The biological role of natural antisense transcripts in zebrafish development

Monica Jessica Piatek

Institute for Cell and Molecular Biosciences

Newcastle University

August 2016

Thesis submitted for the degree of Doctor in Philosophy at Newcastle University

Abstract

Non-coding RNAs (ncRNAs) have recently emerged as important regulators of gene expression and are now appreciated to control development, homeostasis and evolution of higher organisms. Natural antisense transcripts (NATs), a form of long ncRNA, are fully processed mRNA-like transcripts originating from the opposite strand of protein coding genes. NATs are believed to regulate the expression of the corresponding sense transcripts and may lead to transcriptional silencing. A few bi-directionally transcribed genes have been studied extensively. Ectopically expressed NATs have also been linked to diseases such as cancer, Alzheimer's and α-thalassemia. However, many questions remain about the specific biological roles and mechanisms of actions of NATs.

This study examines a specific sense/antisense system during zebrafish embryogenesis where the sense transcript, Slc34a2a, encodes a Na^+-dependent phosphate transporter. The Slc34a2a sense and antisense (Slc34a2a(as)) transcripts demonstrate a reciprocal relationship were both transcripts are expressed at a near equal amount at 48 hours post fertilization. Both transcripts locate to the same tissues: the pharynx, endoderm, primordial midbrain channel and primordial hindbrain channel. No relation was found between Slc34a2a(as) and the paralog Slc34a2b. Slc34a2a(as) shares a bidirectional promoter with Rbpja but does not undergo simultaneous divergent transcription.

Ectopic overexpression of Slc34a2a RNA during early embryogenesis caused failure to develop a cerebellum. Cerebellar loss lead to behavioral impairments including loss of balance and delayed reaction times. Upon investigating the molecular basis for the deleterious effect, overexpression of the protein was ruled out via injection of a frameshift containing Slc34a2a RNA. Injections of RNA fragments from the Slc34a2a locus determined that cerebellar development failure required Slc34a2a/Slc34a2a(as) RNA complementarity. The phenotype was rescued by Dicer knockdown, indicating short interfering RNA (siRNA) formation through RNAi. Mapping and extraction of synthesised and endogenous dsRNA with an anti-dsRNA monoclonal antibody was unsuccessful. Additional experiments are required to determine the exact function of Slc34a2a(as) in zebrafish embryogenesis though

knock-out and methylation studies. Further conditions need be examined to determine the specific circumstances under which J2 antibody would be able to extract dsRNA from whole fish.

Acknowledgments

First and foremost I would like to thank my supervisor, Dr Andreas (Andi) Werner for not only providing me with this opportunity, but for being so supportive and patient for these last few years. Numerous times he has gone out of his way to help and guide me, and for that I am eternally grateful.

I would also like to thank all my colleagues and collaborators; all of you have made this thesis possible. A big thanks goes out to all the zebrafish groups down at the Centre for Life for advising me, sharing your equipment and especially to Dr Bill Chaudry, for use of his zebrafish. I would also like to thank all the support staff of ICAMB, especially Louise Campbell and Stuart Davison for answering all my questions! I cannot forget Dr Tim Cheek, who has been there to aid me through some tough times.

This project was generously funded by Mr and Mrs Jacobson. The non-coding RNA field and I are thankful for your kindness.

My friends both in and outside of Newcastle: Thank you for the emotional support. Sarah, Andrew, Chris; couldn't have asked for better friends. Whether it was our coffee breaks, lunches together, field trips both local and not, it was always a blast to get away from the world of science (or sometimes hotly debate it!). Jaunty: the true gentlemen that was always there to hear me out and provide sound logic. Amy Fearn, you're not just the postdoc that got me through the final stages, but you've become a true friend! Finally, Mr Rory Gissing, thank you for always finding a way to make me smile (and just being super awesome). Thank you all for still always being around no matter how tired, grumpy or asocial my tendencies got.

Of course, my final thanks go out to my family. Mom, dad; without your support, guidance and patience, I would not be where I am today. Thank you for always being there whether it be lending an open ear, popping over for a visit or giving advice. Words cannot describe how thankful I am to have such wonderful parents as yourselves.

Table of contents

List of Figures

Chapter 1.

Chapter 3.

Chapter 4.

Chapter 5.

Chapter 6.

List of Tables

Abbreviations

bp	base pair
Ct	cycle threshold
DNA	deoxyribonucleic acid
dsRNA	double stranded RNA
Endo-siRNA	endogenous small interfering RNA
hpf	hours post fertilization
ISH	*in situ* hybridization
miRNA	microRNA
MO	morpholino oligonucleotide
NaCl	sodium chloride
NaOAc	sodium acetate
NaOH	sodium hydroxide
NATs	natural antisense transcripts
ncRNA	non-coding RNA
Nt	nucleotide
PBS	phosphate buffered saline
PCR	polymerase chain reaction
PTGS	post transcriptional gene silencing
RNA	ribonucleic acid
RNAi	RNA interference
rpm	rotations per minute
RT	reverse transcription

siRNA	small interfering RNA
SSC	saline sodium citrate
ssRNA	single stranded ribonucleic acid
TI	transcriptional interference
UTR	untranslated region
qPCR	quantitative (real time) polymerase chain reaction

Chapter 1. Introduction

1.1 Non-coding RNA

Deoxyribonucleic acid (DNA) is a molecule in which genetic information is stored as sequences of nucleotides. This genetic information holds instructions that are required for growth, development, reproduction and general operations. However, these functions are not directly carried out by the DNA itself but instead by ribonucleic acid (RNA) polymeric molecules and/or proteins. DNA, which is double stranded and organized into chromosomes within the nucleus of a cell, gets transcribed into single stranded RNA by a RNA polymerase. Three RNA polymerases exist in eukaryotes and each transcribes differing types of RNA: RNA polymerase I transcribes ribosomal RNAs (rRNA), RNA polymerase II transcribes small regulatory RNAs and future messenger RNAs (mRNA) whilst RNA polymerase III transcribes small RNAs like transfer RNAs (tRNA) (Clancy, 2008). Primary transcribed RNA is then usually modified in accordance to its future function. Post-transcriptional modification can including capping (addition of a 7-methylguanosine to the 5' end with the help of a phosphatase enzyme), splicing (introns are removed from the primary RNA transcript by a large spliceosome protein complex) and 3' processing (cleavage of the 3' end and addition of roughly 250 adenine residues to form a poly(A)tail) (Day and Tuite, 1998).

Information encoded in DNA that is to become protein is first transcribed into messenger RNA (mRNA) within the nucleus by RNA polymerase II. This pre-mRNA is then subjected to 5' capping, splicing and 3' polyadenylation prior to export into the cytoplasm for translation. All these post-transcriptional modifications ensure the stability of the RNA, prevent degradation and provide it with the necessary components for nuclear export and efficient translation (Day and Tuite, 1998). Translation, which consists of three phases (initiation, elongation and termination), can occur at ribosomes that are free-floating in the cytoplasm, or in the endoplasmic reticulum. During initiation, the ribosome assembles around the target mRNA by interacting with the 5' cap. A transfer RNA (tRNA), a short non-protein coding adaptor molecule, carries an amino acid to the ribosome as directed by the first

three-nucleotide sequence (codon) in the mRNA. During elongation, tRNA transfers an amino acid to the tRNA corresponding to the next codon (Clancy and Brown, 2008). As the ribosomes moves along the mRNA, working with tRNAs, reading the three-nucleotide codons, an amino acid chain is formed. Termination occurs once a stop codon is reached which itself does not code for an amino acid but instead initiates the release of the polypeptide from the ribosome (Day and Tuite, 1998; Clancy and Brown, 2008).

Non-coding RNAs (ncRNAs), otherwise occasionally known as non-messenger RNAs (nmRNAs), are RNA transcripts that do not get translated into proteins. The exact number of ncRNAs encoded within the human genome remain unknown, but bioinformatics and transcriptomic studies have found that they are highly abundant and most likely number in the thousands (Bertone *et al.*, 2004; Washietl *et al.*, 2005; Jeggari *et al.*, 2012). In the latest figures from Ensembl, 22,219 ncRNAs have been found in humans. As only ~1.5% of the mammalian genome encode proteins, it would appear as if the rest is either devoted to RNA-based regulatory circuitry or is just transcriptional noise (Venter *et al.*, 2001; 'Initial sequencing and comparative analysis of the mouse genome,' 2002). ncRNAs come in many forms; combinations of long/short, sense/antisense and functional/non-functional features make up a plethora of ncRNA transcripts as seen in Table 1.1. It is of no surprise that with so many diverse features, ncRNAs have a multitude of functions including involvement in translation, splicing, DNA replication, gene regulation, chromatin structure and genome defence (table 1.1). It has been hypothesised that it is this second tier of gene expression regulation via ncRNA that gives higher eukaryotes their complexity, cogitative ability and evolutionary plasticity (see works by John Mattick including (Mattick, 2001; Barry and Mattick, 2012; Amaral *et al.*, 2013)). Being responsible for many vital biological processes also means that structurally altered or mis-regulated ncRNAs are often implicated in disease. Many ncRNAs have been linked to ailments such as cancer (Matouk *et al.*, 2007), neurodegenerative disorders (review: (Salta and De Strooper, 2012)), Alzheimer's disease (review: (Tan *et al.*, 2013)), Prader-Willi syndrome (Galiveti *et al.*, 2014), autism (Velmeshev *et al.*, 2013) and even hearing loss (Mencia *et al.*, 2009). Even though many ncRNAs remain to have their function validated, it is still possible that many are non-functional products of spurious or pervasive transcription.

Table 1.1 Main features and functions of different types of non-coding RNAs. It is important to note that RNAs can fall in more than one category, for example microRNA can be classified as non-coding RNA and extracellular RNA. Abbreviations; bp: base pairs, nt: nucleotides.

Name	**Acronym**	**Features**	**Main functions**
Long non-coding RNA	lncRNA	Over 200 bp in length, can be transcribed from positive and negative strands	Epigenetic regulation (imprinting, X-chromosome inactivation, telomere maintenance) gene transcription regulation, post-transcriptional regulation (Wilusz *et al.*, 2009).
Transfer RNA	tRNA	79-90 nt in length in eukaryotes	Link between mRNA and amino acids during translation.
Ribosomal RNA	rRNA	Range in size from 121-5070 nt in eukaryotes	Constitute part of the ribosome, catalysing peptide bond formation during translation.
Small nucleolar RNA	snoRNA	60-250 nt in length. Depending on class, contain conserved boxes C, D or H.	Guide chemical modification of other RNAs through methylation and pseudouridylation (Scott and Ono, 2011).
Micro RNA	miRNA	Derive from hairpin structures, about 21-24 nt in length.	Function in RNA silencing and post transcriptional gene expression regulation.
Short interfering RNA	siRNA	Derive from double stranded RNA, 21-23 nt in length	Gene regulation through RNAi, chromatin structure.
Small nuclear RNA	snRNA	Usually around 150 nt in length	Processing pre-messenger RNA, telomere maintenance, regulate transcription factors.
Extracellular RNA	exRNA	Various lengths and orientations possible	Function not fully understood but linked to syntrophy, intercellular communication and cell regulation.
Piwi-interacting RNA	piRNA	26-31 nt in length. Form RNA-protein complexes through piwi proteins	Epigenetic and post transcriptional silencing of retrotransposons. Subspecies include rasiRNAs (Rajan and Ramasamy, 2014).
Small Cajal body RNA	scaRNA	Resemble snoRNAs	RNA maturation of snRNAs in Cajal bodies of the nucleus.

1.2 Natural antisense transcripts

Natural antisense transcripts (NATs) can be classified as a subcategory of long non-coding RNAs. NATs, which are naturally occurring fully processed mRNA-like transcripts originating from the opposite strand of protein coding genes, share sequence complementary with other endogenous RNA transcripts. Two types of NATs occur; cis-NATs and trans-NATs. Cis-NATs are transcribed from the same genomic loci as the cognate sense transcripts whilst trans-NATs are transcribed from a different genomic location with respect to their partnering sense transcript. Another type of closely related antisense transcripts exist; the so-called non-overlapping antisense transcript (NOTs) are transcribed from the same genomic locus as the sense transcript but after processing do not share sequence complementarity with the sense RNA (Figure 1.1). Thus, NOTs can act like NATs at the transcriptional level but no so after splicing. NAT biogenesis involves the same transcriptional machinery as the sense strand partner involving RNA polymerase II (Sigova *et al.*, 2013). Antisense and sense partner strands can be found in three different orientations as exemplified in figure 1.1; in complete overlap, tail to tail where the 3' ends of the transcripts overlap or head to head where the 5' ends align.

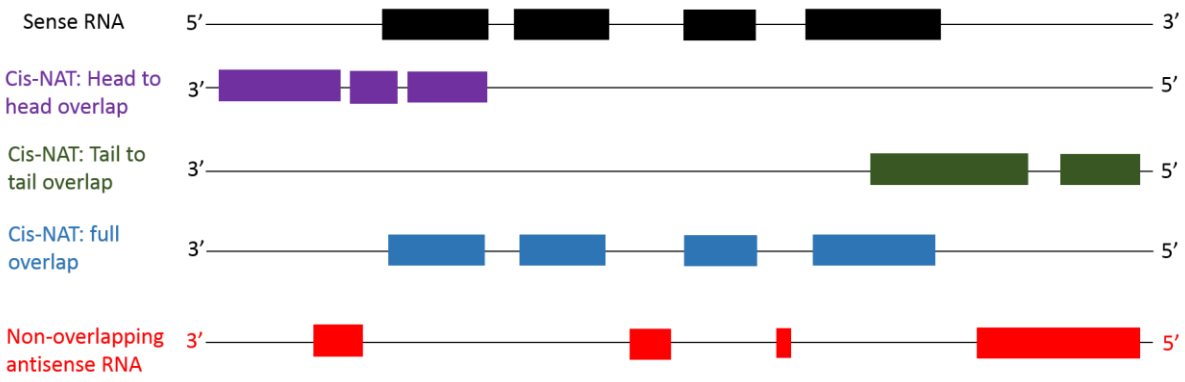


Figure 1.1 Representation of the different arrangements in which cis- natural antisense RNA transcripts (cis-NATs) can be found in relation to their cognate sense transcripts. Boxes represent exons. Antisense transcripts do not always align with the sense RNA after processing and these are known as non-overlapping transcripts (NOTs).

Comprehensive studies predict that in humans and mice, respectively, 40-72% of all transcriptional units show evidence of bi-directional transcription (Chen *et al.*, 2004; Katayama *et al.*, 2005b). A few of these bi-directionally transcribed genes have been studied in detail such as fibroblast growth factor (Baguma-Nibasheka *et al.*, 2012), α-globin (*HBA*) (Tufarelli *et al.*, 2001), β-secretase (Faghihi *et al.*, 2008; Faghihi *et al.*, 2010), low-density lipoprotein receptor-related protein 1 (*Lrp1*) (Yamanaka *et al.*, 2015) and brain-derived neurotrophic factor (*BDNF*) (Modarresi *et al.*, 2012) but the list is quickly growing. Several approaches were used to identify these antisense transcripts, either by annotating cDNA clones to genome sequences or by applying the BLAST algorithm between complete mRNA or expressed sequence tag (EST) sequences. To date, more than 2000 human cis-NATs, 2400 mouse cis-NATs and >1000 *Drosophila* cis-NATs are predicted to exist in the respective genomes (Lehner *et al.*, 2002; Cheng *et al.*, 2005; Katayama *et al.*, 2005a).

1.3 Natural antisense transcript functions

NATs have been shown to control development and homeostasis in bacteria, plants and mammals, and are thought to do so by regulating the expression of the corresponding sense transcript (see reviews (Faghihi and Wahlestedt, 2009; Esteller, 2011; Piatek *et al.*, 2016)). However, the concrete mechanisms by which this occurs is intensely researched and evidence so far has identified a number of possibilities. Current theories as to how NATs regulate gene expression are through transcriptional interference, RNA masking, double-stranded dependent mechanisms and methylation/ chromatin remodeling (figure 1.2). However, as the NAT field is still young, it is possible that unknown mechanism through which NATs function are yet to be determined. Nevertheless, since NATs can regulate gene expression to such an extent, it is not surprising that some ectopically expressed NATs have been linked to diseases such as cancer, Alzheimer's and α-thalassemia (Tufarelli *et al.*, 2003; Guo *et al.*, 2006). Moreover, viral antisense transcripts have been linked to disease progression, for example in HIV pathology (Saayman *et al.*, 2014). Therefore, understanding their individual functions and the mechanisms through which they act, is pivotal in generating new forms of therapeutics in this growing

antibiotic resistant world. A prime example includes that of Davet syndrome, which is a disorder caused by a heterozygous mutation in the SCN1A gene that codes for the pore-forming alpha subunit of the voltage gated sodium channel $Na_v1.1$ in the brain. In mice and non-human primates, knocking down the cis-NAT of SCN1A (SCN1ANAT) with oligonucleotide-based compounds (AntagoNATs), SCN1A levels increased which in turn alleviated the disorders symptoms (improvement in seizure phenotype and excitability of hippocampal interneurons) (Hsiao *et al.*, 2016).

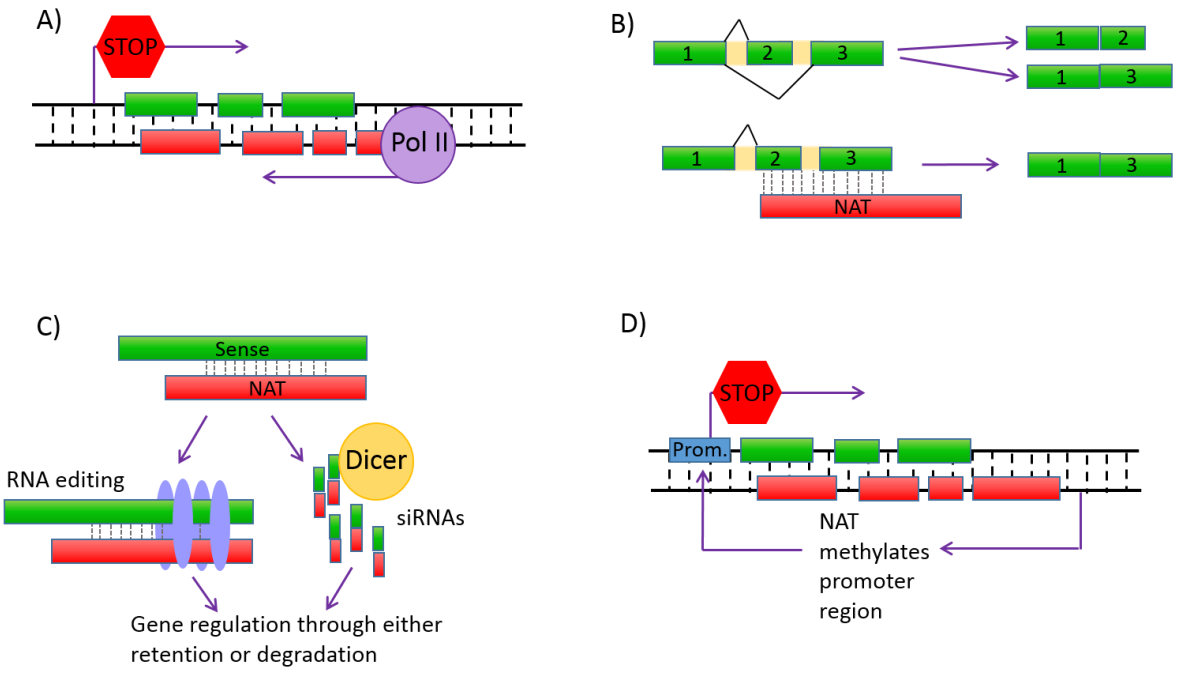


Figure 1.2 Representation of the different mechanisms by which cis-natural antisense transcripts (NATs) can regulate gene expression. A) Transcriptional interference; B) RNA masking; C) double strand dependent mechanisms such as RNA editing and RNA interference; D) antisense-induced methylation and chromatin remodeling. Figure modified from (Lavorgna *et al.*, 2004).

1.3.1 Transcriptional interference

Transcriptional interference (TI) is defined as the suppressive influence of one transcriptional process directly on a second transcriptional process *in cis (Shearwin et al., 2005)*. This does not include processes such as promoter methylation, binding of repressors to operators overlapping a promoter, chromatin remodeling or obstructing the elongation process of RNA polymerase by DNA-bound obstacles. Simply, the idea is that since the transcriptional machinery is large and 'bulky', it is therefore unlikely for two overlapping transcriptional units to be transcribed concurrently. Accordingly antisense transcription would lead to a decrease in transcription of the sense transcript. Potentially NAT induced TI may be widespread but to date very few established examples exist outside of *Neurospora crassa (Xue et al., 2014)*, *Saccharomyces cerevisiae* (Wang et al., 2014) and *Escherichia coli* (Sneppen et al., 2005). To date, two established examples exist in higher eukaryotes. In mice, maternally expressed *Murr1* is regulated by the antisense *U2af1-rs1* (Wang *et al.*, 2004) whilst in resting human peripheral blood T cells, a NAT whose start site is located 450 bases downstream regulates the transcription of the eukaryotic initiation factor (eIF) 2 alpha (Silverman *et al.*, 1992). Alternatively, experimental manipulation of systems, both intentional and unintentional, can result in TI (Proudfoot, 1986; Eszterhas *et al.*, 2002; Prescott and Proudfoot, 2002). Consequently, systems in which TI is visualized must therefore be tightly controlled and validated to confirm that TI is not an unintentional forced artifact of the system being handled.

1.3.2 RNA masking

This mechanism centers on the formation of duplexes between sense and antisense transcript pairs that then can provide a physical barrier against post-transcriptional interactions (Wight and Werner, 2013). For example, binding of an antisense transcript to the sense RNA may 'mask' critical regulatory motifs such as splice sites or miRNA-binding sites. This in turn would inhibit binding of other trans-acting

factors. Therefore any processes such as mRNA splicing, polyadenylation or translation along with RNA stability could be affected.

In higher eukaryotes, many examples of how NATs function in RNA masking have been established and validated. However only two examples have been determined thus far to be involved in alternative splicing. Human thyroid hormone receptor-α gene (*THRα*) has two isoforms, of which, only one is fully active (THRα1). The second spliceform, THRα2, binds the hormone but does not trigger a downstream response. The splicing of the primary transcript is influenced by the antisense transcript *RevErbAα* whose expression correlates with the synthesis of the inactive isoform (Munroe and Lazar, 1991; Hastings *et al.*, 1997). Secondly, the transcriptional repressor Zeb2, which down-regulates epithelial cadherin expression in turn affecting epithelial-mesenchymal transition, has its own expression regulated by a NAT. A large intron at the 5'end of the sense transcript (Zeb2 mRNA) is retained when the NAT is bound (Beltran *et al.*, 2008).

Protein and RNA stability can be both positively and negatively affected by NATs during RNA masking. In the case of Alzheimer's disease (AD), multiple NATs have been linked to disease progression and severity. One of these NATs works via a RNA masking mechanism. The antisense transcript for β-site amyloid precursor protein-cleaving enzyme 1 (*BACE1*), whose levels are elevated in the brains of AD patients, affects *BACE1* stability through two different processes both involving RNA hybridization; RNA duplex formation not only alters the secondary structure of BACE1 to increase stability but also masks a miRNA binding site which prevents BACE1 mRNA degradation (Faghihi *et al.*, 2008; Faghihi *et al.*, 2010). The stability of inducible nitric oxide synthase (*INOS*), an important gene in inflammatory diseases, is affected by its cis-NAT interacting with the 3' end of human antigen R (*ELAV1*) RNA (Matsui et al., 2008). Interestingly, an antisense transcript (aHIF), which is complementary to hypoxia inducible factor (HIF-1α) mRNA, destabilizes the protein-coding sense transcript by exposing AU-rich sequences which consecutively would lower tumours growth and progression (Uchida *et al.*, 2004).

