Validation of CAGE predicted miRNA target sites and target genes in an *in vivo* vertebrate embryo model

By

Christopher Peter Tyrrell

A thesis submitted to the University of Birmingham for the degree of Master of Research in Biomedical Research: Integrative and Translational

Clinical and Experimental Medicine
College of Dental and Medical Science
University of Birmingham
August 2014

ABSTRACT

MicroRNAs (miRNAs) are small, non-coding regulatory RNAs that regulate gene expression. miRNAs exert inhibitory effects on gene expression via complementary binding to cognate messenger RNA (mRNA) transcripts, and subsequent degradation of the targeted transcripts. miRNAs are abundant and have many thousands of potential gene targets – only a few of which are true targets. This project involves testing a potential enhancement in the prediction of miRNA target sites via cap analysis gene expression (CAGE) tags, possibly conferring increased specificity in miRNA predictions. To test this potential prediction tool, unique CAGE predicted miRNA target sites are identified and 1 cell stage *Danio rerio* embryos are injected with RNA constructs containing the predicted miRNA target site linked to a fluorescent probe. Both 'wild type' and 'mutant' target sites are injected, and expression patterns of the target gene observed to confirm the existence of a 'true' predicted miRNA target site. Two CAGE predicted target sites were tested (*cxcr7b* and *nploc4*) and both validated as true target sites, indicated by differential gene expression patterns seen between WT and mutant target sites (significantly reduced expression is seen in WT embryos due to miRNA mediated degradation). These results show a potential for future use of CAGE tags in miRNA prediction.

ACKNOWLEDGEMENTS

I would like to thank everyone of the Mueller lab group for their help during this project. In particular I would like to thank Yavor Hadzhiev for his immense help throughout the project and beyond it.

CONTENTS

ILLUSTRATIONS

TABLES

INTRODUCTION

1.1 MicroRNAs

MicroRNAs (miRNAs) are small (~22nt), non-coding, regulatory RNAs expressed in multicellular organisms that play a functional role in the regulation of gene expression. miRNAs were first discovered during the characterization of *C. elegans* genes controlling larval development, wherein *lin-4* and *let-7* RNAs were found to exhibit temporal expression during *C. elegans* development (Reinhart et al., 2000, Lee et al., 2003).

These miRNAs can modulate a multitude of different processes including (but not limited to) developmental timing, haematopoiesis, organogenesis, apoptosis and cell proliferation (Zhao and Srivastava, 2007), through binding to partially complementary sites within 3' untranslated regions (UTRs) of mRNAs in animals and exerting inhibitory effects on gene translation (Bartel, 2004). miRNAs are estimated to regulate up to two-thirds of the mammalian transcriptome and, as such, miRNA perturbations are thought to be closely linked to many diseases, and may show promise as clinical targets for disease treatment (Chang and Mendell, 2007, Soifer et al., 2007, Sassen et al., 2008, Lu et al., 2008).

1.2 MicroRNA biogenesis

Animal miRNA biogenesis involves, first, transcription of miRNA genes via RNA polymerase II (Pol II) within the nucleus – generating a several kilobase long, 5' capped and 3' polyadenylated primary miRNA transcript (pri-miRNA) containing multiple ~80nt stem loops. Pri-miRNAs are then cropped by a microprocessor (Drosha-DCGR8) complex (Han et al., 2004) to produce a ~70nt pre-miRNA. This pre-miRNA contains a ~2nt overhang that is detected by nuclear export factor exportin 5 (EXP5), which mediates exportation of the pre-miRNA from the nucleus to the cytoplasm (Figure 1). Once exported, cytoplasmic RNase III Dicer catalyses a second processing step in which the pre-miRNA is cleaved ~22nts from the terminal loop of the double stranded RNA to produce a ~22nt miRNA duplex (Park et al., 2011). This duplex is then loaded into an argonaute containing miRNA induced silencing complex (miRISC), wherein one strand of the duplex remains as a mature miRNA and the other (passenger) strand is degraded (based on the thermodynamic stability of the 5' ends of each duplex strand (Khvorova et al., 2003)).

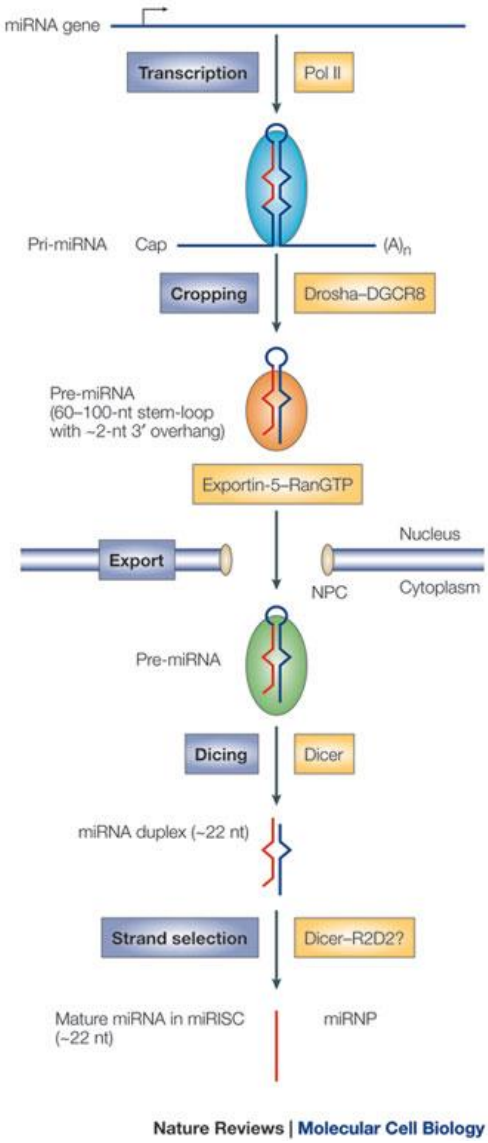


Figure 1: Illustration of the miRNA processing pathway, showing gene transcription via RNA Polymerase II producing a capped and adenylated Pri-miRNA. Further processing by the Drosha-DGCR8 microprocessor complex produces a stem-loop containing pre-miRNA which is transported from the nucleus to the cytoplasm via Exportin-5. Dicer then splices the pre-miRNA, producing a miRNA duplex, one strand of which is degraded, while the other is incorporated into the RISC as a mature miRNA. Source: Kim (2005) MicroRNA biogenesis: coordinated cropping and dicing.

1.3 The RNA induced silencing complex

The miRISC comprises a ~22nt miRNA strand, the endoribonuclease Dicer, the double stranded RNA binding protein TRBP and the argonaute protein Ago2 (Gregory et al., 2005, Rand et al., 2004). It is the miRISC that exerts inhibitory effects on gene expression. RNA interference (RNAi) induced by the RISC can be defined as small interfering RNA (in this case, miRNA) guided, site-specific cleavage of an mRNA target. The miRNA itself is responsible for the specificity of the cleavage, acting as a guide that leads the RISC to its target through complementary Watson-Crick base pairing with cognate mRNA transcripts. Argonaute 2 acts by binding to the transcript and orientating the transcript into a position facilitating target recognition for cleavage or silencing through recruitment of gene-silencing proteins as seen in figure 2 (Pratt and MacRae, 2009). The result of miRISC action is that of post-transcriptional repression of the mRNA target, which may be achieved through multiple proposed mechanisms including: Co-translational protein degradation; inhibition of translation elongation; premature termination of translation; and inhibition of translation initiation (Eulalio et al., 2008). While the exact mechanisms of miRISC action are under debate, the end-result is that of repressing translation of the target mRNA, promoting degradation of the target mRNA, or both (Guo et al., 2010).

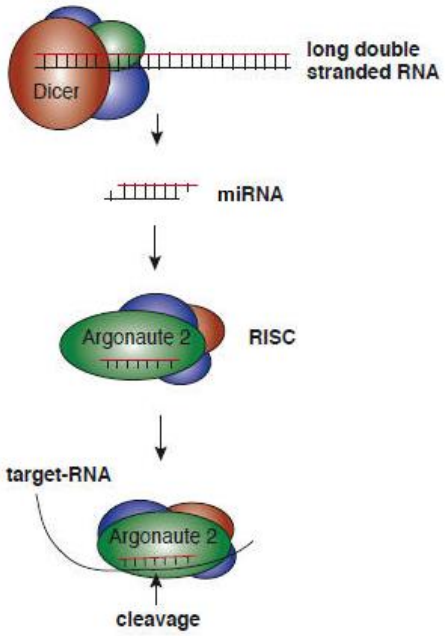


Figure 2: Illustration of RNA induced silencing complex action, from miRNA incorporation, mRNA targeting and Argonaute 2 mediated cleavage. Source: http://www.biologie.uni-regensburg.de/Biochemie1/Research/index_1.htm

1.4 miRNA specificity

Selective miRNA targeting is an important characteristic of the miRISC, with specific conditions governing the selection of targets for translational repression. Base complementarity between nucleotides 2-8 of the 5' region of miRNA (the 'seed' region) and the target mRNA has particular importance in targeting (Jinek and Doudna, 2009). This seed region most commonly binds to target sites within the 3' untranslated region (3' UTR) of mRNAs and complementarity to this region, even by itself, is a strong indicator of potential miRNA targeting. In addition to seed region complementarity, complementarity to the 3' region of the miRNA may also contribute to effective binding of a target mRNA (Shkumatava et al., 2009).

Further to this, both the miRNA seed region and the complementary 3' UTR target site show evolutionary conservation (Friedman et al., 2009, Gaidatzis et al., 2007). In addition, it has been shown that target site accessibility plays a role in miRNA target site recognition (Kertesz et al., 2007), with diminished target accessibility linked to reduced translational repression. The aforementioned factors affecting miRNA targeting can be used to make predictions concerning potential miRNA target sites – and, indeed, multiple tools exist that use existing gene sequence databases to predict potential target sites for specific miRNAs, through the use of an algorithm incorporating parameters linked to seed region complementarity and evolutionary conservation of the target site.

1.5 miRNA target prediction

Predicting miRNA target sites can be a useful tool, with legitimate target site predictions allowing elucidation of miRNA functions at a system wide level, and exploration of the potential for miRNA as therapeutic targets. While exploration of miRNA target sites could be useful for the determination of miRNA functions and contributions to healthy or diseased states, current target site prediction methods are far from flawless. Such miRNA target site prediction tools include TargetScan (Lewis et al., 2005), PicTar (Grun et al., 2005) and RNAhybrid (Rehmsmeier et al., 2004) to list but a few. These prediction tools generally rely on thermodynamically based RNA:RNA duplex binding interactions, optimal free energy calculations, complementarity to the miRNA seed region and orthologous conservation of the mRNA target site to predict miRNA target sites within 3' UTRs

(Mazière and Enright, 2007). Generally, they do not consider binding sites that may appear outside of the 3'UTR (or, if they do – false positive prediction rates are increased) – relying on a perfect or near-perfect Watson-Crick base-pairing between the miRNA seed region and a sequence within the 3'UTR. This does not account for mRNA targets that show imperfect base pairing with the 5' miRNA seed region, yet appear to compensate through additional base pairing to the 3' end of the miRNA (Brennecke et al., 2005). Further, with the potential for miRNA targeting through as little as base pairing between nucleotides 3-9 (inclusive) and allowing for mismatches in nucleotides 1 and 2 of the seed region, this results in 6nt long sequences that have the potential to bind to the 3'UTR of mRNAs. Of course, not every sequence complementary to one specific 6nt miRNA region will be a true miRNA target and, as such, false positive predictions are substantial (generally considered to be ~30%)(Lewis et al., 2003). Couple this with relatively low-throughput miRNA target site validation techniques (the vast majority of predicted targets have not been experimentally validated), and you are left with a large pool of potential miRNA target sites which have not been experimentally validated and show a large number of false positive predictions. With this current state of affairs, improvements to current target site prediction models would be valuable in future work within the field of miRNAs, and much can be gained through the reduction of potential false positive target site predictions.

1.6 Cap Analysis of Gene Expression

Cap analysis of gene expression (CAGE) is a high-throughput sequencing method that allows identification and quantification of RNA transcripts in biological samples (Kodzius et al., 2006, Shiraki et al., 2003) This is achieved through the identification and quantification of unique sequence tags (small ~27 nucleotide (nt) fragments from the start of mRNA transcripts) corresponding to the 5' ends of mRNA present within a sample. In the CAGE method, a cap-trapper full length cDNA library (Carninci et al., 1996) of the total RNA extracted from a biological sample is prepared using oligo dT primers (Figure 3). Following cDNA synthesis, a biotin group is attached to the diol residue of the cap structure of any present cDNA. Once 'capped' by a biotin group, the cDNA is linked with a biotinylated 'linker' at the 5' end that contains recognition sites essential for cloning and endonuclease restriction. The cDNA is then cleaved with EcoP151 (class II) restriction enzyme to produce (27nt) 5' tag fragments, followed by the attachment of a second linker at the 3' end to allow for amplification. Subsequent selection of the capped cDNA fragments can be performed via streptavidin magnetic beads that trap the biotin residue and allow elimination of non-capped and incompletely synthesised cDNA. These 5' tags can then be amplified via PCR, sequenced via the Sanger method (Sanger and Coulson, 1975), identified via comparison to a known genome and quantified. This method is comparable to other tag sequencing methods (SAGE (Velculescu et al., 1995) and MPSS (Reinartz et al., 2002)), offering the advantage of being relatively high throughput and allowing rather precise quantification of present mRNA.

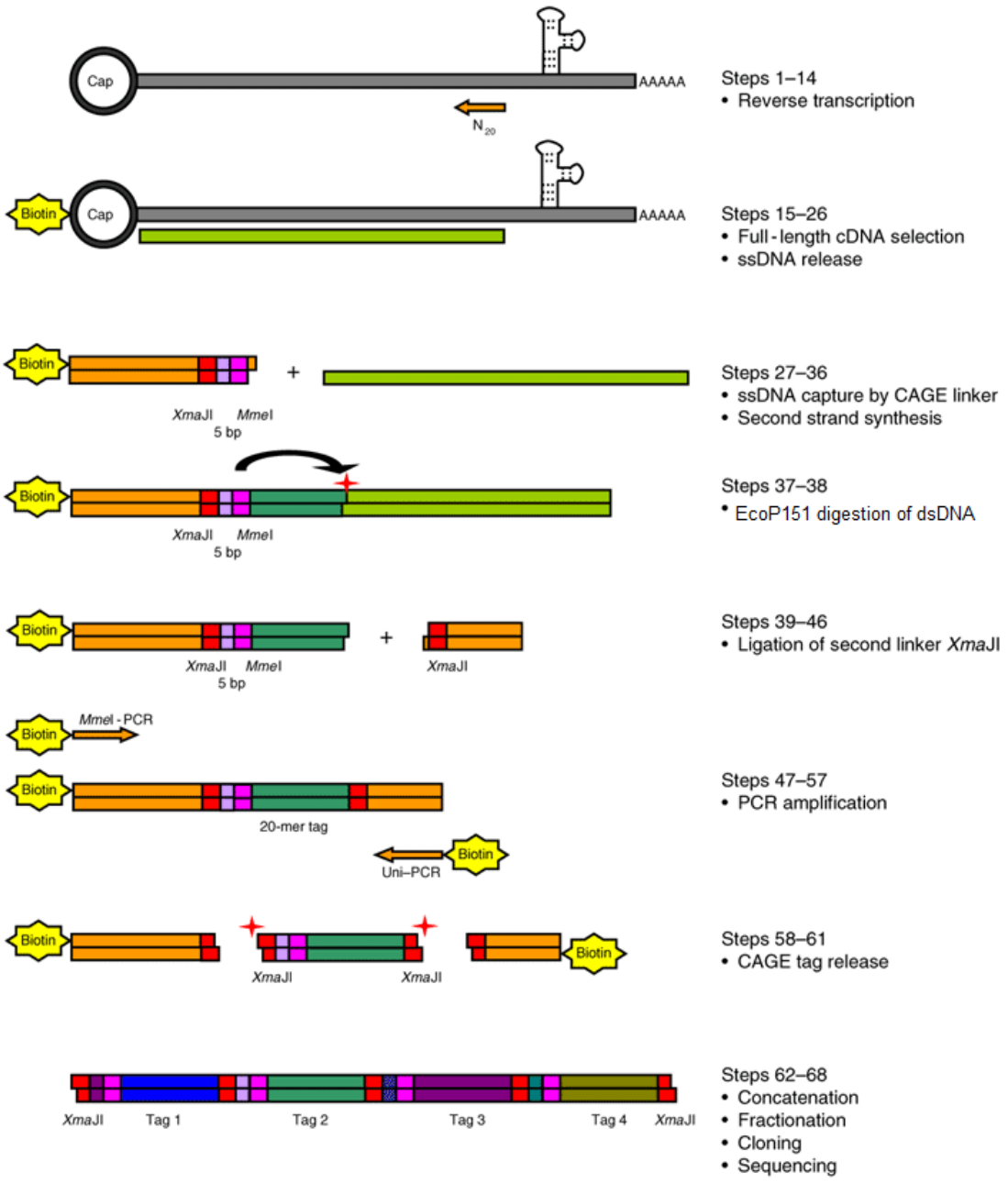


Figure 3. Graphical representation of the CAGE method used in mRNA identification and quantification. Kodzius et al. (2006) CAGE: Cap analysis of gene expression.

1.7 The Project

It is with this in mind that I (under the guidance of the Mueller lab group) undertook a 15 week project with an aim to investigate a potentially improved miRNA target site prediction algorithm based on cap analysis of gene expression (CAGE) data gleaned from research into dynamic core promoter usage during development in a vertebrate (zebrafish) embryo (Nepal et al., 2013). Nepal et al.'s research involved the identification of transcription start sites (TSSs) by CAGE analysis of RNA samples collected throughout embryonic development, and the quantification of temporal TSS usage on a global scale. CAGE allows the high-throughput identification of sequence tags corresponding to the 5' ends of mRNA at biotinylated cap sites (Affymetrix/Cold Spring Harbor Laboratory ENCODE Transcriptome Project, 2011, Kapranov et al., 2007, Project, 2009), and the identification of transcriptional start sites within RNA samples (Shiraki et al., 2003). Figure 4 (below) represents a workflow diagram followed by Chirag Nepal to identify potential miRNA target sites for experimental validation within the early stage zebrafish embryo through the use of CAGE.

