integration allows laborious integration site sequence analysis to be avoided.

Our PhiC31 integrase system facilitates recombination between a single attP site and a single attB site resulting in the integration of the whole donor vector. This has two main advantages over RMCE in zebrafish which has now been published (Hu et al. 2011): recombination between single recognition sites may improve integration efficiency as two recombination events occur in RMCE; and integration of the whole vector may be better suited to integrating very large DNA sequences, including BACs.

The targeted integration of BACs would be especially advantageous for the study of *cis*-regulatory DNA; BACs may eventually allow all surrounding DNA including all *cis*-regulatory elements controlling a target gene to be integrated into a consistent genomic context. If this experiment could be carried out enhancer-promoter interactions could be thoroughly studied using techniques such as mutagenesis of *cis*-regulatory elements to observe changes in reporter gene activity and chromatin conformation capture (3C) to directly study proximity of certain long-range *cis*-regulatory DNA to the core promoter of the target gene. Promoters of target genes could be swapped with promoters of bystander genes that are not thought to be driven by the same *cis*-regulatory elements to better understand the mechanisms of enhancer-promoter specificity. Our system provides a good basis on which to develop these kinds of methods in zebrafish.

6.2 Reducing mosaicism in transient transgenic zebrafish embryos and germline transmission rates

Injection of zebrafish embryos with DNA constructs, results in only a subset of cells in the embryo containing the construct. This is a common problem found with injections in zebrafish (Patil et al. 1994), including injections of enhancer-promoter-reporter constructs. This results in mosaic expression of the reporter gene and can lead to the enhancer function not being revealed in F0 embryos. This mosaicism also reduces the efficiency of germline transmission as either none or only a subset of the gametes will contain the transgene. The utilisation of Tol2 transposase to facilitate integration of a transgene into the genome led to a reduction in mosaicism and an increase in the number of founders transmitting the transgene to their offspring (from ~3 % using linearised DNA to 50 % using Tol2 transgenesis (Kawakami, Takeda, et al. 2004). If mosaicism in injected zebrafish embryos could be further reduced so that an even higher proportion of cells contain the transgene, it could improve efficiency of making stable transgenic lines used in applications such a cell-labelling and disease modelling as well as the study of enhancers.

The next aim of this project was to try to reduce mosaicism in injected zebrafish embryos by integrating the transgene as early as possible after fertilisation. In order to do this, PhiC31 integrase protein was provided maternally by making a transgenic line using the zebrafish *zp3* promoter to drive expression of PhiC31 integrase specifically in the oocyte. Transgenesis methods have already been developed using maternal

transgenic lines to express integrase proteins in zebrafish (Liu et al. 2008); however, they did not test for the presence of Cre before 2 hpf. At the 2 hpf stage, a zebrafish embryo already has 64 cells so if the protein has only just been produced, the embryo will still show mosaic reporter activity. They do not comment on mosaicism of Cre-mediated excision, however previous data from the same lab showed GFP expression in the adult ovaries and the one cell stage embryo (Liu et al. 2006). Due to the obvious presence of functional GFP protein at the one cell stage, the *zp3* promoter looked ideal to provide functional PhiC31 integrase protein at the time when the donor construct is injected. This was expected to result in the earlier integration (within the first few cell divisions) of the transgene and therefore show a reduction in mosaicism. Although targeted integration into the zebrafish genome was shown to be facilitated when using a maternal PhiC31 female founder, there was no evidence to suggest that providing functional PhiC31 integrase protein at the time of injection of the donor vector resulted in earlier integration to reduce mosaicism.

Other researchers have begun to use PhiC31 integrase mRNA with a UTR from the *nanos1* gene which targets the mRNA to the germline precursor cells (Hu et al. 2011) in an attempt to boost germline transmission of the transgene. This has showed some promising results increasing the percentage of founders which transmit the transgene to their offspring from an average of 4.92 % using PhiC31 integrase mRNA (from 4 experiments) to 50.06 % using PhiC31 integrase mRNA with *nanos1* UTR (from 2 experiments). From the experiment I described in Chapter 5.2.4 I show a transgenesis

rate of 15.38 % (Table 5-1) using a higher concentration of PhiC31 integrase mRNA in injections. Together, these data suggest that the addition of a *nanos1* UTR region to the PhiC31 integrase mRNA is the best way to increase transgenesis rates; however this does not improve mosaicism in somatic cells of the injected embryo and therefore does not allow the study of enhancers in the F0 embryos.