1.3.3 Double stranded RNA-dependent mechanisms

Once transcribed and processed, NATs can form dsRNA with their sense partner. Depending on the cellular location, dsRNA can affect gene expression through RNA editing or gene silencing pathways such as RNA interference (RNAi). During RNA editing, long duplexes with near perfect complementarity are hyper-edited so that up to 50% of adenosines are deaminated by an enzyme of the ADAR protein family. Deamination can lead to either retention in the nucleus by the RNA binding protein p54nrb multiprotein complex or degradation by a cytoplasmic endonuclease (Kumar and Carmichael, 1997; Zhang and Carmichael, 2001). However, since RNA editing seems to preferentially target intronic stem loops, this suggests RNA editing occurs predominantly co-transcriptionally (Neeman *et al.*, 2005). Therefore, as majority of NATs are processed, mRNA/ endogenous NAT hybridization would occur post-transcriptionally (Bass, 2002). Consequently, only one example can be provided; mouse ubiquitin carboxyl-terminal hydrolase L1 (*Uchl1*), a gene involved in brain function and neurodegenerative disease, is controlled by antisense *Uchl1* which is found in a head to head orientation and contains two embedded repetitive sequences of SINEB1 and SINEB2 (Carrieri *et al.*, 2012). Various other reports fail to provide evidence of RNA editing of sense/antisense complementary regions in mammals (see review: (Lavorgna *et al.*, 2004)).

In RNAi, cytoplasmic dsRNA is recognized by the enzyme Dicer which is responsible for cleaving the dsRNA into short 21-23 nucleotide duplexes which in conjunction with a protein of the Argonaute family exhibit a regulatory function. RNAi was first discovered in *Caenorhabditis elegans* in 1998 (Fire *et al.*, 1998) but has since been reported in plants, fungi and certain higher eukaryotes (nematodes and fruit fly) (Cogoni and Macino, 2000; Waterhouse *et al.*, 2001). RNAi and Dicer dsRNA products, along with their pathways, are discussed in more detail in section 1.4.

1.3.4 Chromatin remodelling

NATs are also involved in epigenetic regulation of transcription through mechanisms such as DNA methylation and chromatin modification, which may in turn induce monoallelic expression. The exact mechanisms of action as to how NATs induce methylation or cause chromatin changes are still unclear, but several established occurrences have been discovered in higher eukaryotes including mice and humans.

In the case of imprinted genes, which tend to be found in clusters in human and mouse genomes, certain antisense RNAs recruit repressor complexes to inactivate chromatin. This is how insulin-like growth factor type 2 receptor (*Igf2r*) (by the antisense RNA *Airn*) (Sleutels *et al.*, 2002) and the imprinting control region of *Kcnq1* are regulated (by the antisense RNA *Kcnq1ot1*) (Pandey *et al.*, 2008). In mouse studies, when *Airn* was shortened, both alleles from both parents were active resulting in small sized offspring being born showing the importance of the antisense transcript (Sleutels *et al.*, 2002). Tumour suppressor gene p15 is too affected by a cis-NAT; p15-NAT expression led to heterochromatin formation and transcriptional silencing of the sense p15 transcript (Yu *et al.*, 2008). A NAT has also been implicated in X chromosome inactivation; *Tsix* (X inactive specific transcript, antisense) silences *Xist* (X inactive specific transcript) through chromatin structure remodelling in the *Xist* promoter region (Ohhata *et al.*, 2008).

As mentioned previously, many NATs have been implicated in disease progression and severity including Alzheimer's disease (AD). Low-density lipoprotein receptor-related protein (LRP) 1 and its spliced cis-NAT, LRP1-AS, regulate one another negatively. LRP1 is important in a variety of physiological processes such as cellular transport of cholesterol, transcytosis across the blood-brain barrier and in the systemic clearance of AD amyloid-beta. LRP1-AS, which is higher in AD patients, directly binds to high-mobility group box 2 (HMGB2) and inhibits the activity of HMBG2 to enhance SREBP1A-dependent transcription of LRP1. Therefore, the increase levels of LRP1-AS in AD patients cases lower transcription levels in LRP1, consequently progressing the disease (Yamanaka *et al.*, 2015). In certain cases of α-thalassemia, an inherited autosomal recessive anaemia, a chromosomal deletion disturbs the normal production of the α-globin gene (HBA2); an antisense HBA_2 RNA transcript is produced which in turn causes the methylation of the HBA2 CpG island consequently silencing the intact gene (Tufarelli *et al.*, 2003).

There too remain examples that are not fully understood, such as in the case of the antisense *Paupar* and Interleukin 1 antisense transcripts, which are known to affect or act as regulatory elements for protein-coding genes. The central nervous system expressed lncRNA *Paupar*, is located antisense and upstream to Pax6 which is required for eye and diencephalon specification. Knockdown of *Paupar* disrupts the normal cell cycle of neuroblastoma cells and induces neural differentiation. *Paupar*, however not only acts in cis to regulate Pax6 RNA levels but has also been found to regulate genes on multiple chromosomes by in part interacting with the Pax6 protein. *Paupar* binding sites have been located next to several protein-coding gene promoters and therefore it remains to be determined how the antisense *Paupar* works as a regulatory element (Vance *et al.*, 2014). In the case of Interleukin 1 (IL-1α), which is lowly expressed in resting macrophages, IL-1α and its antisense are found in cis and have a concordant relationship. IL-1α levels need to be tightly regulated since high levels can cause tissue damage. However, in the case of infection, IL-1α levels need to go up and this is done via the antisense transcript which, through a yet undermined pathway, acts by recruiting RNA polymerase II to the initiation site of IL-1α (Chan *et al.*, 2015).

1.4 RNA interference

RNA interference (RNAi), is a process that leads to post-transcriptional gene silencing (PTGS) where gene expression is inhibited in a sequence-specific manner. The RNAi pathway is found in most eukaryotes including zebrafish (Schyth, 2008), mice (Watanabe *et al.*, 2008) and humans (Schwarz *et al.*, 2002) but was first demonstrated in the worm, *C elegans* (Fire *et al.*, 1998). Essential to this process are the type III ribonucleases Drosha and/or Dicer which cleave the dsRNA precursors along with the endonuclease Argonaute 2 (Ago2) which is part of the RNA-Induced Silencing Complex (RISC) (Matranga *et al.*, 2005). The RNAi pathway is affected and slightly differs depending on the source of dsRNA and in which organism RNAi is taking place. Double stranded RNA can come from three places. It can be endogenous like the transcription of inverted repeats which in turn get processed into microRNA. Secondly, via Watson-Crick base pairing, endogenous

long dsRNA can form between RNA sharing sequence complementarity, such as in the case of sense/antisense transcripts which then get processed into endogenous short interfering RNAs (endo-siRNAs). Thirdly, dsRNA can be exogenous such as in the case of viral infection. RNAi plays a critical role in plant and insect resistance against viruses, however, in mammals, RNAi as an antiviral defence mechanism has yet to be validated (Burgess, 2013).

MicroRNA (miRNA) biogenesis involves RNA polymerase II transcribing long RNA precursors called pri-miRNAs. Majority of pri-miRNAs tend to come from introns of their pre-mRNA host genes (Lee *et al.*, 2004). The pri-miRNAs get post-transcriptionally processed in such that a 5' cap and poly(A) tail are added before processing by the microprocessor complex consisting of the RNAse III enzyme Drosha and double stranded RNA binding protein Pasha / DGR8 (Han *et al.*, 2006). The result is pre-miRNAS that are roughly 70 nucleotides long that fold perfectly back on themselves to form stem loops (hairpin like structures). Exportin 5 and RAN-GTP complex export the processed pre-miRNAs into the cytoplasm for further processing by the type III ribonuclease Dicer (Han *et al.*, 2006). Dicer produces miRNAs by cleaving the hairpin like structures into 21-23 nucleotide double stranded fragments, with two nucleotide overhangs on the unphosphorylated 3' ends (Billy *et al.*, 2001; Lee *et al.*, 2003). Dicer also activates the RISC ribonucleoprotein complex which contains proteins from the Argonaute (Ago) family. The dsRNA duplex is bound by the endonuclease Argonaute 2 (Ago2) and once within the active site of Ago2, one strand, the passenger strand, gets cleaved and released (Matranga *et al.*, 2005). The other single strand, known as the guide strand, remains bound to the RISC complex. It is hypothesized that strand selection is based on either the position of the stem loop in Dicer and/or the thermodynamic instability and weaker base pairing on the 5' end relative to the opposing strand (Schwarz *et al.*, 2003). The guide strand is then used to find its mRNA target through either complete or incomplete base pairing which in turn can then lead to either mRNA degradation, translational repression or deadenylation (Martin *et al.*, 2014). For example, by complete complementarity, miR16 causes the degradation of many unstable mRNAs like TNF alpha, microtubule-associated protein 7 (MAP7) and PR domain containing 4 (PRDM4); the miRNA/RISC complex binds the AU-rich element in the unstable mRNAs (Jing *et al.*, 2005; Yan *et al.*, 2013).

The endo-siRNA RNAi pathway is slightly different from miRNA RNAi. Firstly Drosha, Pasha and DGR8 are not involved. Dicer works to cleave both endogenous and exogenous long dsRNA into short ~21 nucleotide endo-siRNA duplexes. Ago2 and the RISC complex bind to the duplex and, just like with miRNAs, the passenger strand gets cleaved and released whilst the guide strand remains in the complex. The siRNA within the RISC complex hybridises to its target messenger RNA via complete sequence complementarity which in turn initiates mRNA cleavage by the Ago2 within the complex (Matranga *et al.*, 2005). However, unlike miRNAs which do not require perfect complementarity to their target mRNA to induce cleavage or translational repression, siRNAs do require perfect binding to their target to induce cleavage (Chu and Rana, 2006).

There is still speculation about where in the cell the endo-siRNA RNAi pathway occurs. In cell line studies, such as those using Hela cells, RISC and Ago proteins were found to be both in the cytoplasm and the nucleus (Berezhna *et al.*, 2006). Contradictory, *in vivo* studies claim that endo-siRNAs, just like microRNAs, get processed by Dicer and get incorporated into RISC only in the cytoplasm (Much *et al.*, 2016). Therefore, unlike the fairly well understood miRNA pathway, much work is left to be done in the field of endo-siRNA generation and function. However, endogenously occurring siRNA existence and importance can no longer be disputed. High levels of endo-siRNAs have been found in developing haploid sperm cells where NATs are predominantly expressed. These NATs form long dsRNA with their sense partners and are processed via Dicer into these endo-siRNAs where they function in genomic quality control (Wight and Werner, 2013; Yuan *et al.*, 2016).

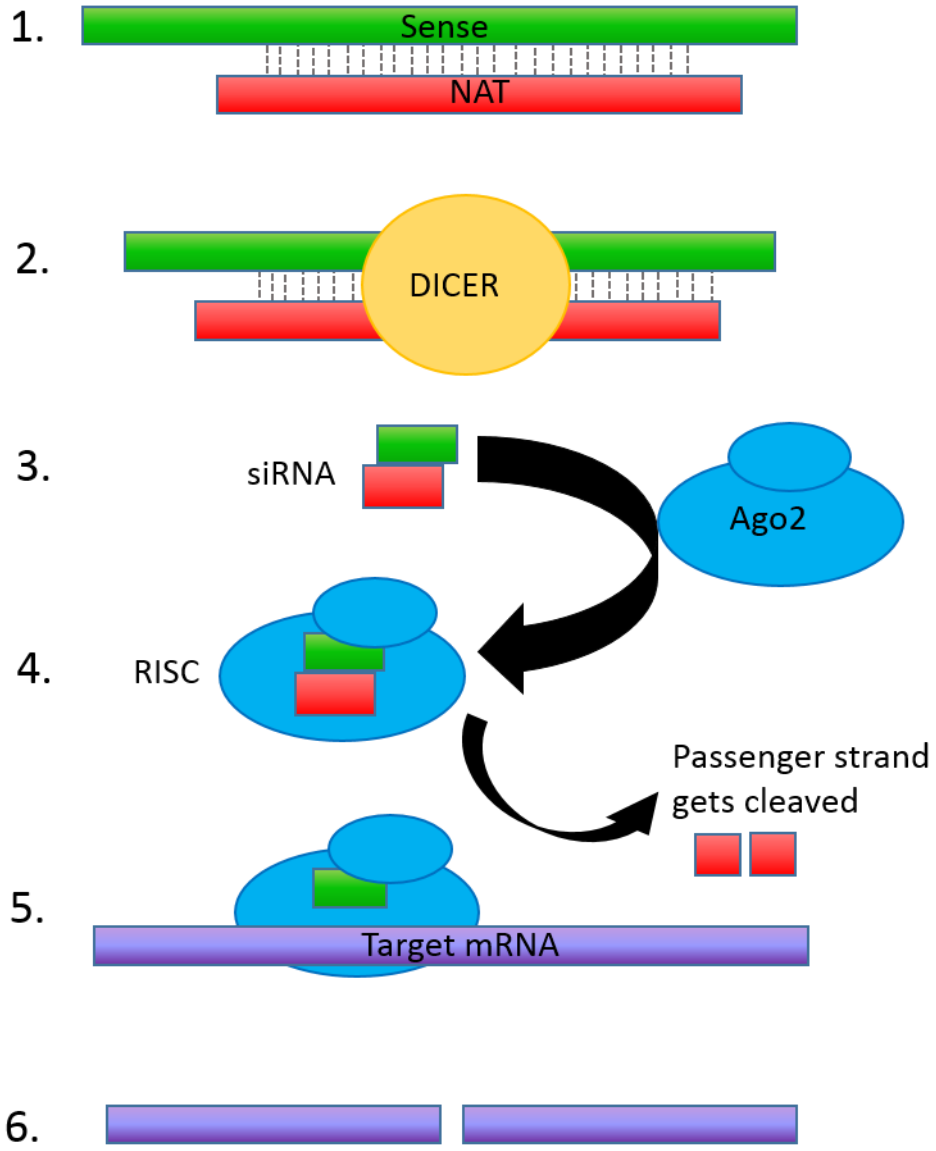


Figure 1.3 Representation of the main steps of RNA interference (RNAi) relevant to this thesis. Dicer recognises double strands formed between sense and antisense RNA pairs (1) (2) and cleaves the dsRNA into short interfering RNAs (siRNAs) of 21-23 nucleotides (3). The siRNAs are incorporated into the RNA-Induced Silencing Complex (RISC) (4) where the passenger strand gets cleaved. The guide strand is used to locate the target RNA (5) for degradation (6).

1.4.1 Interferon immune response

To fight infection, cells have developed a sensitive and intricate defence system that comprises of adaptive and innate immune responses which can ultimately lead to programmed cell death (apoptosis). A group of signalling proteins, interferons (IFN), are made and released by a host cell in response to invading pathogens including viruses, bacteria, parasites and even cancerous cells (Perry *et al.*, 2005; Fuertes *et al.*, 2013). During viral infection and in some cases bacterial infection, a major by-product of replication is double stranded RNA (Kawashima *et al.*, 2013; Hoffmann *et al.*, 2015). Once cytoplasmic dsRNA is detected by protein kinase R, a chain of events is activated including inhibition of protein synthesis and transcriptional induction of interferons, consequently leading to apoptosis of the infected cell. Cytokines, including interferons, communicate to neighbouring cells to increase their protective defence mechanisms. Thus, dsRNA-activated immune pathways are crucial to a healthy immune system (Kawashima *et al.*, 2013).

The role of RNAi in plants, nematodes and arthropods as an antiviral response is well established. However, the purpose of RNAi in higher eukaryotes such as mammals, remains controversial. For example, infection of mammalian somatic cells with different viruses followed by deep sequencing of small RNAs failed to find evidence of virus-derived siRNAs (Parameswaran *et al.*, 2010). Nevertheless, the RNAi pathway has become a powerful tool in discovering gene function in higher eukaryotes. Studies have found that siRNAs can induce the interferon response pathway both *in vivo* and *in vitro* (Hornung *et al.*, 2005; Judge *et al.*, 2005; Sioud, 2005). Therefore, if RNAi is to be further used within molecular and medical sciences, understanding how dsRNA can initiate or bypass the immune system is vital. To date, it appears as if siRNA activation of immune cells depends on a combination of siRNA sequence and cell type. For example, toll-like receptor 3 (TLR3) is involved in a siRNA induced interferon response, but not all immune cells express TLR3 (Karikó *et al.*, 2004). It also appears that siRNA under 30 bp fail to activate protein kinase R (PKR). When PKR is activated by dsRNA, it leads to a global inhibition of protein synthesis (Judge *et al.*, 2005; Sioud, 2005). However, with the recent establishment of endogenous dsRNA existence, such as sense/antisense RNA pairing, it begs to questions how not all dsRNA activates an innate immune response.

1.5 The sodium phosphate cotransporter family

The extracellular concentration of phosphate (P_i) needs to be tightly regulated and maintained in mammals as it is important for healthy body function; from bone formation to intercellular cell signalling and DNA formation. There are different sodium-phosphate cotransporters; the Slc17 family also known as Type I sodium-phosphate cotransporters, the Slc34 family known as Type II and Slc20 family which is Type III (Alexander *et al.*, 2015). There are also other families of cotransporters that can influence P_i concentration such as members of the Slc25 and Slc37 family (Alexander *et al.*, 2015). The RNA transcripts of interest for this study come from Slc34 family genes. Topological modelling of proteins from this family have been found to have eight transmembrane domains with both C- and N- termini being in the cytoplasm (Murer *et al.*, 2004). The Slc34 sodium phosphate cotransporter family in humans and mice comprises of NaPi-IIa, NaPi-IIb and NaPi-IIc (table 1.2) (Alexander *et al.*, 2015). In the zebrafish genome, there are four Slc34 genes; Slc34a1a (chromosome 14), Slc34a1b (chromosome 21), Slc34a2a (chromosome 1) and Slc34a2b (chromosome 23). Slc34a2a and Slc34a2b are occasionally otherwise known as NaPi-IIb1 and NaPi-IIb2 respectively. All these members transport exclusively phosphate ions in a sodium-dependent manner and are located at the apical surfaces of epithelia, predominantly in the intestine and renal proximal tubules (Murer *et al.*, 2004).

Table 1.2 Members of the Slc34 Type II sodium phosphate cotransporter family.

Protein name	NaPi-IIa	NaPi-IIb	NaPi-IIc
Human gene name	SLC34A1	SLC34A2	SLC34A3
Substrates	Inorganic phosphate/ 3 Na ions	Inorganic phosphate /3 Na ions	Inorganic phosphate /2 Na ions
Charge transfer	Electrogenic	Electrogenic	Electroneutral
Location of expression	Kidney, osteoclasts,	Small intestine, lung, testis, liver, secreting mammary gland	Kidney

The gene of interest for this PhD is Slc34a2, which is well conserved in many organisms including *C. elegans*, *Xenopus laevis*, mice and humans (Werner and Kinne, 2001). The model organism for this study, zebrafish (*Danio rerio*), a stenohaline freshwater teleost, expresses two paralogs of the Slc34a2 cotransporter; Slc34a2a and Slc34a2b (Nalbant *et al.*, 1999; Graham *et al.*, 2003). Repetition of the two loci as a result of a partial genome duplication was confirmed by the recent sequencing of the zebrafish genome. At the amino acid level, the two proteins are 66% identical (Graham *et al.*, 2003). However, the main interest of this work lies in the natural antisense transcript that is transcribed from the Slc34a2a gene locus. Remarkably, even though the genomic structure of the Slc34a2a sense transcript is conserved amongst species, the antisense is not. The Slc34a2a antisense transcript has been found in zebrafish, flounder, mice and humans, but in each case, the NAT is spliced differently giving various lengths and areas of sequence complimentary with the sense strand (figure 1.4) (Piatek *et al.*, 2016).

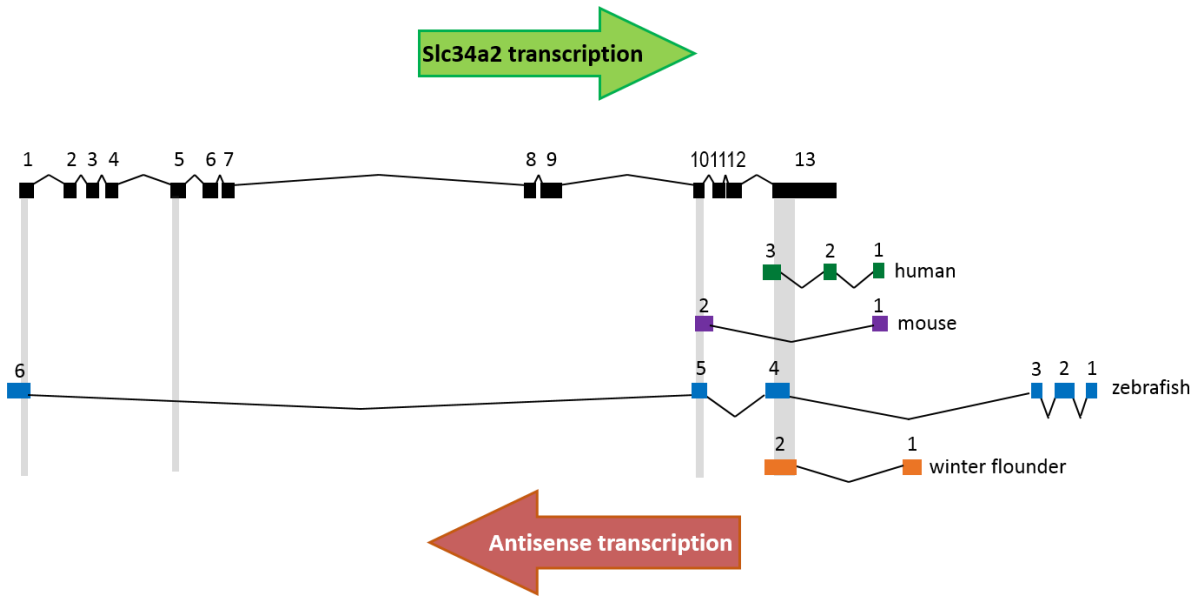


Figure 1.4 Schematic representation of the sodium phosphate cotransporter gene Slc34a2 and its antisense transcript in various vertebrates. Transcript length and exon sizes (boxes) are not to scale. Even though the sense transcript is conserved, the antisense is not. For each antisense transcript, there is at least one region of sequence complementarity (highlighted in grey). Figure modified from (Carlile *et al.*, 2009).

Even though transcribed from the same region of chromosome 1, the sizes of Slc34a2a and Slc34a2a(as) transcripts differ significantly. The sense transcript contains 2,607 bases (encoding a protein of 631 amino acids) whilst the antisense is about half the size at 1,351bp. The Slc34a2b transcript is 2,048bp encoding a protein of 642 amino acids. The Slc34a2a(as) transcript has exonic regions spanning all the way to the adjacent protein Rbpja but they themselves do not overlap (discussed further in section 1.8.1) (figure 1.5). There are 229 base pairs between the first exon of Slc34a2a(as) and the first exon of Rbpja.