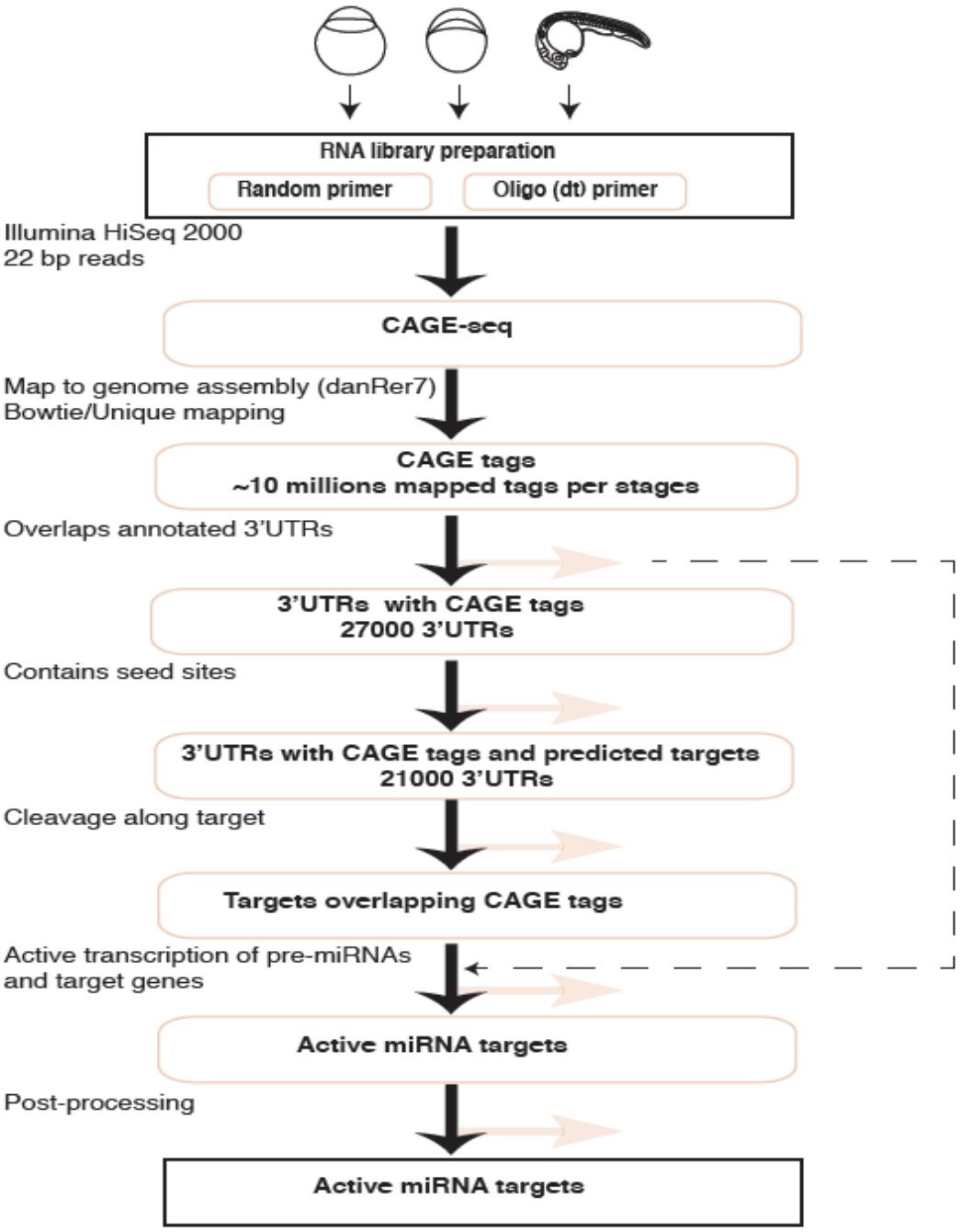


Figure 4: Diagram representing the workflow of Chirag Nepal for identifying active miRNA targets within the zebrafish for examination as CAGE predicted 'true' miRNA target sites. Once an RNA library is prepared, CAGE tags are mapped to potential miRNA seed target sites within 3' UTRs.

This examination was prompted after a secondary finding within the CAGE data of the Nepal 2013 paper showed CAGE tag enrichment within the 3'UTRs of mRNA sequences, potentially corresponding to mRNA cleavage sites (Table 1 & Figure 5, below).

	3'UTR with CAGE tags	3' UTR without Cage Tags
miRNA target	17497	8164
Non miRNA Target	566	1610

Table 1: Data from Nepal et al. 2013 shows enrichment of CAGE tags at predicted miRNA target sites, suggesting a potential for CAGE assisted miRNA prediction.

Figure 5: A Schematic representation of CAGE tag enrichment within an mRNA 3'UTR indicating the presence of potential miRNA target sites.

Graphical representation of CAGE tag enrichment within 3' UTRs and surrounding maternal miRNA seed sites can be seen below, in Figure 6:

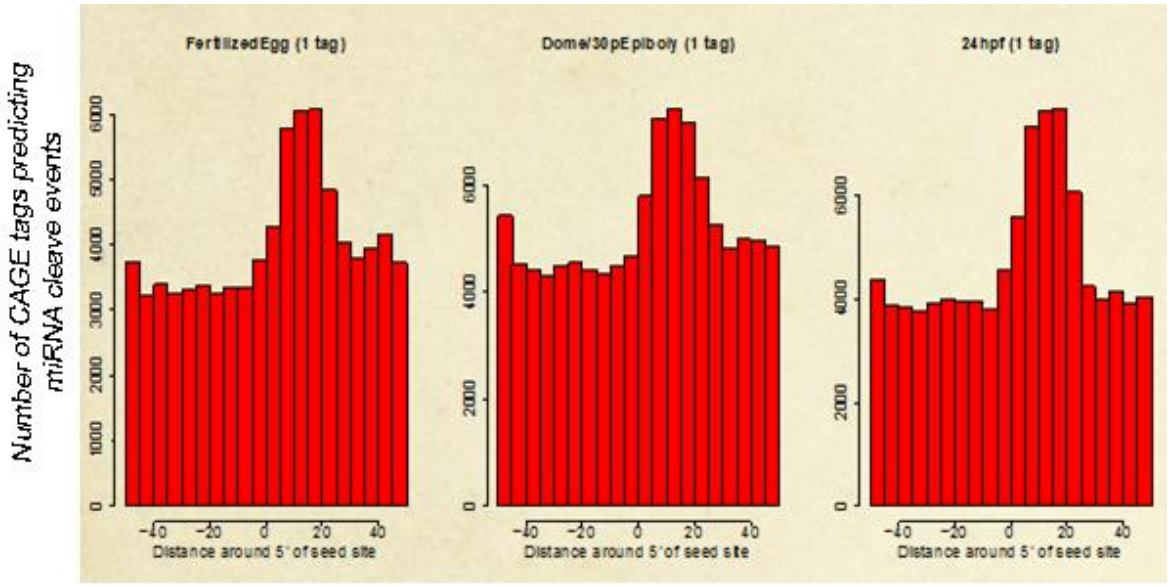


Figure 6: Graphical representation of CAGE tag enrichment within *Danio rerio* 3' UTRs at early embryonic stages – showing number of CAGE tags against relative position of 5' miRNA seed site. Produced by Chirag Nepal.

Figure 6 shows peak CAGE tag enrichment at 15-20bp downstream of the miRNA seed sites, potentially linked to miRNA cleavage events (miRNA themselves being 22bp in length).

It is plausible to suggest (hypothetically, not-supported by prior published data) that this CAGE tag enrichment is indicative of functional miRNA-mRNA interactions and, if so, there is a potential for CAGE tag enrichment to be used in conjunction with current miRNA target prediction methods to improve prediction specificity. Coupled with the bioinformatics data above suggesting the potential for CAGE tags to be indicative of miRNA-mRNA interactions, there is theoretical plausibility supporting such a scenario: essentially, CAGE capping 'could' occur at miRNA induced cleavage sites within mRNA sequences, as a by-product of the cleavage process itself (theoretical conjecture) which could act to delay

degradation at the point of RISC interaction, allowing recapping of the RNA and production of the CAGE tag. In addition to this, comparison of CAGE predicted target site data in wild-type zebrafish embryos against MZ-Dicer mutant zebrafish embryos (lacking the Dicer enzyme and as such, not exhibiting miRNA processing) shows a predicted lack of CAGE tag enrichment at suspected target sites – suggesting that CAGE tag enrichment may indeed be indicative of miRNA cleavage events.

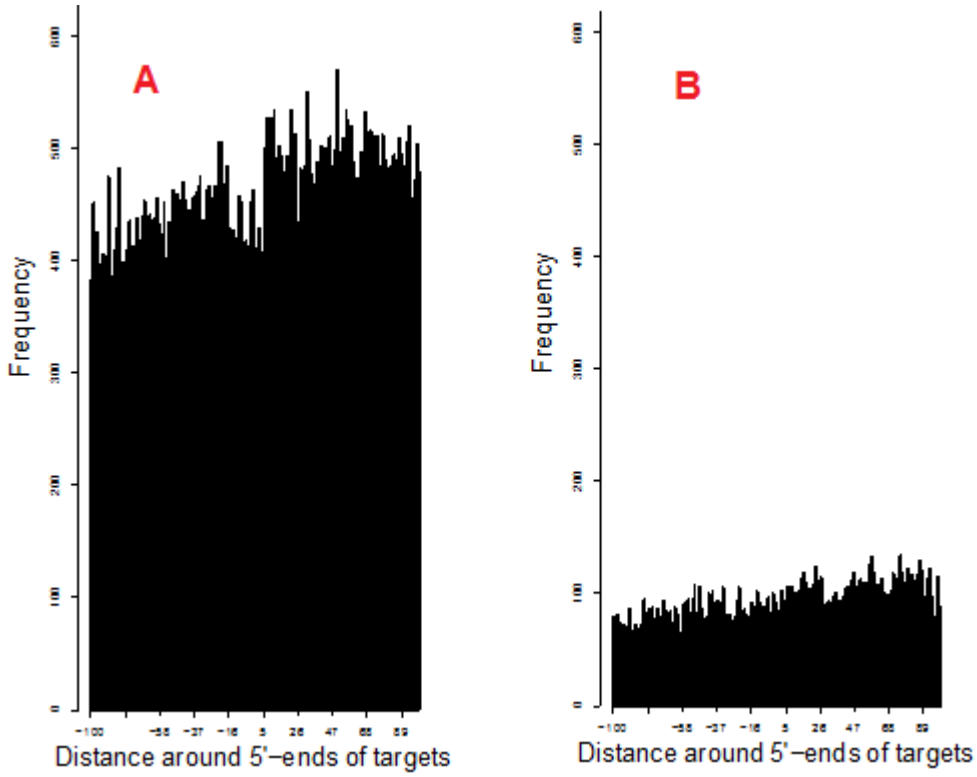


Figure 7: Represents CAGE tag enrichment within the zebrafish genome at 24 hpf in two conditions - A: WT embryos at 24 hours, B: MZ Dicer knockout embryos at 24hpf. X axis represented as nucleotide distance from seed site (Produced by Chirag Nepal).

Figure 7 shows CAGE tag enrichment seen in wild type *Danio rerio* embryos at 24hpf as compared to a lack of enrichment in MZ Dicer knockout embryos at

24hpf. MZ Dicer knockout embryos lack Dicer, a critical component in maternal miRNA mediated degradation and results in a lack of any miRNA function (Abrams and Mullins, 2009). This somewhat supports the suggestion that CAGE tag enrichment within 3' UTRs may indicate a miRNA cleavage event and miRNA target site.

These findings however, are preliminary – and it remains unclear as to whether CAGE tags seen at miRNA target sites within 3' UTRS are the result of miRNA mediated degradation, a predictor of miRNA mediated direction, or even as an artefact of computational analysis. In an effort to shed some light on this finding, it is the goal of my project to identify an ideal CAGE predicted miRNA target site and experimentally validate the site as a 'true' target site. As such, my project can be somewhat divided into three parts; First, computational analysis of RNA sequencing data and CAGE data of the *Danio rerio* (zebrafish) genome database to determine 'true' miRNA target site predictions (performed almost entirely by Chirag Nepal of the University of Copenhagen), and an ideal predicted target for experimental validation. Second, the design and trial of a protocol to successfully validate a predicted miRNA target site (using pre-validated target genes). Finally, experimental validation of a previously unverified miRNA target site predicted to be a 'true' target based on CAGE prediction data, and creation of the mRNA constructs that allow such validation.

1.8 *Danio rerio*: Zebrafish in genomics

The proposed method for the validation of bona fide miRNA target sites involves following the expression patterns of predicted target genes during zebrafish embryo development of both wild type (WT) and mutant (MUT) embryos. Zebrafish function well as a model vertebrate organism for the study of development, with a small size and optical transparency (Figure 8) allowing high resolution imaging of live embryos at various stages. This is coupled with a relatively high embryo production rate, and relatively fast embryo growth, allowing the investigation into multiple stages of development over short periods of time (Lieschke and Currie, 2007).

Figure 8: Image showing zebrafish embryo transparency, and easily achievable microscopic resolution

(http://www.zebrafishlab.be/sites/default/files/styles/media_gallery_large/public/embryos-7.jpg).

Such characteristics should enable the use of fluorescent proteins to act as markers linked to specific genes for quantification of gene expression (Villefranc et al., 2007, Finley et al., 2001). This is a core concept of my project; with the proposed mechanism for validation of CAGE predicted miRNA target sites relying on the specific binding characteristics of miRNAs to their cognate mRNA transcripts. As mentioned previously, many factors affect miRNA binding to target mRNA transcripts, an important factor of which is strict Watson-Crick base-pairing (Valencia-Sanchez et al., 2006). As such, if a specific miRNA target is predicted on a certain gene, one would expect miRNA mediated suppression of this gene, and degradation of mRNA transcripts. If, however, the cognate mRNA sequence complementary to the miRNA is manipulated to contain base mismatches against the RNA, then presumably (considering the site a 'true' miRNA target), miRNA mediated degradation shall not occur, and accumulation of the mutated gene shall arise (Doench and Sharp, 2004). As such, the rationale behind this project is to identify unique CAGE predicted miRNA target sites within the early zebrafish developmental stages and to design and create fluorescent protein linked RNA constructs incorporating two versions of such a predicted miRNA target site – A wild type unmodified target site, and a mutant target site containing base mismatches. Assuming natural production of the targeting miRNA in early zebrafish stages, one could inject the aforementioned constructs into newly fertilised zebrafish embryos and measure gene expression (via linked fluorescent probe) throughout development (Giraldez et al., 2006). If the target site incorporated into the injected RNA construct is legitimate, then one would expect

miRNA mediated suppression and degradation of gene transcripts (and relatively low expression of fluorescent marker) (Bartel, 2004, Flynt and Lai, 2008, Wakiyama et al., 2007). The mutant construct on the other hand, by way of induced base mismatches within the identified target site should show (relative) accumulation of gene transcripts (and relative abundance of fluorescent marker) via lack of miRNA mediated repression. On the other hand, if the predicted target site has, in fact, been wrongly predicted, one would expect no difference in the expression of the injected RNA construct, and similar levels of fluorescence in both the wild type and mutant treatment groups. In addition to wild type and mutant RNA probes, a control fluorescent probe shall have to be used in conjunction with the probe linked to a predicted miRNA target. This probe should be linked to an invariant gene, one not targeted for miRNA mediated degradation (at the least, not in early development stages up to 72hpf) (Stürzenbaum and Kille, 2001). As such, this probe would act as a control for the amount of construct injected into individual embryos (to which the treatment probe can be normalised against), and as a potential indicator of non-specific / non-miRNA mediated degradation (if such results arose that may suggest this).

OBJECTIVES

The main and principal goal of this short research project was to determine if CAGE data of the *Danio rerio* genome has the potential for use in the prediction of true miRNA target sites within zebrafish. The determination of the usefulness of CAGE data in this capacity was performed through multiple steps. In chronological order, these steps comprised:

- The identification of maternally inherited and early-stage active miRNAs in the zebrafish.
- Selection of prospective *Danio rerio* genes targeted by identified maternal miRNAs for use in the experimental validation of CAGE predicted miRNA target sites.
- Experimental validation of a novel CAGE predicted miRNA target site as a ‘true’ miRNA target site.
- Evaluation of CAGE as an assistive tool in the prediction of bona fide miRNA target sites within zebrafish.

MATERIALS & METHODS

3.1 Selection of candidate target genes for experimental validation of CAGE predicted miRNA targeting

The selection of candidate genes for experimental validation of CAGE predicted miRNA target sites was performed through the use of multiple RNA libraries and RNA sequencing data gained from previous studies into RNA expression within zebrafish, both temporal and spatial (Yao et al., 2014, Wei et al., 2012). These libraries, containing expression data of both early developmental stage mRNA and miRNA expression were used in conjunction with the CAGE data gleaned from Nepal et al.'s 2013 research paper. The data from these projects were uploaded into custom tracks on the UCSC Genome Browser (http://genome.ucsc.edu/) for analysis and comparison side by side.

3.1.1 Candidate target gene criteria

The selection of *Danio rerio* candidate genes for experimental validation of CAGE predicted miRNA target sites is determined by these main criteria:

- The gene must be expressed maternally (i.e. Pre-MBT; unfertilised egg up to 64 cell stage and ≥5 transcripts per million (tpm)).
- The gene must have at least one CAGE predicted maternal miRNA target site present within the 3'UTR pre-MBT (preferably only one).

- As far as possible, this predicted miRNA target site should not overlap with any other predicted or known miRNA target sites present pre-MBT (i.e. have physically overlapping target recognition sites) in an attempt to ensure that any change in expression is due to the miRNA in question and not due to a different, potentially confounding miRNA .
- The candidate gene should show an expression pattern consistent with that of a gene affected by the maternal miRNA predicted to be targeting it.

3.1.2 Identification of maternal miRNAs

To identify maternal miRNAs and potential candidate genes for the validation of predicted miRNA target sites, certain filters were applied to RNA library data: miRNAs that are inherited maternally and significantly active pre-MBT are identified, using RNA expression data from previous work into RNA expression during early zebrafish development (Yao et al., 2014, Wei et al., 2012).

3.1.3 Identification of potential gene targets

Identification of potential miRNA gene targets involves the identification of any CAGE predicted target site of miRNAs identified as candidates for experimental validation in 2.1.2. These CAGE predicted target sites are reviewed to identify only target sites showing an expression pattern consistent with maternal miRNA degradation according to the RNA library expression data. Specifically, target genes were filtered to include only those that were expressed in all three maternal

stages (≥5 tpm), and whose expression levels were higher (at least ≥1.5 fold in at least two pre-MBT stages) during Pre-MBT stages as compared to MBT stages. Any potential maternal miRNA gene targets meeting the criteria for use are then manually viewing in the UCSC genome browser to both confirm expression patterns throughout early development, and to identify ideal candidates with no or few conflicting CAGE predicted miRNA target sites during early development.

3.1.4 Proof of concept for experimental miRNA target site validation

To validate the proposed experimental method for confirming the presence of a 'bona fide' miRNA target site, a proof of concept experiment is performed on a known and previously validated miRNA target site. The *cxcr7b* gene was chosen as a previously verified miRNA target of miR-430, a maternally inherited miRNA (Staton et al., 2011). Experimental validation of *cxcr7b* as a miR-430 target required multiple steps:

- The design, creation and purification of two gene constructs; one construct containing the wild type *cxcr7b* seed region to which miR-430 is targeted, and a second, mutant construct, containing base mismatches within the seed region to prevent miR-430 targeting and miR-430 mediated mRNA degradation. Both constructs are designed with an mCherry reporter for mRNA degradation and a CFP reporter for control of RNA injection amounts and retrospective normalisation of mCherry values.