6.3 Targeted integration of an enhancer-promoter-reporter construct reduces ectopic reporter expression thought to be due to position effects

As shown in Figure 3-1, the study of the EL18-*krt4-yfp* construct using Tol2 transgenesis resulted in variable, broad YFP expression patterns that do not correlate with known Errγ expression data (Bardet et al. 2004; Thisse et al. 2004). This is thought to be due to the influences of the flanking genomic DNA which, as a result of random integration, is different for each injected embryo. To investigate whether targeting the EL18-*krt4-yfp*-attB-*mCherry* construct to the same genomic locus results in a more restricted and reproducible YFP expression pattern, it was co-injected with PhiC31 integrase mRNA into recipient line embryos. This resulted in a more restricted pattern but expression was too low (due to mosaicism) to suggest whether EL18 was an enhancer of *errγ* (Figure 5-6). When non-mosaic F1 embryos from three founders were screened for YFP expression, they were both more reproducible and more restricted compared with F1 from lines made using Tol2 (Figure 5-8). They also shared several similar expression domains with endogenous Errγ (Figure 5-7). These data provide evidence that EL18

could be an enhancer that interacts with the *errγ* gene in zebrafish. In order to confirm this *in situ* hybridisation experiments will be carried out (see section 6.4). The *cis*-regulatory elements have not previously been studied and so this data provides evidence that a predicted *errγ* enhancer functions *in vivo*.

As well as showing *in vivo* function of a predicted enhancer, these experiments also provide evidence that the PhiC31 integrase system can be utilised in experiments requiring the study of small changes in gene expression. For example, this system will now be used by another PhD student to study changes in reporter expression due to point mutations in human enhancers that are thought to be involved in type II diabetes.

6.4 Future work

Due to the time constraints of this PhD I was unable to carry out some important experiments that would have enabled stronger conclusions to be made from this work. These experiments are currently being carried out by another PhD student as part of the revisions to the publication submitted to Development (see section 9). Southern blots are being carried out on the two recipient lines to establish that only a single copy of the integration site is present. A further two enhancers will be tested using the PhiC31 system developed here to ascertain whether position effects will be reduced when studying other putative enhancers. *In situ* hybridisation experiments are also being carried out using a YFP probe and an endogenous errγ probe to confirm that the YFP expression domains overlap with the endogenous errγ pattern.

In order to optimise the system further, more recipient lines could be tested to ensure maximum efficiency of PhiC31 mediated recombination and to establish whether the YFP pattern would look consistent in five or more different genomic loci.

7 APPENDICES

7.1 Appendix 1

Primer sequences

PCR product	Forward primer (FP) and reverse primer (RP)
PhiC31 coding sequence	FP - ATGGATACCTACGCCGGAG
	RP - TCACACTTTCCGCTTTTTCTTAG
attB-linker	FP - GTCGACGATGTAGGTCACG
	RP - CAGCCCGACCGCTGCGCCTTGTCGACATGCCCGCCGT
Linker-*mCherry*-pA	FP - AAGGCGCAGCGGTCGGGCTGGCCGCCACCATGGTGAGCAAGGGCGA
	RP - TGC AAA GTC TGT TCA AGC ATG TG
attB-linker-*mCherry*-pA	FP - GTCGACGATGTAGGTCACG
	RP - TGC AAA GTC TGT TCA AGC ATG TG
Integration junction	FP - GGTTCATCGCATGCAGACAG
	RP - CTCGCCCTTGCTCACCAT
γ-crystallin-attP-*gfp*	FP - GGTTCATCGCATGCAGACAG
	RP - CTGCTAGTTGAACGCTTCCAT
attB(83)-linker	FP - TGACGGTCTCGAAGCCGC
	RP - CAGCCCGACCGCTGCGCCTTGGACCAGATGGGTGAGGTG
attB(83)-linker-*mCherry*-pA	FP - TGACGGTCTCGAAGCCGC
	RP - TGC AAA GTC TGT TCA AGC ATG TG
ebfp2 coding sequence	FP - AAATAGGATCCATGGTGAGCAAGGGCGAGG
	RP - AAATTCTCGAGTTACTTGTACAGCTCGTCCATGC
zp3-PhiC31-pA	FP - AAAAAGGGCCCGCCCTTAAAATCCCCATGACATG
	RP - AAAAAGCATGCGAATTAAAAAACCTCCCACACCTC
zp3-ebfp2-pA	FP - AAAAAGCATGCGCCCTTAAAATCCCCATGACATG
	RP - AAAAAAGATCTGAATTAAAAAACCTCCCACACCTC

7.2 Appendix 2

PhiC31 integrase recognition site sequences

attP 84 bp	AGAAGCGGTTTTCGGGAGTAGTGCCCCAACTGGGGTAACCTttgAGTTCTCTCAGTTGGGGGCGTAGGGTCGCCGACATGACAC
attB (original) 285 bp	GTCGACGATGTAGGTCACGGTCTCGAAGCCGCGGTGCGGGTGCCAGGGCGTGCCCttgGGCTCCCCGGGCGCGTACTCCACCTCACCCATCTGGTCCATCATGATGAACGGGTCGAGGTGGCGGTAGTTGATCCCGGCGAACGCGCGGCGCACCGGGAAGCCCTCGCCCTCGAAACCGCTGGGCGCGGTGGTCACGGTGAGCACGGGACGTGCGACGGCGTCGGCGGGTGCGGATACGCGGGGCAGCGTCAGCGGGTTCTCGACGGTCACGGCGGGCATGTCGAC
attB (shorter) 83 bp	TGACGGTCTCGAAGCCGCGGTGCGGGTGCCAGGGCGTGCCCttgGGCTCCCCGGGCGCGTACTCCACCTCACCCATCTGGTCC

8 LIST OF REFERENCES

Abhyankar, M.M., Urekar, C. & Reddi, P.P., 2007. A novel CpG-free vertebrate insulator silences the testis-specific SP-10 gene in somatic tissues: role for TDP-43 in insulator function. *The Journal of Biological Chemistry*, 282(50), pp.36143–36154.