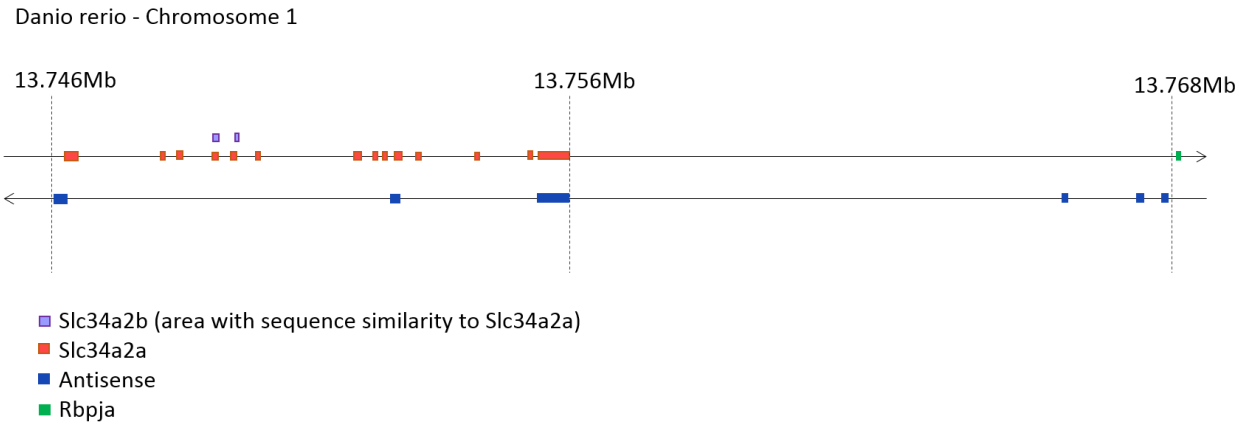


Figure 1.5 Representation of the zebrafish Slc34a2a exons (red) and Slc34a2a(as) exons (blue). Slc34a2b sequence regions that are highly similar to Slc34a2a are shown in purple above the Slc34a2a sense strand; these do not overlap with Slc34a2a(as). Polymerase II uses the bi-directional Rbpja promoter to transcribe Slc34a2a(as).

1.5.1 Slc34a2a and Slc34a2a(as)

Previous work focusing on the Slc34a gene has been done by Dr Andreas Werner and Dr Mark Carlile here at Newcastle University. Zebrafish Slc34a2a sense and antisense transcripts were found to be differentially expressed during early development by end point RT-PCR (figure 1.6, A) (Carlile *et al.*, 2008). The

antisense transcript is expressed from fertilisation up to 48 hours post fertilisation (hpf). The sense transcript, however, is expressed from 48 hpf onwards. When the zebrafish sense/antisense RNA were injected into the cytoplasm of *Xenopus* oocytes in different combinations, the RNA remained intact. However, if injected into the nucleus, the transcripts were processed to short RNAs of ~23 nucleotides (Carlile *et al.*, 2008). The same was found when the Slc34a1 sense/antisense pair from mouse were injected into *Xenopus* oocytes (Carlile *et al.*, 2009). Work completed with mouse Slc34a1 sense and antisense transcripts found that at least 29 base pairs were required to induce processing (Carlile *et al.*, 2009). Moreover, northern blot on RNA preparations from 48 hpf zebrafish embryos (the phase of sense/antisense co-expression) found ~23 nucleotide long endogenous small interfering RNAs (endo-siRNAs) (Carlile *et al.*, 2008). The conclusions from this work was that these small ~23 nucleotide RNAs were likely to be produced by Dicer and feed into a RNA interference related mechanism.

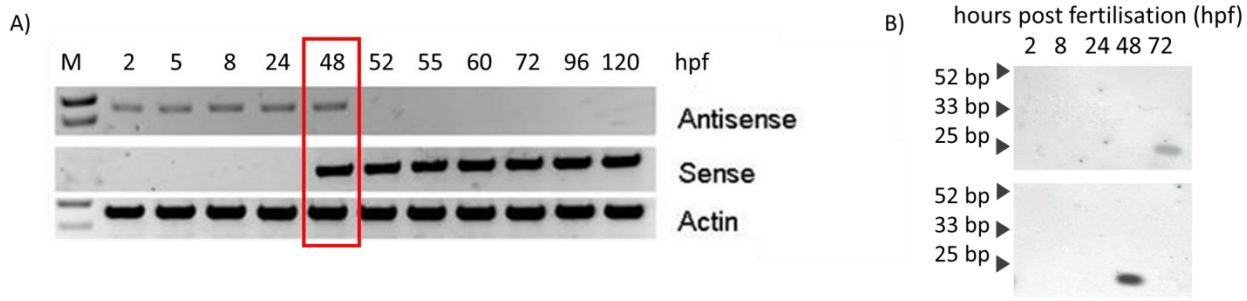


Figure 1.6 Findings from Carlile, M and Werner, A on Slc34a2a sense and antisense zebrafish transcripts. (A) Slc34a2a sense and antisense RNA are differentially expressed during early zebrafish development as demonstrated by end point RT-PCR. (B) Northern blot experiments found that the generation of endogenous small RNAs occurs during co-expression of sense/antisense transcripts at 48 hpf and also at 72 hpf in a strand specific form; top panel, sense probe and bottom panel, antisense probe. These endo-siRNAs are roughly 23 nucleotides in length.

1.6 Zebrafish

1.6.1 Zebrafish as a model organism

Danio rerio, a tropical freshwater minnow otherwise known as zebrafish due to its horizontal stripes, belongs to the Cypriniformes order (Saitoh *et al.*, 2006). It was first established as a vertebrate model organism by George Streisinger and has been used in countless studies since. Many advantages exist to using zebrafish as a model organism in animal research. Zebrafish are small (2 cm in length in early adult/breeding stages) and do not require much in terms of replicating a natural environment (Nusslein-Volhard and Dahm, 2002). Overall they are easier to house and care for than other vertebrate research animals. Secondly, as fish, the number of offspring is remarkably high and mating can occur often. Pairings can be set up one-two times a week, with each pair laying 200-300 eggs per mating. Thus, there is more offspring for examination and a ready supply of future breeding partners. Oocytes are laid external to the female prior to fertilization by the male and embryos are completely transparent prior to 32 hpf (Nusslein-Volhard and Dahm, 2002). Consequently, it is possible to examine the development and any impact of genetic manipulation or medical/chemical treatment with neither invasive assessment nor impacting the welfare of the animal. Thus, reproducible results can be collected easier and more accurately in comparison to many other model organisms.

Another benefit to using externally developing zebrafish embryos is that it is easy to induce genetic changes. Embryos readily absorb chemicals added to the water which can influence their development and gene expression. For example, the addition of 1-phenyl-2-thiourea (PTU) to the water of the developing embryos prior to 24 hpf inhibits melanisation; the embryos stay transparent which facilitates examination of 32 hpf and older embryos (Nusslein-Volhard and Dahm, 2002). In comparison to rodents, zebrafish can tolerate higher levels of chemical mutagens, therefore a higher density of mutations in the genome is possible (Lieschke and Currie, 2007; Augustine-Rauch *et al.*, 2010). Mircoinjection of RNA or morpholino oligonucleotides to manipulate the expression of specific genes is also easily done in fertilized oocytes. Transparency allows for localization of the needle and visualization of the injected material. Material injected into the yolk reaches the first

cells by diffusion and natural cytoplasmic flow if injection is completed prior the 4-cell stage (Nusslein-Volhard and Dahm, 2002; Carmany-Rampey and Moens, 2006).

Importantly, zebrafish are an excellent model organism due to their genetic similarity to humans. *Danio rerio* contain functional homologs for approximately 70% of human disease genes and develop similarly to higher vertebrates (Barbazuk *et al.*, 2000; Langheinrich, 2003). Even though there are downsides to modelling human disease in zebrafish, such as absence of certain organs (example: lungs, mammary gland) and gene duplications resulting in subfunctionalization, they remain highly used in medical investigations (Postlethwait *et al.*, 2004). With the aim of treating human heart ailments, zebrafish are used in regeneration studies (Poss *et al.*, 2002; Lepilina *et al.*, 2006). They are also used in oncology (Li *et al.*, 2012; Yen *et al.*, 2014), embryogenesis (Pauli *et al.*, 2012; Dash *et al.*, 2014), and the study of neurodegenerative disease (Panula *et al.*, 2010) and behavioural traits (Cachat *et al.*, 2013; Tran *et al.*, 2016).

Furthermore, as zebrafish rise in popularity in biological/medical studies, numerous resources have been developed. The main source for information on zebrafish is the Zebrafish Information Network (ZFIN), an online database that lists transgenic lines, mutants and knockdowns generated, gene expression profiles, antibodies, an atlas of zebrafish anatomy and publications. Overall, ZFIN constitutes a comprehensive knowledgebase on all aspects of zebrafish research (http://zfin.org/).

1.6.2 Zebrafish brain anatomy

No brain development occurs prior to 12 hpf (six somite stage) but by 18 hpf (18 somite stage), ten neuromeres have developed; the telencephalon and diencephalon in the forebrain, mesencephalon in the midbrain, and the seven rhombomeres of the hindbrain (Kimmel *et al.*, 1995). By 24 hpf (prim-5 stage), the ventral diencephalon has expanded to include what will become the hypothalamus and the epiphysis is visible in the midline of the diencephalic roof (figure 1.7). In addition, by the 24 hpf stage, the tectum (dorsal midbrain) separates from the tegmentum (ventral midbrain) and a cerebellum is now present behind the midbrain-hindbrain boundary as shown

in figure 1.7 (Kimmel *et al.*, 1995). Circulation begins between 24-26 hpf and thus a beating heart is visible during visual examination. Angiograms of embryo heads between 24-28 hpf show that the major arterial vessels (lateral dorsal aorta and primitive internal carotid artery) and venous vessels (the primordial midbrain channel and the primordial hindbrain channel) have formed (Isogai *et al.*, 2001). The same brain divisions visible within the 24 hpf embryo, are visible at 48 hpf (figure 1.8). By 48 hpf, nearly all morphological features are fully developed outside of the jaw, gill arches and pectoral fins (Kimmel *et al.*, 1995; Nusslein-Volhard and Dahm, 2002). Prior to 72 hpf, the developing zebrafish is known as an embryo, thereafter, whether it has hatched or not from its chorion, it is considered to be a larvae (Nusslein-Volhard and Dahm, 2002).

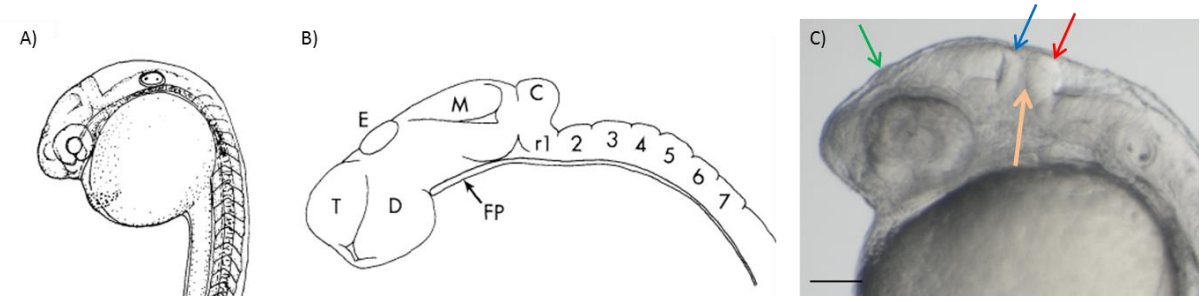


Figure 1.7 Lateral view of zebrafish brain anatomy at 24 hours post fertilization. A) Sketch showing the different features visible in a whole embryo. B) Figure depicting the different brain segments; telencephalon (T), diencephalon (D), epiphysis (E), mesencephalon (M), cerebellum (C), rhombomeres (r1-r7) and floor plate (FP). C) Photograph of embryo pointing out the epiphysis (green arrow), mesencephalon (blue arrow), cerebellum (red arrow) and midbrain-hindbrain boundary (orange arrow). Scale bar at 0.1 mm. Figures for A) and B) taken from (Kimmel *et al.*, 1995).

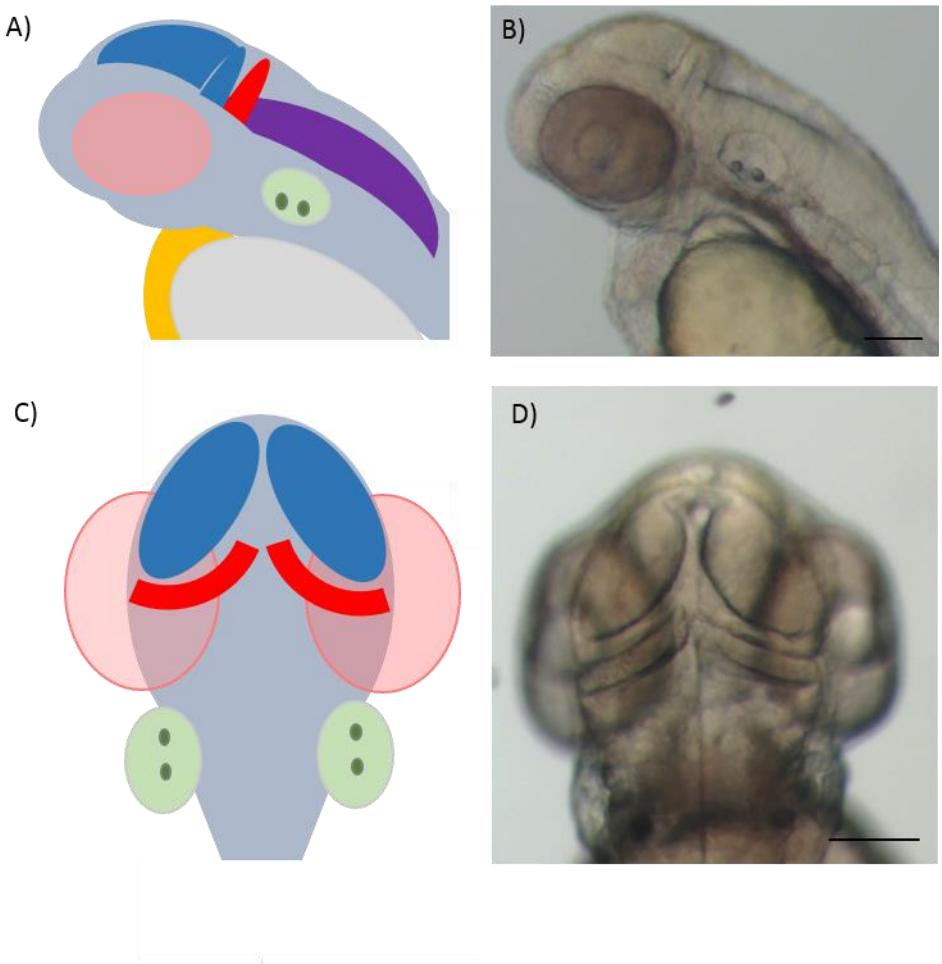


Figure 1.8 Zebrafish head anatomy at 48 hours post fertilization. Lateral (A) and dorsal (C) representations indicating the different visible structures where navy, mesencephalon; red, cerebellum; purple, rhombomeres (r1-r7); green, otic vesicle (ear); pink, eye; orange, heart and cream, yolk. Lateral (B) and dorsal (D) view photographs of an embryo visualized by light microscope. Scale bar at 0.1 mm.

1.7 Cerebellar agenesis

Cerebellar agenesis is an extremely rare condition in which the brain develops either with an incomplete or absent cerebellum, irrespective of the etiology (Altman *et al.*, 1992; Romaniello and Borgatti, 2013). The condition was first reported in humans in 1831 and since 1982, only 9 living cases of complete cerebellar agenesis have been

reported (Yu *et al.*, 2014). Understanding pathogenesis is still undergoing but certain genetic causes have been determined such as single gene mutations in FOXC1, CASK or ZIC (Grinberg and Millen, 2005; Najm *et al.*, 2008), chromosomal copy number aberrations, i.e. trisomies 9, 13 and 18 (Chen and Shih, 2005) and certain sequence deletions or duplications (Ballarati *et al.*, 2007; Boland *et al.*, 2007). The cerebellum, an area in the hindbrain, is comprised of three lobes; the anterior lobe receives information from the spinal cord, the posterior lobe that receives information from the cerebral cortex and the flocculonodular lobe receives information from the vestibular nerve. The cerebellum is involved in many important functions ranging from movement coordination, balance, equilibrium and muscle tone. Thus, people born with this condition or those that have had cerebellar damage, experience severe developmental delays, impaired motor control, language deficits and neurological abnormalities (Yu *et al.*, 2014). Children born with the disorder regain some control over their movements with age as the normally developed areas of the brain partially compensate, but can never do so completely (Taylor *et al.*, 2010). By itself the condition is not fatal, but can be when coinciding with other malformations like anencephaly (neural tube defect leading to smaller and misshapen head and brain), holoprosencephaly (no brain division leading to median cleft lip and cyclopia) and microencephaly (abnormal smallness of the head) (Velioğlu *et al.*, 1998).

Cerebellar agenesis is not unique to humans since incomplete cerebellar development has been noted in many other species including primates, mice and zebrafish. Genes of the ZIC family that are linked to cerebellar agenesis in humans have orthologos in mice which produce the same phenotypic effect when mutated (Aruga *et al.*, 1998). Functional orthologues for PTF1A, a pancreatic transcription factor, have been found in frogs, zebrafish and mice; mutation resulted in the incomplete development of the midbrain hindbrain boundary and cerebellar agenesis (Sellick *et al.*, 2004). The repercussions of cerebellar agenesis in animals is similar to that in humans; impaired balance, decomposition of movement, involuntary tremors and compromised working memory (Moens and Prince, 2002; Lalonde and Strazielle, 2003; Morton and Bastian, 2004).

1.8 Zebrafish genes examined within this study

Throughout the course of this study, multiple zebrafish genes will be referred to including, Engrailed homeobox 2a (Eng2), Presenilin2 (Psen2), Wingless-type MMTV integration site family member 4b (Wnt4b) and RAS-like, family 11, member B (Rasl11b). All relevant information about each gene will be located in the corresponding section in which the gene is examined.

1.8.1 Recombination Signal Binding Protein for Immunoglobulin Kappa J Region A

Recombination Signal Binding Protein For Immunoglobulin Kappa J Region A, abbreviated as Rbpja, is also occasionally known as recombining binding protein Suppressor of Hairless (Su(H)). Orthologs are present in humans (RBPJ), mice (Rbpj) and in Drosophilia (Su(H)). Located on chromosome 1 in *Danio rerio*, it is expressed from fertilization (Echeverri and Oates, 2007). As seen in figure 1.5, *Rbpja* is the gene adjacent and downstream to Slc34a2a. Found within the nucleus, Rbpja has transcription factor activity; by interacting selectively and non-covalently, Rbpja assists sequence specific DNA binding and modulation for transcription by RNA polymerase II (ZFIN, 2002). Rbpja is also involved in numerous cell biological processes including the notch signalling pathway (Sieger *et al.*, 2003), left/right symmetry determination (Echeverri and Oates, 2007; Lopes *et al.*, 2010), neurogenesis (Echeverri and Oates, 2007), somite specification (Sieger *et al.*, 2003; Echeverri and Oates, 2007), and the development of many organs like the adrenal gland (O'Brien *et al.*, 2011), glomerulus (O'Brien *et al.*, 2011) and heart (Echeverri and Oates, 2007).

1.9 Aim

The overall aim of this project is to gain insight into the fundamental mechanisms of how NATs regulate gene expression during zebrafish embryogenesis. This study focuses on the well-defined bi-directionally transcribed Slc34a2a gene. To achieve the main goal of this project, three specific questions were investigated in detail. First, a temporal spatial expression pattern for all RNA transcripts of interest within this project was determined by in situ hybridization and RT-qPCR. A potential co-expression and co-localisation would give first indications towards a putative mechanistic interaction between sense and antisense. Second, the sense/antisense balance was perturbed by microinjection. Thorough phenotypic characterization determined how Slc34a2a and Slc34a2a(as) affected one another and characterised the consequences of ectopic expression of the transcripts. Third, experimental evidence was sought to establish an involvement of RNA interference in the processing of co-expressed sense and antisense transcripts.

Chapter 2. Materials and methods

Fine chemicals were purchased from Sigma Aldrich, Poole, UK, unless otherwise stated. Consumables and plastic ware were purchased from Starlab, Milton Keynes, UK.

2.1 General molecular biology techniques and buffers

2.1.1 Small scale DNA preparation

All small scale plasmid preparations were done in accordance to the Qiagen protocol with the Qiagen Spin miniprep kit (Qiagen). DNA and RNA samples were quantified by spectrophotometry at 260 nm (Thermo Scientific Nanodrop 1000). The purity of preparations was estimated by the ratio of the readings at 260 nm and 280 nm. Reading values for OD_{260}/OD_{280} of 1.8 and 2.0 were considered pure for DNA and RNA respectively.

2.1.2 Ethanol precipitation

Ethanol precipitation was carried out to either concentrate or purify DNA by adding 3 M NaOAc pH 5.5 to a final concentration of 0.3 M, followed by 2.5 times the volume of cold 100% ethanol. This was mixed by inversion and placed either at -80ºC for at least half an hour or at -20ºC for at least 2 hours but typically overnight. Centrifugation at top speed (14000 rpm; Juan BR4i centrifuge) for 20 minutes at 4ºC was completed and the DNA/RNA pellet was washed with 1 ml 70% ethanol. This was followed by another centrifugation at top speed (14000 rpm; Juan BR4i centrifuge) for 15 minutes. Ethanol was removed with water suction and the pellet was air dried for 10 minutes prior to re-suspension with distilled water.

2.1.3 Transformation of competent bacteria

Single use HB101 E. coli competent cells (Promega) were thawed on ice for roughly 20-30 minutes. In an Eppendorf tube, 1-5 µl of DNA (10 ng-1 µg) was added to 20-50 µl of the cells. These were gently mixed and left on ice for half hour. Heat shock took place at 42ºC for 60 seconds and transformations were placed back on ice for 5 minutes. 1 ml of LB medium without antibiotic was added and left shaking (550 rpm; Eppendorf® Thermomixer® R) for one hour at 37ºC. Depending on the original concentration of DNA used and transformation efficiency, 50-100 µl of transformed bacteria was spread on pre-warmed LB agar plates containing relevant antibiotic. Plates were left to incubate overnight at 37ºC. Plates were stored in at 4 ºC, sealed with parafilm.

2.1.3.1 Lysogeny broth (LB) medium

To 800 ml of distilled water, 10 g bacto-tryptone, 5 g yeast extract and 10 g NaCl were added. pH was adjusted to 7.5 with NaOH. The final volume of 1 litre was made up and the LB broth was sterilized by autoclaving.

2.1.3.2 LB agar

To 800 ml of distilled water, 10 g bacto-tryptone, 5 g yeast extract, 10 g NaCl and 15 g agar were added. The final volume of one litre was made up and the solution was sterilized by autoclaving. Prior to pouring plates 1 ml of Ampicillin (50 mg/ml) was added to the warmed solution.