- Microinjection of the RNA constructs into 1 cell stage *Danio rerio* embryos of two groups; wild type (WT), or mutant (MUT) RNA.
- Growth of the embryos to 72h.p.f.
- Random selection of embryos for automated fluorescent imaging and quantification of mRNA via the measurement of both mCherry and CFP fluorescence.
- Optional: Further quantification of mRNA via RT-qPCR
- Comparison of the relative *cxcr7b* expression between WT and MUT groups.

3.1.5 Design of mRNA constructs for use in experimental validation

Note: The design, production and implementation of both the *cxcr7b* WT/MUT and the *nploc4* WT/MUT RNA sequences in the laboratory for microinjection into zebrafish embryos follow identical procedures and, as such, only the design and production of the *nploc4* constructs (being the more important component of this project) shall be detailed.

Design of the RNA constructs for use in the validation experiments was performed in Serial Cloner 2.5, a tool allowing graphical representation of DNA constructs with virtual construction, fragmentation, sequence alignment etc. (http://serialbasics.free.fr/Serial_Cloner.html). All constructs are designed within a pCS2 vector; owing to its high level of transient expression in vertebrate cells and ability for *in vitro* RNA synthesis of sequences cloned into one of the available polylinker sites (http://www.xenbase.org/reagents/vectorAction.do?method=displayVectorSummary&vectorId=1221270):

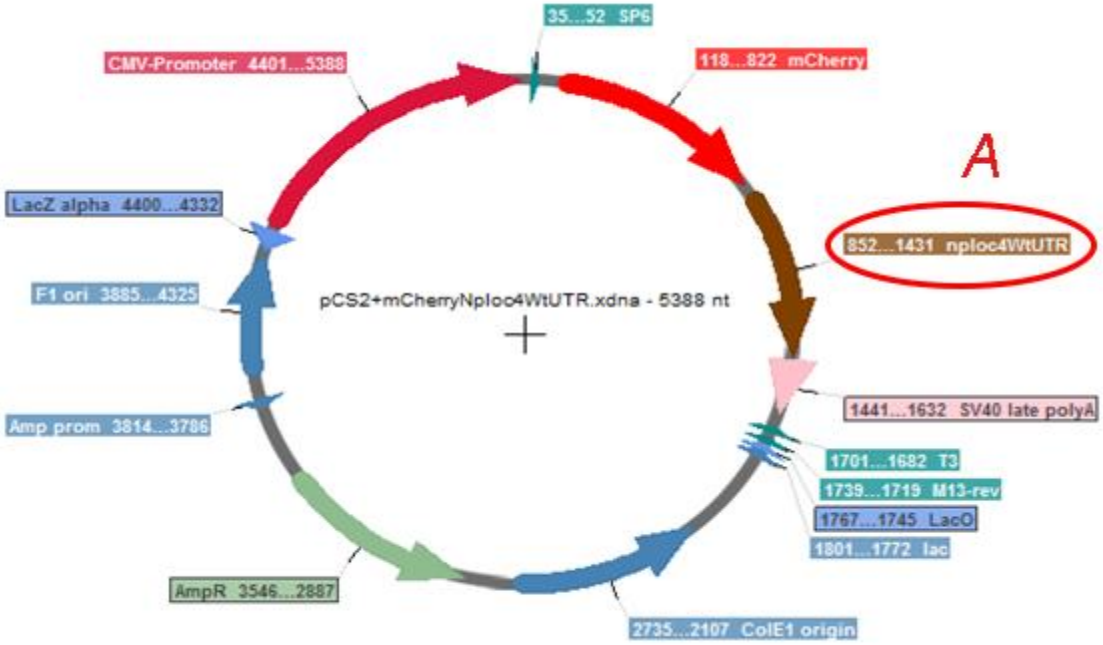


Figure 9: Serial Cloner graphical map of the *nploc4* wild type RNA construct, designed within a pCS2 vector. 'A' represents the WT *nploc* UTR sequence.

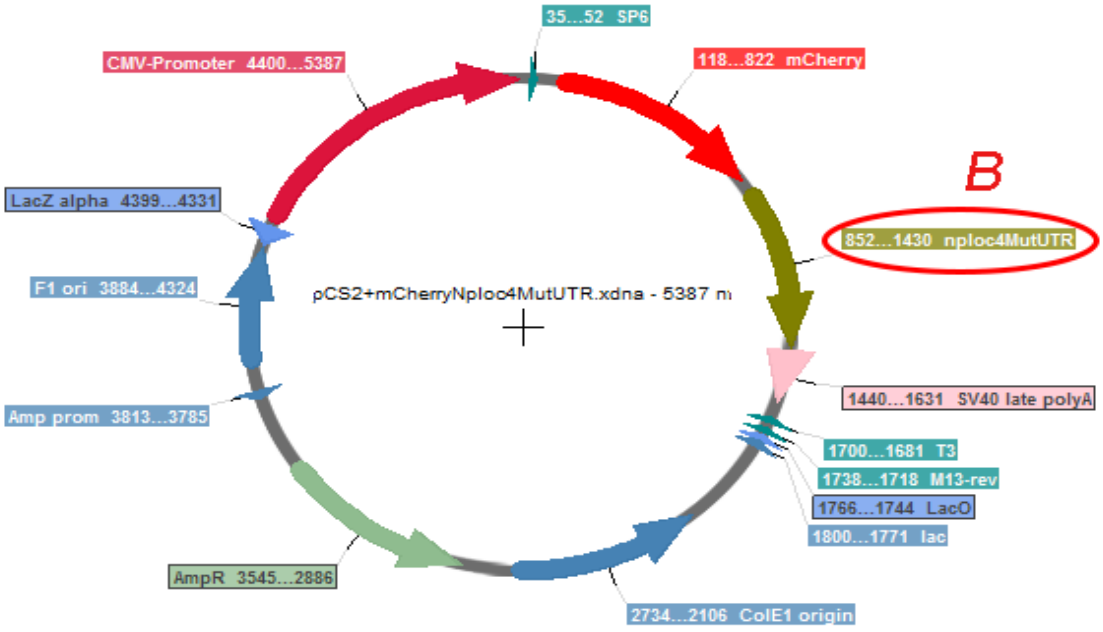


Figure 10: Serial Cloner graphical map of the *nploc4* mutant RNA construct, designed within a pCS2 vector. 'B' represents the MUT *nploc4* UTR sequence.

Figures 9 & 10 represent graphical maps of the designed pCS2 plasmid vectors, containing multiple key features. Importantly, the vector contains an mCherry coding sequence for the production of red fluorescent protein as a marker for gene expression (Villefranc et al., 2007). Additionally (though not shown on the sequence map) this pCS2 vector contains multiple endonuclease restriction sites for linearization of the plasmid at specific sites for incorporation of the *nploc4* gene sequence. Further, the vector contains CMV and SP6 promoter sequences for relatively high expression of the *nploc4* WT/MUT gene (Schmidt et al., 1990) (Melton et al., 1984) and an SV40 polyA site for polyadenylation and stability of the mRNA (Connelly and Manley, 1988) .On top of this, the vector contains ampicillin resistance genes for selection of correctly transfected cells during transformation (Glover, 2013). Both the WT and MUT constructs contain all of these key features, and only differ in the *nploc4* 3'UTR, in which the WT construct contains the original sequence complementary to the miR-430 seed region (allowing miR-430 mediated degradation), and the MUT 3'UTR contains base mismatches introduced into this sequence (to prevent miR-430 mediated degradation), as seen in Figure 11 (which results in a 1 bp plasmid size difference, 5388bp (WT)[A in figure 9] compared to 5387bp (MUT)[B in figure 10]).

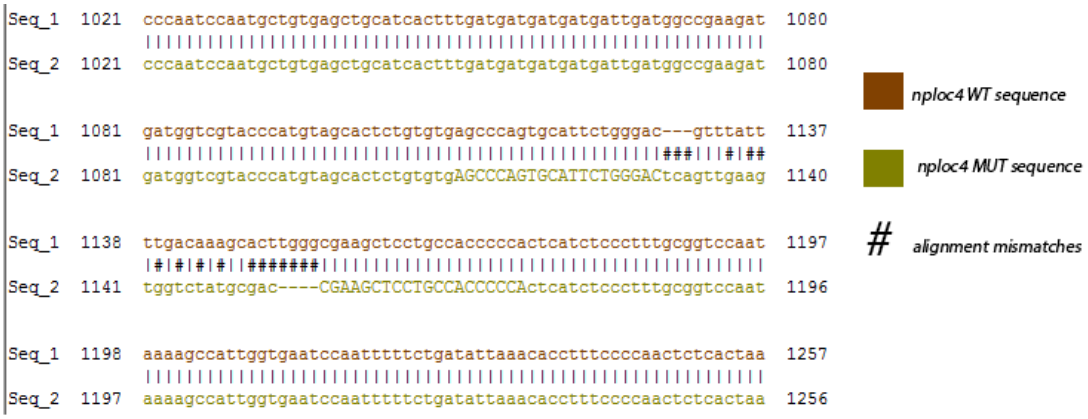


Figure 11: A local sequence alignment representing a section of the *nploc4* 3'UTR contained within the pCS2 vector. The only difference between the WT and MUT constructs is that of base mismatches between bases 1131 and 1154 within the *nploc4* sequence, introduced to prevent predicted miR-430 mediated degradation of the mutated sequence.

Both the *nploc4* WT and *nploc4* MUT 3'UTR sequences were ordered as oligonucleotide forward (FP) and reverse (RP) primers for use in the production of *nploc4* WT / MUT RNA via In-Fusion HD Cloning (Table 4).

Primer Name	MW	Tm(°C)	µg	nmol	GC%	µl for 20 µM
*nploc4*_3UTR_FP	11043	79.7	322.2	29.1	50.0	1459
*nploc4*_3UTR_RP	12475	75.6	325.3	26	42.5	1304
*nploc4*_m430DL_FP	13129	89.5	366	27.8	58.1	1394
*nploc4*_m430DL_RP	13177	85.2	420.6	31.9	53.4	1596

Primer Name	Sequence (5-3)
*nploc4*_3UTR_FP	AGTGAGTCGTQTTQCAACCCCAATGAACCCGGGCTA
*nploc4*_3UTR_RP	CATGTCTGGATCTACGTAAGAAAGTAGATGTGGTGATGT G
*nploc4*_m430DL_FP	TCAGTTGAAGTGGTCTATGCGACCGAAGCTCCTGCCACC CCCA
*nploc4*_m430DL_RP	GTCGCATAGACCACTTCAACTGAGTCCCAGAATGCACTG GGCT

Table 2: Properties of the In-Fusion primers ordered for the *nploc4* WT / MUT fragment generation. GC% content within 40-60% and primer pair Tm within 5°C of each other.

3.2 Production of mRNA constructs for use in experimental validation

The first step towards creation of an injectable RNA construct involved linearization of the pCS2 vector via restriction digestion with SnaBI (digesting immediately before the SV40 promoter sequence). A 50 µl reaction mixture was made containing 10 µl pSC2 plasmid, 5ul 10x reaction buffer, 2 µl SnaBI restriction enzyme (1 unit/µl) and 33 µl dH_2O, gently pipetted to thoroughly mix and incubated at 37°C overnight (~20 hours) to produce a 4806bp linearized plasmid (correct digestion and lack of contamination confirmed through gel electrophoresis of sample).

3.2.1 pCS2 Vector Linearization

Materials:

10 µl pSC2 plasmid

5 µl 10x buffer

2 µl SnaBI enzyme (1 unit/µl)

33 µl dH_2O

1 ml centrifuge tube

37°C Incubator

3.2.2 PCR amplification of desired WT / MUT fragments

Materials:

2 µl template DNA

15 µl HD Buffer

6 µl dNTP mixture

0.75 µl HD polymerase (2.5 units/µl)

47.5 µl dH_2O

1.25 µl *nploc4* WT 3'UTR FP

1.25 µl *nploc4* WT 3'UTR RP

0.625 µl *nploc4* MUT 3' UTR FP

0.625 µl *nploc4* MUT 3' UTR RP

Geneflow Thermocycler

Following plasmid linearization, PCR amplification is performed to amplify the insert DNA of both WT and MUT 3'UTR constructs. All reagents are thawed on ice before a main reaction mixture is prepared comprising 2 µl template DNA, 15 µl HD Buffer, 6 µl dNTP mixture, 0.75 µl HD polymerase and 47.5 µl dH_2O (total 71.25 µl) as the base for 3 x 25 µl reaction mixtures.
23.75 µl of this reaction mixture was taken for each of three amplification reactions:

1. 0.625 µl WT 3'UTR FP + 0.625 µl WT 3'UTR RP (+23.75 µl mix) [WT]
2. 0.625 µl WT 3'UTR FP + 0.625 µl MUT 3'UTR RP (+23.75 µl mix) [F1]
3. 0.625 µl MUT 3' UTR FP + 0.625 µl WT 3'UTR RP (+23.75 µl mix) [F2]

Each reaction mixture was pipetted into separate PCR tubes, mixed and centrifuged briefly (10s) before undergoing the following PCR protocol:

Repeat cycles	Temp (°C)	Times (s)	Cycles
	94	120	1
	94	20	30
	55	15	30
↻	72	45	30
	72	30	1
	10	∞	1

Once the PCR protocol is completed, a sample of each product is taken and run through gel electrophoresis, followed by spectrophotometric analysis for confirmation of desired product amplification and purity (end product consists of 20ul amplified fragment samples).

3.2.3 Gel Electrophoresis

Materials:

0.3g pure agarose powder

30ml TAE buffer

3 µl 100bp ladder marker

2 µl product sample (x3)

1 µl ethidium bromide (x3)

1 µl loading dye (x3)

1% agarose gels (30ml TAE, 0.3g Agarose powder) containing 1 µl ethidium bromide were used for electrophoresis. A 2 µl sample of each product was taken and loaded with 1 µl loading dye and 1 µl nuclease free H_2O after thorough mixing. This was compared against 3 µl of 100bp ladder as a marker. The gel was run at 80V for 40 minutes before visualisation under UV light to confirm purity of expected products (Figure 12).

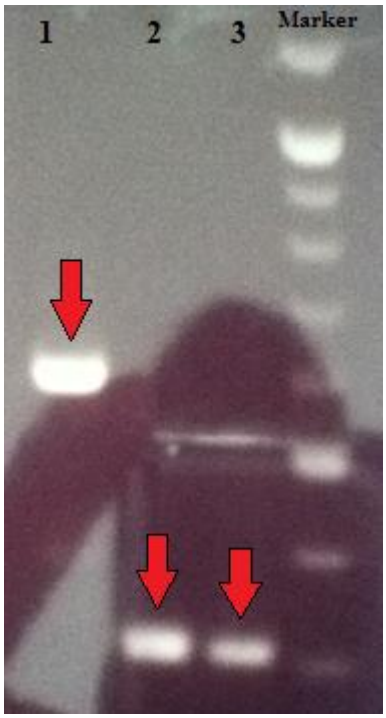


Figure 12: UV visualisation of 1% agarose gel electrophoresis to determine correct plasmid amplification, insert ligation, and to confirm lack of contamination. Lane 1: F1. Lane 2: F2, Lane 3: WT. 100bp ladder was used as a marker.

3.2.4 Spin column purification

Materials:

In-Fusion HD cloning kit including:

80 µl dH_2O (making up to 100 µl) per sample

200 µl NTI Binding Buffer per sample

750 µl NT3 washing buffer per sample

Following this, the products of the previous PCR amplification (F1, F2 and WT) were purified via spin-column purification (plasmid DNA clean up) – each product was made up to 100 µl with dH_2O, followed by the addition of 200 µl binding buffer NTI in a spin-column. Each sample was then centrifuged at 11000g for 30s before being washed with 750 µl NT3 washing buffer. Each sample is spun for a further 30s at 11000g to remove the NT3 buffer, further spun at 11000g for 30s for improved removal of any ethanol that may remain from the NT3 buffer, and finally the DNA for each sample is eluted with 30 µl NE buffer via a 60s 11000g spin. The purified product for each sample is then kept and used for the In-Fusion cloning procedure.

Taking the purified samples of WT, F1 and F2, three reactions are set up for in fusion cloning to produce an *nploc4* WT 3'UTR incorporating plasmid, an *nploc4* MUT 3'UTR incorporating plasmid and a Control sample. 2 µl linearized vector, 6 µl HD polymerase enzyme and 13 µl dH_2O is mixed to produce a 21 µl reaction mixture to be split 3 ways for each reaction:

7 µl reaction mixture + 1 µl *nploc4* WT 3'UTR + 2 µl dH_2O (WT)

7 µl reaction mixture + 1.5 µl *nploc4* F1 MUT 3'UTR + + 1.5 µl *nploc4* F2 MUT 3'UTR (MUT)

7 µl reaction mixture + 3 µl dH_2O (control)

Each reaction mixture is pipetted into separate 1ml centrifuge tubes and incubated for 15 minutes at 50°C before being stored at -20°C until transformation.

3.2.5 Transformation of electrocompetent E. coli cells with purified PCR fragments

Materials:

Electrocompetent E. coli cells

In-Fusion transformation reaction mixture

LB agar

Ampicillin (100 µg/µl)

Pre-prepared electrocompetent E. coli cells are used for the transformation of the three purified products: *nplco4* WT, *nploc4* MUT and Control plasmids. 1µl of each reaction product is placed into reaction tubes with 2.5 µl in-fusion reaction mixture and made to 100 µl with electrocompetent E. coli cells incubated on ice for 5 minutes. Following this, each reaction mixture was spread onto ampicillin containing agar plates and incubated at 37°C for selection of ligated plasmids.

Following incubation, 5 colonies from each of WT and MUT reactions were taken (as expected, control plasmids results in few colonies) and transferred into falcon tubes containing 5ml of LB agar and 5 µl of ampicillin. These mixtures were then further incubated at 34.5°C in an agitating incubator for growth and further selection.

3.2.6 Selection of desired plasmids

Materials (per PCR reaction):

1.6 µl dNTP mix

4 µl HD Buffer

0.2 µl TAC polymerase

14 µl dH_2O

0.1 µl T6 Primer

0.1 µl P7 Primer

1 µl reaction mixture of each colony to be tested for insert

Following incubation and bacterial growth of transformed cells, 1 ul of each reaction mixture is taken for selection of bacteria with desired plasmid insert incorporation (5 WT and 5 MUT). TAC polymerase rather than high fidelity polymerase was used in the PCR amplification of each colony:

Repeat cycles	Temp (°C)	Time (s)	Cycles
	94	300	1
	94	30	30
	55	40	30
↻	72	90	30
	72	300	1
	12	∞	1

Following amplification, each PCR product was run on an electrophoresis gel (as per previous protocol – 1% agarose, 80v, 40 minutes) to identify colonies with correctly incorporated WT and MUT UTR inserts within the vector.