Agulnick, A.D. et al., 1996. Interactions of the LIM-domain-binding factor Ldb1 with LIM homeodomain proteins. *Nature*, 384(6606), pp.270–272.

Akbari, O.S. et al., 2008. A novel promoter-tethering element regulates enhancer-driven gene expression at the bithorax complex in the DrosoPhila embryo. *Development (Cambridge, England)*, 135(1), pp.123–131.

Alberts, B. et al., 2008. *Molecular Biology of the Cell* 5Rev Ed., Garland Publishing Inc,US.

Allende, M.L. et al., 2006. Cracking the genome's second code: enhancer detection by combined phylogenetic footprinting and transgenic fish and frog embryos. *Methods (San Diego, Calif.)*, 39(3), pp.212–219.

Altschul, S.F. et al., 1990. Basic local alignment search tool. *Journal of molecular biology*, 215(3), pp.403–410.

Amano, T. et al., 2009. Chromosomal dynamics at the Shh locus: limb bud-specific differential regulation of competence and active transcription. *Developmental Cell*, 16(1), pp.47–57.

Antoch, M.P. et al., 1997. Functional identification of the mouse circadian Clock gene by transgenic BAC rescue. *Cell*, 89(4), pp.655–667.

Ariazi, E.A. & Jordan, V.C., 2006. Estrogen-related receptors as emerging targets in cancer and metabolic disorders. *Current topics in medicinal chemistry*, 6(3), pp.203–215.

Bach, I. et al., 1997. A family of LIM domain-associated cofactors confer transcriptional synergism between LIM and Otx homeodomain proteins. *Genes & Development*, 11(11), pp.1370–1380.

Bardet, P.-L. et al., 2004. Cloning and developmental expression of five estrogen-receptor related genes in the zebrafish. *Development genes and evolution*, 214(5), pp.240–249.

Bateman, J.R., Lee, A.M. & Wu, C. -tin., 2006. Site-Specific Transformation of DrosoPhila via {Phi}C31 Integrase-Mediated Cassette Exchange. *Genetics*, 173(2), pp.769–777.

Becker, Kleinsmith, L.J. & Hardin, J., 2002. *The World of the Cell, 5th Edition* 5th ed., Benjamin Cummings.

Bedell, V.M. et al., 2012. In vivo genome editing using a high-efficiency TALEN system. *Nature*, 491(7422), pp.114–118.

Bell, A.C., West, A.G. & Felsenfeld, G., 1999. The protein CTCF is required for the enhancer blocking activity of vertebrate insulators. *Cell*, 98(3), pp.387–396.

Belteki, G. et al., 2003. Site-specific cassette exchange and germline transmission with mouse ES cells expressing [phgr]C31 integrase. *Nat Biotech*, 21(3), pp.321–324.

Berman, B.P. et al., 2002. Exploiting transcription factor binding site clustering to identify cis-regulatory modules involved in pattern formation in the DrosoPhila genome. *Proceedings of the National Academy of Sciences of the United States of America*, 99(2), pp.757–762.

Berman, J., Hsu, K. & Look, A.T., 2003. Zebrafish as a model organism for blood diseases. *British Journal of Haematology*, 123(4), pp.568–576.

Beuling, E. et al., 2012. GATA6 is required for proliferation, migration, secretory cell maturation, and gene expression in the mature mouse colon. *Molecular and cellular biology*, 32(17), pp.3392–3402.

Bischof, J. et al., 2007. An optimized transgenesis system for DrosoPhila using germ-line-specific C31 integrases. *Proceedings of the National Academy of Sciences*, 104(9), pp.3312–3317.

Bonn, S. & Furlong, E.E.M., 2008. cis-Regulatory networks during development: a view of DrosoPhila. *Current Opinion in Genetics & Development*, 18(6), pp.513–520.

Bulger, M. & Groudine, M., 1999. Looping versus linking: toward a model for long-distance gene activation. *Genes & Development*, 13(19), pp.2465–2477.

Butler, J.E. & Kadonaga, J.T., 2001. Enhancer-promoter specificity mediated by DPE or TATA core promoter motifs. *Genes & Development*, 15(19), pp.2515–2519.

Butler & Kadonaga, 2002. The RNA polymerase II core promoter: a key component in the regulation of gene expression. *Genes & Development*, 16(20), pp.2583–2592.

Cai & Levine, M., 1995. Modulation of enhancer-promoter interactions by insulators in the DrosoPhila embryo. *Nature*, 376(6540), pp.533–536.

Cai & Shen, P., 2001. Effects of cis arrangement of chromatin insulators on enhancer-blocking activity. *Science (New York, N.Y.)*, 291(5503), pp.493–495.