2.1.4 Plasmids

Listed in table 2.1 below are all the plasmids used within this study. Plasmids used were generated prior to commencing this work either in the Dr A. Werner lab (Newcastle University) or were received from the Dr C. Houart group (Kings College

London). However all sequences were confirmed with DNA sequencing prior to any experimentation.

Table 2.1 Plasmids used within the course of this study.

Plasmid name	**Vector**	**Resistance**	**Sequence (NCBI Reference/ Accession number)**
Slc34a2a	pSport1	Ampicillin	NM_131624.1
Slc34a2a(as)	pSport1	Ampicillin	NR_002876.2
Slc34a2b	pSport1	Ampicillin	NM_182877.2
Engrailed2a	pBSK	Ampicillin	NM_131044.2
Sonic Hedge Hog (SHH)	pSport1	Ampicillin	NM_131063.3

2.1.5 Restriction enzyme digestion of plasmids

Restriction enzyme digestion of plasmids was completed using commercially supplied restriction endonucleases (Promega) for approximately two hours at the supplier recommended temperature in the corresponding buffer. No more than 10% of the reaction volume comprised of the restriction enzyme.

Table 2.2 The different uses of restriction endonucleases and their corresponding buffers for this study. All enzymes and buffers were purchased from Promega unless stated. Buffers were chosen using the Promega Restriction Enzyme Tool (Promega, 2016). Abbreviations: SHH, Sonic Hedgehog; Eng2, Engrailed-2a.

Restriction enzyme	**Corresponding buffer**	**Use**
*Age*I	Buffer K	• Slc34a2a RNA synthesis • Frameshift Slc34a2a RNA synthesis

		• Slc34a2a(as) RNA synthesis
*BamH*I	Buffer E	• To confirm insert in clone (pSport1 vector)
*EcoR*I	Buffer H	• Site at which inserts for Slc34a2a, Slc34a2b and Slc34a2a(as) occurred • Slc34a2b RNA synthesis
*Hind*III	Buffer E	• SHH RNA synthesis
*Kpn*I	Buffer J	• Determining Slc34a2a insert presence • SHH RNA synthesis • Eng2 RNA synthesis
*Nde*I	Buffer D	• To confirm insert in clone (pSport1 vector)
*Not*I	Buffer D	• Site at which inserts for Slc34a2a, Slc34a2b and Slc34a2a(as) occurred • Slc34a2b and Slc34a2a(as) RNA synthesis
*Sma*I	Buffer J	• To confirm clone has Slc34a2b insert as it does not cut in Slc34a2a or antisense
*Xba*I	Buffer D	• Slc34a2a RNA synthesis • Frameshift Slc34a2a RNA synthesis
*Xho*I	Buffer D	• To determine size of SHH • Eng2 RNA synthesis

2.1.6 DNA sequencing

DNA (~1 µg) was commercially sequenced (GENEIUS Labs, previously known as Gene Vision, Newcastle, UK). Generic primers such as T7 and SP6 were supplied by the company; otherwise 10 µl of 10 µM stock of the specific primers was sent in a

separate tube. Once sequencing confirmed the integrity of the gene of interest, glycerol stocks were generated for long term storage.

2.1.6.1 Glycerol stocks of E. coli containing plasmids

Glycerol stocks of E.coli containing the corresponding plasmid were generated by adding 15 µl of fresh pre-grown culture to 1.5 ml of LB medium containing the corresponding antibiotic and left to incubate at 37°C for 4 hours at 1400 rpm (Eppendorf Thermomixer Comfort). Cultures were then mixed by inversion in a 1:1 ratio with 50% glycerol and placed immediately into a -80°C freezer.

2.1.7 Agarose gel electrophoresis

Agarose BioReagent (Sigma Aldrich; molecular biology grade) was dissolved in 1x TAE (Tris acetate EDTA) electrophoresis running buffer to either 1, 2, or 2.5% w/v depending on size of DNA fragments to be separated. The solution was heated by microwave until the agarose was fully dissolved. The agarose solution was left to cool for 5-10 minutes at room temperature with occasional shaking. Either 1 µl of 10 mg/ml ethidium bromide (Sigma Aldrich) or 1:10,000 GelRed™ Nucleic Acid Gel stain (Biotium) was added prior to pouring the solution into a gel tray with a comb.

2.1.8 General buffers

Table 2.3 Recipes of general buffers used within this study

Solution	**Method**	**pH**	**Autoclaved**	**Notes**
4% PFA	40 g of paraformaldehyde into 1 L of sterilised 1x PBS over small heat until solution is clear.		No	Stored at -20°C up to a year.
20x saline sodium citrate (SSC)	150 mM NaCl and 15 mM trisodium citrate	7	No	
Maleic Acid Buffer pH 7.5 (MAB, 5x Stock)	150 mM NaCl and 100 mM Maleic Acid	7.5	Yes	Working concentration is 1x. Tween could be added subsequently if required.
10 mM PBS	137 mM NaCl, 2.7 mM KCl, 10 mM Na_2HPO_4 and 2mM KH_2PO_4	7.4	Yes	Working concentration is a 1x solution.
PBST (PBT)	Add 0.1% (v/v) Tween 20 in 1x PBS.		No	
1x TAE electrophoresis running buffer	0.04 M Tris-Acetate (Fisher Scientific), 0.001 M EDTA		No	

2.2 Animals

Danio rerio zebrafish were maintained in aerated aquarium conditions in dechlorinated filtered water at 27.5 ± 1 ºC. Adult fish were fed three times daily; twice with commercial feed (Tetra) and once with fresh brine shrimp larvae. Husbandry and all other procedures were performed to the regulations stated by the UK Home Office (Westerfield, 1995).

Three different strains were used throughout this thesis: AB, Golden and Islet1. The AB strain (Zebrafish International Resource Centre) is a wildtype strain that has been specifically outbred over successive generations to maximise genetic variation. Embryos from this strain were used for microinjection studies and collection for RT-qPCR. The Golden strain is like AB wildtype outside of the fact that the melanocytes do not produce the black pigment. Therefore, Golden fish appear clear which makes them perfect for in situ hybridization studies (ISH). The final strain, Islet1, was used for microinjection studies; a transgenic strain which expresses green fluorescent *Islet1* protein that labels central nervous system cranial motor neurons (Higashijima, 2008).

2.2.1 Embryo collection

Danio rerio female and male fish were paired the evening prior to the day eggs were required. Paired males and females were placed into tanks together but were physically separated by a mesh until the following morning. Eggs were usually laid and fertilised after the lights were turned on (at 8:30 am regularly). Laying would commence upon removal of the physical barrier between male and female fish. Embryos were collected in blue water (2 ml of 0.1% methylene blue to 1 L aquarium tank water) shortly after being laid. Embryos were raised from the day of collection to the required stage at 28ºC in E3 (5.0 mM NaCl, 0.17 mM KCl, 0.33 mM $CaCl_2$, 0.33 mM $MgSO_4$ and pH adjusted to 7.0 with 1 M NaOH) or E3 PTU (0.2 mM 1-phenyl-2-thiourea (PTU) added too 1x E3). The solution was changed every 24 hours to

prevent fungal growth. Staging of embryos was done in accordance to morphological criteria provided by Kimmel *et al.* (Kimmel *et al.*, 1995).

Embryos to be used for ISH were collected at the appropriate stage and fixed overnight at 4ºC in 4% PFA. If the embryos were older than 24 hours, they were first dechorionated with pronase (see below) (Westerfield, 1993; Brand, 1995). Embryos younger than 24 hours were fixed and then dechorionated manually using forceps. Embryos were dehydrated using increasing concentrations of methanol in water (25%, 50%, 75%, and 100%) prior to storage at -20ºC in 100% methanol.

Embryos to be used for RT-qPCR were placed directly into 100% Trizol (Ambion, Life Technologies) and stored at -80ºC until needed.

Embryos to be used in dsRNA immunoprecipitation were flash frozen. 50 embryos at the correct time point were placed on dry ice for half an hour after all liquid was removed. Samples were stored -80ºC.

2.2.1.1 Pronase dechorionation

Embryos were batch dechorionated in 2 mg/ml pronase. 5 ml pronase/E3 was added to approximately 60 embryos and left to incubate for 1-2 minutes at 28ºC followed by rinsing the embryos in E3 solution.

2.3 Microinjection

2.3.1 In vitro synthesised full length mRNA

Plasmids were linearized as described in section 2.1.4 with the relevant restriction enzyme. Full linearization was confirmed by running a sample of uncut and cut plasmid on a 1% TAE agarose gel. Capped mRNA synthesis was completed by following manual specifications of mMESSAGE mMACHINE kit (Applied Biosystems, Thermo Fisher Scientific). Non-capped mRNA was made using either the T7 RNA Polymerase or SP6 RNA Polymerase MEGAscript Transcription kit (Thermo Fisher

Scientific) by following manual instructions included. Both capped and non-capped RNA was DNase treated prior to use in accordance to the kits instructions. The product was purified using SigmaSpin™ Post-Reaction Clean-Up Columns (Sigma-Aldrich) following the manufacturer's instructions. RNA concentration was measured using a nanodrop (Thermo Scientific Nanodrop 1000). Size and integrity of the RNA was confirmed by running an aliquot on a 2% TAE agarose gel. RNA was stored at -80ºC until required. Prior to injection, RNA was diluted with a 10% Phenol Red solution, which served as a visual marker.

2.3.1.1 In vitro synthesised frameshift Slc34a2a RNA

The frameshift was created by site-directed mutagenesis (QuikChange Lightning Site-Directed Mutagenesis Kit, Agilent) by removing a single nucleotide 18 base pairs downstream from the start codon. The following primers were used in creation of the frameshift via PCR:

Forward: 5' GGC ACC ACG TCC AAA GCA GAG CAT GAA TC 3'

Reverse: 5' GAT TCA TGC TCT GCT TTG GAC GTG GTG CC 3'

Frameshift mutation was confirmed via DNA sequencing (section 2.1.6). Use of the online tool 'Translate' by ExPASy confirmed the transcript was now non-protein coding. Multiple stop codons are present in the frameshift sequence. The longest possible amino acid chain that could be produced would be 28 amino acids and it shares no similarity with the full Slc34a2a protein.

Capped mRNA synthesis was completed by following the manual protocol from the mMESSAGE mMACHINE kit (Applied Biosystems, Thermo Fisher Scientific).

2.3.2 In vitro synthesised RNA fragments

End-point PCR was utilized to create the short DNA fragments. All primers were ordered from Integrated DNA Technologies (IDT) and are listed below. Primers 1+2 and 3+4 were used on a plasmid containing the Slc34a2a sequence whilst primers 5+6 were used in a reaction with plasmid containing the Slc34a2a(as) sequence. All primers contained a T7 or SP6 RNA polymerase binding site so that the fragments could be directly used for in vitro RNA synthesis.

Table 2.4 List of primers used to create Slc34a2a fragments via end-point PCR.

Primer name	**Sequence**
1.17/10/13 2a (T7)	5’ – GACAATTAATACGACTCACTATAGGGGCACCACGTC CAAAGCATGAGC – 3’
2.17/10/13 2a (SP6)	5’ – CAGAGGCATACGATTTAGGTGACACTATAGCAGCAG CACCAGGACTGAGAG – 3’
3.17/10/13 2a (T7)	5’ – GGCATTGTAATACGACTCACTATAGGGCGGTTGGCG ATCCCGAAGCC – 3’
4.17/10/13 2a (SP6)	5’ – AGCACTTCATACGATTTAGGTGACACTATAGGCAGCA ACAGCAACAGCGGG – 3’
5.17/10/13 2a (T7)	5’ – GTCCACTAATACGACTCACTATAGGGCCAGGAGTCC CAGCACACTGC – 3’
6.17/10/13 2a (SP6)	5’ – GAAGGAACATACGATTTAGGTGACACTATAGGGAAA CCCGAGGACTGAGGACC – 3’

Each PCR reaction consisted of:

Reagent	**Volume per reaction**	**Concentration**
Forward primer	1 µl	10 µM
Reverse primer	1 µl	10 µM
2x Red PCR Master Mix (Rovalabs)	10 µl	
Plasmid	1 µl	0.002 µg/µl
Water	7 µl	

The cycling conditions were as follows:

Step	Temperature	Length of time
Initial denaturing	95°C	2 minutes
Amplification (30 cycles)	95°C 60°C 72°C	15 seconds 25 seconds 10 seconds
Final hold	4°C	∞

An aliquot of amplified product was analysed on a 2% TAE agarose gel to confirm purity and length of the fragment. The product was purified up using SigmaSpin™ (Sigma-Aldrich) following product instructions. RNA probes were synthesised by incubating the following reactions at 37°C for two and a half hours:

Reagent	Volume per reaction
DNA template	1 µg
10x Transcription Buffer (Thermo Scientific)	2 µl
ATP/GTP/CTP/UTP Mix (Thermo Scientific)	2 mM final Concentration
RiboLock™ RNase inhibitor (Thermo Scientific)	1 µl (50 U)
Polymerase (T7 or SP6) (Thermo Scientific)	2 µl (20 U)
Water	To make up to 20 µl

The DNA template was degraded by adding 2 µl RNase free DNase (Thermo Scientific) to the reaction and incubating the mix for 5 minutes at 37 °C. The digestion was stopped by adding 1 µl of 0.5 M EDTA pH 8.0. SigmaSpin™ Reaction Clean-Up Columns (Sigma-Aldrich) were used to purify the RNA. Thermo Scientific Nanodrop 1000 was used to determine the RNA concentration. The integrity was confirmed by running a small aliquot on a 2% TAE agarose gel. The RNA was stored

at -80 ºC. Prior to injection, RNA was diluted with a 10% Phenol Red solution, which served as a visual marker.

2.3.3 Morpholino oligonucleotides

A morpholino oligonucleotide targeting pre-mRNA splicing of the Slc34a2a(as) transcript was designed and ordered from Gene Tools LLC (Philomath, US), along with a mismatch morpholino as a control. Both Dicer fluorescein tagged morpholinos were a kind gift from Prof. JG Patton (Department of Biological Sciences, Vanderbilt University, Nashville, USA) who had them designed and ordered from Gene Tools LLC. No Dicer mismatch control morpholino was received. A standard control morpholino, with no biological target, was utilized instead. A tested p53 morpholino designed by Gene Tools LLC was purchased for co-injection studies and always injected at a concentration of 4 ng as per Gene Tools recommendation.

Table 2.5 Morpholino oligonucleotide sequences and their targets. All morpholinos were designed by and ordered from Gene Tools LLC.

Morpholino	Sequence	Target
Block AS	5’ – GCCATCTGGTGAAAAGACAGAGTTT – 3’	Antisense pre-mRNA third exon
AS mismatch	5’ – GACATCTCGTCAAAACACAGACTTT – 3’	No target
Dicer 5’ UTR	5’ – CTGTAGGCCAGCCATGCTTAGAGAC – 3’	5’ UTR of *dicer1*
Dicer start	5’ – TCTTTCTCTTCATCTTCCTCCGATC – 3’	Translational start site of *dicer1*
P53	5’ – GCGCCATTGCTTTGCAAGAATTG – 3’	P53
Standard control	5' - CCTCTTACCTCAGTTACAATTTATA - 3'	No target

2.3.4 Hairpin RNA

Ultramers to generate hairpin RNA for each of the candidate genes were designed incorporating T7 and universal M13 primer sites (table 5.1). The designed oligonucleotides were amplified by end-point PCR using T7/ M13 primers prior to purification with a PCR clean-up kit (GeneJet, Thermo Scientific). Restriction digest to remove the M13 primer site was completed with *Xba*I as stated in section 2.1.4. The product was purified using SigmaSpin™ (Sigma-Aldrich) following product instructions. RNA probes were synthesised by incubating the following reactions at 37°C for two and a half hours:

Reagent	**Volume per reaction**
DNA template	1 µg
10x Transcription Buffer (Thermo Scientific)	2 µl
ATP/GTP/CTP/UTP Mix (Thermo Scientific)	2 mM final Concentration
RiboLock™ RNase inhibitor (Thermo Scientific)	1 µl (50 U)
Polymerase (T7) (Thermo Scientific)	2 µl (20 U)
Water	To make up to 20 µl

Thermo Scientific Nanodrop 1000 was used to determine the RNA concentration and integrity was confirmed by running a small aliquot on a 2% TAE agarose gel. The RNA was stored at -80 ºC. Prior to injection, RNA was diluted with a 10% Phenol Red solution, which served as a visual marker.

2.3.5 Injection set-up and calibration

Injection needles were prepared by pulling filament containing 3.3 borosilicate glass capillaries (product code: 1403517 from Hilgenberg GmbH, Malsfeld, Germany) with a horizontal pipette puller (Sutter Instrument Co) with the following settings: Heat- 635, Pull- 50, Velocity- 80, and Time- 200. The tip was opened by snapping off the tip of the needle using a pair of sharp forceps. Morpholino or RNA was rear-loaded using Eppendorf loading tips The droplet size was measured by injecting into oil on a 1 mm graticule. A femtojet compressor (Eppendorf) with the settings 280 hPa injection pressure, 0.3 sec injection time and 17 hPa compensation pressure was used to consistently deliver a droplet volume of 2 nl (diameter of 160 µm). Compressed air flow was triggered by a foot pedal. The dilution of RNA or morpholino was altered to give the desired dose since the injected droplet size remained constant.

2.3.6 Embryo loading (Injection)

1 to 4- cell stage embryos (10-30 minutes hours post fertilization) were lined up against a glass slide within a petri dish. 2 nl injection was done into the yolk sac as close as possible to the first cell and never later than the 4-cell stage (Nasevicius and Ekker, 2000). Dual or triple RNA/morpholino injections were completed as a mixture in the same needle to ensure equal dosing. From each clutch that underwent injection, 50 uninjected embryos were kept as a fitness control group for the entire clutch. Both uninjected and injected embryos, where average number of embryos per petri dish was 50, were incubated in the dark at 28ºC within E3 solution. Roughly eight hours after collection, all unfertilised and dead eggs from each petri dish were removed and discarded. The following morning, at 24 hpf, all embryos that died overnight were counted and logged. If percent survival was below 80% in uninjected control groups, all embryos from clutch were discarded and data not used.

2.3.7 Anaesthetisation and photomicroscopy

Embryos were anaesthetised in 0.2 mg/ml Tricaine ® (Sigma) for 3-5 minutes prior to any microscopy. Embryos were orientated using fine tipped forceps and either photographed in E3 solution or in 3% methylcellulose in E3. Imaging was completed using a Leica dissection light microscope (Leica Application Suite version 4.2.0). For fluorescent images a GFP filter was applied. For live imaging, a Chameleon digital camera (model CMLN-13S2M, Chameleon) was utilized and images were processed with FlyCap2 (Point Grey) software. Prior to returning embryos to the incubator, they were washed and left in fresh E3 for five minutes.

2.4 Whole mount in situ hybridization (ISH)

2.4.1 Probe generation

Plasmids were linearized as mentioned in section 2.1.4 with the relevant restriction enzymes in matching buffers. Full linearization was confirmed by running a sample of uncut and cut plasmid on a 1% TAE agarose gel. Cut plasmid was purified using QIAprep Spin Miniprep kit (Qiagen) according to the manufacturer's instructions. In vitro transcription reactions were assembled according to the scheme below and incubated at 37°C for 60 minutes for T7/T3 polymerases or 90 minutes for the SP6 polymerase:

Reagent	Volume per reaction
DNA template	1 µg
10x Transcription Buffer (Thermo Scientific)	2 µl
10x DIG-NTP mix (Roche)	2 µl
RiboLock™ RNase inhibitor (Thermo Scientific)	1 µl (50 U)
Polymerase (T7/T3 or SP6) (Thermo Scientific)	1.5 µl (20 U)
Water	To make up to 20 µl

2 µl of RNase-free DNase I (Promega) was added at the end of the reaction and left for a further 30 minutes at 37°C. The RNA was purified using SigmaSpin™ Post-Reaction Clean-Up Columns (Sigma-Aldrich) following manufacturer's instructions. The presence of RNA was confirmed by agarose gel electrophoresis. The probes were stored at -20°C in 1x the probe final volume of 20x SSC and 2x the probe final volume of formamide.

2.4.2 Embryo paraformaldehyde fixation

Live embryos were placed into sterile 20 ml bijou tubes. Pronase dechorionation was completed as described in section 2.2.1.1. The pronase solution was removed and replaced with 0.2 mg/ml Tricaine ® (Sigma) in E3 and the embryos were left for five minutes. Tricaine® was replaced with chilled 4% paraformaldehyde/PBS (PFA) and stored at 4°C overnight. The following morning, PFA was removed and the embryos were dehydrated using a gradient of methanol/water at room temperature (five minutes each in 25%, 50%, 75% and 100%) prior to being stored in fresh 100% methanol at -20°C.

2.4.3 Whole mount RNA in situ hybridization

Embryos, which were stored in methanol, were progressively rehydrated on ice by adding 2-3 drops of PBST at a time. Once hydrated, two washes in 100% PBST were completed at room temperature whilst lightly shaking for five minutes. 10-20 embryos were placed into 1.5 ml Eppendorf tubes. 10 mg/ml Proteinase K (Sigma) was diluted 1:1000 in PBST and added to the embryos to permeabilize the tissue. Treatment length differed with age; 1 minute for embryos between 1-8 somites, 3 minutes for embryos that were 9-18 somites, 10 minutes for those between 18-24 hpf and 30 minutes for 24-48 hpf embryos. For 56 hpf and 75 hpf embryos, proteinase K was diluted 1:500 and embryos were incubated for 20 and 30 minutes, respectively. Proteinase K treatment was completed at room temperature whilst slightly shaking. The permeabilised embryos were theny washed twice in PBST followed by a second fixation in 4% PFA for 20 minutes at room temperature (not shaking). Embryos were washed twice in PBST before adding pre-warmed 65°C hybridization mix (100 ml: 1 g Boehringer Blocking Reagent (Roche), 50 ml formamide, 25 ml 20x SSC, 100 mg Torula yeast RNA (Sigma), 10 mg Heparin (Sigma), 0.1% Tween20, 0.1% CHAPS made up with distilled water). Samples were placed onto 65°C heat block for five minutes. From this moment onwards, embryos were never allowed to cool and remained at 65°C. Pre-hybridization was done for 2-4 hours with one change of the hybridisation solution. For each sample, diluted probe (amount dependent on concentration) was mixed with 300 µl hybridization mix and pre-warmed to 65°C for 5-10 minutes. The hybridization/ probe mix was added and embryos were incubated overnight at 65°C.

The following morning, the hybridisation/probe solution was recovered and stored at -20°C as it could be used multiple times. Pre-warmed 65°C hybridization mix was added to the embryos and incubated 30 minutes. The subsequent washes were completed at 68°C. Embryos were washed in 2x SSC for 5 minutes, twice in 2x SSC for 30 minutes, twice in 0.2x SSC for 30 minutes and a final wash was done in 0.2x SSC for five minutes on a shaker at room temperature.

Embryos were equilibrated in MAB (0.1% Tween20) for one hour with a minimum of four changes of solution prior to a blocking step in MAB/Blocking 2% (0.1% Tween20) for 2 hours (Blocking Reagent from Roche). Anti-DIG Antibody (Roche)

was diluted 1:4000 in MAB/Blocking and embryos were incubated overnight at 4°C rocking slowly.

The following day, the embryos were washed in MAB for 3-6 hours, changing the solution multiple times. Successive washes with PBST were completed for 3-6 hours were the solution was replaced every half hour. Embryos were transferred into 12 well plates and left in PBST overnight shaking at 4°C.