3.2.7 Isolation and purification of DNA from bacterial cells

Purification of the WT 3' UTR and MUT 3' UTR plasmids was performed through the use of the Qiagen plasmid mini-kit following this protocol:

- Cells thawed on ice and collected by centrifugation for 15 minutes at 6000g and 4°C
- Cells suspended in 0.5ml P1 resuspension buffer (50mM Tris-Cl, 10mMEDTA, 100µg/mL RNase A)
- 0.5ml of P2 lysis buffer (200mM NaOH, 1% SDS) added, with immediate mixture via tube inversions and incubation at room temperature for 5 minutes
- 0.5ml P3 neutralisation buffer (3M potassium acetate) added, with immediate mixture via tube inversions and incubation on ice for 5 minutes
- Centrifugation of each mixture for 10 minutes at 11000g

- Application of the centrifugation supernatant to a Qiagen purification tip for complete filtering
- Addition of 2 x 2ml QC wash buffer (1M NaCl, 50mM MOPS, 15% isopropanol) for complete filtering
- Addition of 0.8ml QF elution buffer (1.25M NaCl, 50mM Tris-Cl, 15% isopropanol) with the eluate collected
- DNA within eluate precipitated with the addition of 0.56ml isopropanol and centrifugation at 11000g for 30 minutes at 4°C

With the DNA purification complete, both the pCS2 WT UTR and pCS2 MUT UTR solutions are further tested for correct plasmid sequences via restriction digestion and gel electrophoresis for identification of expected restriction fragments.

3.3 Confirmation of pCS2 WT UTR and pCS2 MUT UTR plasmids via restriction digestion and gel electrophoresis

Materials (per reaction):

0.5 µl BamHI

0.5 µl XbaI

1 µl SnaBI

2 µl CutSmart buffer (50mM Potassium Acetate, 20mM Tris-acetate, 10mM Magnesium Acetate, 100µg/ml BSA)

11 µl dH_2O

5 µl DNA sample

Samples of both the pCS2 WT UTR and pCS2 MUT UTR purified DNA were digested with SnaBI, BamHI and XmaI restriction endonucleases for 40 minutes at 37°C followed by gel electrophoresis on a 1% agar gel (as per previous protocol). Assuming correct plasmid restriction, 3 fragments were expected to be obtained for both WT and MUT plasmids.

pCS2 *nploc4* WT UTR expected restriction digestion:

```
<Serial Cloner V2.5>

Restriction analysis of pCS2+mCherryNploc4WtUTR.xdna [Circular]
Incubated with BamHI + SnaBI + XmaI

3 fragments generated.

   1:    4,036 bp  -  From SnaBI[1431]   To  BamHI[79]

   2:      785 bp  -  From BamHI[79]     To  XmaI[864]

   3:      567 bp  -  From XmaI[864]     To  SnaBI[1431]
```

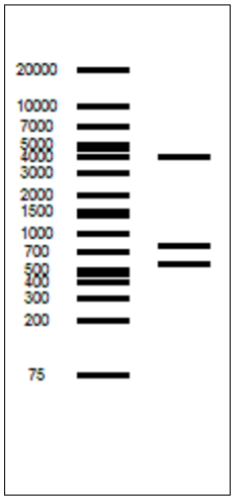


Figure 13: Serial cloner virtual restriction digestion of the pCS2 *nploc4* WT UTR containing plasmid with BamHI, SnaBI and XmaI resulting in the generation of three fragments.

pCS2 *nploc4* MUT UTR expected restriction digestion:

```
<Serial Cloner V2.5>

Restriction analysis of pCS2+mCherryNploc4MutUTR.xdna [Circular]
Incubated with BamHI + SnaBI + XmaI

3 fragments generated.

   1:    4,036 bp  -  From SnaBI[1430]   To  BamHI[79]

   2:      785 bp  -  From BamHI[79]     To  XmaI[864]

   3:      566 bp  -  From XmaI[864]     To  SnaBI[1430]
```

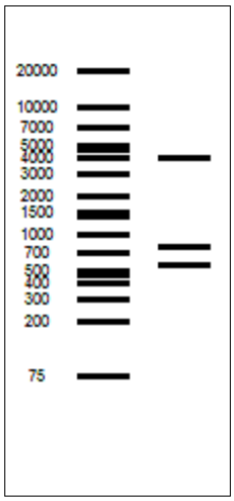


Figure 14: Serial cloner virtual restriction digestion of the pCS2 *nploc4* MUT UTR containing plasmid with BamHI, SnaBI and XmaI resulting in the generation of three fragments.

The fragments generated shall differ only in 1bp of one fragment (566 bp vs 567 bp of the XbaI to SnaBI cut) representing the 1 bp WT / MUT fragment difference (Figures 13 & 14).

3.4 RNA synthesis of purified wild type and mutant pCS2 plasmids

Materials (per reaction):

Linearization:

30 µl plasmid

1 µl NotI

10 µl CutSmart buffer

9 µl H_2O

Transcription:

10 µl 2x NTP/CAP buffer solution

2 µl 10x Reaction Buffer

2 µl Enzyme mix (RNA polymerase, RNase inhibitor)

2 µl nuclease free H_2O

4 µl linearized plasmid

An mMESSAGE mMACHINE transcription kit (http://tools.lifetechnologies.com/content/sfs/manuals/cms_055516.pdf) is used for in vitro synthesis of MUT / WT capped RNAs. First the purified plasmids are

digested with NotI restriction endonuclease for 3 hours at 37°C in preparation for transcription.

Following linearization, the transcription reaction mix is assembled at room temperature and mixed thoroughly, followed by brief 10s centrifugation. The reaction mixture is then incubated at 37°C for one hour.

Subsequent to incubation and RNA synthesis, recovery of the RNA is performed through phenol:chloroform extraction and isopropanol precipitation.

Materials:

115 µl nuclease free H_2O

15 µl ammonium acetate stop solution

20 µl transcription reaction mix

150 µl phenol:chloroform

150 µl chloroform

150 µl isopropanol

115 µl nuclease free H_2O and 15 µl ammonium acetate stop solution is added to the 20 µl transcription reaction mixture following incubation and mixed thoroughly. This is followed by phenol:chloroform extraction with equal volume phenol:chloroform and centrifugation at 10,000 g for 5 minutes at room temperature. The aqueous phase is collected, transferred to a fresh tube and further extracted with equal volume chloroform at 10,000 g for 5 minutes. Again, the aqueous phase is collected and transferred to a fresh tube before being precipitated with 1 volume isopropanol and chilled on ice for 15 minutes. Finally,

the mixture is centrifuged at 15,000 g for 15 minutes at 4°C and the purified RNA pellet is dried and resuspended in 60 µl of DNase free water for storage of the RNA at -20°C.

3.5 Confirmation of RNA purity via gel electrophoresis and spectrophotometry

One final gel electrophoresis gel is run to confirm purified RNA products and to ensure lack of contamination (as per previous protocol). In addition, spectrophotometric analysis of *nploc4* WT / *nploc4* MUT RNA samples is undertaken to ensure lack of contamination and confirm RNA concentration. Both RNA mixtures are made up to 150 ng/µl via dilution in nuclease free H_2O and stored at -20° until use in embryo microinjection.

3.6.1 Embryo collection and RNA injection

Materials:

Breeding tanks

Breeding tank dividers

100mL petri dishes

Pre-prepared RNA for miRNA target of study (*nploc4 / cxcr7b*)

CFP RNA

MINJ-1 Tritech Research microinjection system

E3 + Gentamycin media

Prior to embryo collection, dividing barriers within each breeding tank are removed and the tanks placed at an angle (~30°) to simulate a shallow/deep water gradient. The fish are then left to lay and fertilise the embryos before collection (~20 minutes).

Upon collection, embryos are counted, divided equally into treatment groups and pipetted onto petri dishes for RNA injection. Excess fish water is removed via thin-tipped pipette and the embryos arranged into a one cell high layer in preparation for injection (note: RNA injection is best done at the 1 cell stage, and as such, embryo collection, preparation and injection should be done as quickly as possible). The embryos are then micro-injected with ~2nl of ~30ng/ µl RNA (Wild Type + CFP / Mutant + CFP, dependent upon group). Microinjection is performed using a MINJ-1 Tritech Research microinjection system.

First, a few microlitres of the relevant RNA construct is loaded into the microinjection needle, the needle is attached to the needle holder and fastened in place. The micro-injector gas valves are then loosened to allow gas-powered microinjection, followed by cutting the tip of the needle with a scalpel to allow injection (requires a good cut to produce a needle tip that is not too long and flexible, and yet not too short and thick). Viewing the embryos under a microscope (approximately 15x magnification), the needle is manually controlled to deliver a single ‘burst’ (~2nl) of injected RNA (via foot pedal) into the yolk of the embryo (with a smooth and ‘clean’ stabbing motion through the chorion). This is repeated

for all embryos within the petri dish, with water tension preventing the embryos from moving. Once completed, the embryos are resubmerged in fish water and stored in a 27.5°C incubator while the next RNA construct is delivered to the next treatment group with a repeat of the above procedure.

Once all injections are complete, embryos are submerged in solution containing E3 media and gentamycin, and stored within a transparent petri dish in an incubator at 28.5°C. The embryos are periodically checked throughout growth to replace media, remove dead or deformed cells, to perform dechorionation and for the addition of phenylthiourea (PTU) to remove pigmentation before examination and imaging (if applicable).

3.6.2 Fluorescence imaging and mRNA quantification

Materials:

Olympus Scan^R IX 81 microscope

Custom made 96 well plate

Tricaine mesylate

Broad (cut) tipped pipette

Embryos undergoing automated fluorescent imaging for determination of RNA quantification are dechorionated and treated with PTU at around 8hpf (to avoid potential developmental defects that may be induced if applied earlier) to clear pigmentation and allow visualisation of tissue specific fluorescence.

Fluorescence imaging is undertaken using an Olympus Scan^R IX 81 fluorescent microscope, capable of automated imaging of 96 well plates at multiple wavelengths. In this case, embryos are imaged through brightfield (BF), red fluorescent protein (RFP) and cyan fluorescent protein (CFP) filters. Following injection and growth to the relevant stage (72 hpf), embryos from each treatment group (be it *cxcr7b* WT / MUT or *nploc4* WT / MUT) are screened pre-visualisation to ensure uptake and dissemination of injected RNA via CFP screening. Of those embryos positive for injected RNA incorporation, 48 embryos of each treatment group are randomly selected for anaesthetisation with a drop of tricaine mesylate (MS-222) within a petri-dish of E3 media and gentamycin. Following anaesthesia, each embryo is pipetted into a separate well within a 96-well plate (using a cut-tipped pipette, and along with 200 µl of media) and orientated in a lateral position (with the aid of a custom pin-prick induced hole at the bottom of each well to accommodate the yolk sac bulge) at the centre of each well (Figure 15).

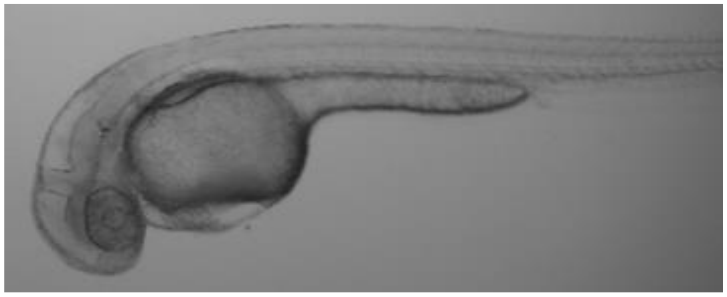


Figure 15: An example of embryo orientation with a well of a 96-well plate. Each embryo is orientated in a lateral position, as close to the centre of the plate well as possible to reduce variability between each embryo when imaged automatically.

It is important to correctly orientate each embryo in a lateral position and as similarly as possible to reduce variation in imaging due to improper orientation. The Scan^R automated imaging system can somewhat adjust to minor deviations

in orientation, but this is best minimised for consistent imaging and results.

Once all 96 well plates are loaded (48 WT / 48 MUT), and the embryos correctly orientated, the plate can be loaded on the Scan^R microscope for automated imaging. The Scan^R acquisition software enables the use of multiple channels for simultaneous BF, RFP and CFP imaging at multiple z planes for each embryo. Automated adjustments are made for each embryo to maintain focus and provide images along the z plane. These images are then compiled to create a projected ‘3d’ image of each embryo (for each channel) (Figure 16).

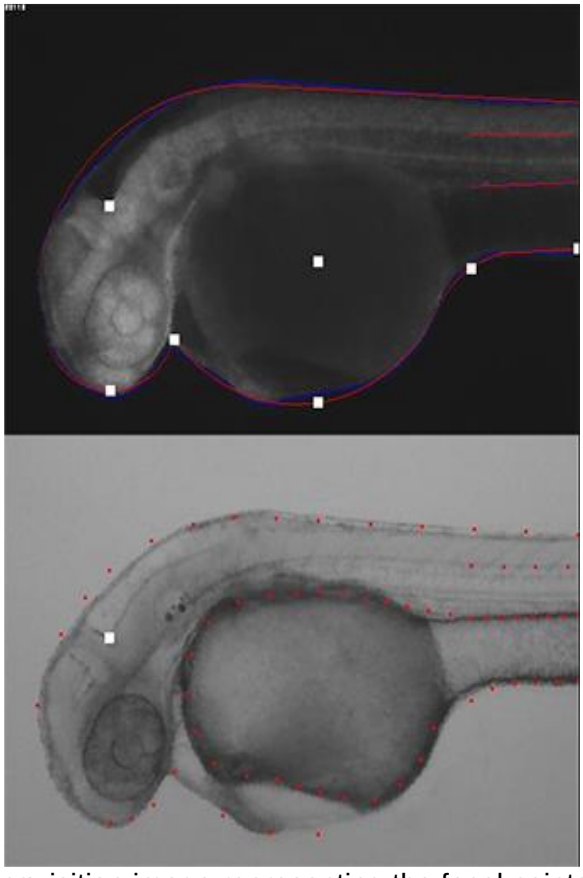


Figure 16: A Scan^R acquisition image representing the focal points used to identify the embryo and in conjunction with zMiner, assign specific sectors to various compartments of the embryo (Yolk, Eye, Skin, Brain, Cerebellum, Heart, Notochord and Spinal Cord).

Following imaging, the Scan^ acquired images are processed through zMiner: a computer program designed to both identify distinct regions within each embryo (Yolk, Eye, Skin, Brain, Cerebellum, Heart, Notochord and Spinal Cord), and to calculate relative fluorescent expression in each of these regions (Gehrig et al., 2009). As such, relative values for tissue specific fluorescent expression of both mCherry and CFP are obtained for each embryo for use in quantitative analysis of RNA expression between treatment groups as a representation of miRNA mediated degradation. mCherry expression is representative of relative target gene expression changes between treatment groups (expression is relative to degree of miRNA mediated mRNA repression and degradation) (Giraldez et al., 2006). CFP meanwhile, acts as a normalisation control for deviances in RNA injection amounts between embryos (expression is relative to injection amount due to no degradation within the embryo) (Gong et al., 2001), acting as an alternative to GFP.

3.6.3 RT-qPCR and mRNA quantification

3.6.3a RNA extraction and purification

Materials:

Shield stage RNA injected embryos

Trizol reagent RNA mini kit

Chloroform

100% ethanol

70% ethanol

Homogenizing kit

RNase free centrifuge tubes

As an alternative to fluorescent imaging, mRNA quantification of embryos can be performed through the use of real time reverse transcription quantitative polymerase chain reaction (Gehrig et al., 2009). For this procedure, the same RNA constructs are used for injection of the zebrafish embryos, and the breeding, injection and growth procedures are all the same as previously used, up to ~6hpf (shield stage) of embryo growth. Following dechorionation and clearing of dead and deformed cells, embryos at shield stage (~6hpf, if embryo stages vary between treatment groups, incubation temperature can be adjusted in an effort to normalise stage progression, with colder temperatures slowing progression and vice versa) are anaesthetised in MS-222 in preparation for RNA extraction via Trizol RNA purification.

First, the embryos are homogenised in 1mL of Trizol reagent and a tissue homogeniser and allowed to incubate at room temperature for 5 minutes. Next, RNA is isolated from the homogenate via the addition of 0.2 ml of chloroform and incubation at room temperature for 3 minutes after mixing. The solution is then centrifuged for 15 minutes at 12,000 g and 4°C, and ~600 µl of the colourless aqueous phase containing RNA is collected and transferred to a new centrifuge tube. Following this, 600 µl of 70% ethanol is added to complete phase separation.

Following phase separation, 700 µl of the RNA / 35% ethanol mixture is transferred to a spin cartridge within a centrifuge tube and further centrifuged at 12000 g for 15 minutes at 4°C. Subsequently, 700 µl of Trizol wash buffer I is added to the spin cartridge and centrifuged for 15s at 12000g and room temperature. The wash buffer I flow-through is discarded and 500 µl Trizol wash buffer II added with further centrifugation at 12,000 g and room temperature for 15s. Further centrifugation at 12,000 and room temperature is performed for 1 minutes to dry the spin cartridge membrane before elution of the column with 3 x 100 µl DNase free water and collection of the flow-through in the centrifuge tube. The tube is then centrifuged one final time at 15,000 g for 2 minutes at room temperature to obtain a final solution containing the isolated RNA.
The isolated RNA is then stored in 75% ethanol at -20°C until further use.

3.6.3b RT-qPCR

Materials:

BioRad iQ5 Light Cycler

RT Buffer

dNTP Mix

RNase Inhibitor

Reverse Transcriptase

DEPC-H_2O

PCR Buffer

PCR primers

TAQ polymerase

A reverse transcription quantitative polymerase chain reaction was used as a secondary measurement of *nploc4* mCherry expression differences in wild type and mutant embryos using a standard RT-qPCR protocol. First, multiple primers were designed and created for effective quantification of the gene of interest (mCherry linked *nploc4*). Primer pairs were designed targeting (a) *nploc4* mCherry 1, (b) *nploc4* mCherry 2 and (c) SF3A2 (relative control) (Table 5). The mCherry primer pairs are for use in quantifying *nploc4* gene expression, while the SF3A2 primer pair is for use as a relative control – targeting an invariant exogenous gene for comparison of relative expression change. Considerations need to be made alluding to the design of suitable RT-PCR primers. Suitable primer pairs should be designed so that the target amplicon is no longer than 200bp in length. A Tm of ~60°C is preferable, but importantly, each primer pairs Tm should not differ by more than ~3°C. In addition to these factors, it is best if primers span an exon-exon junction, to prevent amplification of contaminating genomic DNA, and best if the primer GC content is ~50-60% for product stability. Finally, the lower the self complementarity between primers the better, to minimise primer-dimer formation and lack of efficiency (Brownie et al., 1997).

Primer Name	Sequence (5-3)
mCherry 1 FP	ACCACCTACAAGGCCAAGAAG
mCherry 1 RP	ACTGTTCCACGATGGTGTAGTC
mCherry 2 FP	ACGGCGAGTTCATCTACAAGG
mCherry 2 RP	AGCCCATGGTCTTCTTCTGC
SF3A2 FP	GCGTTAGAGACCATCGACATCAATAA
SF3A2 RP	AGTGTGTGCAAGATAACTCCCCT

Table 3: Sequence structures of primer pairs for use in RT-qPCR quantification of *nploc4*

3.6.3c Reverse transcription

Reaction mixture assembled containing 1 µl *nploc4* WT/MUT RNA, 2 µl random decamers, 2 µl RT buffer, 4 µl dNTP mix, 1 µl RNase inhibitor, 1 µl Reverse transcriptase and 9 µl DEPC-H_2O. Mixture spun briefly and incubated in thermocycler at 44^oC for 60 minutes, and 92^OC for 10 minutes > Storage prepared cDNA.