Calhoun & Levine, 2003. Long-range enhancer-promoter interactions in the Scr-Antp interval of the DrosoPhila Antennapedia complex. *Proceedings of the National Academy of Sciences of the United States of America*, 100(17), pp.9878–9883.

Calhoun, V.C., Stathopoulos, A. & Levine, M., 2002. Promoter-proximal tethering elements regulate enhancer-promoter specificity in the DrosoPhila Antennapedia complex. *Proceedings of the National Academy of Sciences of the United States of America*, 99(14), pp.9243–9247.

Chen et al., 2005. Multiple Promoter Targeting Sequences exist in Abdominal-B to regulate long-range gene activation. *Developmental Biology*, 286(2), pp.629–636.

Chen & Manley, J.L., 2003. Core Promoter Elements and TAFs Contribute to the Diversity of Transcriptional Activation in Vertebrates. *Molecular and Cellular Biology*, 23(20), pp.7350–7362.

Chrast, R., Scott, H.S. & Antonarakis, S.E., 1999. Linearization and purification of BAC DNA for the development of transgenic mice. *Transgenic research*, 8(2), pp.147–150.

Crawford, G.E. et al., 2006. DNase-chip: a high-resolution method to identify DNase I hypersensitive sites using tiled microarrays. *Nature methods*, 3(7), pp.503–509.

Dahlem, T.J. et al., 2012. Simple methods for generating and detecting locus-specific mutations induced with TALENs in the zebrafish genome. *PLoS genetics*, 8(8), p.e1002861.

Davidson, A.E. et al., 2003. Efficient gene delivery and gene expression in zebrafish using the Sleeping Beauty transposon. *Developmental Biology*, 263(2), pp.191–202.

DiBiase, A. et al., 2006. Piloting the zebrafish genome browser. *Developmental dynamics: an official publication of the American Association of Anatomists*, 235(3), pp.747–753.

Duret, L. & Bucher, P., 1997. Searching for regulatory elements in human noncoding sequences. *Current Opinion in Structural Biology*, 7(3), pp.399–406.

Ellingsen, S. et al., 2005. Large-scale enhancer detection in the zebrafish genome. *Development*, 132(17), pp.3799–3811.

Fantes, J. et al., 1995. Aniridia-associated cytogenetic rearrangements suggest that a position effect may cause the mutant phenotype. *Human Molecular Genetics*, 4(3), pp.415–422.

Fisher, S. et al., 2006. Evaluating the biological relevance of putative enhancers using Tol2 transposon-mediated transgenesis in zebrafish. *Nature Protocols*, 1(3), pp.1297–1305.

Flicek, P. et al., 2012. Ensembl 2012. *Nucleic acids research*, 40(Database issue), pp.D84–90.

Foster, J.W. et al., 1994. Campomelic dysplasia and autosomal sex reversal caused by mutations in an SRY-related gene. *Nature*, 372(6506), pp.525–530.

Gama Sosa, M.A., De Gasperi, R. & Elder, G.A., 2012. Modeling human neurodegenerative diseases in transgenic systems. *Human genetics*, 131(4), pp.535–563.

Gehrig, J. et al., 2009. Automated high-throughput mapping of promoter-enhancer interactions in zebrafish embryos. *Nat Meth*, 6(12), pp.911–916.

Gilmour, D.T., Maischein, H.-M. & Nüsslein-Volhard, C., 2002. Migration and function of a glial subtype in the vertebrate peripheral nervous system. *Neuron*, 34(4), pp.577–588.

Giresi, P.G. et al., 2007. FAIRE (Formaldehyde-Assisted Isolation of Regulatory Elements) isolates active regulatory elements from human chromatin. *Genome research*, 17(6), pp.877–885.

Glass, A.S. & Dahm, R., 2004. The Zebrafish as a Model Organism for Eye Development. *Ophthalmic Research*, 36(1), pp.4–24.

Goode, D.K., Snell, P. & Elgar, G., 2003. Comparative analysis of vertebrate Shh genes identifies novel conserved non-coding sequence. *Mammalian Genome: Official Journal of the International Mammalian Genome Society*, 14(3), pp.192–201.

Gopaul, D.N., Guo, F. & Van Duyne, G.D., 1998. Structure of the Holliday junction intermediate in Cre-loxP site-specific recombination. *The EMBO Journal*, 17(14), pp.4175–4187.

Groth et al., 2000. A phage integrase directs efficient site-specific integration in human cells. *Proceedings of the National Academy of Sciences of the United States of America*, 97(11), pp.5995–6000.

Groth, A.C. et al., 2004. Construction of transgenic DrosoPhila by using the site-specific integrase from phage PhiC31. *Genetics*, 166(4), pp.1775–1782.

Gurnett, C.A. et al., 2007. Two novel point mutations in the long-range SHH enhancer in three families with triphalangeal thumb and preaxial polydactyly. *American Journal of Medical Genetics. Part A*, 143(1), pp.27–32.

Hardison, R.C., Oeltjen, J. & Miller, W., 1997. Long Human–Mouse Sequence Alignments Reveal Novel Regulatory Elements: A Reason to Sequence the Mouse Genome. *Genome Research*, 7(10), pp.959–966.