To develop, embryos were equilibrated 2x 5 minutes in fresh NTMT buffer (for 50 ml: 1 ml 5 M NaCl, 2.5 ml 2 M Tris HCl pH 9.5, 1.25 ml 2 M $MgCl_2$, 5 ml 10% Tween 20 and made up with distilled water to the final volume) before the developing solution was added. The developing solution consisted of 3.3 µl NBT (nitro blue tetrazolium chloride, Roche) and 3.5 µl BCIP (5-bromo-4-chloro-3-indolyl-phosphate, Roche) per 1.5 ml NTMT buffer. Embryos were incubated in the dark and monitored for staining every 30 minutes. To stop the reaction, embryos were washed 3 x 5 minutes in PBST and progressively (drop-wise) transferred into 70% glycerol/PBS for imaging and storage (at 4°C).

2.4.4 Photomicrography

High power images of fixed embryos in 70% glycerol/PBS were obtained using a Leica dissection light microscope (Leica Application Suite version 4.2.0). Images were assembled in Microsoft PowerPoint.

2.5 Quantification of gene expression by real time polymerase chain reaction (qPCR)

2.5.1 Oligonucleotides

Primers for qPCR were either designed by and ordered from PrimerDesign Ltd or designed in house and ordered from Sigma Aldrich or IDT (Integrated DNA Technologies).

Table 2.6 List of oligonucleotides used in real time PCR reactions.

Target transcript	**Sequence**	**Product Length**	**Other information**
Actin	Company reference gene Information not provided	87	Design and purchased from PrimerDesign Used for relative quantification in non-injected wildtype embryos (chapter 3)
Slc34a2a	Forward: 5'-AACACAGATTTCCCT TTTCCATTT- 3' Reverse: 5'-CAAGAGGAGTTATAGC CGATGT-3'	121	Design and purchased from PrimerDesign Used for quantification in non-injected wildtype embryos (chapter 3)
Slc34a2b	Forward: 5'-ATCTGGCGGTGGGT TTGA- 3' Reverse: 5'-CAGCATGGAGTTCAG GAGTTT -3'	92	Design and purchased from PrimerDesign Used for quantification in non-injected wildtype embryos (chapter 3)
Rbpja	Forward: 5'-CTGTCTCTCTTTGTT GTGGTTCT- 3'	96	Design and purchased from PrimerDesign Used for quantification in non-injected wildtype embryos (chapter 3)

	Reverse: 5’-GCATTGATCCTGTGG TATCTTGA -3’		
Antisense	Forward: 5’-GAGAGTGGGGACG GACGGTTG- 3’ Reverse: 5’- CAGGAGTCCCAG CACACTGCAGA -3’	144	Designed in house and purchased from Sigma Aldrich Used in all experiments
Actin	Forward: 5’-CCGTCTTCCCCTCCA TCGTTGG -3’ Reverse: 5’-GTTGGTAATGATGCC GTCC TCGATG -3’	149	Designed in house and purchased from Sigma Aldrich Used for relative quantification in injected embryos and dsRNA IP studies (chapter 4-6)
Slc34a2a	Forward: 5’-CCTGGGAAATGATGG AGCTGCAGG-3’ Reverse: 5’-CTCCAACAAGCTGGA AAGCTGAGC-3’	174	Designed in house and purchased from Sigma Aldrich Used for quantification in injected embryos and dsRNA IP studies (chapter 4-6)
Slc34a2b	Forward: 5’-CCACTGGAGACCCTGC TGCC-3’ Reverse: 5’-AGCACAGGGCGTCCG GAGTG-3’	126	Designed in house and purchased from Sigma Aldrich Used for quantification in injected embryos and dsRNA IP studies (chapter 4-6)

2.5.2 RNA extraction

Five embryos from multiple developmental stages were selected and placed into 50 µl Trizol ® (Life Technologies). The five embryos at stages 75% epiboly (8 hpf),

prim-5 (24 hpf), (32 hpf), long pec (48 hpf), (56 hpf) and protruding mouth (72 hpf) were crushed with a pestle in 50 µl of Trizol ® in a 1.5 ml Eppendorf tube. Samples were incubated for five minutes at room temperature. 30 µl of chloroform (Sigma Aldrich) was added and samples were rocked for 10 seconds prior to a two minute incubation at room temperature. Samples were centrifuged at full speed (14,000 rpm; MSE MicroCentaur centrifuge) for 15 minutes to aid phase separation; a lower red phenol-chloroform, an interphase and a colourless upper aqueous phase. The upper phase, which contains the RNA, was moved into fresh 1.5 ml Eppendorf tube without disturbing the interphase. RNA was precipitated with 50 µl of isopropanol and 0.5 µl of Glycoblue TM coprecipitant (Life Technologies). Samples were incubated on ice for half an hour prior to being centrifuged at 14,000 rpm for 10 minutes (Juan BR4i centrifuge). RNA pellets were seen as blue due to Glycoblue coprecipitant. Pellets were washed with 1 ml cold 70% ethanol followed by another centrifugation at 14,000 rpm for 10 minutes (Juan BR4i centrifuge). The supernatant was removed and tubes were left inverted for ten minutes at room temperature to air dry. RNA was resuspended in 16 µl of RNAse free water (vortexed for 10 second intervals with alternative intervals being on ice).

DNase I (Thermo Scientific) treatment to remove all genomic DNA contamination was completed in two steps. In the same tube where RNA was resuspended, DNase I treatment was first completed using a $MgCl_2$ containing buffer for half an hour at 37°C; $MnCl_2$ was then added and left for a further 20 minutes at 37°C. To stop the reaction, 1 µl 0.5M EDTA was added and the sample was then further incubated at 37°C for 10 minutes. This was followed by purification using SigmaSpin Reaction Clean-up Columns (Sigma Aldrich) as described in corresponding protocol.

2.5.3 Analysis of RNA concentration and quality

RNA concentrations were determined using a NanoDrop spectrophotometer (NanoDrop Technologies Inc). The quality of the RNA was predicted using the 260/280 reading. Running 2 µl of the RNA sample on a 2% agarose gel confirmed the integrity of the RNA.

2.5.4 cDNA synthesis

Total RNA (~ 1 µg) was reverse transcribed to produce cDNA using a Qiagen Omniscript RT kit by following the corresponding protocol. Random nonamers (Sigma Aldrich) were used as primers and reactions were run for 1 h at 37°C .

Reagent	**Volume per reaction**
RNA	~ 1 µg
10x Reverse Transcription Buffer	2 µl
Random nonmers (Sigma)	2 µl
dNTP Mix (5 mM each dNTP)	2 µl
RiboLock™ RNase inhibitor (Thermo Scientific)	1 µl (10 U)
Omniscript Reverse Transcriptase	1 µl (4 U)
Water	To make up to 20 µl

2.5.5 Semi-quantitative real-time RT-PCR

The assembly of the qPCR samples was completed on ice. The cDNA produced by reverse transcription was diluted 1:4 in RNAse free water whilst RNA for control purposes was diluted 1:10. Master-mixes comprising of LightCycler ® 480 SYBR Green I Master, primers (PrimerDesign or Sigma Aldrich) and water were made prior to pipetting into 96-well plates. Each well contained 5 µl of LightCycler ® 480 SYBR Green I Mastermix, 1 µl forward primer of 10 µM stock, 1 µl reverse primer of 10 µM stock, 1 µl water and 2 µl cDNA,RNA or water. The plates were centrifuged twice for 20 second intervals at full speed prior to loading into a LightCycler ® 480 (Roche).

The qPCR program was as follows:

Step	**Target (°C)**	**Acquisition Mode**	**Time (mm:ss)**	**Ramp rate (°C/s)**	**Acquisitions (per °C)**
Activation	95	none	10:00	4.40	
Cycling	95	none	00:05	4.40	
	53 (PrimerDesign) 59 (Sigma Aldrich)	None	00:20	2.20	
	72	Single	00:01	4.40	
Melt Curve	95	none	00:01	4.40	
	65	none	00:30	2.20	
	95	Continuous		0.11	5
Cooling	40	none	00:10	1.5	

2.5.6 Analysis

Calculated Ct values (Absolute Quantification/ Second Derivative Maximum) were produced for each gene along with the melting temperatures via the LightCycler ® 480 software (release 1.5.0 SP3). Following the removal of outliers based on Ct values and melting curves, Ct values were analyzed in Microsoft Excel. Each sample was run in triplicate on a plate and a minimum of three samples were collected from different clutches. A Ct difference of a minimum of three cycles had to be present between cDNA and RNA samples to be included into the analysis. Alternatively, the melting temperature and curve between cDNA and RNA samples had to differ sizably. Averages and standard deviations were taken from a sample triplicate. Using these values, ΔCt was calculated by subtracting the Ct of the reference gene (β-actin) from the Ct value of the gene of interest. The ΔCt values from a minimum of three different clutches was averaged for a final value displayed in graphical format. The standard error was calculated according to the formula:

SE= √((standard deviation of β-actin)^2+(standard deviation of gene of interest)^2)

P-values were not determined due to small N of three. ‘Practical Statistics for Environmental and Biological Scientists’ (Townend, 2002) recommends at least six replicates to justify T-tests or non-parametric tests such as the Wilcoxon (Mann-Whitney) test (Townend, 2002). Normal distribution of data cannot be assumed or proven with a N number of three.

2.6 dsRNA mapping and extraction using the J2 monoclonal antibody

2.6.1 Immunofluorescent staining

PFA-fixed zebrafish (4% PFA in PBS, overnight) were embedded in OCT (Cell Path) compound, snap-frozen in isopentane over liquid nitrogen and then stored at -80°C. Cryo-preserved zebrafish were sectioned at 5 µm using a Leica cryostat. Sections were collected onto Superfrost plus microscopic slides (Thermo Scientific), dried at room temperature (RT) for 4-6 hours and then either stored at -80°C or used for immunostaining immediately. Sections stored at -80°C were thawed to RT for 1 hour before staining. Slides were washed once in 1x PBS for 5 minutes at RT with agitation. Non-specific binding of the secondary antibody was blocked by the addition of 20% goat serum in 1x PBS followed by an incubation for 1 hour at RT. The J2 monoclonal antibody diluted 1:500 in 1x PBS containing 20% goat serum was added and incubated overnight in a humidified chamber. The next day, sections were washed three times each for 5 minutes, with 1x PBS before incubating in the dark at RT for 1 hour with the secondary antibody. Antibodies used were goat anti-mouse IgG (H+L) alexafluor 488 F(ab’)2 or goat anti-mouse IgG TRITC-conjugated secondary antibody (Molecular Probes A11070), diluted 1:200 (stock = 2 mg/ml) in 1x PBS. Slides were washed four times in 1x PBS (in the dark), each for 5 minutes with agitation, before mounting the coverslips with FluorSave reagent (Calbiochem). Slides were stored in the dark at room temperature overnight, then transferred to 4°C. Immuno-stained tissues were visualised using a Zeiss AxioImager fluorescent microscope.

2.6.2 dsRNA extraction

Slc34a2a and Slc34a2a(as) uncapped RNA was *in vitro* synthesized as mentioned above. Double stranded RNA (dsRNA) was generated by mixing equal quantities of Slc34a2a and Slc34a2a(as) RNA in 1x PBS prior to leaving at 72°C for 10 minutes which was followed by cooling on ice. Control samples included ssRNA Slc34a2a(as) in 1xPBS and a mix of Slc34a2a RNA and Slc34a2a(as) RNA in water. Unless stated otherwise, experiment were completed on ice or in the cold room. RNA extraction was performed either using the Trizol ® (Life Technologies) method mentioned above or the Bioline Isolate II RNA Kit as per the manufacturer's instructions. After RNA extraction (or no extraction for control), the J2 monoclonal antibody (mAb) IgG2a (Scicons) was added (1:200) and left rotating overnight at 4°C. The following morning, 5 µl of goat anti-mouse IgG Magnetic Beads (New England Biolabs) were added and left rotating for another hour at 4°C. MACS columns were rinsed three times with 1 ml 1x PBS prior to adding the beads bound to dsRNA. Samples were washed on the columns twice with 1 ml of ice-cold 1x PBS whilst on the magnetic stand. Samples in columns were removed from the magnetic stand and the stand was heated for 15 minutes in a boiling water bath. Columns were then placed back onto hot magnetic stand, and RNA was eluted using 2x 100 µl 90°C water. The RNA eluted from J2/beads was reverse transcribed (Qiagen Omniscript RT kit) as described above. End-point PCR was completed utilizing primers designed in house (table 2.6) with the following cycling conditions:

Step	**Temperature**	**Length of time**
Initial denaturing	95°C	2 minutes
Amplification (30 cycles)	95°C 57°C 72°C	30 seconds 30 seconds 90 seconds
Final elongation	72°C	180 seconds
Final hold	4°C	∞

Final products from the end-point PCR were separated on a 2% TAE agarose gel alongside 100 bp ladder (New England Biolabs) to determine correct band size.

Chapter 3. Temporal expression pattern of Slc34a2a sense and antisense, Slc34a2b and Rbpja transcripts in wildtype zebrafish embryos

The expression pattern and quantity of a transcript can be a key indicator as to its possible biological function and mechanism. This is particularly important within this project since the interplay of at least three transcripts from two genes (Slc34a2a, Slc34a2a(as) and Rbpja) form a regulatory network during zebrafish development. Carlile (Carlile *et al.*, 2008) found that the Slc34a2a transcript and its antisense are reciprocally expressed in early zebrafish embryos by RT-PCR. This study lacked the specialist detail and quantification required to draw any proper conclusions. Therefore mapping and more extensive quantification was further sought-after to determine temporal expression patterns for each of the genes of interest.

3.1 Temporal expression levels of Slc34a2 sense, antisense and Slc34a2b transcripts determined by RT-qPCR

To quantify Slc34a2a and related transcripts during zebrafish development, embryos were collected at 8, 24, 32, 48, 56 and 72 hours post fertilization (hpf) as mentioned in Chapter 2; Materials and methods. Five embryos from the same clutch were stored in 50 µl of Trizol prior to RNA extraction and DNase I treatment. Samples for each time point were taken from multiple clutches to account for the minor genetic variation between parents such as naturally occurred individual point mutations. A minimum of three different samples of five embryos from different parents were utilised for semi-quantitative expression analysis. Samples were run in triplicate on the same qPCR plate, always including negative controls (RNA without reverse transcription and water instead of cDNA) and alongside the reference house-keeping gene, β- actin (Sequence information: http://www.ncbi.nlm.nih.gov/nuccore/NM_131031.1).

3. Wildtype expression

SYBR Green (Roche) was used within this study to quantify cDNA. SYBR Green which is highly specific for double-stranded DNA (dsDNA), fluoresces when bound to the minor-groove of DNA (Dragan *et al.*, 2012). This fluorescence is proportional to the amount of dsDNA within the reaction. Cycle threshold (Ct) is the number of cycles required for the fluorescent signal to be higher than background level fluorescence. However, as SYBR Green shows no particular sequence specificity, it can detect dimerised primers and unrelated fragments to give false positive readings. To rule out such contaminations, a dissociation curve was completed after each run. As each fragment has a characteristic melting temperature, the dissociation curve can discriminate between a specific amplicon and unspecific product.

Alongside Slc34a2a and its antisense transcript (Slc34a2a(as)), mRNA levels of the paralog Slc34a2b were investigated. As mentioned in chapter 1, Slc34a2a and Slc34a2a(as) are transcribed from chromosome 1 whilst Slc34a2b is transcribed from chromosome 23.

All results from RT-qPCR experiments were normalised to the house-keeping gene β-actin, which maintained a consistent Ct value of about 21 throughout all experimental time-points. After careful consideration, a ΔCt value (average Ct of three individual reactions of the gene of interest – average Ct of β-actin triplicate) of over 13 considered the transcript level to be too low for detection. A ΔCt value over 13 with β-actin having a Ct of 21 meant individual Ct values were ~35 or over. Individual Ct values ± 10 of the positive gene (β-actin) are considered reliable whilst individual Ct values above 37 are considered negative (due to low reliability as this is near background levels) within SYBR Green studies and according to MIQE Guidelines (MIQE: minimum information for publication of quantitative real-time PCR experiments) (Bustin *et al.*, 2009; Chugh *et al.*, 2010; Jiang *et al.*, 2014). This cut off point is indicated by the dotted line in figure 3.1 and any values above are considered non-detectable.

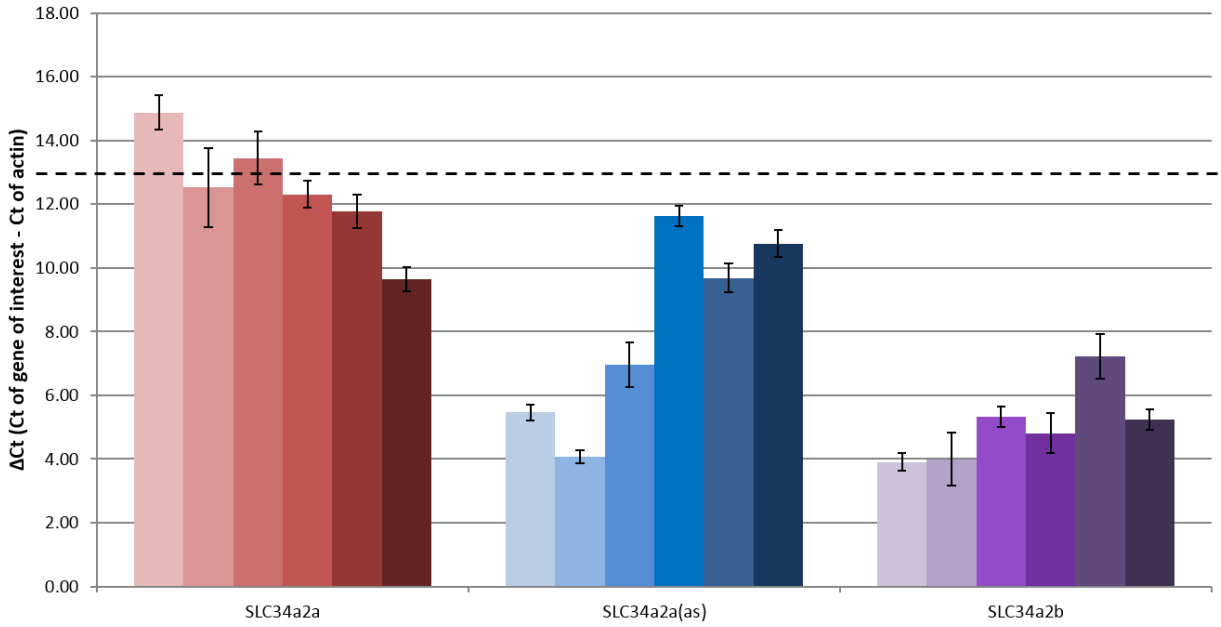


Figure 3.1 Temporal expression profiles of Slc34a2a, its antisense transcript Slc34a2a(as) and paralog Slc34a2b quantified by RT-qPCR. Expression profiles are in relation to the house-keeping gene β-actin which maintained a stable Ct value. The dotted line indicates a ΔCt of 13; levels above are considered undetectable transcript levels. A decrease in bar size designates an increase in transcript level. Sample age, 8, 24, 32, 48, 56 and 72 hpf, is represented as lightest to darkest colour of bars for each transcript. N=3

Throughout the early stages of development, Slc34a2a was not detectable till approximately 48 hpf as seen in figure 3.1. ΔCt values were above 13 for time points 8, 24 and 32 hpf, with specific ΔCt values of 14.87±0.54, 12.52±1.24 and 13.45±0.82 respectively (individual Ct values = 37.37±0.38, 33.97±0.81 and 34.19±0.82 correspondingly). ΔCt values decreased from 12.31±0.42 at 48 hpf to 11.78±0.52 at 56 hpf and 9.56±0.39 at 72 hpf, indicating an increased expression (individual Ct = 34.60±0.33, 31.78±0.26 and 30.40±0.34, respectively).

Contrariwise, Slc34a2a(as) was present from fertilization with levels decreasing from 24 hpf but the transcript remained detectable at all time points. ΔCt values increased to 11.62±0.33, 9.69±0.46 and 10.77±0.43 at 48, 56 and 72 hpf time points, respectively (individual Ct values = 30.88±0.32, 28.26±0.37 and 30.14±0.40). With comparably low ΔCt values of 5.46±0.26, 4.08±0.20 and 6.96±0.71 for 8, 24 and 32

hpf time points (individual Ct values = 27.71±0.25, 27.15±0.17 and 27.49±0.62 correspondingly), antisense expression was found to be high.

As Slc34a2a and Slc34a2a(as) are transcribed from the same genomic region and share sequence complementarity, the hypothesis was that an expression correlation would emerge. From the RT-qPCR data, a reciprocal expression pattern between the two transcripts is visible, where at 48 hpf, both transcripts are present at a near equal level (ΔCt Slc34a2a 11.78±0.52, ΔCt Slc34a2a(as) 11.62±0.33).

Slc34a2b was found in higher abundance than its paralog Slc34a2a. Slc34a2b levels showed a slight decrease over time with ΔCt values of 3.91±0.28, 3.99±0.82, 5.34±0.32, 4.82±0.62 and 5.23±0.32 for time points 8, 24, 32, 48 and 72 hpf (individual Ct values of 26.40±0.07, 25.44±0.11, 27.19±0.19, 27.11±0.60 and 25.98±0.20 respectively). At 56 hpf, ΔCt increased to 7.22±0.70 indicating a decrease in transcript level (Ct=27.22±0.40).

The paralogs Slc34a2a and Slc34a2b share 68.4% nucleotide identity as seen in Appendix A. As highlighted in figure 3.2, an e!Ensembl BLAST nucleotide alignment (EMBL-EBI, 2015), Slc34a2b and Slc34a2a(as) do not share significant regions of complimentary (sequence alignment in Appendix A). A sequence alignment between Slc34a2b and Slc34a2a(as) finds a maximal complementarity of 14 nucleotides between the two transcripts. In addition, the genes are transcribed from different chromosomes. Therefore a direct co-regulation of Slc34a2b and Slc34a2a(as) seemed unlikely and the RT-qPCR data supported this hypothesis.

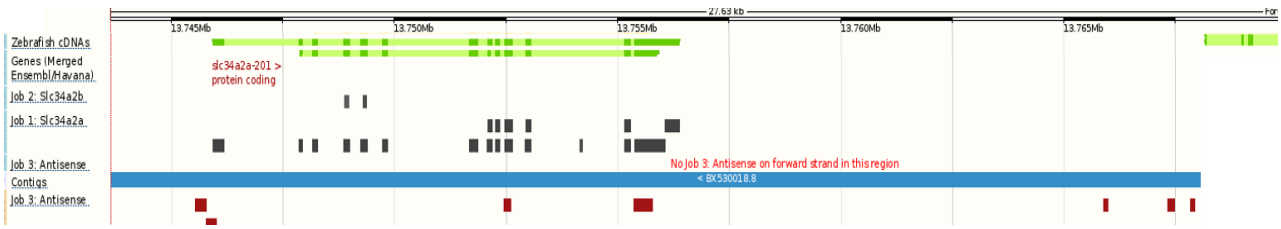


Figure 3.2 Sequence alignment comparing Slc34a2b, Slc34a2a sense and antisense transcripts using the BLAST algorithm (EMBL-EBI, 2015).