3.6.3d RT-qPCR

PCR mixture made up containing (for each reaction): 2 µl cDNA product, 2.5 µl 10x PCR buffer, 1.25 µl relevant forward primer, 1.25 µl relevant reverse primer, 1.25 µl dNTP mix, 0.2 µl TAQ polymerase and finally dH_2O up to 20 µl. Reaction mixtures are assembled in triplicate for each primer pair (mCherry 1, mCherry 2 and SF3A2) and run in lidded PCR tube strips with the following PCR cycle:

PCR Cycle:	Temp (°C)	Time (s)
Cycle 1		
Step 1:	50	120
Cycle 2 (1x)		
Step 1	95	600
Cycle 3 (40x)		
Step 1	95	15
Step 2	60	60
Cycle 4 (81x)		
Step 1	55-95	30

Table 4: PCR protocol for RT-qPCR quantification of *nploc4*

Melting curve data was automatically collected and recorded in real-time throughout the PCR.

RESULTS

4.1 Selection of *nploc4* as a novel candidate for experimental validation of CAGE predicted miRNA targeting

Using *Danio rerio* RNA expression libraries created by Yao et al. (2014), and Wei et al. (2012) (detailed in methods), ideal candidates for experimental validation of CAGE predicted miRNA targeting were selected. Only genes that are expressed maternally, possess at least one CAGE predicted maternal miRNA target site and show an expression pattern consistent with a gene targeted by maternal miRNA are selected.

4.1.1 Identification of maternal miRNAs

RNA expression data was filtered to identify only miRNAs predominantly active during pre-MBT stages. This step was performed by academic collaborator and computational biologist Chirag Nepal of the University of Copenhagen. Six maternal miRNAs were identified as being predominantly active in maternal stages; let-7a, miR-1, miR-17a, miR-22a, miR-93 and miR-206 (table 2 & 3, below).

miRNA	miRNA expression (RPKM)							
	1 cell	16 cell	Other late stages					
dre-let-7a	**613**	**1014**	1207	1029	2363	2595	3100	25732
dre-miR-1	**855**	**268**	326	283	1080	674	228134	503222
dre-miR-206	**2772**	**100**	112	126	7605	71080	229877	142202
dre-miR-17a	**27**	**20**	11	17	240	816	2333	924
dre-miR-93	**46**	**44**	7	12	177	600	2070	1246
dre-miR-22a	**2886**	**3744**	1216	1442	2758	6409	20148	33976

Table 5: *Danio rerio* expression data by stage for maternally inherited miRNAs derived from the library of Yao et al. (2014).

miRNA	miRNA expression (RPKM)			
	256 cell	Other late stages		
dre-let-7a	**2373**	415	835	2852
dre-miR-1	**215**	584	620	46776
dre-miR-206	**365**	5979	749	120989
dre-miR-17a	**852**	659	46	4235
dre-miR-93	**1933**	3036	414	38836
dre-miR-22a	**14865**	6339	485	25130

Table 6: *Danio rerio* expression data by stage for maternally inherited miRNAs derived from the library of Wei et al. (2012).

4.1.2 Identification of candidate miRNA potential gene targets

Identification of potential gene targets of the six mentioned maternally active miRNAs was also performed by Chirag Nepal. The RNA library is filtered to select for genes showing an expression pattern consistent with maternal miRNA degradation (Figure 17).

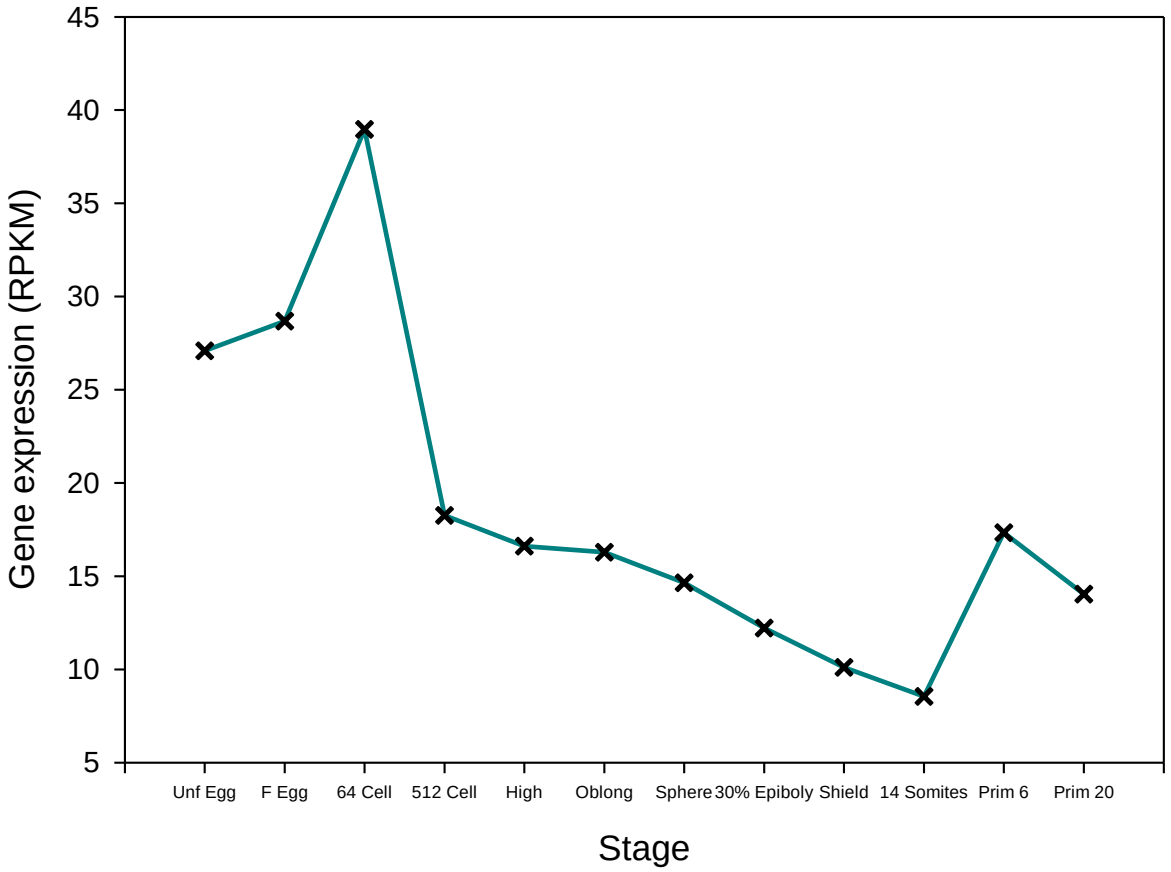


Figure 17: A graph representing an ideal expression pattern for a candidate maternal miRNA target gene (i.e. The SLC41A2 gene, targeted by miR-22). RPKM: Reads per kilobase per million ([# of mapped reads]/([length of transcript]/1000)/([total reads]/10^6).

Filtering of the library resulted in a potential 54 target genes that were expressed in all three maternal stages (≥5 tpm), and whose expression levels were higher (at least ≥1.5 fold in at least two pre-MBT stages) during Pre-MBT stages as compared to MBT stages (representing degradation during MBT). These 54 target genes were manually viewed (by me) on the UCSC Genome Browser to both confirm expression patterns throughout early development, and to identify ideal candidates with no or few conflicting CAGE predicted target sites (Figure 18).

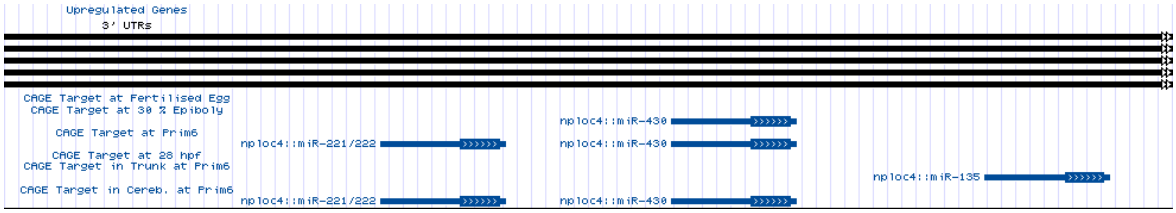


Figure 18: UCSC Genome Browser image showing *nploc4* 3'UTR and CAGE predicted miRNA target sites during early development. *nploc4* is a good candidate with relatively few miRNA targets in its 3'UTR at early stages. A: *nploc4* 3'UTR. B: miRNA CAGE predicted target sites.

4.1.3 *nploc4* as a CAGE predicted miRNA target for experimental validation

The *nploc4* gene on chromosome 12 of the zebrafish shows both a CAGE predicted miR-430 target site (Figure 18, above), and an expression pattern consistent with miR-430 degradation (Fig 19, below) and, As such, *nploc4* was chosen as the most suitable predicted miRNA target for experimental validation as a 'true' target site, as predicted by CAGE data.

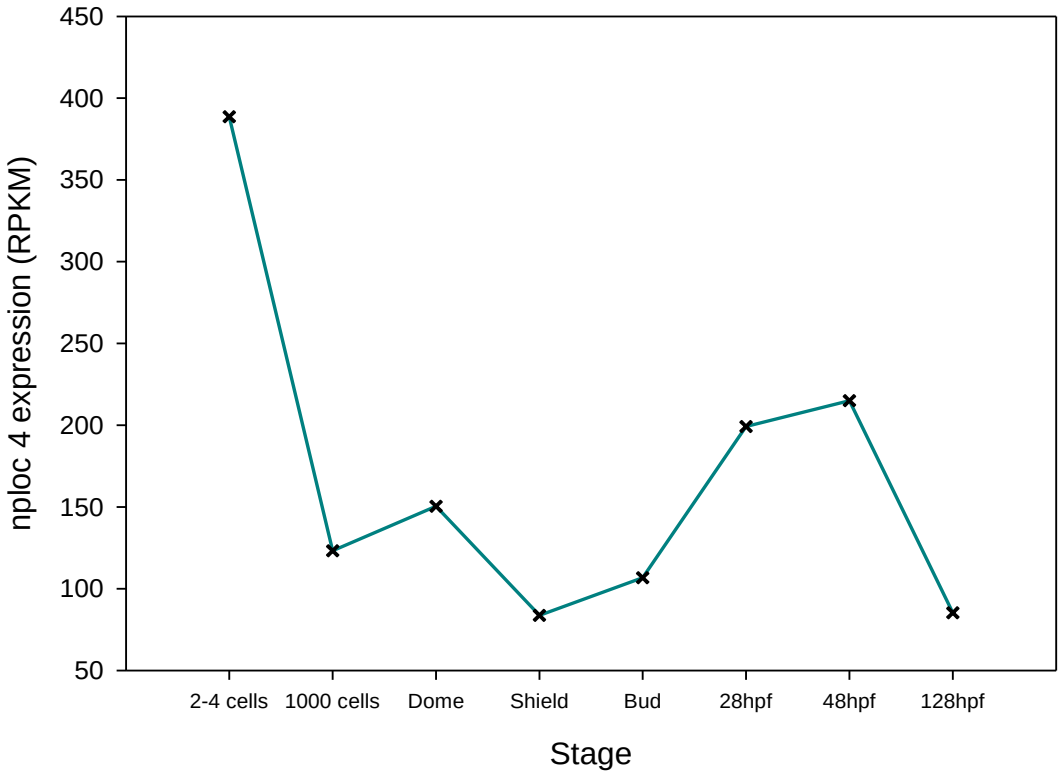
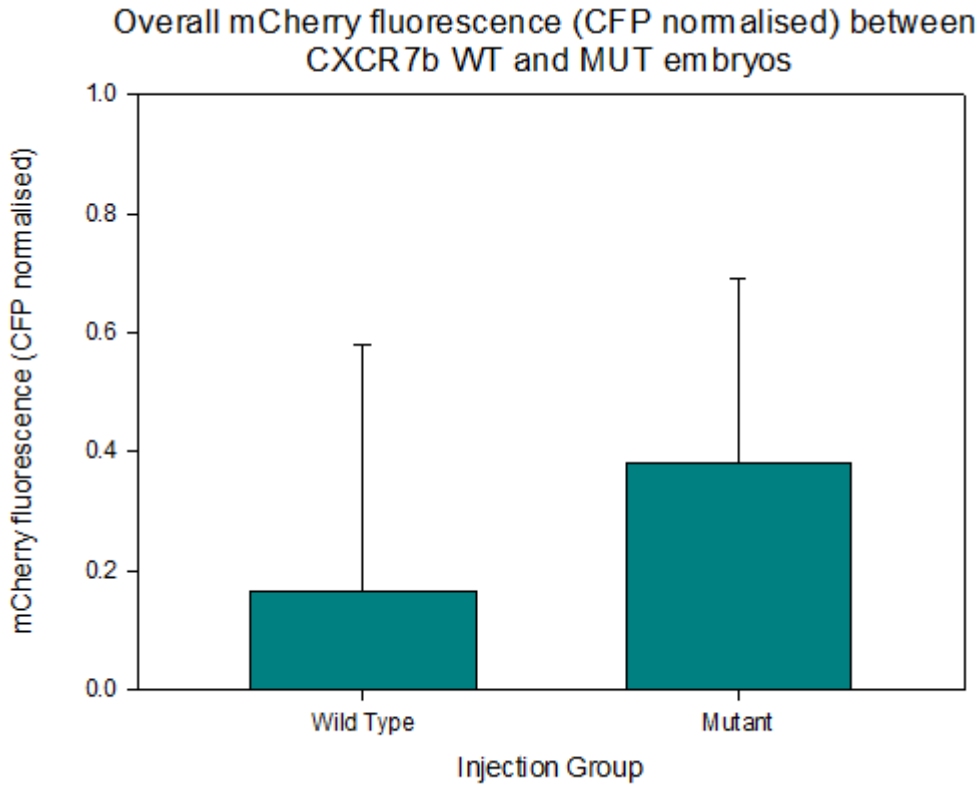


Figure 19: Graph representing zebrafish *nploc4* expression pattern at early developmental stages, derived from UCSC Genome Browser RefSeq data (Nepal et al. 2013). *nploc4* shows sharp degradation in the MBT transition, followed by a slow rise at later stages, consistent with miR-430 degradation.

4.2 Experimental validation of the *cxcr7b* gene as a miR-430 target via fluorescent imaging assay (proof of concept)

4.2.1 Overall *cxcr7b* mCherry expression

Prior to experimental validation of *nploc4* as a true CAGE predicted target site, *cxcr7b* was identified as an ideal previously experimentally validated miR-430 target to use as a proof of concept test to validate our own experimental method for determining a 'true' CAGE predicted miRNA target. Zebrafish embryos in *cxcr7b* WT and MUT treatment groups (injected with a functional miR-430 *cxcr7b* seed site and a mismatched non-functional *cxcr7b* seed site respectively) are examined at 72hpf for measurement of both mCherry and CFP fluorescence levels via automatic fluorescent microscopy as a measurement of mRNA expression (Figure 20).



Injection Group	n	Mean mCherry fluorescence	Std Dev
cxcr7b WT	54	0.17	0.42
cxcr7b MUT	40	0.38	0.31

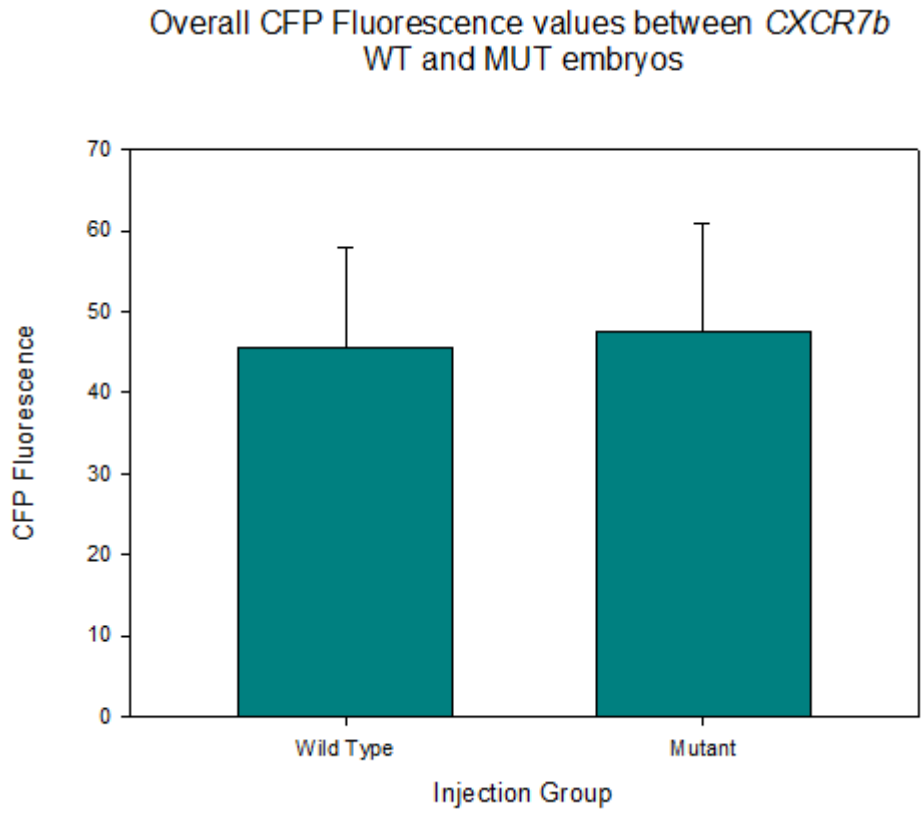
Figure 20: Bar chart and table representing differences and standard deviations in overall mCherry fluorescence throughout the whole embryo (sum of each segment average) between *cxcr7b* Wild Type injected (n=54), and *cxcr7b* mutated (n=40) embryos. mCherry expression is normalised against average CFP values.

Figure 20 represents relative overall mCherry fluorescence (i.e. *cxcr7b* expression) of both *cxcr7b* wild type injected and *cxcr7b* mutant injected embryos. Overall mCherry expression was calculated via the summation of each separate embryo segments (Yolk, Eye, Skin, Brain, Cerebellum, Heart, Notochord and Spinal Cord (Figures 27 & 28 in appendices)) relative average mCherry expression values as determined via RFP channel fluorescent imaging and zMiner quantification of expression (Gehrig et al., 2009). All values are expressed after normalisation against individual embryo CFP fluorescence values (RFP/CFP) to

account for variations in RNA injection amounts between embryos (CFP linked to invariant gene). The relative expression of *cxcr7b* is significantly increased (p = 0.007, Mann-Whitney U test) in the mutated embryos (those with mismatches induced within *cxcr7b* 3'UTR seed site to prevent miR-430 mediated suppression of translation and mRNA degradation) in comparison to the wild type (functioning miR-430 seed site) embryos.