Heintz, 2001. BAC to the future: the use of bac transgenic mice for neuroscience research. *Nature reviews. Neuroscience*, 2(12), pp.861–870.

Herendeen, D.R., Kassavetis, G.A. & Geiduschek, E.P., 1992. A transcriptional enhancer whose function imposes a requirement that proteins track along DNA. *Science (New York, N.Y.)*, 256(5061), pp.1298–1303.

Hoess, R.H. & Abremski, K., 1985. Mechanism of strand cleavage and exchange in the Cre-lox site-specific recombination system. *Journal of molecular biology*, 181(3), pp.351–362.

Hogan, Costanini & Lacy, 1982. Manipulating the mouse embryo. A laboratory manual. *Cold Spring Harbor Laboratory*.

Hu, G., Goll, M.G. & Fisher, S., 2011. ΦC31 Integrase Mediates Efficient Cassette Exchange in the Zebrafish Germline. *Developmental Dynamics*, 240(9), pp.2101–2107.

Iklé, J.M., Artinger, K.B. & Clouthier, D.E., 2012. Identification and characterization of the zebrafish pharyngeal arch-specific enhancer for the basic helix-loop-helix transcription factor Hand2. *Developmental biology*, 368(1), pp.118–126.

Ivics, Z. et al., 1997. Molecular reconstruction of Sleeping Beauty, a Tc1-like transposon from fish, and its transposition in human cells. *Cell*, 91(4), pp.501–510.

Jeong, Y. et al., 2006. A functional screen for sonic hedgehog regulatory elements across a 1 Mb interval identifies long-range ventral forebrain enhancers. *Development*, 133(4), pp.761–772.

Jessen, J.R., Willett, C.E. & Lin, S., 1999. Artificial chromosome transgenesis reveals long-distance negative regulation of rag1 in zebrafish. *Nature Genetics*, 23(1), pp.15–16.

Jiang, N. et al., 2009. Genome-wide mapping of boundary element-associated factor (BEAF) binding sites in DrosoPhila melanogaster links BEAF to transcription. *Molecular and Cellular Biology*, 29(13), pp.3556–3568.

Jin, F. et al., 2011. Enhancers. *Transcription*, 2(5), pp.226–230.

Kawakami, Takeda, H., et al., 2004. A transposon-mediated gene trap approach identifies developmentally regulated genes in zebrafish. *Developmental Cell*, 7(1), pp.133–144.

Kawakami, Imanaka, K., et al., 2004. Excision of the Tol2 transposable element of the medaka fish Oryzias latipes in Xenopus laevis and Xenopus tropicalis. *Gene*, 338(1), pp.93–98.

Kawakami, K. et al., 1998. Excision of the tol2 transposable element of the medaka fish, Oryzias latipes, in zebrafish, Danio rerio. *Gene*, 225(1-2), pp.17–22.

Kawakami, K., 2007. Tol2: a versatile gene transfer vector in vertebrates. *Genome biology*, 8 Suppl 1, p.S7.

Kawakami & Noda, T., 2004. Transposition of the Tol2 element, an Ac-like element from the Japanese medaka fish Oryzias latipes, in mouse embryonic stem cells. *Genetics*, 166(2), pp.895–899.

Kazuki, Y. et al., 2003. A new mouse model for Down syndrome. *Journal of neural transmission. Supplementum*, (67), pp.1–20.

Kenyon, E.J. et al., 2011. Functional analysis of conserved non-coding regions around the short stature hox gene (shox) in whole zebrafish embryos. *PloS one*, 6(6), p.e21498.

Kimmel et al., 1995. Stages of embryonic development of the zebrafish. *Developmental dynamics: an official publication of the American Association of Anatomists*, 203(3), pp.253–310.

Koga, A. et al., 1996. Transposable element in fish. *Nature*, 383(6595), p.30.

Kuhn, E.J. et al., 2003. A test of insulator interactions in DrosoPhila. *The EMBO Journal*, 22(10), pp.2463–2471.

Kuhn, E.J. & Geyer, P.K., 2003. Genomic insulators: connecting properties to mechanism. *Current Opinion in Cell Biology*, 15(3), pp.259–265.

Kuhstoss, S. & Rao, R.N., 1991. Analysis of the integration function of the streptomycete bacteriophage Phi C31. *Journal of Molecular Biology*, 222(4), pp.897–908.

Langenau, D.M. et al., 2005. Cre/lox-regulated transgenic zebrafish model with conditional myc-induced T cell acute lymphoblastic leukemia. *Proceedings of the National Academy of Sciences of the United States of America*, 102(17), pp.6068–6073.

Langheinrich, U. et al., 2002. Zebrafish as a Model Organism for the Identification and Characterization of Drugs and Genes Affecting p53 Signaling. *Current Biology*, 12(23), pp.2023–2028.

Langheinrich, U., 2003. Zebrafish: A new model on the pharmaceutical catwalk. *BioEssays*, 25(9), pp.904–912.