3.2 Mapping of Slc34a2 sense, antisense, and Slc34a2b transcripts by in situ hybridization in wildtype zebrafish embryos

Whole mount in situ hybridization (ISH) was completed on wildtype zebrafish embryos which were PFA fixed at numerous time-points in development; 8, 24, 48, 56 and 72 hpf. Digoxigenin (DIG) labelled probes for Slc34a2a, Slc34a2a(as) and Slc34a2b were utilized to map the expression of each transcript. Sonic Hedgehog (SHH) ISH staining was included as a control. SHH staining, as shown in bottom row of figure 3.3, was reproducibly found in the central nervous system, floor plate and brain consistent with published work (Sumanas *et al.*, 2005; Thisse, 2005; Jensen *et al.*, 2012; Chatterjee *et al.*, 2014).

Slc34a2a sense and antisense transcripts showed the same temporal expression pattern as seen in the RT-qPCR data; Slc34a2a staining was consistently absent in early stage embryos and became only detectable at 48 hpf. At later stages Slc34a2a localised predominantly to the endoderm and the pharynx (future digestive tract and swim bladder) as well as to the primordial midbrain and hindbrain channel.

Slc34a2a(as) was present from fertilization but was dispersed throughout the head region and did not localise to specific structures until 48 hpf. From 48 hpf onwards, Slc34a2a(as) was found in the pharynx, endoderm, primordial midbrain channel and primordial hindbrain channel. Both the primordial midbrain and hindbrain channels only begin to develop at 24 hpf (prim-5 stage).

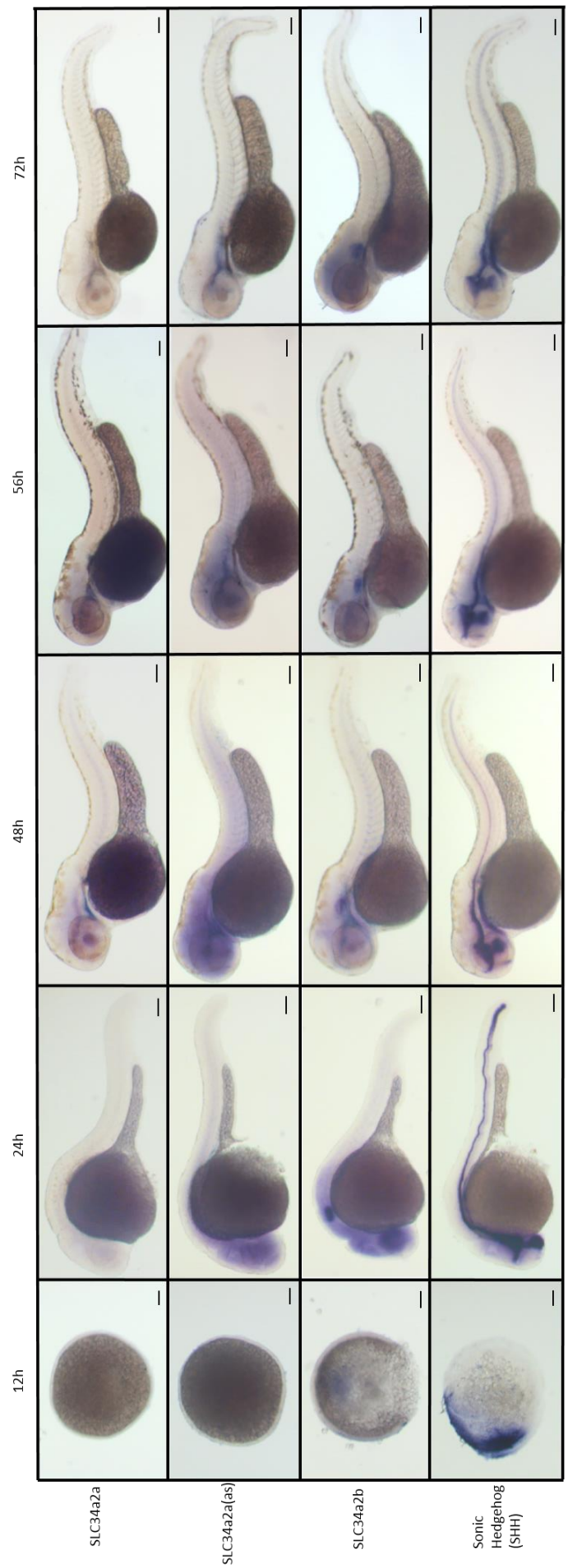


Figure 3.3 Localization of Slc34a2a related transcripts at different time points throughout early zebrafish development. DIG-labelled probes for endogenously expressed Slc34a2a, Slc34a2a(as), Slc34a2b and Sonic Hedgehog (SHH) transcripts were used. SHH was used as a positive control for the in situ hybridization protocol. Negative controls included the DIG labeled complementary probes where possible (not shown). A minimum of 15 embryos at each time point for every transcript were examined. Scale bar at 0.1 mm.

Slc34a2b was found to be expressed from fertilization and was present at all time points examined. At the 8 hpf (75%-epiboly stage), Slc34a2b expression did not seem to be restricted to any specific group of cells. However, by 24 hpf (prim-5), the transcript was visualised throughout the head region with specific localisation to the optic vesicle (eye) and the otic vesicle (ear). From 48 hpf onwards (hatching period), localisation was specific in the otic vesicle and the intestinal bulb (endoderm which is to become the gut) with some very weak but specific staining in the midbrain/hindbrain region. All of these data on Slc34a2b localisation, including the weak brain staining, is in agreement with the ISH data submitted by Thisse, B and Thisse, C to the ZFIN database (Thisse, 2004b).

The presented ISH experiments are in line with the RT-qPCR data in terms of temporal expression profiles; the Slc34a2a(as) staining appeared to decrease at later time points whilst Slc34a2a only appeared at 48 hpf onwards. However, whereas the RT-qPCR data was quantified relative to β-actin, ISH staining only provided non quantitative information about the location of a specific transcript. Nevertheless, the intensity of an ISH signal depends on the quantity of the transcript and the development time; for example, for SHH which is highly expressed during embryogenesis (Yu *et al.*, 2006), staining developed in under an hour whereas staining took approximately 15 hours for the lowly expressed Slc34a2a.

In Slc34a2a(as) probed embryos, as well as in Slc34a2b stained samples at earlier time points, the signal appeared to be unspecific. For several reasons, however, this is highly unlikely and the dispersed staining is indeed transcript specific but not localised to precise cells/organs. Within the final steps of the ISH protocol, stained embryos undergo gradient washes in glycerol prior to being stored permanently in 70% glycerol. The benefit of glycerol washes is that unspecific purple stain is washed out. This would also lessen the purple staining if the embryos were over-stained. An added benefit of the glycerol is that it whitens the yolk for imaging purposes. The high number of embryos assessed by ISH with each probe and the consistent staining suggest that the pattern is specific. A minimum of 15 embryos from three batches of fish were used. The findings reported here are all backed up by the current published work that found comparable dispersed staining (Thisse, B & Thisse, C and Siger et al).

3. Wildtype expression

3.3 Slc34a2a(as) and Rbpja share a bidirectional promoter

Rbpja is located downstream of Slc34a2a and is in a head to head orientation with Slc34a2a(as). To date there is no published literature stating where the promoter of Rbpja is located. Therefore, by using three different online bioinformatics tools and searching for specific DNA sequences manually, the 229 base pairs (bp) spanning between the first exon of Slc34a2a(as) and the first exon of Rbpja, along with 1500 bp either way, was examined to pinpoint a possible promoter location.

Certain DNA sequences, such as the TATA box (sequence TATAAA) which is usually located 25-35bp upstream to the start site, can be indicative for where transcription commences (NatureEducation, 2014). However, looking at the region in and around the first exons of Rbpja and Slc34a2a(as), no TATA box was found. A TATA box was found in the 3rd intron of Rbpja but there was no B recognition element in the vicinity. Downstream or upstream B recognition elements are DNA sequences that allow contact with the transcription factor TFIIB which promotes the assembly of eukaryotic RNA polymerase II (Deng and Roberts, 2006). When searching for other known DNA sequences that typically bind transcriptions factors, only one E-box (also occasionally called a G-box: sequence CACGTG) was located close by to the TATA box. Taken together, this was insufficient evidence to pinpoint a TATA box promoter sequence to either Slc34a2a(as) or Rbpja.

Bidirectional promoters tend not to have a TATA box but instead may have a CCAAT box (Orekhova and Rubtsov, 2013). There are three CCAAT boxes in the relevant area but none of them lie directly in-between the two reported start sites; two boxes are in the second intron of Slc34a2a(as) and one is in the first intron of Rbpja. Bidirectional promoters for non-coding transcripts, especially those which have 5' ends facing one another, have CpG islands in 90% of cases and are significantly enriched with CCG and CGG repeats (Imamura *et al.*, 2004; Uesaka *et al.*, 2014). Both of these qualities were found within the region between Slc34a2a(as) and Rbpja. The CpG island is highlighted in grey and the 12 CCG/CGG regions are bolded in figure 3.4.

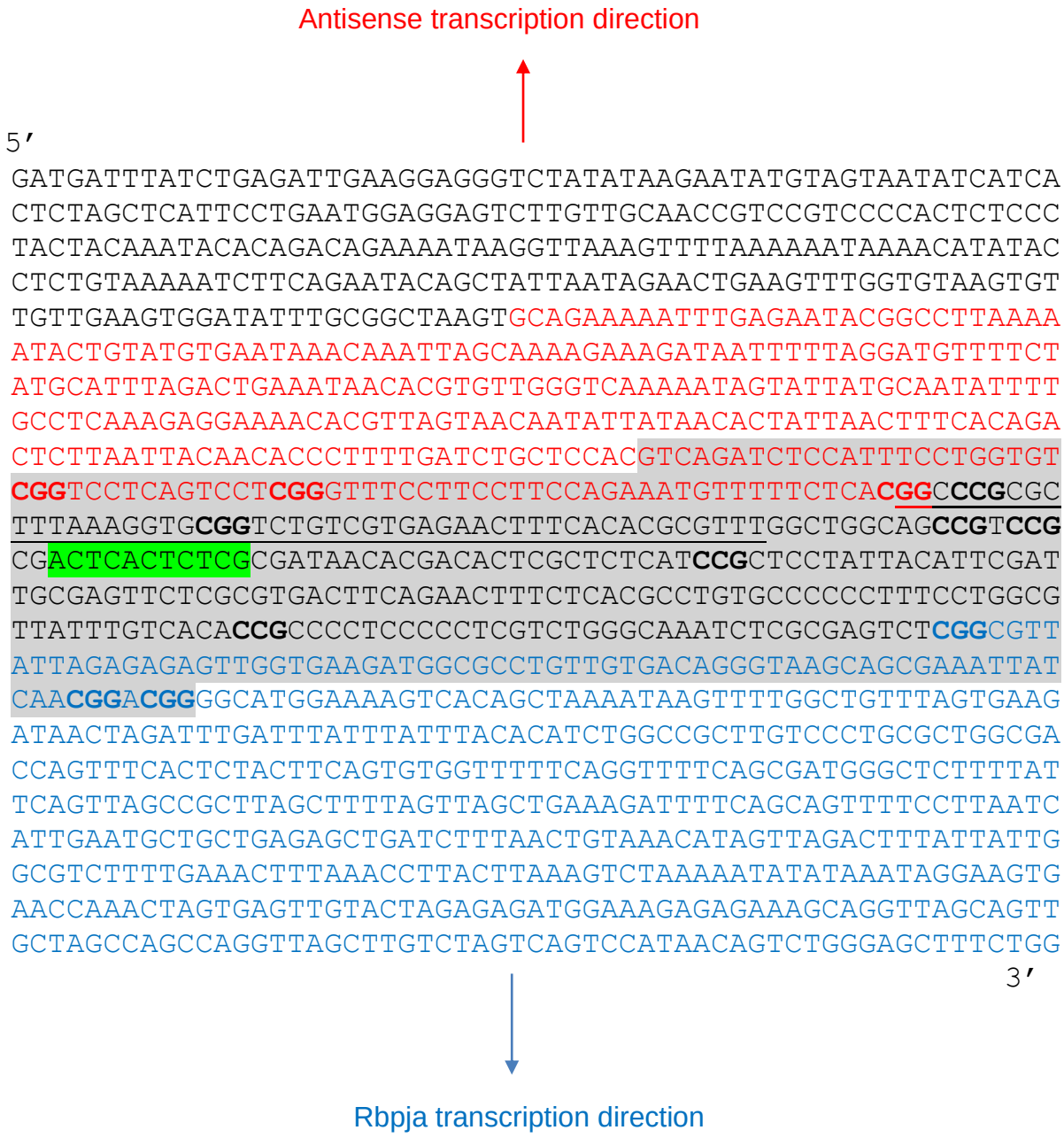


Figure 3.4 Genome sequence from *Danio rerio* showing the first exon of Slc34a2a antisense (red) and the first exon of Rbpja (blue). CCG and CGG repeats, which are enriched in zones around bidirectional promoters, are in bold. The CpG island is highlighted in grey. Eukaryotic promoter predication software has determined the regions which are underlined or highlighted in green to be promoters.

3. Wildtype expression

CpGFinder software confirmed the presence of a single CpG island which was located in between the Slc34a2a(as) and Rbpja sequences (highlighted in grey in figure 3.4; GC%= 54.9) (SoftBerry, 2016). Promoter2.0 prediction software from the Center for Biological Sequence Analysis at the Technical University of Denmark DTU pinpointed a more specific area as the promoter sequence; with a high score of 1.082, the sequence is highlighted in green in figure 3.4 (position score likelihood above 1.0 indicated the predication is highly likely, while below 0.5 is to be ignored) (Knudsen, 2013). The online tool 'Neural Network Promoter Prediction' predicated a promoter region which is underlined in figure 3.4 (Reese, 2015). With a perfect high score of 1.0 on a scale 0-1 for 'minimum promoter score', and both located on forward and reverse strands, this sequence is within the region denoted by Softberry and right next to the predicated site of Promoter2.0.

Both Promoter2.0 and Neural Network Promoter Prediction predict promoter regions in genomes by searching for eukaryotic polymerase II binding sites. The output may differ slightly between the two programs as the relative weighting given to different elements differed within the coding of each online tool. Both programs combine the use of neural networks with algorithms that identify a set of patterns with various distances in the DNA sequence such as TATA-boxes, GC-boxes, and CAAT-boxes. When tested with well documented promoters, Promoter2.0 had a correlation coefficient of 0.63 whilst Neural Network Promoter Prediction had a correlation coefficient of 0.61 (Knudsen, 2013; Reese, 2015).

3.4 Temporal expression levels of Rbpja determined via RT-qPCR

Genes that share a bidirectional promoter are often functionally related and may be co-regulated since promoter modifications, such as DNA methylation or histone modification, would impact both. Therefore, Rbpja transcript levels were examined alongside that of Slc34a2a(as) using RT-qPCR and in situ hybridisation experiments.

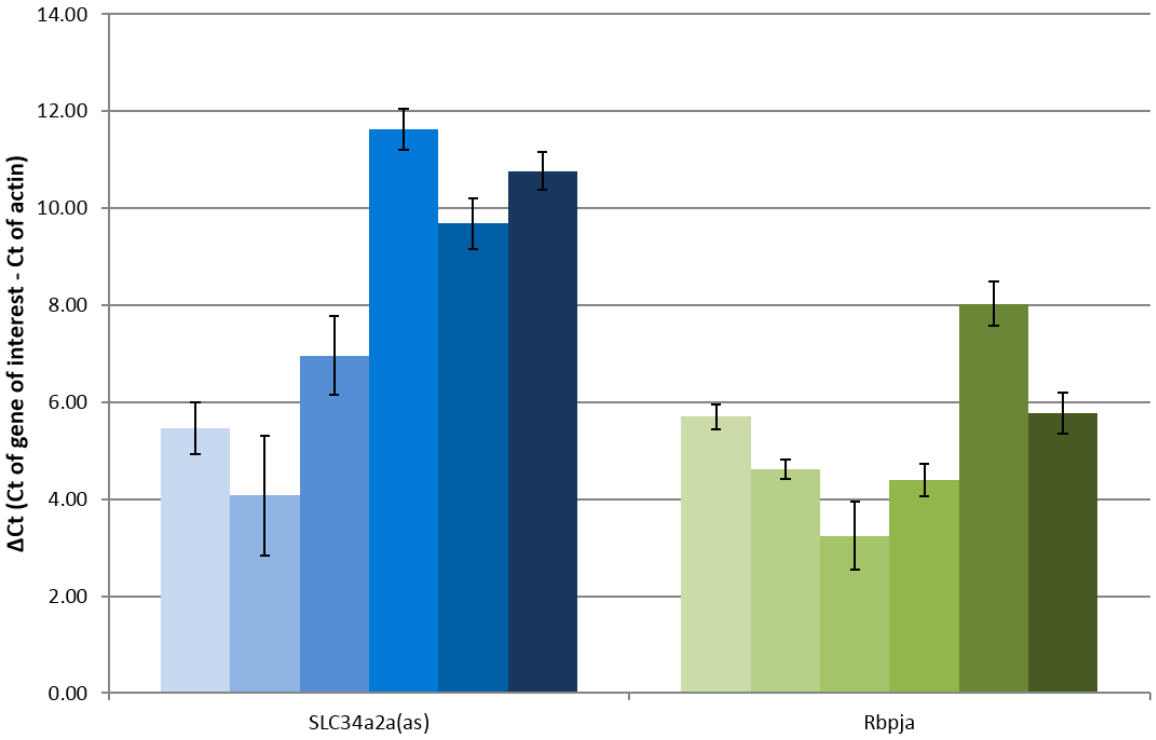


Figure 3.5 Comparison of the Rbpja temporal expression profile to Slc34a2a(as) quantified by RT-qPCR. Expression profiles are in relation to the house-keeping gene β-actin which maintained a stable Ct value. A decrease in bar size designates an increase in transcript level. Sample age, 8, 24, 32, 48, 56 and 72 hpf, is represented as lightest to darkest colour of bars for each transcript. Results are from three independent quantifications from three different clutches are shown. N=3

Detectable from fertilization, Rbpja expression showed a U-shaped pattern with highest relative quantity recorded at 32 hpf. Thus, from 8 hpf to 32 hpf, a decrease in ΔCt values from 5.70±0.33 to 4.61±0.97 (24 hpf) and 3.25±0.49, indicated an increase in transcript levels (individual Ct values of 28.20±0.016, 26.05±0.47, 25.10±0.38 respectively). From 48 hpf, ΔCt values increased from 4.39±0.31 at 48 hpf, 8.03±0.54 at 56 hpf and 5.77±0.52 at 72 hpf, meaning a decrease in transcript quantity (Ct values respectively: 26.69±0.24, 28.03±0.24, 26.52±0.39). These results suggest that Rbpja temporal expression levels, shown in figure 3.5, do not relate to that of Slc34a2a(as).

3. Wildtype expression

3.5 Mapping of Rbpja in comparison to Slc34a2a(as) transcripts by in situ hybridization

Rbpja is expressed from fertilisation with undefined localization in 8 hpf and 24 hpf embryos (figure 3.6). By 48 hpf and onwards, specific staining was found within the otic vesicle (ear) and outlining the posterior of the mesencephalon (midbrain). Rbpja is involved in the Notch signalling pathway and has many functions including adrenal gland, artery and heart development (Echeverri and Oates, 2007; O'Brien *et al.*, 2011; Sacilotto *et al.*, 2013), determining left/right symmetry and polarity during embryogenesis (Sieger *et al.*, 2003; Echeverri and Oates, 2007), angiogenesis (Siekmann and Lawson, 2007) and neurogenesis (Echeverri and Oates, 2007), and therefore was not expected to be localised specifically in the midbrain and ear.

As seen previously with Slc34a2a(as) and Slc34a2b probed embryos, Rbpja staining in embryos at early developmental stages appears to be unspecific. Nevertheless, several lines of evidence confirm that the signal is transcript specific. As mentioned previously, the extensive washes and glycerol storage reduced background unspecific staining. Moreover, the high number of embryos from different clutches that underwent ISH and showed a consistent pattern also supported that this staining was specific. Importantly, current published work on Rbpja (also known as Suppressor of Hairless) confirmed the findings reported here (Echeverri and Oates, 2007).

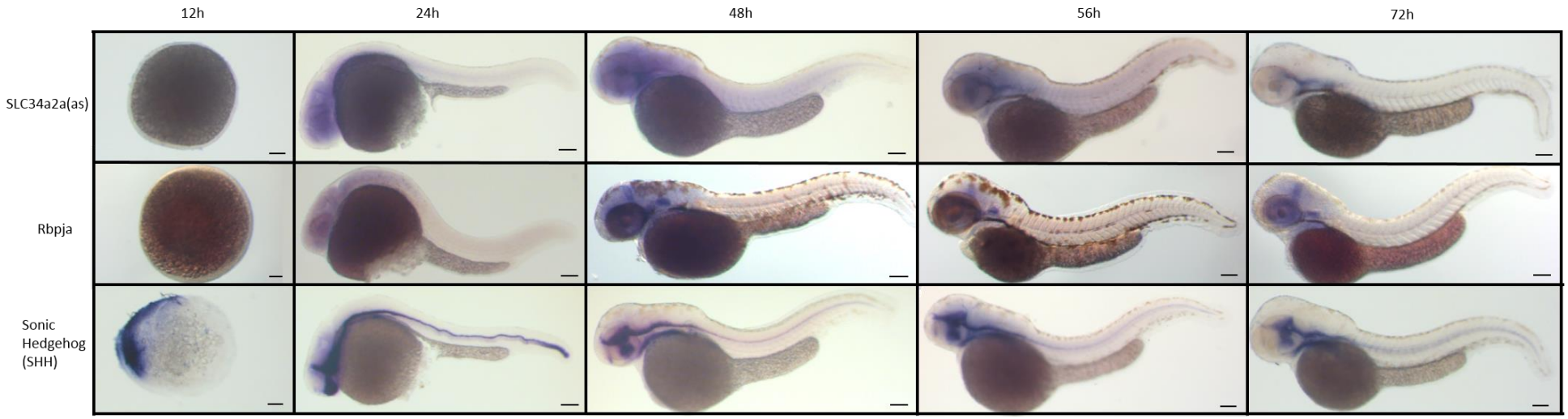


Figure 3.6 Localization of the Rbpja transcript in comparison to Slc34a2a(as) at different time points throughout early zebrafish development. DIG-labelled probes for endogenously expressed Slc24a2a(as), Rbpja and Sonic Hedgehog (SHH) transcripts were used. SHH was used as a positive control for the in situ hybridization protocol. Slc34a2a(as) and SHH samples are the same as those seen in figure 3.3. Negative controls included the DIG labeled complementary probes where possible (not shown). A minimum of 15 embryos at each time point for every transcript were examined. Scale bar at 0.1 mm.

3. Wildtype expression

3.6 Discussion

This chapter gives a comprehensive expression analysis of Slc34a2a and related genes during zebrafish embryogenesis. Previously, Carlile et al. showed that Slc34a2a(as) expression occurs from fertilization up to 48 hpf and a significant decrease thereafter by end point RT-PCR data. The sense transcript (Slc34a2a), on the other hand, was only expressed from 48 hpf onwards. Both transcripts were co-expressed at 48 hpf (Carlile *et al.*, 2008). Expression levels determined with the more sensitive RT-qPCR strategy in this study confirmed that both Slc34a2a and Slc34a2(as) transcripts were present in 48 hpf embryos and a reciprocal expression pattern was established with a gradual shift in expression during the examined 72 hours of development.