4.2.2 Overall *cxcr7b* CFP expression as a normalisation control

As mentioned previously, CFP was used as a fluorescent marker linked to an invariant gene (SF3A2) co-injected with *cxcr7b* and unaffected by maternal mRNA degradation. This allowed its use as a normalisation control, to control against variations in RNA injection amounts between individual embryos. Representing mCherry expression as a ratio against CFP expression allowed for normalisation against any differences in amount of RNA injected. A comparison of CFP expression values between WT and MUT treatment groups was made, as any significant variations between the two groups may confer experimental flaws (Figure 21).



Injection group	n	Mean CFP fluorescence	Std Dev
cxcr7b WT	54	45.48	12.49
cxcr7b MUT	40	47.48	13.44

Figure 21: Bar chart and table representing differences in overall CFP expression values between *cxcr7b* wild type (n=54) and *cxcr7b* mutated (n=40) embryos.

As seen in Figure 21, CFP expression levels are relatively consistent between treatment groups, represented by mean CFP fluorescent levels of 45.48 in the *cxcr7b* wild type embryos and 47.48 in the *cxcr7b* mutant embryos. These results do not significantly differ from each other statistically (*p>0.05,* Mann-Whitney U test), reflecting relatively consistent RNA injections between the two treatment groups.

4.3 Experimental validation of the *nploc4* gene as a miR-430 target via fluorescent imaging assay and RT-qPCR analysis

4.3.1 Overall mCherry expression determination via fluorescence assay

Given the preliminary findings supporting our methodology for determining gene expression differences in the *cxcr7b* trial run, the methodology was repeated, with the use of a novel, CAGE predicted miR-430 target - *nploc4.* The same methodology was followed as with previous *cxcr7b* validation, but with *nploc4* WT and MUT injections as opposed to *cxcr7b*.

A comparison of mCherry fluorescence (normalised against CFP expression) between both *nploc4* WT and *nploc4* MUT injected embryos was made at 72hpf (Figure 22) to compare relative *nploc4* expression as a measurement of miR-430 mediated degradation (Figure 23).

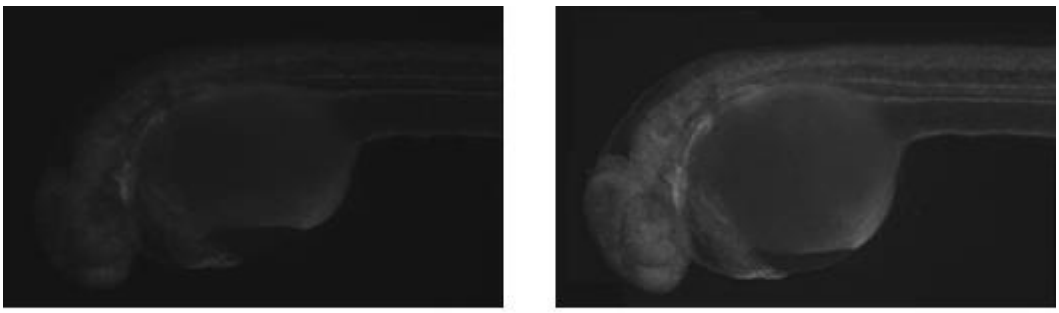
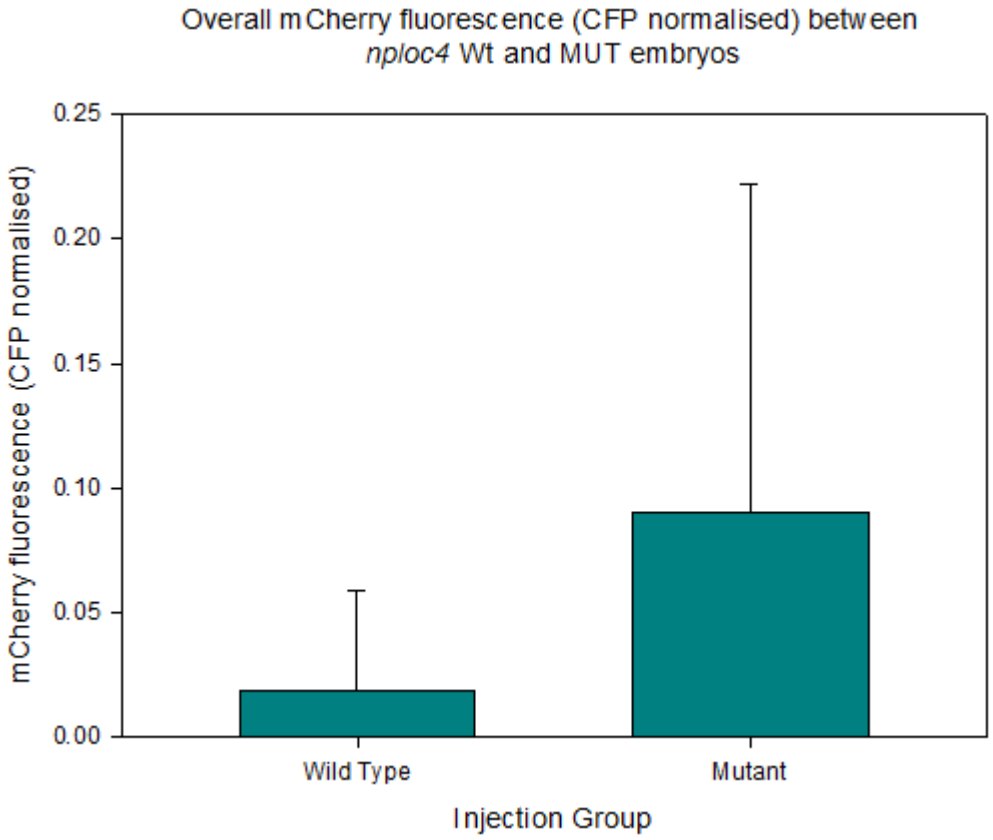


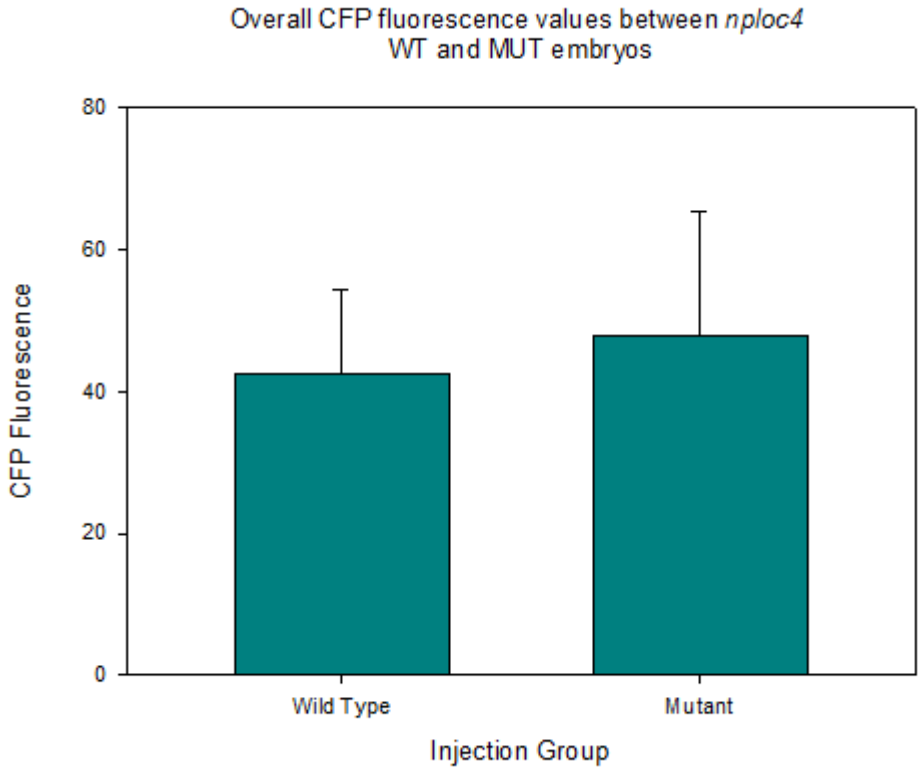
Figure 22: Olympus Scan^R fluorescent image comparing *nploc4* WT (left) and MUT (right) embryos, showing mCherry fluorescence at 72hpf.

The results obtained in the fluorescence assay of *nploc4* as a CAGE predicted target of miR-430 are similar to those of the *cxcr7b* assay. There is a statistically significant (p=0.002, Mann-Whitney U test) difference in mCherry fluorescence values (and consequently, *nploc4* expression) between *nploc4* WT and *nploc4* MUT embryos (Figure 23), with statistically significant reduced expression seen in the wild type embryos. CFP expression between WT and MUT *nploc4* embryos does not differ by a statistically significant amount (p>0.05) as seen in Figure 24.



Injection Group	n	Mean mCherry fluorescence	Std Dev
nploc4 WT	36	0.019	0.04
nploc4 MUT	41	0.091	0.13

Figure 23: Bar chart and table representing differences and standard deviations in overall mCherry fluorescence throughout the whole embryo (sum of each segment average) between *nploc4* Wild Type injected (n=36), and *nploc4* mutant (n=41) embryos. mCherry expression is normalised against average CFP values.



Injection Group	n	Mean CFP fluorescence	Std Dev
nploc4 WT	36	42.58	11.81
nploc4 MUT	41	47.91	17.59

Figure 24: Bar chart and table representing differences in overall CFP expression values between *nploc4* wild type (n=36) and *nploc4* (n=41) mutated embryos.

4.3.2 Overall *nploc4* expression quantification via RT-qPCR

As a follow up to the quantification of relative nploc4 expression between WT and MUT treatment groups via fluorescence assay, quantitative reverse transcription polymerase chain reaction (RT-qPCR) was subsequently performed as a secondary measurement to support the fluorescent assay findings. Two (successful) RT-qPCR runs were performed on *nploc4* WT and MUT RNA samples to determine relative *nploc4* expression between the two groups. mCherry linked *nploc4* expression was normalised against the invariant Beta-actin gene (Figures 25 & 26).

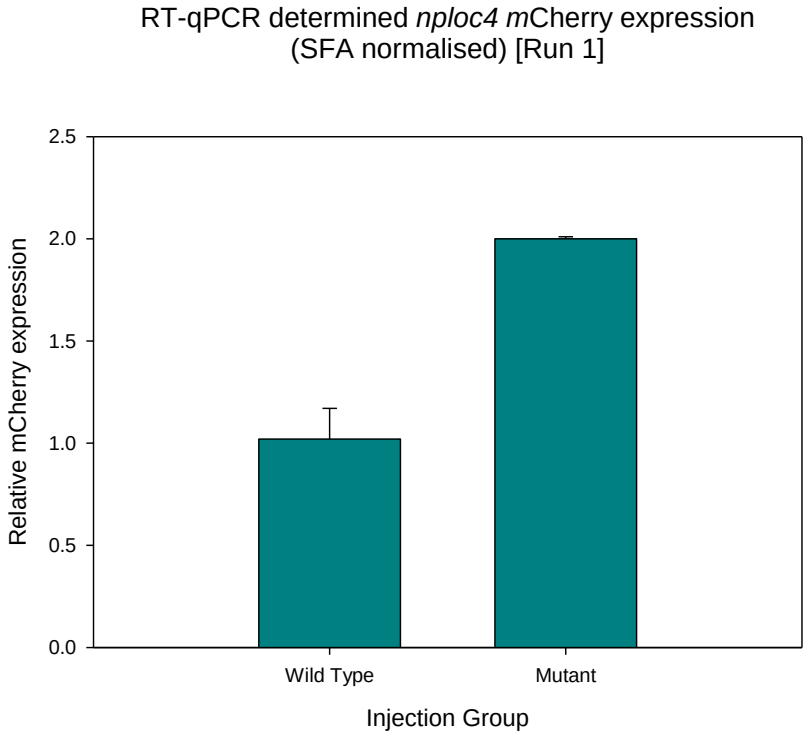


Figure 25 Bar chart representing mean and SD values for relative mCherry expression in both *nploc4* wild type (1.02, 0.15) and mutant (2.00, 0.01) embryos as determined by RT-qPCR (Run 1). mCherry expression values normalised against invariant exogenous gene detection (SFA).

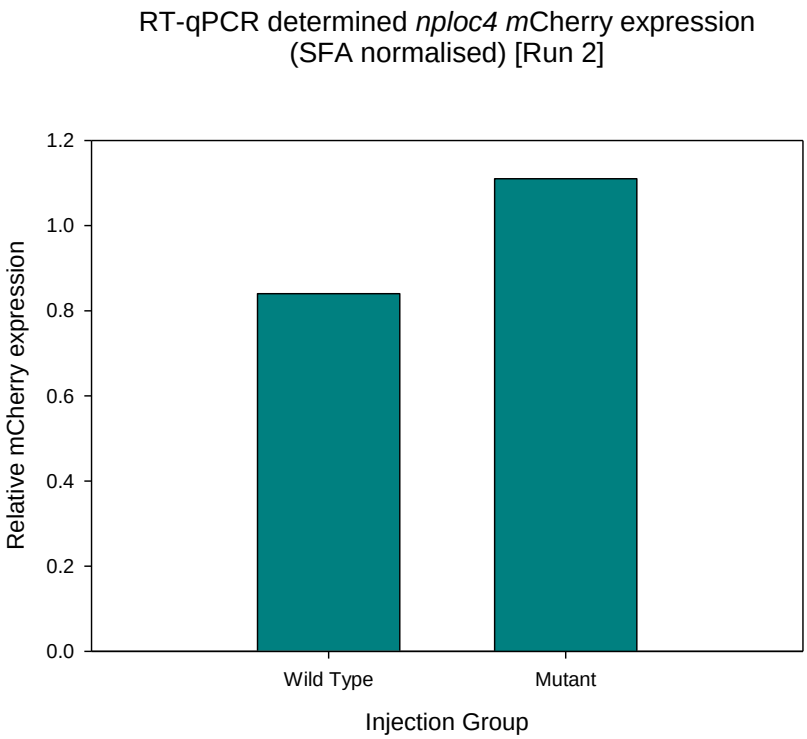


Figure 26: Bar chart representing mean and SD values for relative mCherry expression in both *nploc4* wild type (0.84, 0.00) and mutant (1.11, 0.00) embryos as determined by RT-qPCR (Run 2). mCherry expression values normalised against invariant exogenous gene detection (SFA).

The two RT-qPCR runs were performed to confirm that no drastic differences are seen in the direct measurement of *nploc4* expression in an embryonic RNA sample and as such, in-depth analysis of the PCR data was not performed (including CFP analysis). Figures 25 & 26 however, show results similar to those found in the *nploc4* fluorescence assay, with increased *nploc4* expression seen in the MUT group as compared to the WT group seen in both runs (each run consists of 15 WT and 15 MUT injected embryos), supporting the fluorescence assay findings.

Discussion & Conclusions

The main aim of this short laboratory project was to somewhat elucidate the potential for CAGE data to be used as a predictor for genuine miRNA target sites within the *Danio rerio* genome – with the possibility that, if successful, CAGE may be used in conjunction with current methods to reduce false-positive target site identification. Using CAGE data of early stage (72hpf) zebrafish generated by Chirag Nepal, we set out to identify a CAGE predicted maternal miRNA target site, and to confirm its authenticity as a true target experimentally. In the context of this project, our initial findings appear to support the potential for CAGE to do just that. *nploc4*, our identified target site candidate gene showed significantly reduced expression (mean fluorescence values of 0.019 WT vs 0.91 MUT (p<0.05 – Figure 22, and relative expression values of 1.02 WT vs 2.00 - Figure 25, and 0.84 WT vs 1.11 MUT – Figure 26 (p<0.05) via RT-qPCR) at 72hpf after injection into 1-cell zebrafish embryos when the wild type and intact miR-430 seed site was present, in comparison to the relatively increased expression seen in the mutant seed site. Coupled to this, expression values of the CFP linked SF3A2 invariant gene did not show a significant difference between WT and MUT treatment groups (mean CFP fluorescence of 42.58 WT vs 47.91 MUT (p>0.05) – Figure 24), giving some confidence in degradation differences being attributable to the seed site mismatches between the treatment groups and subsequent variation in miR-430 degradation, as opposed to being due to uncontrolled exogenous degradation.

These findings could suggest that *nploc4* was indeed a 'bona fide' miR-430 miRNA target, as predicted by Chirag Nepal's CAGE data. For this to be true however, we need to make a number of assumptions. Firstly, we must assume that miR-430 was the sole acting miRNA in causing *nploc4* degradation during the experiment. The consistent CFP fluorescence in a control supports this, but it is still possible that some other confounding miRNA, or a non-specific exogenous factor acted to affect the changes seen in the *nploc4* expression. This shows one problem of this project – testing just one potential miRNA target site (out of the thousands of candidates) is a very small first step to testing the usefulness of CAGE based miRNA target site prediction, especially considering the pre-existing problems with both false-positive and false-negative target site prediction. To show the value of CAGE (if any exists) based predictions, hundreds, and ideally thousands of novel CAGE identified miRNA target sites would have to be tested and experimentally validated as true sites. Only with a large sample size could you overcome the inherent chance involved with miRNA target site prediction – one result in and of itself could not be used meaningfully.

Further to the small sample size, it appears, in retrospect, that the computational data provided by Chirag Nepal as the basis for this project – namely, apparent CAGE tag enrichment relating to miRNA seed sites within 3' UTRs was flawed. When running further computational bioanalysis in an attempt to narrow parameters for identifying potential miRNA targets, Chirag found confounding results. The CAGE data analysis now showed CAGE tag enrichment for genes that should not be targeted by any currently identified maternal miRNAs in the

maternal stages of embryo development (source data unavailable). If this was not an artefact of incorrect analysis of the CAGE data, then it would suggest that the CAGE tags are not actually related to miRNA cleavage events or mi-RNA-mRNA interactions – considering that the genes showing enrichment shouldn't be undergoing miRNA degradation in the first place. Unfortunately, this contradicting data was only uncovered near to the end of the project, and any attempt to uncover the cause for such conflicting results was not possible for Chirag at the time. To further this, it is unfortunate indeed that near the end of the project the very data used in justification for testing CAGE as a predictive tool in miRNA targeting was shown to be flawed by Chirag, as this appears to be a novel hypothesis based on a finding unique within his own research – and not supported by other literature in the field of RNA genomics and computational biology.

In regards to the methodology of this project, previous work using CAGE in the mapping of promoter-enhancer interactions in zebrafish embryos (Gehrig et al., 2009), showed the effectiveness of high throughput automated zebrafish embryo imaging via the use of fluorescent microscopy and zMiner software to determine overall and segmented gene expression of particular genes of interest. The combination of automated embryo detection, automated spatial orientation of the embryo, automated 3d imaging of the identified embryo under multiple wavelength filters and the use of zMiner software to quantitatively determine and compare segmented fluorescence values from those images allowed the high-throughput comparison of hundreds of individual embryos in an objective manner. While this method does allow high-throughput imaging and fluorescent quantification of gene

expression, it does have some downsides. The custom software for detection and automated orientation of embryos within 96 well plates does not have much margin of error in the original orientation of embryos – it can only compensate for deviances from its 'model' embryo orientation (Figure 15) to a very small degree – which can result in 'failed detection' of up to 1/3 of a 96 well plate of embryos in some cases, whereby embryos simply are not 'detected' by the software. Further, in quantifying fluorescence expression within the embryos, and equating this to gene expression, relatively large standard deviations were seen, both in the *cxcr7b* and *nploc4* fluorescence experiments (i.e. mean fluorescence of 0.019 and SD of 0.04 in *nploc4* WT, and 0.091 mean fluorescence and SD of 0.13 in MUT – Figure 23). This suggests large variations in the fluorescent values quantified in individual embryos of the same group furing fluorescence assays, and may reflect a weakness in the methodology of fluorescence measurement via automated imaging. It should be noted that while only a perfunctory RT-qPCR was run on the embryos, much smaller SD's and variation was measured between samples (Figure 33, appendices) as compared to measurement of gene expression via the fluorescent assay.