Lettice, L.A. et al., 2003. A long-range Shh enhancer regulates expression in the developing limb and fin and is associated with preaxial polydactyly. *Human Molecular Genetics*, 12(14), pp.1725–1735.

Lettice, L.A. et al., 2002. Disruption of a long-range cis-acting regulator for Shh causes preaxial polydactyly. *Proceedings of the National Academy of Sciences of the United States of America*, 99(11), pp.7548–7553.

Li, X. & Noll, M., 1994. Compatibility between enhancers and promoters determines the transcriptional specificity of gooseberry and gooseberry neuro in the DrosoPhila embryo. *The EMBO Journal*, 13(2), pp.400–406.

Lister, J.A., 2011. Chapter 11 - Use of Phage PhiC31 Integrase as a Tool for Zebrafish Genome Manipulation. In M. W. and L. I. Z. H. William Detrich, ed. *Methods in Cell Biology*. Academic Press, pp. 195–208. Available at:

http://www.sciencedirect.com/science/article/pii/B9780123748140000112 [Accessed December 23, 2012].

Lister, J.A., 2010. Transgene excision in zebrafish using the PhiC31 integrase. *genesis*, 48(2), pp.137–143.

Liu et al., 2008. Generation of oocyte-specifically expressed <I>cre</I> transgenic zebrafish for female germline excision of loxp-flanked transgene. *Developmental Dynamics*, 237(10), pp.2955–2962.

Liu, J. et al., 2009. PhiC31 integrase induces a DNA damage response and chromosomal rearrangements in human adult fibroblasts. *BMC Biotechnology*, 9(1), p.31.

Liu, Wang, H. & Gong, Z., 2006. Tandem-repeated Zebrafish zp3 genes possess oocyte-specific promoters and are insensitive to estrogen induction. *Biology of Reproduction*, 74(6), pp.1016–1025.

Lo, K.A. et al., 2011. Genome-wide profiling of H3K56 acetylation and transcription factor binding sites in human adipocytes. *PloS one*, 6(6), p.e19778.

Luo, G. et al., 1998. Chromosomal transposition of a Tc1/mariner-like element in mouse embryonic stem cells. *Proceedings of the National Academy of Sciences of the United States of America*, 95(18), pp.10769–10773.

Maas, S.A., Suzuki, T. & Fallon, J.F., 2011. Identification of spontaneous mutations within the long-range limb-specific Sonic hedgehog enhancer (ZRS) that alter Sonic hedgehog expression in the chicken limb mutants oligozeugodactyly and silkie breed. *Developmental Dynamics: An Official Publication of the American Association of Anatomists*, 240(5), pp.1212–1222.

Maksimenko, O., Golovnin, A. & Georgiev, P., 2008. Enhancer-promoter communication is regulated by insulator pairing in a DrosoPhila model bigenic locus. *Molecular and Cellular Biology*, 28(17), pp.5469–5477.

Maston, G.A., Evans, S.K. & Green, M.R., 2006. Transcriptional regulatory elements in the human genome. *Annual review of genomics and human genetics*, 7, pp.29–59.

Merika, M. et al., 1998. Recruitment of CBP/p300 by the IFN beta enhanceosome is required for synergistic activation of transcription. *Molecular Cell*, 1(2), pp.277–287.

Merli, C. et al., 1996. Promoter specificity mediates the independent regulation of neighboring genes. *Genes & Development*, 10(10), pp.1260–1270.

Michel, D. et al., 1993. H-DNA can act as a transcriptional insulator. *Cellular & Molecular Biology Research*, 39(2), pp.131–140.

Min, C. et al., 2012. Molecular cloning and partial functional characterization of a proliferation inducing ligand (APRIL) in zebrafish (Danio rerio). *Developmental and comparative immunology*, 37(1), pp.202–206.

Morcillo, P. et al., 1997. Chip, a widely expressed chromosomal protein required for segmentation and activity of a remote wing margin enhancer in DrosoPhila. *Genes & Development*, 11(20), pp.2729–2740.

Morcillo, P., Rosen, C. & Dorsett, D., 1996. Genes regulating the remote wing margin enhancer in the DrosoPhila cut locus. *Genetics*, 144(3), pp.1143–1154.

Muller, F. et al., 1999. Intronic enhancers control expression of zebrafish sonic hedgehog in floor plate and notochord. *Development*, 126(10), pp.2103–2116.

Müller, F. et al., 1999. Intronic enhancers control expression of zebrafish sonic hedgehog in floor plate and notochord. *Development (Cambridge, England)*, 126(10), pp.2103–2116.

Nasevicius, A. & Ekker, S.C., 2000. Effective targeted gene “knockdown” in zebrafish. *Nature Genetics*, 26(2), pp.216–220.

Norton, W. & Bally-Cuif, L., 2010. Adult zebrafish as a model organism for behavioural genetics. *BMC Neuroscience*, 11(1), p.90.

Ohtsuki, S., Levine, M. & Cai, H.N., 1998. Different core promoters possess distinct regulatory activities in the DrosoPhila embryo. *Genes & Development*, 12(4), pp.547–556.