Different types of relationships can exist between *cis* and *trans* NAT pairs in eukaryotes; concordant, reciprocal (discordant) and occasionally they are permanently co-expressed (Knee *et al.*, 1994; Vanhée-Brossollet *et al.*, 1995; Lehner *et al.*, 2002; Nakaya *et al.*, 2007). These publications suggest that antisense transcript mediated gene regulation is likely to occur when sense and antisense RNAs are reciprocally expressed. For example, the sense transcript for infected-cell protein number zero (ICP0) is only found in humans suffering with an acute infection of herpes simplex virus type 1, whilst the presence of the antisense ICP0 transcript is believed to play a role in inducing latency (Croen *et al.*, 1987). In a patient with inherited α-thalassemia a genomic deletion repositioned the constitutively active LUC7 gene closely downstream of the HBA2 gene. The resulting antisense transcript of the α-globin gene (HBA2) transcriptionally silences the HBA2 gene by inducing the methylation of a CpG island in the HBA2 promoter (Tufarelli *et al.*, 2003). Eukaryotic transcription of initial factor 2 alpha (elF2α) and its cis-NAT have a discordant relationship at different cell cycle stages and if the NAT promoter is deleted or mutated, a sharp increase in sense RNA is detected in T cells (Silverman *et al.*, 1992). In chick embryos, a large (>10 kb) cis-NAT is responsible for the down-regulation of the alpha 1 collagen gene in chondrocytes (Farrell and Lukens, 1995).

The qPCR data in this chapter confirms a reciprocal temporal expression pattern for Slc34a2a and Slc34a2a(as) transcripts. Just as in some of the examples mentioned above, this may suggest that antisense-mediated Slc34a2a gene regulation is occurring via RNA hybrid formation during zebrafish embryogenesis. The possibility

of antisense-mediated Slc34a2a transcript control is further supported by the finding that both transcripts localise to the same tissues including the pharynx, endoderm, primordial midbrain channel and primordial hindbrain channel in roughly the same quantities (ΔCt Slc24a2a 11.78±0.52, Slc24a2a(as) 11.62±0.33). No literature published to date has demonstrated the localisation of Slc34a2a or its antisense transcript in developing zebrafish. However, as Slc34a2a is found in healthy digestive tract and kidney tissues (Graham *et al.*, 2003), it was as hypothesised that Slc34a2a would be found in the endoderm.

β-actin was chosen as the reference gene for the above relative quantitative PCR experiments for its high and steady expression level throughout zebrafish embryo development (Casadei *et al.*, 2011). Quantitative analysis of RNA extraction from 23-24 zebrafish embryos has shown that β-actin quantity remains at a stable level of 54.4± 23.9 ng throughout embryo/larvae stages (Casadei *et al.*, 2011). Initially, use of one embryo for the above RT-qPCR was attempted but this proved difficult as the obtained values were close to the detection limit, making results unreproducible. Consequently five embryos were used. Therefore, based on the above results, for five embryos at 48 hpf, Slc34a2a, Slc34a2a(as), Slc34a2b and Rbpja levels are considerably less than 6.49 to 16.66 ng range. Fold change or ΔΔCt could not be calculated as samples were taken only from one condition (wildtype embryos maintained under normal developmental conditions).

Slc34a2b, which is transcribed from chromosome 23, has a nucleotide sequence identity of 68.4% to Slc24a2a, which is transcribed from chromosome 1 (EMBL-EBI, 2016). No relationship between Slc34a2b and Slc24a2a expression was expected as the transcripts are transcribed from different chromosomes and therefore are driven by different promoters. No published works have reported a correlation between Slc34a2a and Slc34a2b expression levels and the above qPCR data endorses independent expression of the two genes.

The nucleotide sequence identity between Slc34a2b and Slc24a2a(as) is 37.7% and the longest stretch of complementary base pairs encompasses 11 bp. Therefore, hybrid formation between Slc34a2b and Slc24a2a(as) would not support processing into endo-siRNAs nor recognition by ADARs (Carlile *et al.*, 2008; Barraud and Allain, 2012). Thus no expression pattern relation was expected and none was found in RT-qPCR. In situ hybridization, which found that the Slc34a2b and Slc24a2a/

3. Wildtype expression

Slc34a2a(as) transcripts do not locate to the same tissue in developing embryos, further supports the idea that Slc34a2b and Slc34a2a(as) RNA transcripts are unlikely to interact.

Eukaryotic promoters are difficult to characterise since there are many diverse elements to take into account. Gene promoters tend to be 10-40 base pairs away from the transcriptional start site but regulatory elements can be several kilobases apart making matters considerably more complicated. In the human genome, nearly 11% of genes share a bidirectional promoter (Trinklein *et al.*, 2004). These promoters are usually short (less than 1000 base pairs) and are found between the 5' ends of two genes (Orekhova and Rubtsov, 2013). Even though more than 70% of the human genome is transcribed, only 2% is protein coding, indicating that most transcription is non-protein coding (Consortium, 2007; Consortium, 2012). Recent studies have shown that non-coding RNA transcripts are often linked to the promoters of protein-coding genes (Wei *et al.*, 2011).

Accordingly, many indicators for bidirectional promoter activity were found in the 229 bp flanked by Rbpja and Slc24a2a(as), including a CpG island, an accumulation of CCG and CGG repeats, and a CCAAT box. The region between the protein-coding Rbpja gene and the non-coding transcript Slc34a2a(as) can be confidently proposed as a bidirectional promoter. This statement could be validated with the completion of reporter gene studies.

Genes that share a bidirectional promoter are often functionally related and tend to be co-regulated since any modification, such as DNA methylation or histone modifications, to their shared promoter would impact both (Shu *et al.*, 2006). Rbpja RNA levels were examined by qPCR and a U-shaped temporal expression pattern was found. This Rbpja expression pattern, however, did not correlate with the steady decline of Slc24a2a(as) transcript levels.

Even though certain genes that share bidirectional promoters undergo simultaneous divergent transcription and are co-expressed, such as many of those that are involved in DNA repair, this is not the rule for all bidirectional promoters (Xu *et al.*, 2012). Nucleosomes can shift transcription from one strand to the other, or distant control elements and cognate transcription factors can affect the direction based on being expressed or repressed (Trinklein *et al.*, 2004). This is in line with the RT-

qPCR results which indicate that Rbpja and Slc24a2a(as) are not simultaneously present in equal amounts.

In situ hybridization confirmed Rbpja expression from fertilisation with undefined localization in 8 hpf and 24 hpf embryos (figure 3.6). Comparable dispersed staining was reported by Sieger et al. when mapping Rbpja in the early stages of development (Sieger *et al.*, 2003). The experiments reported in this chapter established specific staining within the otic vesicle (ear) and it the posterior of the mecencephalon (midbrain) 48 hpf and onwards. However, considering Rbpja's many functions, the localisation in the ear was unexpected. It was hypothesised that Rbpja would rather localise to the heart region, the endoderm and the brain due to its involvement in adrenal gland, artery and heart development (Echeverri and Oates, 2007; O'Brien *et al.*, 2011; Sacilotto *et al.*, 2013), determining left/right symmetry and polarity during embryogenesis (Sieger *et al.*, 2003; Echeverri and Oates, 2007), angiogenesis (Siekmann and Lawson, 2007) and neurogenesis (Echeverri and Oates, 2007). It may be possible that Rbpja has yet another role in zebrafish development that has not yet been characterised which is why it localises to the otic vesicle or mayhap Rbpja is stored in otic vesicle cells until required. No published data is available for Rbpja expression past 24 hpf (prim-5 stage). Nevertheless, in the presented experiments Slc34a2a(as) and Rbpja probes never localised to the same regions within the embryos. These findings indicate that even though the two genes share a bidirectional promoter, simultaneous, divergent transcription is not occurring within the same cells; their individual transcription is tissue specific. Thus, nucleosome placement or additional control elements may have a role in driving the expression of Slc34a2a(as) and Rbpja.

The RT-qPCR experiments show a significant decrease of Rbpja and Slc34a2b expression in zebrafish at 56 hpf. These embryos are within the hatching period where growth rate has slowed minimally but organ rudiment morphogenesis has neared completion outside of gut and related organs (Kimmel *et al.*, 1995). Consequently, there is no specific embryonic developmental event that would predict the decrease seen in Slc34a2b and Rbpja RNA levels. As this decrease is not visible with Slc34a2a or Slc34a2a(as) transcript levels, the incident is unlikely due to the experimental procedures.

3. Wildtype expression

3.7 Conclusion

The aforementioned RT-qPCR and ISH data are consistent in temporal expression patterns for all transcripts examined. Even though Slc34a2a(as) and Rbpja share a bidirectional promoter, the two do not undergo simultaneous divergent transcription as they do not localise to the same embryonic structures. Slc34a2a and Slc34a2a(as) transcripts demonstrate a reciprocal relationship and were found within the same developing organs. It is conceivable that the two complimentary transcripts control each other's expression like the examples of infected-cell protein number zero (ICP0) and initial factor 2 alpha (eIF2α) in humans. Whether Slc34a2a(as) is controlling Slc34a2a levels, or vice versa, by generating dsRNA and initiating RNAi requires interference studies affecting the natural balance of endogenous Slc34a2a/ Slc34a2a(as) expression levels.

Chapter 4. Interfering with the natural Slc34a2a sense and antisense balance leads to cerebellar loss

In the previous chapter characterising wildtype expression, it was found that Slc34a2a and its antisense transcript showed a temporal reciprocal expression pattern. At time-points when both were expressed, they located to the same regions of the developing embryo: the pharynx, endoderm (future digestive tract and swim bladder), primordial midbrain channel and the primordial hindbrain channel. For genes where the natural antisense transcripts show reciprocal expression, it has been widely demonstrated that the complementary transcripts control each other's expressions levels (Kraus *et al.*, 2013; Zong *et al.*, 2016) . Therefore this chapter aims to investigate the consequences of a disturbed sense/ antisense balance on embryonic development. The time frame of embryogenesis is important because of the significant expression changes to both sense and antisense transcripts occurring as demonstrated in chapter 3.

4.1 Phenotype classification of zebrafish embryos

The following subsections describe experiments where endogenous RNA levels of embryos were manipulated via microinjection. A simple classification system, with a scale from 1-5, was created to consistently log the phenotypic changes for a high number of injected embryos during early development.

Level 1, as seen in figure 4.1, describes wildtype embryos as observed in non-injected fitness controls or in water/phenol red injected control embryos. These embryos have the correct length (24 hpf length ~1.9 mm, 48 hpf length ~3.1 mm) and anatomical features as stated by Kimmel et al (Kimmel *et al.*, 1995).

Level 2 classified embryos appear like wildtype fish but on closer inspection of the morphology, a single developmental defect can be identified. In the case of the level 2 24 hpf embryo pictured in the top row of figure 4.1, the 'back' is bent inwards whilst

4. Cerebellar loss

the 48 hpf embryo (bottom row) has not yet straightened out which in turn shortens its overall length. Other examples of individual defects that can lead to level 2 classification include heart jogging in the wrong direction, incorrect heart rate, blunt tail tip or the lack of fin buds (at 48 hpf).

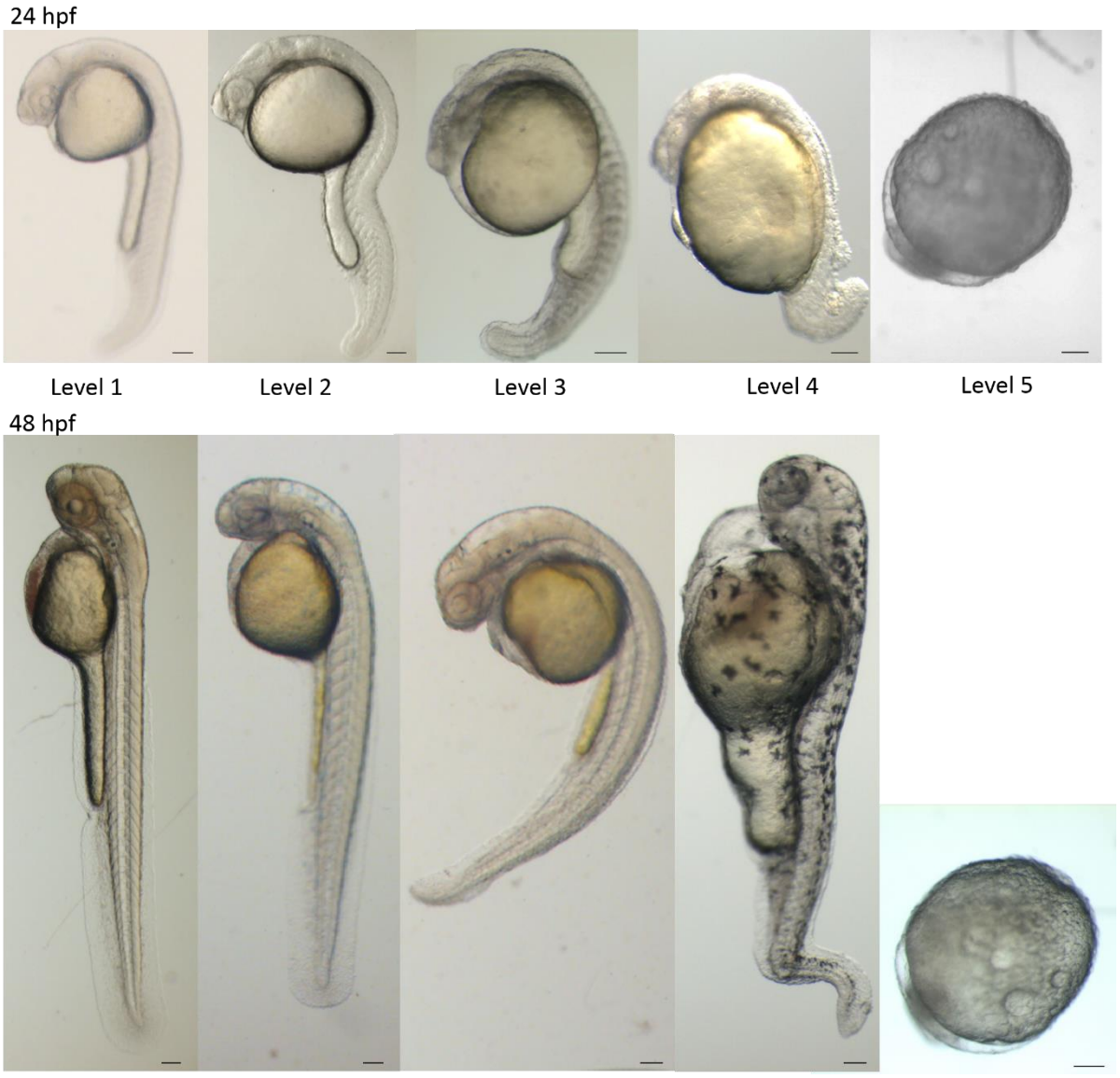


Figure 4.1 Danio rerio embryos at 24 hpf (prim-5 stage) (top row) and at 48 hpf (long-pec stage) (bottom row) separated into a level classification system based on phenotype severity. Level 1; wildtype. Level 2; one developmental defect. Level 3; two or more developmental defects but still recognizable as a fish embryo. Level 4; severely disfigured and barely recognizable as a fish embryo. Level 5; failure to develop passed the blastula period. Scale bar at 0.1 mm.

Embryos classified as *level 3*, have multiple developmental defects but are still recognizable as zebrafish embryos. In figure 4.1, the level 3 24 hpf embryo in the top row, has neither initiated eye nor brain development and has a significantly shortened and curved body. The level 3 48hpf embryo (bottom row figure 4.1) has a significantly shortened and curved body. In addition, it shows an enlarged area above the rhombomeres and a flattened face.

Level 4 embryos are barely recognizable as fish embryos and therefore lack numerous features and organs whilst *level 5* embryos never developed past the blastula period (figure 4.1).

In addition to the morphological phenotype, heart rate and basic behaviour were taken into account. The heart rate for a healthy wildtype embryo within the first three days of development is 120-160 beats per minute (De Luca *et al.*, 2014). A reaction test was completed by prodding the embryo with forceps to screen for abnormal behaviour. A wildtype embryo would not tolerate being touched and would quickly swim away to find safety in either the edge of the petri dish or in a group of embryos on the other side of the dish. Fish with an abnormal behavioural phenotype could be poked either numerous times and would only swim away after a prolonged delay and/or would escape only a short distance or swim in a small circle only to end up in the original area (in danger of another poke).

4.2 Microinjection of full length capped Slc34a2a RNA causes developmental deformities

Full length capped Slc34a2a, Slc34a2b and Slc34a2a(as) *in vitro* synthesized transcripts were created as stated in Chapter 2. Different dilutions of each RNA were injected into wildtype AB or Golden 1-4 cell stage embryos. Phenol red was used as a marker to visualize the injection droplet. Therefore, phenol red mixed with water was used as negative control for the injection procedure. Morphological phenotypes and behaviour were logged at both 24 and 48 hpf. Data were only collected from clutches whose fitness non-injected control group (min 50 embryos) exceeded 80%

4. Cerebellar loss

survival in the first 24 hours of development. Examined embryos came from a minimum of three clutches for each injected sample.

Based on current published work, anywhere between 25-500 pg of RNA material has been microinjected into zebrafish embryos to produce a specific effect (Person *et al.*, 2010; Kim *et al.*, 2011; Lu *et al.*, 2011; Dahlem *et al.*, 2012). Therefore, in a first set of experiments, 287 embryos were injected with 165 pg of Slc34a2a RNA. The sense transcript Slc34a2a is not expressed within the first 48 hours of development. Thus it was hypothesised that injection of exogenous Slc34a2a RNA into fertilized eggs would mimic ectopic co-expression of complementary sense and antisense Slc34a2a transcripts during early development.

All 165 pg Slc34a2a RNA injected 24 hpf embryos appeared developmental delayed upon first inspection. Although brain sections are fully formed by 24 hpf in wild type fish (Kimmel *et al.*, 1995), the embryos injected with Slc34a2a RNA showed no sign of brain development since no segment division was visible. Alongside indistinguishable brain features, the majority of embryos had developmental defects of the heart, yolk, eyes, body and/or tail. Due to the apparent overall developmental delay at 24 hpf, most of the data presented was based on 48 hpf embryos since all organs and features were discernible (figure 4.2).

At 48 hpf (long-pec stage) 60 (20.9%) injected embryos were classified as wildtype as seen in figure 4.2. The majority of embryos were classified as level 2 (68 embryos, 23.6%) or level 3 (101 embryos, 35.1%). Phenotype deformities amongst these level 2/3 embryos varied considerably from shortened body length, smaller heads, improperly sized and/or shaped eyes, blunt tails, improper yolk shape, and/or incorrect heart development (wide-ranging heart rates). If any brain formation was present, proportions were incorrect and/or midbrain -hindbrain regions were missing. A significant proportion of 165 pg Slc34a2a RNA injected embryos were severely affected (stage 4; 39 embryos, 13.5%) including embryos that were completely missing upper bodies, cyclops with a twisted curly tail and even a two headed embryo was found. A large number of embryos did not progress past the blastula stage (19 embryos, 6.6%). All embryos that were stage 2 to 4, were noted to be significantly less active and did not react appropriately to stimuli, i.e. did not swim away when prodded (Appendix B, videos). Especially those with tail or yolk defects could not swim straight or in an upright position.

The severity and variety of the phenotypic deformities and behavioural changes seen in embryos injected with 165 pg Slc34a2a RNA suggested that the effects were mostly unspecific. Hence the amount of injected RNA was reduced to 110 pg and 82.5 pg, respectively (figure 4.2).

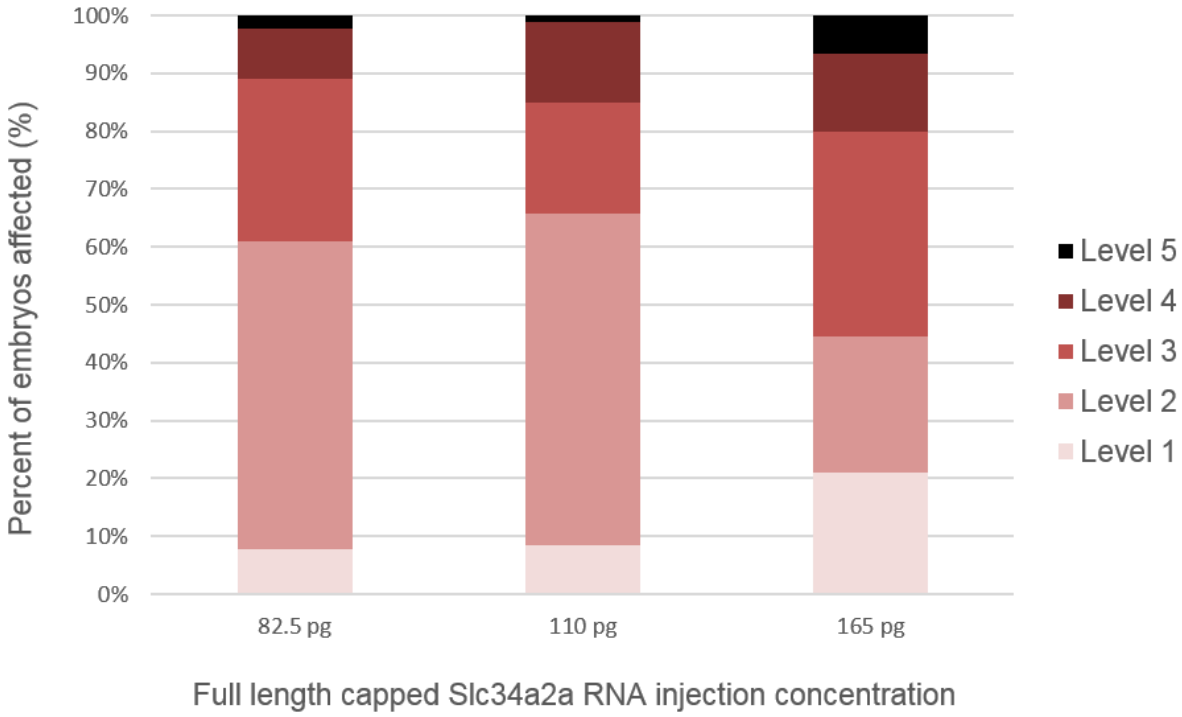


Figure 4.2 Compiled results of microinjections of full-length capped Slc34a2a RNA transcript into fertilised zebrafish eggs. Three different amounts of RNA were injected. The embryos were analysed at 48 hpf and scored according to the 1-5 classification introduced above. The percentage of embryos with a specific degree of deformation (1-5) are presented as a bar graph. A minimum of 247 embryos from three clutches were analysed for each injected amount of RNA.

247 48 hpf embryos were injected with 110 pg Slc34a2a RNA and were scored as follows: 21 embryos (8.4%) were classified as level 1 (wildtype) as shown in figure 4.2. The majority were classified as level 2 (142 embryos, 57.5%) or level 3 (47 embryos, 19.2%). 34 embryos (13.8%) were severely disfigured and thus classified

4. Cerebellar loss

as level 4 and 3 did not develop past blastula stage (level 5, 1.2%). In comparison to the embryos injected with 165 pg Slc34a2a RNA, embryos at level 2/3 of the 110 pg injected group appeared to experience less severe developmental malformations. Deformations included improper segmentation of the brain, complete loss of cerebellum and shortened, bent or kinked tail. These embryos too had slower reaction times to stimuli and seemed overall less active in comparison to their non-injected or water injected control groups. Those with bent or kinked tails were unable to swim straight or in an upright position.