In conclusion, it would be fair to say that the biggest benefit of this research project was in the practical experience gained in the laboratory setting – the introduction to zebrafish and their use in genomics (Schier, 2013, Cifuentes et al., 2010), and experience in the many procedures outlined in the methods section. In terms of advancing CAGE data, and 3' UTR CAGE tag enrichment as a predictor for miRNA target sites, this is hard to justify without concrete and stringent

computational analysis of the data to produce clear indications of true miRNA activity at a particular site, and importantly, at a particular time. It is only after such data can be reliably and consistently produced that it should be pursued as a potential predictive tool.

APPENDICES

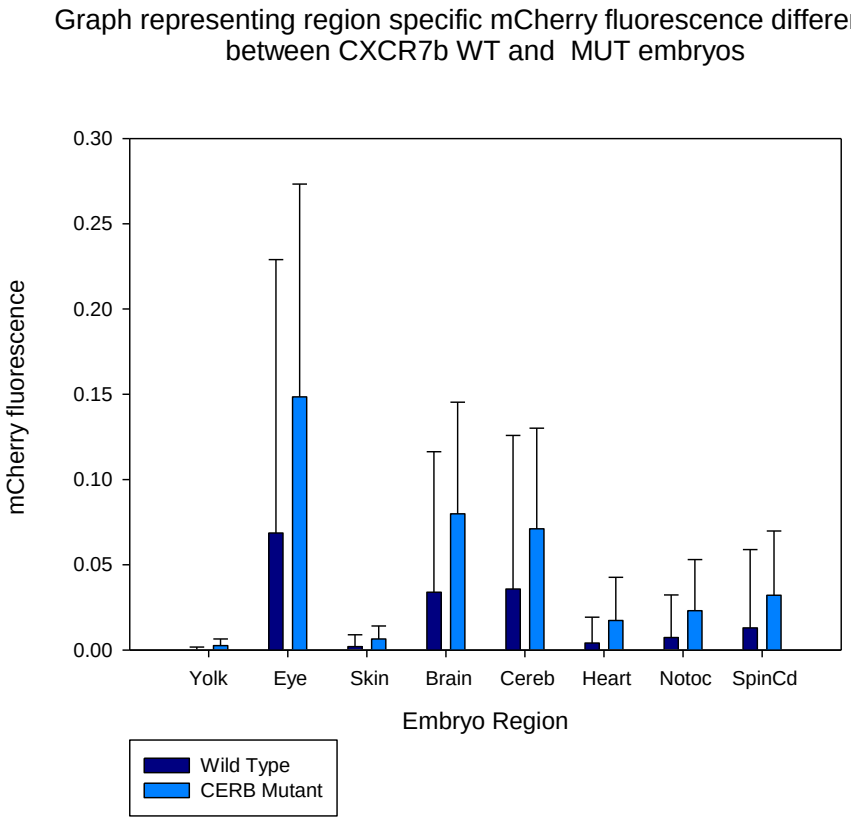


Figure 27: Bar chart representing average mCherry fluorescence (CFP normalised) values for individual segments within the zebrafish embryo. *cxcr7b* wild type and *cxcr7b* mutant embryos are compared after imaging at 72hpf.

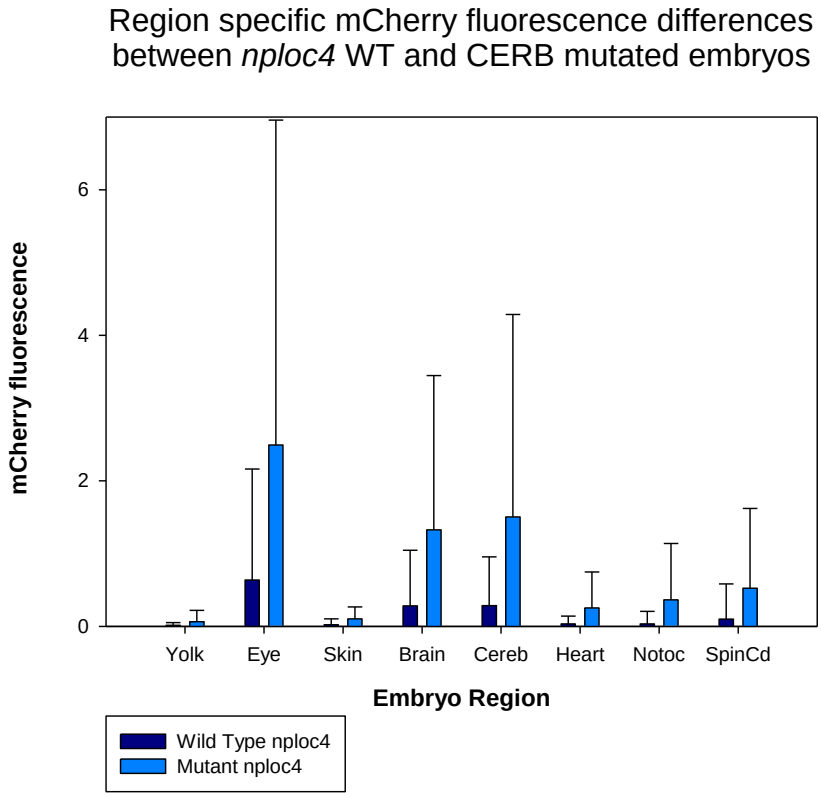


Figure 28: Bar chart representing average mCherry fluorescence (CFP normalised) values for individual segments within the zebrafish embryo. *nploc4* wild type and *nploc4* mutant embryos are compared after imaging at 72hpf.

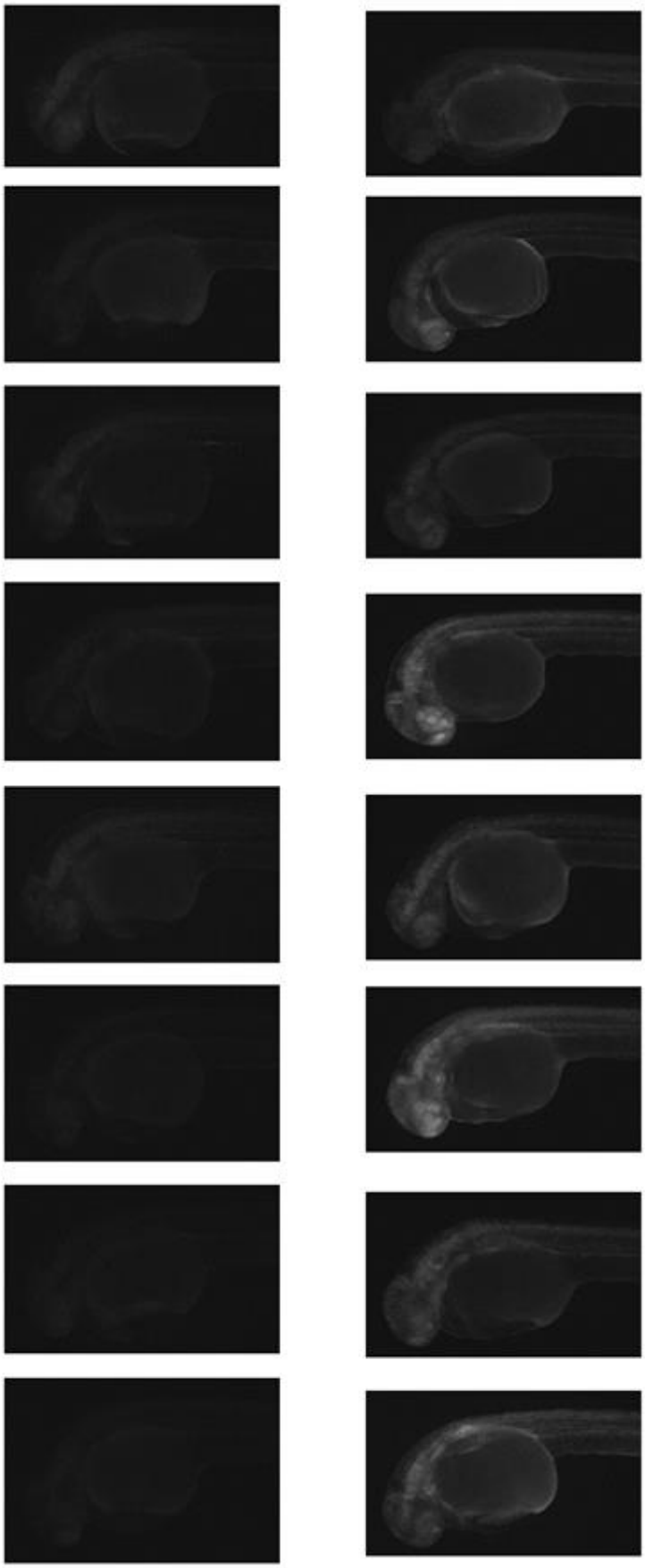


Figure 29: Side by side comparison of multiple *Cxcr7b* WT (left) and *Cxcr7b* MUT (right) embryos showing mCherry expression under an RFP filter at 72hpf.

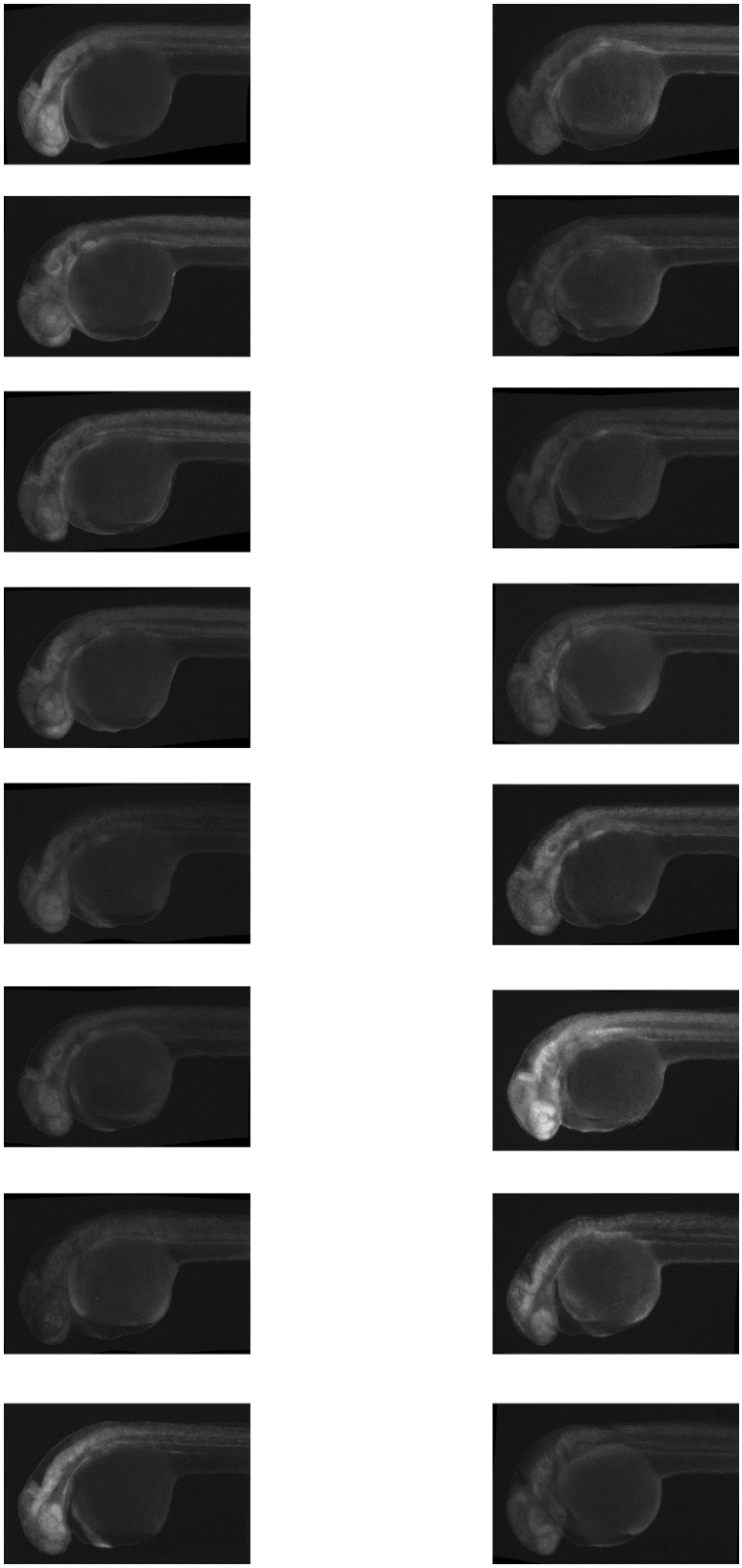


Figure 30: Side by side comparison of multiple *Cxcr7b* WT (left) and *Cxcr7b* MUT (right) embryos showing CFP expression under a CFP filter at 72hpf.

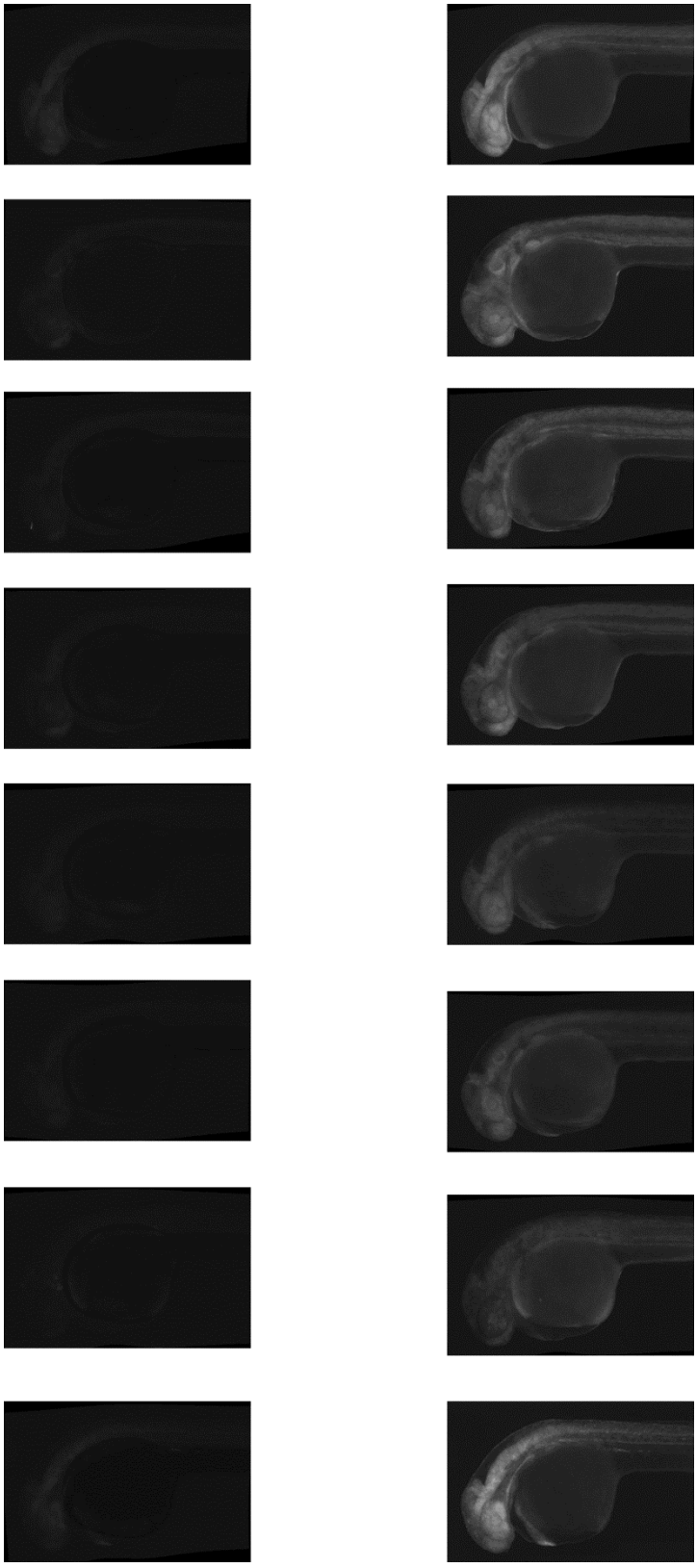


Figure 31: Side by side comparison of multiple *Cxcr7b* WT embryos showing both mCherry expression (left) and CFP expression (right) at 72hpf.

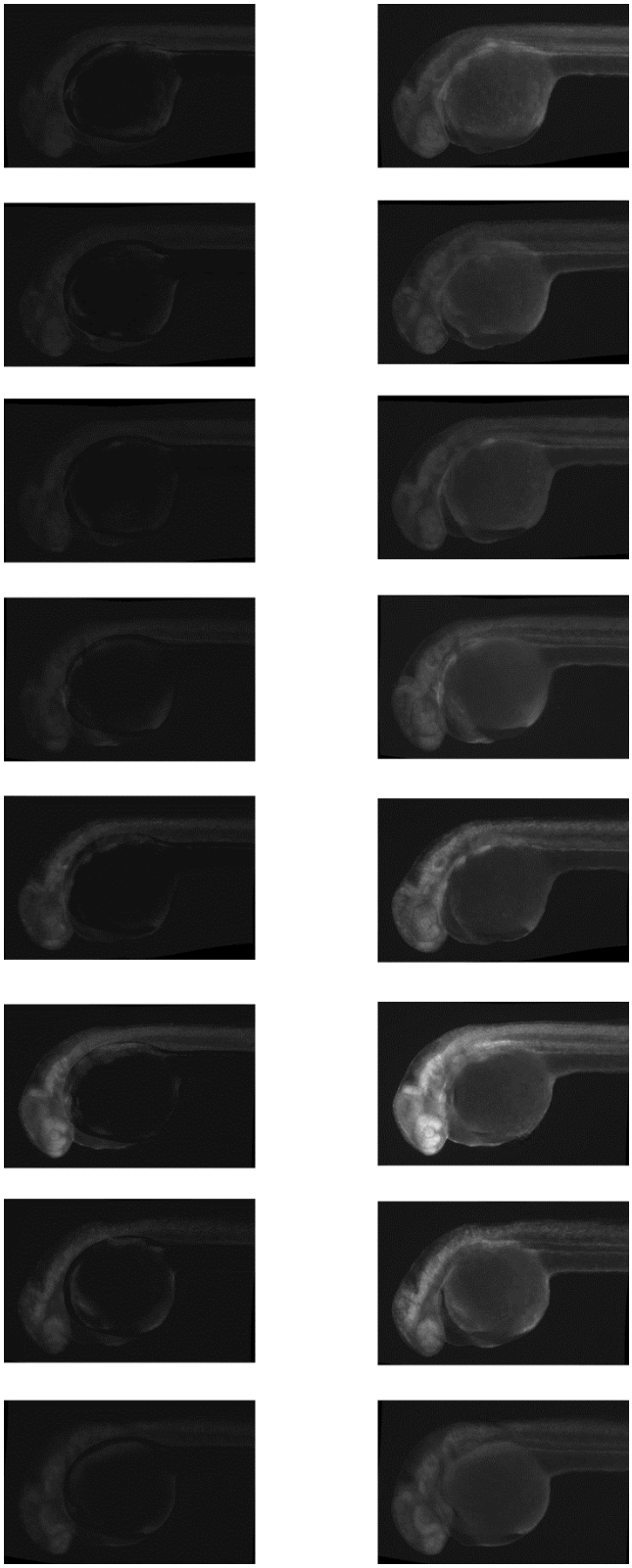


Figure 32: Side by side comparison of multiple *Cxcr7b* MUT embryos showing both mCherry expression (left) and CFP expression (right) at 72hpf.