Ossipow, V. et al., 1995. A mammalian RNA polymerase II holoenzyme containing all components required for promoter-specific transcription initiation. *Cell*, 83(1), pp.137–146.

Pan, Y. & Nussinov, R., 2011. The Role of Response Elements Organization in Transcription Factor Selectivity: The IFN-β Enhanceosome Example. *PLoS Computational Biology*, 7(6). Available at: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3116919/ [Accessed December 4, 2012].

Panne, D., 2008. The enhanceosome. *Current Opinion in Structural Biology*, 18(2), pp.236–242.

Patil, J.G., Wong, V. & Khoo, H.W., 1994. Assessment of pMTL construct for detection in vivo of luciferase expression and fate of the transgene in the zebrafish, Brachydanio rerio. *Zoological science*, 11(1), pp.63–68.

Petrascheck, M. et al., 2005. DNA looping induced by a transcriptional enhancer in vivo. *Nucleic Acids Research*, 33(12), pp.3743–3750.

Pfeifer, D. et al., 1999. Campomelic dysplasia translocation breakpoints are scattered over 1 Mb proximal to SOX9: evidence for an extended control region. *American Journal of Human Genetics*, 65(1), pp.111–124.

Rada-Iglesias, A. et al., 2011. A unique chromatin signature uncovers early developmental enhancers in humans. *Nature*, 470(7333), pp.279–283.

Raymond, C.S. & Soriano, P., 2007. High-efficiency FLP and PhiC31 site-specific recombination in mammalian cells. *PloS One*, 2(1), p.e162.

Rubin & Spradling, A.C., 1982. Genetic transformation of DrosoPhila with transposable element vectors. *Science (New York, N.Y.)*, 218(4570), pp.348–353.

Sagai, T. et al., 2009. A cluster of three long-range enhancers directs regional Shh expression in the epithelial linings. *Development (Cambridge, England)*, 136(10), pp.1665–1674.

Sandelin, A. et al., 2004. Arrays of ultraconserved non-coding regions span the loci of key developmental genes in vertebrate genomes. *BMC Genomics*, 5(1), p.99.

Sangiorgi, E., Shuhua, Z. & Capecchi, M.R., 2008. In vivo evaluation of PhiC31 recombinase activity using a self-excision cassette. *Nucleic acids research*, 36(20), p.e134.

Schmid, 2008. Introduction to sequence motif discovery and search. Available at: http://www.ch.embnet.org/CoursEMBnet/PagesTR08/slides/CS_mot_search.pdf.

Shakes, L.A. et al., 2008. Context dependent function of APPb enhancer identified using enhancer trap-containing BACs as transgenes in zebrafish. *Nucleic Acids Research*, 36(19), pp.6237–6248.

Sharpe, J. et al., 1998. Selectivity, sharing and competitive interactions in the regulation of Hoxb genes. *The EMBO Journal*, 17(6), pp.1788–1798.

Shin, J.T. et al., 2005. Human-zebrafish non-coding conserved elements act in vivo to regulate transcription. *Nucleic Acids Research*, 33(17), pp.5437–5445.

Siegal, M.L. & Hartl, D.L., 1996. Transgene Coplacement and high efficiency site-specific recombination with the Cre/loxP system in DrosoPhila. *Genetics*, 144(2), pp.715–726.

Sims, R.J., Belotserkovskaya, R. & Reinberg, D., 2004. Elongation by RNA polymerase II: the short and long of it. *Genes & Development*, 18(20), pp.2437–2468.

Sipos & Gyurkovics, 2005. Long-distance interactions between enhancers and promoters. *FEBS Journal*, 272(13), pp.3253–3259.

Sipos, L. et al., 1998. Transvection in the DrosoPhila Abd-B domain: extensive upstream sequences are involved in anchoring distant cis-regulatory regions to the promoter. *Genetics*, 149(2), pp.1031–1050.

Smith et al., 2004. Synapsis and DNA cleavage in φC31 integrase-mediated site-specific recombination. *Nucleic Acids Research*, 32(8), pp.2607–2617.

Spicuglia, S. & Vanhille, L., 2012. Chromatin signatures of active enhancers. *Nucleus (Austin, Tex.)*, 3(2), pp.126–131.

Sprague, J. et al., 2001. The Zebrafish Information Network (ZFIN): a resource for genetic, genomic and developmental research. *Nucleic acids research*, 29(1), pp.87–90.

Stuart, G.W., McMurray, J.V. & Westerfield, M., 1988. Replication, integration and stable germ-line transmission of foreign sequences injected into early zebrafish embryos. *Development (Cambridge, England)*, 103(2), pp.403–412.

Sun, F.L. & Elgin, S.C., 1999. Putting boundaries on silence. *Cell*, 99(5), pp.459–462.

Suster, M.L. et al., 2009. Transgenesis in zebrafish with the tol2 transposon system. *Methods in Molecular Biology (Clifton, N.J.)*, 561, pp.41–63.

Tanimoto, K. et al., 2008. A single nucleotide mutation in the mouse renin promoter disrupts blood pressure regulation. *The Journal of Clinical Investigation*, 118(3), pp.1006–1016.

Thisse et al., 2004. Spatial and temporal expression of the zebrafish genome by large-scale in situ hybridization screening. *Methods in Cell Biology*, 77, pp.505–519.

Thomas & Chiang, 2008. The General Transcription Machinery and General Cofactors. Available at: http://www.informahealthcare.com/doi/abs/10.1080/10409230600648736 [Accessed August 12, 2009].

Thorpe & Smith, 1998. In vitro site-specific integration of bacteriophage DNA catalyzed by a recombinase of the resolvase/invertase family. *Proceedings of the National Academy of Sciences of the United States of America*, 95(10), pp.5505–5510.

Thyagarajan, B. et al., 2001. Site-Specific Genomic Integration in Mammalian Cells Mediated by Phage {Phi}C31 Integrase. *Mol. Cell. Biol.*, 21(12), pp.3926–3934.

Urasaki, A., Asakawa, K. & Kawakami, K., 2008. Efficient transposition of the Tol2 transposable element from a single-copy donor in zebrafish. *Proceedings of the National Academy of Sciences of the United States of America*, 105(50), pp.19827–19832.

Urasaki, A., Morvan, G. & Kawakami, K., 2006. Functional dissection of the Tol2 transposable element identified the minimal cis-sequence and a highly repetitive sequence in the subterminal region essential for transposition. *Genetics*, 174(2), pp.639–649.

Vardhanabhuti, S., Wang, J. & Hannenhalli, S., 2007. Position and distance specificity are important determinants of cis-regulatory motifs in addition to evolutionary conservation. *Nucleic Acids Research*, 35(10), pp.3203–3213.

Vintersten, K. et al., 2008. Bacterial artificial chromosome transgenesis through pronuclear injection of fertilized mouse oocytes. *Methods in molecular biology (Clifton, N.J.)*, 415, pp.83–100.

Visel, A. et al., 2009. ChIP-seq accurately predicts tissue-specific activity of enhancers. *Nature*, 457(7231), pp.854–858.

Visel, A. et al., 2007. VISTA Enhancer Browser--a database of tissue-specific human enhancers. *Nucleic acids research*, 35(Database issue), pp.D88–92.

Wagner, T. et al., 1994. Autosomal sex reversal and campomelic dysplasia are caused by mutations in and around the SRY-related gene SOX9. *Cell*, 79(6), pp.1111–1120.

Weber, F., de Villiers, J. & Schaffner, W., 1984. An SV40 “enhancer trap” incorporates exogenous enhancers or generates enhancers from its own sequences. *Cell*, 36(4), pp.983–992.

Wendt, K.S. & Peters, J.-M., 2009. How cohesin and CTCF cooperate in regulating gene expression. *Chromosome Research: An International Journal on the Molecular, Supramolecular and Evolutionary Aspects of Chromosome Biology*, 17(2), pp.201–214.

Westerfield, M., 2000. *The Zebrafish Book* 4th edition.,

Wilson, C. et al., 1989. P-element-mediated enhancer detection: an efficient method for isolating and characterizing developmentally regulated genes in DrosoPhila. *Genes & Development*, 3(9), pp.1301–1313.

Wolner, B.S. & Gralla, J.D., 2000. Roles for Non-TATA Core Promoter Sequences in Transcription and Factor Binding. *Mol. Cell. Biol.*, 20(10), pp.3608–3615.

Woolfe, A. et al., 2007. CONDOR: a database resource of developmentally associated conserved non-coding elements. *BMC Developmental Biology*, 7(1), p.100.

Woolfe, A. et al., 2004. Highly Conserved Non-Coding Sequences Are Associated with Vertebrate Development. *PLoS Biol*, 3(1), p.e7.

Wunderle, V.M. et al., 1998. Deletion of long-range regulatory elements upstream of SOX9 causes campomelic dysplasia. *Proceedings of the National Academy of Sciences of the United States of America*, 95(18), pp.10649–10654.

Yáñez-Cuna, J.O., Kvon, E.Z. & Stark, A., 2013. Deciphering the transcriptional cis-regulatory code. *Trends in genetics: TIG*, 29(1), pp.11–22.

Yang et al., 2006. Modified bacterial artificial chromosomes for zebrafish transgenesis. *Methods (San Diego, Calif.)*, 39(3), pp.183–188.

Zhao, Z. et al., 2001. Double-Stranded RNA Injection Produces Nonspecific Defects in Zebrafish. *Developmental Biology*, 229(1), pp.215–223.

Zhou & Levine, 1999. A novel cis-regulatory element, the PTS, mediates an anti-insulator activity in the DrosoPhila embryo. *Cell*, 99(6), pp.567–575.

Zhou & Snider, W.D., 2006. Intracellular control of developmental and regenerative axon growth. *PhilosoPhical Transactions of the Royal Society B: Biological Sciences*, 361(1473), pp.1575–1592.

9 SUBMITTED PUBLICATIONS

Roberts et al (n.d) Targeted transgene integration overcomes position effect variation in zebrafish *Submitted to Development June 2013*