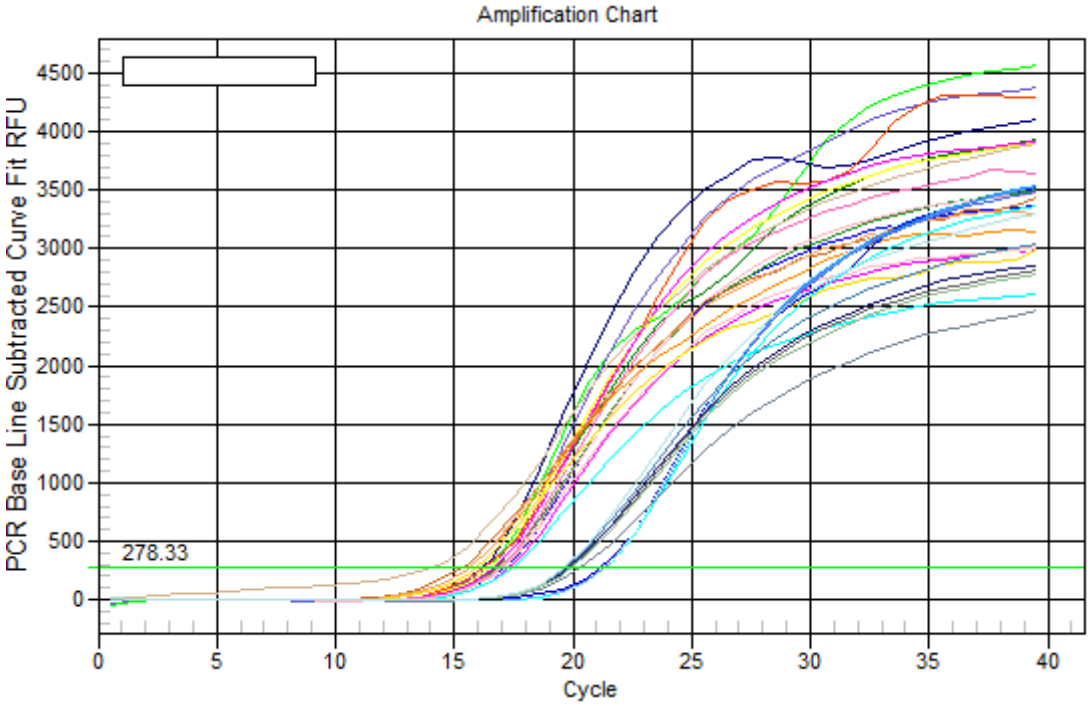


Figure 33: RT-qPCR standard curve for run 1 of the *nploc4* mCherry quantification. SYBR green fluor was used. Samples were loaded in triplicates with the following primers: mCherry 1, mCherry 2, Beta-actin (relative control) and SFA (input control).

6. REFERENCES

ABRAMS, E. W. & MULLINS, M. C. 2009. Early zebrafish development: It's in the maternal genes. *Current opinion in genetics & development,* 19**,** 396-403.

BARTEL, D. P. 2004. MicroRNAs: Genomics, Biogenesis, Mechanism, and Function. *Cell,* 116**,** 281-297.

BRENNECKE, J., STARK, A., RUSSELL, R. B. & COHEN, S. M. 2005. Principles of microRNA-target recognition. *PLoS biology.,* 3.

BROWNIE, J., SHAWCROSS, S., THEAKER, J., WHITCOMBE, D., FERRIE, R., NEWTON, C. & LITTLE, S. 1997. The elimination of primer-dimer accumulation in PCR. *Nucleic Acids Research,* 25**,** 3235-3241.

CARNINCI, P., KVAM, C., KITAMURA, A., OHSUMI, T., OKAZAKI, Y., ITOH, M., KAMIYA, M., SHIBATA, K., SASAKI, N., IZAWA, M., MURAMATSU, M., HAYASHIZAKI, Y. & SCHNEIDER, C. 1996. High-Efficiency Full-Length cDNA Cloning by Biotinylated CAP Trapper. *Genomics,* 37**,** 327-336.

CHANG, T. C. & MENDELL, J. T. 2007. microRNAs in vertebrate physiology and human disease. *Annu Rev Genomics Hum Genet,* 8**,** 215-39.

CIFUENTES, D., XUE, H., TAYLOR, D. W., PATNODE, H., MISHIMA, Y., CHELOUFI, S., MA, E., MANE, S., HANNON, G. J., LAWSON, N. D., WOLFE, S. A. & GIRALDEZ, A. J. 2010. A Novel miRNA Processing Pathway Independent of Dicer Requires Argonaute2 Catalytic Activity. *Science,* 328**,** 1694-1698.

CONNELLY, S. & MANLEY, J. L. 1988. A functional mRNA polyadenylation signal is required for transcription termination by RNA polymerase II. *Genes & development,* 2**,** 440-452.

DOENCH, J. G. & SHARP, P. A. 2004. Specificity of microRNA target selection in translational repression. *Genes Dev,* 18**,** 504-11.

EULALIO, A., HUNTZINGER, E. & IZAURRALDE, E. 2008. Getting to the Root of miRNA-Mediated Gene Silencing. *Cell,* 132**,** 9-14.

FINLEY, K. R., DAVIDSON, A. E. & EKKER, S. C. 2001. Three-color imaging using fluorescent proteins in living zebrafish embryos. *Biotechniques,* 31**,** 66-70, 72.

FLYNT, A. S. & LAI, E. C. 2008. Biological principles of microRNA-mediated regulation: shared themes amid diversity. *Nat Rev Genet,* 9**,** 831-842.

FRIEDMAN, R. C., FARH, K. K., BURGE, C. B. & BARTEL, D. P. 2009. Most mammalian mRNAs are conserved targets of microRNAs. *Genome Res,* 19**,** 92-105.

GAIDATZIS, D., VAN NIMWEGEN, E., HAUSSER, J. & ZAVOLAN, M. 2007. Inference of miRNA targets using evolutionary conservation and pathway analysis. *BMC Bioinformatics,* 8**,** 69.

GEHRIG, J., REISCHL, M., KALMAR, E., FERG, M., HADZHIEV, Y., ZAUCKER, A., SONG, C., SCHINDLER, S., LIEBEL, U. & MULLER, F. 2009. Automated high-throughput mapping of promoter-enhancer interactions in zebrafish embryos. *Nat Meth,* 6**,** 911-916.

GIRALDEZ, A. J., MISHIMA, Y., RIHEL, J., GROCOCK, R. J., VAN DONGEN, S., INOUE, K., ENRIGHT, A. J. & SCHIER, A. F. 2006. Zebrafish MiR-430 promotes deadenylation and clearance of maternal mRNAs. *Science,* 312**,** 75-9.

GLOVER, D. M. 2013. *Genetic Engineering - Cloning DNA*, Chapman and Hall.

GONG, Z., JU, B. & WAN, H. 2001. Green fluorescent protein (GFP) transgenic fish and their applications. *Genetica,* 111**,** 213-225.

GREGORY, R. I., CHENDRIMADA, T. P., COOCH, N. & SHIEKHATTAR, R. 2005. Human RISC Couples MicroRNA Biogenesis and Posttranscriptional Gene Silencing. *Cell,* 123**,** 631-640.

GRUN, D., WANG, Y. L., LANGENBERGER, D., GUNSALUS, K. C. & RAJEWSKY, N. 2005. microRNA target predictions across seven Drosophila species and comparison to mammalian targets. *PLoS Comput Biol,* 1**,** e13.

GUO, H., INGOLIA, N. T., WEISSMAN, J. S. & BARTEL, D. P. 2010. Mammalian microRNAs predominantly act to decrease target mRNA levels. *Nature,* 466**,** 835-840.

HAN, J., LEE, Y., YEOM, K. H., KIM, Y. K., JIN, H. & KIM, V. N. 2004. The Drosha-DGCR8 complex in primary microRNA processing. *Genes Dev,* 18**,** 3016-27.

HOSKINS, R. A., LANDOLIN, J. M., BROWN, J. B., SANDLER, J. E., TAKAHASHI, H., LASSMANN, T., YU, C., BOOTH, B. W., ZHANG, D., WAN, K. H., YANG, L., BOLEY, N., ANDREWS, J., KAUFMAN, T. C., GRAVELEY, B. R., BICKEL, P. J., CARNINCI, P., CARLSON, J. W. & CELNIKER, S. E. 2011. Genome-wide analysis of promoter architecture in Drosophila melanogaster. *Genome Res,* 21**,** 182-92.

JINEK, M. & DOUDNA, J. A. 2009. A three-dimensional view of the molecular machinery of RNA interference. *Nature,* 457**,** 405-12.

KAPRANOV, P., CHENG, J., DIKE, S., NIX, D. A., DUTTAGUPTA, R., WILLINGHAM, A. T., STADLER, P. F., HERTEL, J., HACKERMULLER, J., HOFACKER, I. L., BELL, I., CHEUNG, E., DRENKOW, J., DUMAIS, E., PATEL, S., HELT, G., GANESH, M., GHOSH, S., PICCOLBONI, A., SEMENTCHENKO, V., TAMMANA, H. & GINGERAS, T. R. 2007. RNA maps reveal new RNA classes and a possible function for pervasive transcription. *Science,* 316**,** 1484-8.

KERTESZ, M., IOVINO, N., UNNERSTALL, U., GAUL, U. & SEGAL, E. 2007. The role of site accessibility in microRNA target recognition. *Nat Genet,* 39**,** 1278-1284.

KHVOROVA, A., REYNOLDS, A. & JAYASENA, S. D. 2003. Functional siRNAs and miRNAs Exhibit Strand Bias. *Cell,* 115**,** 209-216.

KIM, V. N. 2005. MicroRNA biogenesis: coordinated cropping and dicing. *Nat Rev Mol Cell Biol,* 6**,** 376-385.

KODZIUS, R., KOJIMA, M., NISHIYORI, H., NAKAMURA, M., FUKUDA, S., TAGAMI, M., SASAKI, D., IMAMURA, K., KAI, C., HARBERS, M., HAYASHIZAKI, Y. & CARNINCI, P. 2006. CAGE: cap analysis of gene expression. *Nat Meth,* 3**,** 211-222.

LEE, Y., AHN, C., HAN, J., CHOI, H., KIM, J., YIM, J., LEE, J., PROVOST, P., RADMARK, O., KIM, S. & KIM, V. N. 2003. The nuclear RNase III Drosha initiates microRNA processing. *Nature,* 425**,** 415-9.

LEWIS, B. P., BURGE, C. B. & BARTEL, D. P. 2005. Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets. *Cell,* 120**,** 15-20.

LEWIS, B. P., SHIH, I. H., JONES-RHOADES, M. W., BARTEL, D. P. & BURGE, C. B. 2003. Prediction of Mammalian MicroRNA Targets. *Cell,* 115**,** 787-798.

LIESCHKE, G. J. & CURRIE, P. D. 2007. Animal models of human disease: zebrafish swim into view. *Nat Rev Genet,* 8**,** 353-367.

LU, M., ZHANG, Q., DENG, M., MIAO, J., GUO, Y., GAO, W. & CUI, Q. 2008. An Analysis of Human MicroRNA and Disease Associations. *PLoS ONE,* 3**,** e3420.

MAZIÈRE, P. & ENRIGHT, A. J. 2007. Prediction of microRNA targets. *Drug Discovery Today,* 12**,** 452-458.

MELTON, D., KRIEG, P., REBAGLIATI, M., MANIATIS, T., ZINN, K. & GREEN, M. 1984. Efficient in vitro synthesis of biologically active RNA and RNA hybridization probes from plasmids containing a bacteriophage SP6 promoter. *Nucleic Acids Research,* 12**,** 7035-7056.

NEPAL, C., HADZHIEV, Y., PREVITI, C., HABERLE, V., LI, N., TAKAHASHI, H., SUZUKI, A. M., SHENG, Y., ABDELHAMID, R. F., ANAND, S., GEHRIG, J., AKALIN, A., KOCKX, C. E., VAN DER SLOOT, A. A., VAN IJCKEN, W. F., ARMANT, O., RASTEGAR, S., WATSON, C., STRAHLE, U., STUPKA, E., CARNINCI, P., LENHARD, B. & MULLER, F. 2013. Dynamic regulation of the transcription initiation landscape at single nucleotide resolution during vertebrate embryogenesis. *Genome Res,* 23**,** 1938-50.

PARK, J.-E., HEO, I., TIAN, Y., SIMANSHU, D. K., CHANG, H., JEE, D., PATEL, D. J. & KIM, V. N. 2011. Dicer recognizes the 5[prime] end of RNA for efficient and accurate processing. *Nature,* 475**,** 201-205.

PRATT, A. J. & MACRAE, I. J. 2009. The RNA-induced Silencing Complex: A Versatile Gene-silencing Machine. *Journal of Biological Chemistry,* 284**,** 17897-17901.

PROJECT, A. C. S. H. L. E. T. 2009. Post-transcriptional processing generates a diversity of 5'-modified long and short RNAs. *Nature,* 457**,** 1028-32.

RAND, T. A., GINALSKI, K., GRISHIN, N. V. & WANG, X. 2004. Biochemical identification of Argonaute 2 as the sole protein required for RNA-induced silencing complex activity. *Proceedings of the National Academy of Sciences of the United States of America,* 101**,** 14385-14389.

REHMSMEIER, M., STEFFEN, P., HOCHSMANN, M. & GIEGERICH, R. 2004. Fast and effective prediction of microRNA/target duplexes. *RNA,* 10**,** 1507-17.

REINARTZ, J., BRUYNS, E., LIN, J. Z., BURCHAM, T., BRENNER, S., BOWEN, B., KRAMER, M. & WOYCHIK, R. 2002. Massively parallel signature sequencing (MPSS) as a tool for in-depth quantitative gene expression profiling in all organisms. *Brief Funct Genomic Proteomic,* 1**,** 95-104.

REINHART, B. J., SLACK, F. J., BASSON, M., PASQUINELLI, A. E., BETTINGER, J. C., ROUGVIE, A. E., HORVITZ, H. R. & RUVKUN, G. 2000. The 21-nucleotide let-7 RNA regulates developmental timing in Caenorhabditis elegans. *Nature,* 403**,** 901-6.

SANGER, F. & COULSON, A. R. 1975. A rapid method for determining sequences in DNA by primed synthesis with DNA polymerase. *J Mol Biol,* 94**,** 441-8.

SASSEN, S., MISKA, E. A. & CALDAS, C. 2008. MicroRNA--implications for cancer. *Virchows Arch,* 452**,** 1-10.

SCHIER, A. F. 2013. Genomics: Zebrafish earns its stripes. *Nature,* 496**,** 443-444.

SCHMIDT, E. V., CHRISTOPH, G., ZELLER, R. & LEDER, P. 1990. The cytomegalovirus enhancer: a pan-active control element in transgenic mice. *Molecular and Cellular Biology,* 10**,** 4406-4411.

SHIRAKI, T., KONDO, S., KATAYAMA, S., WAKI, K., KASUKAWA, T., KAWAJI, H., KODZIUS, R., WATAHIKI, A., NAKAMURA, M., ARAKAWA, T., FUKUDA, S., SASAKI, D., PODHAJSKA, A., HARBERS, M., KAWAI, J., CARNINCI, P. & HAYASHIZAKI, Y. 2003. Cap analysis gene expression for high-throughput analysis of transcriptional starting point and identification of promoter usage. *Proceedings of the National Academy of Sciences,* 100**,** 15776-15781.

SHKUMATAVA, A., STARK, A., SIVE, H. & BARTEL, D. P. 2009. Coherent but overlapping expression of microRNAs and their targets during vertebrate development. *Genes Dev,* 23**,** 466-81.

SOIFER, H. S., ROSSI, J. J. & SAETROM, P. 2007. MicroRNAs in Disease and Potential Therapeutic Applications. *Mol Ther,* 15**,** 2070-2079.

STATON, A. A., KNAUT, H. & GIRALDEZ, A. J. 2011. miRNA regulation of Sdf1 chemokine signaling provides genetic robustness to germ cell migration. *Nat Genet,* 43**,** 204-211.

STÜRZENBAUM, S. R. & KILLE, P. 2001. Control genes in quantitative molecular biological techniques: the variability of invariance. *Comparative Biochemistry and Physiology Part B: Biochemistry and Molecular Biology,* 130**,** 281-289.

VALENCIA-SANCHEZ, M. A., LIU, J., HANNON, G. J. & PARKER, R. 2006. Control of translation and mRNA degradation by miRNAs and siRNAs. *Genes Dev,* 20**,** 515-24.

VELCULESCU, V. E., ZHANG, L., VOGELSTEIN, B. & KINZLER, K. W. 1995. Serial analysis of gene expression. *Science,* 270**,** 484-7.

VILLEFRANC, J. A., AMIGO, J. & LAWSON, N. D. 2007. Gateway compatible vectors for analysis of gene function in the zebrafish. *Developmental Dynamics,* 236**,** 3077-3087.

WAKIYAMA, M., TAKIMOTO, K., OHARA, O. & YOKOYAMA, S. 2007. Let-7 microRNA-mediated mRNA deadenylation and translational repression in a mammalian cell-free system. *Genes Dev,* 21**,** 1857-62.

WEI, C., SALICHOS, L., WITTGROVE, C. M., ROKAS, A. & PATTON, J. G. 2012. Transcriptome-wide analysis of small RNA expression in early zebrafish development. *Rna,* 18**,** 915-29.

YAO, Y., MA, L., JIA, Q., DENG, W., LIU, Z., ZHANG, Y., REN, J., XUE, Y., JIA, H. & YANG, Q. 2014. Systematic characterization of small RNAome during zebrafish early developmental stages. *BMC Genomics,* 15**,** 117.

ZHAO, Y. & SRIVASTAVA, D. 2007. A developmental view of microRNA function. *Trends in Biochemical Sciences,* 32**,** 189-197.