When the amount of Slc34a2a RNA was reduced to 82.5 pg, 25 (7.8%) of 307 embryos examined were classified as level 1 and over half were level 2 (163 embryos, 53.1%). 147 of the level 2 embryos shared the same unique malformation, a complete loss of the cerebellum. Only 16 embryos differed in their one developmental malformation including missing eyes, heart jogging to the right instead of the left, or a bent or shortened tail. The majority of level 3 classified embryos (86, 28.1%), shared the loss of cerebellum alongside another deformation like bent tail, curved body or shortened tail length. The remaining embryos were either classified as level 4 (26 embryos, 8.6%) or level 5 (7 embryos, 2.3%) as visualised in figure 4.2. All embryos classified as level 2-4 showed slower reaction times in the poke test as visualized in the attached videos (Appendix B).

The highest percentage of embryos classified as level 1, phenotypically ‘normal’, were found in the group injected with the highest quantity of Slc34a2a RNA (20.85% of 165 pg injected) as compared to the 8.38% of the 110 pg injected and 7.81% of 82.5 pg injected groups, as depicted in light pink in figure 4.2. This result was unexpected as the remaining 165 pg injected embryos had severe malformations that were easily recognizable. The majority of embryos injected with 82.5 pg or 110 pg Slc34a2a RNA were classified as level 2. Upon first inspection of embryos injected with 110 pg, in comparison to the 165 pg injected groups, the embryos looked affected but were not as severely malformed. In contrast, a first general assessment of embryos injected with 82.5 pg of Slc34a2a RNA, majority appeared normal. Further detailed inspection of individual fish revealed that they shared the same phenotypic imperfection with the cerebellum missing as shown in figure 4.3.

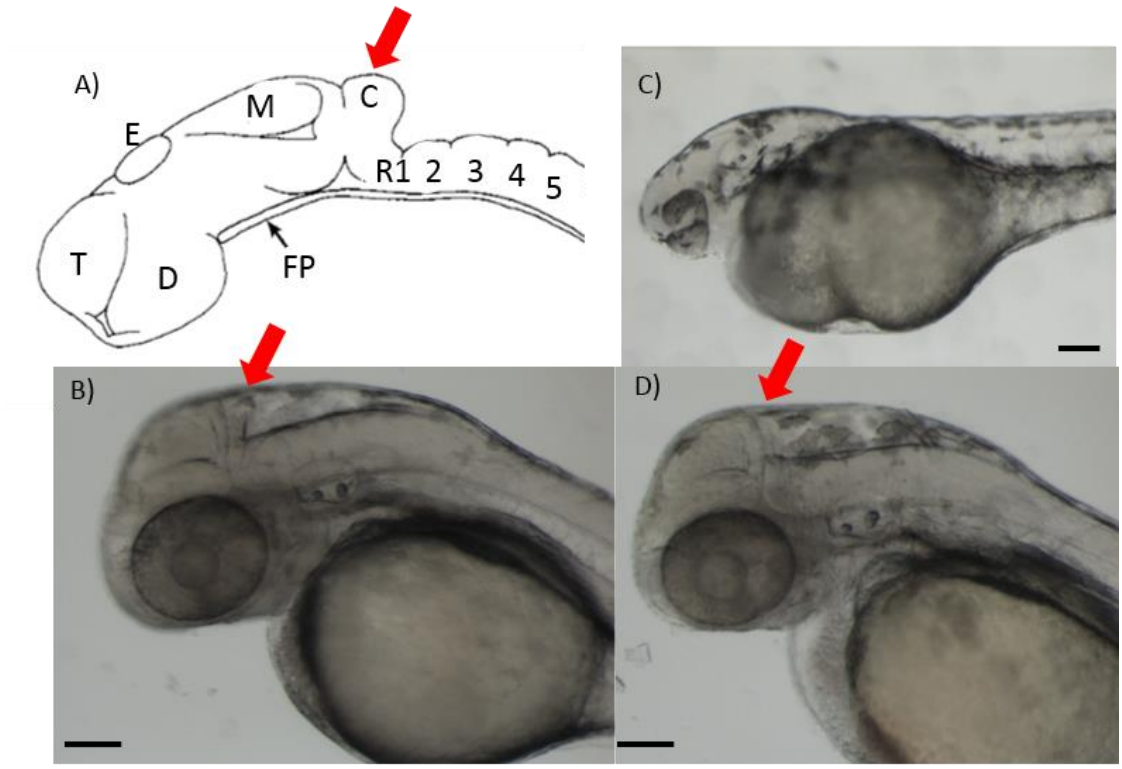


Figure 4.3 Zebrafish embryos at 48 hpf (long- pec stage). A) Representation of the different segments of the zebrafish brain taken from Kimmel et al (1995); FP- floor plate, D- diencephalon, T- telencephalon, E- epiphysis, M- mesencephalon, C- cerebellum and R (1-7) are the seven hindbrain rhombomeres. B) Fitness control: the cerebellum is clearly visible (red arrow). C) 110 pg Slc34a2a transcript injected embryos showing severe brain and eye abnormalities. D) 82.5 pg Slc34a2a injected embryos lacking cerebellar development (red arrow). Scale bar at 0.1 mm.

Slc34a2b is a paralog of Slc34a2a and as the two share 68.4% nucleotide sequence identity, Slc34a2b RNA microinjections were completed alongside Slc34a2a. From the examined 291 embryos injected with 165 pg of Slc34a2b RNA, 223 were classified as wildtype level 1 (76.6%), 44 were level 2 (15.0%), 16 were level 3 (5.6%), 3 were severely deformed at level 4 (0.9%) whilst 5 never developed past the blastula stage (level 5, 1.9%). These values are represented as percentages in stack column format in figure 4.4. In contrast to Slc34a2a injected embryos, the 165 pg Slc34a2b controls that were classified as level 2 or 3, did not have consistent

4. Cerebellar loss

malformations. Deformations included body curvature, tail length and shape, eye size and form, size of ‘nose’ and heart rates. When the amount of Slc34a2b RNA was reduced to 110 pg, the bulk of the 247 embryos assessed were classified as level 1 (194 embryos, 78.6%). Of the remaining embryos, 14 were level 2 (5.7%), 14 were level 3 (5.7%), 5 were level 4 (2.1%) and 20 were level 5 (7.9%). No consistent phenotypic deformations were visible and the malformations noted were comparable to the 165 pg Slc34a2b RNA injected group. As these two experiments did not produce a specific phenotype, the injected amount of Slc34a2b was not further reduced.

Capped full length Slc34a2a(as) RNA microinjection was also completed to determine whether overexpression of the antisense transcript would interfere with normal embryonic development. Slc34a2a(as) RNA injection was completed with 110 pg of RNA into fertilised oocytes from three clutches. Of the 94 embryos examined, not a single one was found to be affected; thus 100% at level 1 as depicted in the 100% stacked column in figure 4.4.

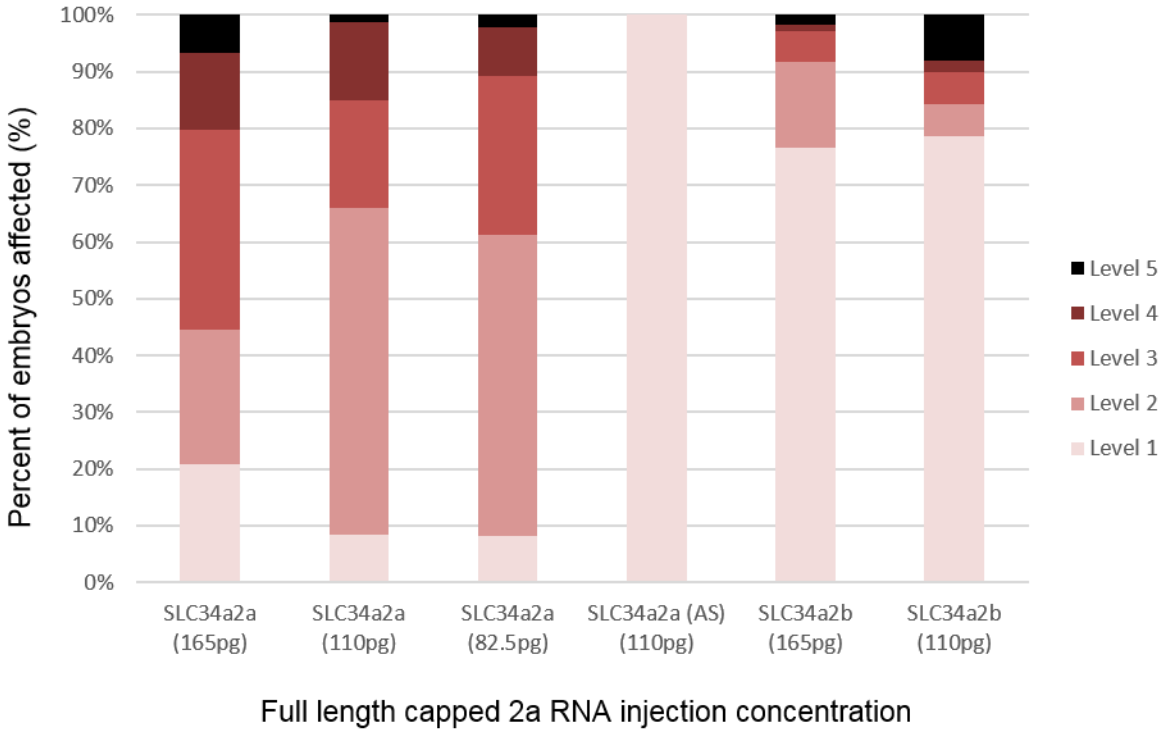


Figure 4.4 Compiled results of microinjections of all full-length capped RNA transcripts into fertilised zebrafish eggs. The resulting embryos were analysed at 48 hpf and the number of embryos with a specific degree of deformations (1-5) are presented as a 100% stacked bar graph. The bars represent the percentage of a specific phenotype. A minimum of ninety embryos from three clutches were used for these calculations. A total of 1186 embryos were analysed; 554 injected with Slc34a2a, 538 injected with Slc34a2b and 94 with Slc34a2a(as) RNA.

4.3 Capping of microinjected RNA transcripts is not necessary

Capped and uncapped *in vitro* synthesized RNA transcripts were produced for the sense transcript Slc34a2a. Upon microinjection, comparable phenotypic changes were observed at 48 hpf (figure 4.5). The majority of embryos injected with capped and uncapped 82.5 pg Slc34a2a RNA showed the same characteristic loss of cerebellum and were classified as level 2.

4. Cerebellar loss

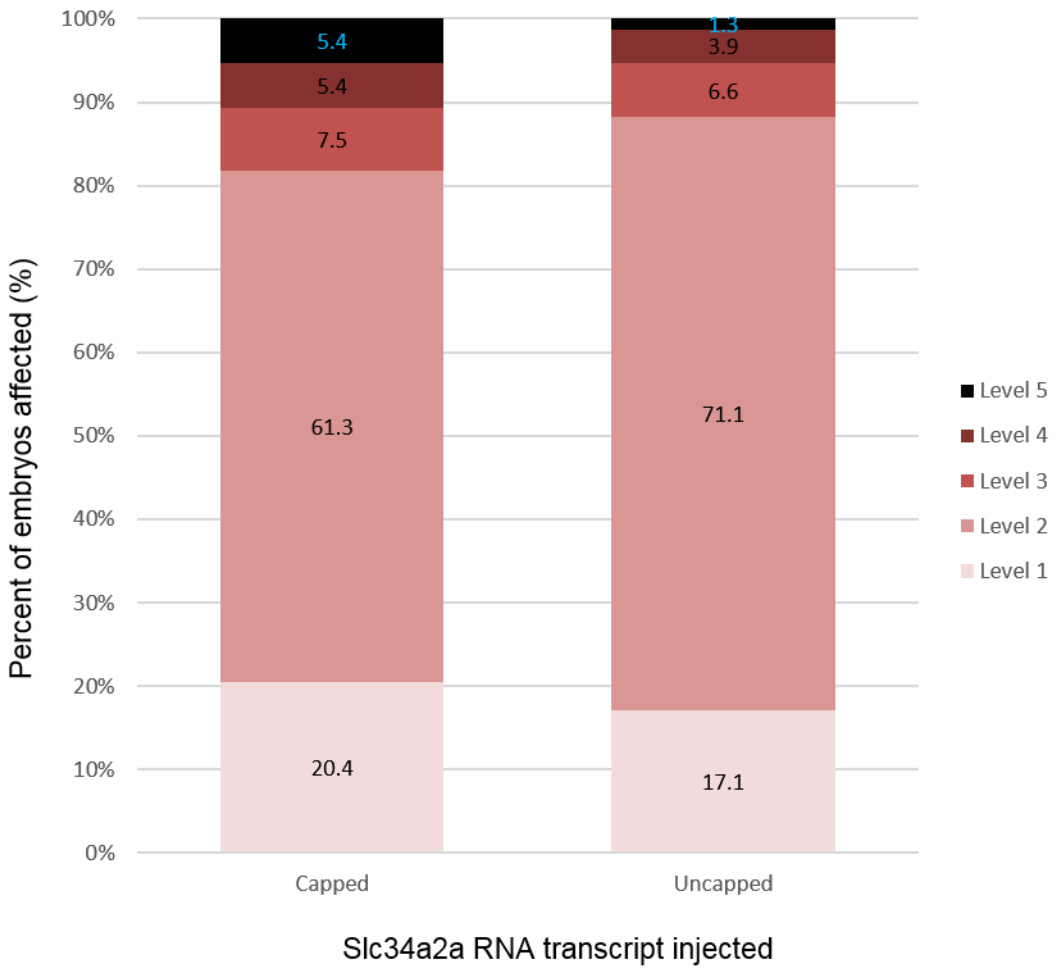


Figure 4.5 Capping of *in vitro* synthesized Slc34a2a RNA transcript is not required to induce phenotypic malformations. At 48 hpf, loss of cerebellum was detected in the majority of injected embryos whether capped or uncapped Slc34a2a RNA was microinjected. 76 embryos from three clutches were tested with uncapped Slc34a2a RNA.

4.4 In situ hybridization with the specific marker Eng2 confirms the loss of cerebellum

To confirm the loss of the cerebellum, in situ hybridization (ISH) was completed using a specific marker for the cerebellar region of the brain. A plasmid containing the cDNA sequence for the homeobox protein Engrailed-2 (Eng2) was obtained from Prof. C. Houart (Kings College London) from which both forward (T7) and reverse

probes (T3) were synthesized as mentioned in Chapter 2. No staining was visible with the T7 control probes and results for T3 probes are presented in figure 4.6.

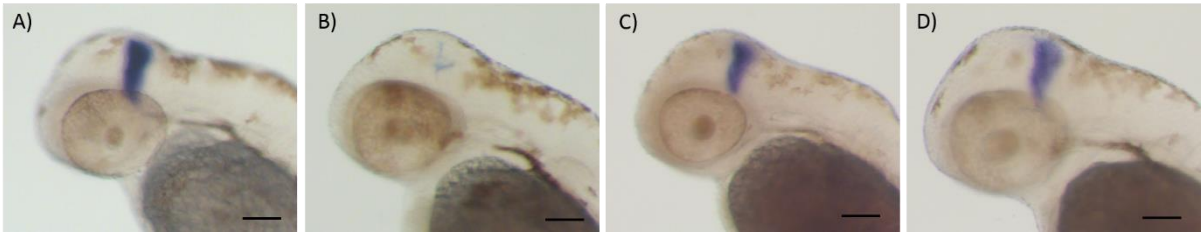


Figure 4.6 Engrailed-2 probe used in whole-mount in situ hybridization to show the loss of cerebellum at 48 hpf after full length RNA transcript microinjection. A) wildtype (non-injected control) B) 110 pg Slc34a2a injected C) 110 pg Slc34a2a(as) injected D) 165 pg Slc34a2b injected. Scale bar at 0.1 mm.

Zebrafish Engrailed-2a (Eng2) is transcribed from chromosome 7 and the protein is located within the nucleus. Expression commences during the blastula stage (~5-10 hpf) and continues throughout embryogenesis. As it is vital for proper midbrain-hindbrain boundary development, expression can be found in the midbrain-hindbrain furrow (midbrain-hindbrain boundary) and the cerebellum (Fjose *et al.*, 1992; Thisse, 2005; D'Aniello *et al.*, 2013). Orthologs in fruit flies, mice and humans also regulate pattern formation during central nervous system development and a knock-out of Eng2 is lethal (Hanks *et al.*, 1995; Hanks *et al.*, 1998).

Eng2 was therefore used to corroborate the effect of Slc34a2a RNA microinjections on cerebellar development. As seen in figure 4.6 A, the midbrain-hindbrain region, including the entire cerebellum, stained as a thick purple wedge in the wildtype non-injected sample. Comparable staining was visible within embryos injected with 110 pg Slc34a2a(as) RNA (Fig 4.6, C) and 165 pg Slc34a2b RNA (Fig 4.6, D). However, in embryos injected with 110 pg Slc34a2a RNA (Fig 4.6, B), the cerebellar Eng2 staining was almost completely absent indicating a complete loss of cerebellum and a widening of the midbrain-hindbrain boundary furrow (indicated by the sliver of purple pointing backwards).

4. Cerebellar loss

4.5 Loss of cerebellum does not alter cranial motor neuron development (in Islet1-GFP transgenic zebrafish)

In order to test whether Slc34a2a RNA expression in the first 48 hours of embryogenesis affected other less evident brain structures, Slc34a2a RNA microinjection was complete in Islet1-GFP embryos. Islet1-GFP transgenic fish express green fluorescent labelled Islet1 (Isl1), which is located in all post mitotic central nervous system motor neurons (Higashijima, 2008). Islet1-GFP zebrafish fertilized eggs were used to establish if the loss of cerebellum altered cranial motor neuron development. As demonstrated in figure 4.7, the loss of cerebellum in 82.5 pg Slc34a2a RNA injected embryos (D) caused neither a change in motor neuron development nor arrangement as compared to the non-injected controls (C). However, it is important to note that Isl1 containing neurons are neither expressed in the mesencephalon nor the cerebellum but localise below these brain segments within the rhombomeres (Fig 4.7, A).

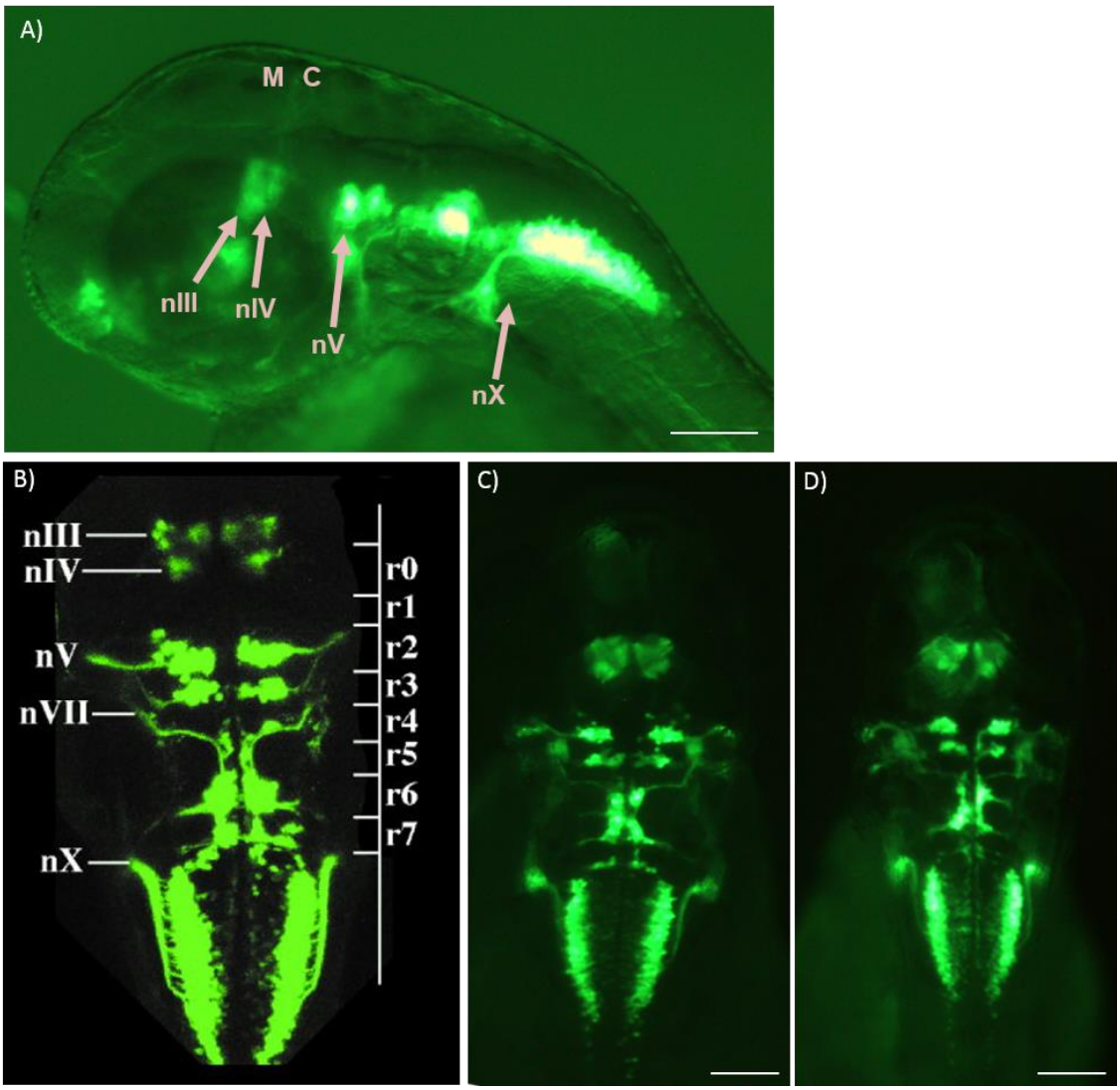


Figure 4.7 Loss of cerebellum due to Slc34a2a RNA microinjection had no impact on CNS motor neuron development in 48 hpf Islet1-GFP transgenic zebrafish. A) Lateral view of a wildtype embryo to show the cranial nerve fluorescence relative to the mesencephalon (M) and cerebellum (C). B) Dorsal view of wildtype embryos depicting cranial nerve arrangement in relation to brain segmentation where r0-7 indicate the rhombomeres. Figure taken from (Moens and Prince, 2002). C) Dorsal view of wildtype non-injected control embryo. D) Dorsal view of an embryo injected with 82.5 pg Slc34a2a RNA. Scale bar at 0.1 mm.

4. Cerebellar loss

4.6 Expression profiles of Slc34a2b, Slc34a2a sense and antisense in microinjected embryos

Due to the possibility of interaction and cross regulation between all the transcripts of interest, endogenous and exogenous levels of Slc34a2a, Slc34a2a(as) and Slc34a2b were monitored by semi quantitative RT-qPCR and in situ hybridization. RNA extraction was performed on 10 and 24 hpf samples for RT-qPCR analysis due to the likelihood of biological interference occurring early after injection.

During the microinjection process, from each clutch of fertilized eggs, a minimum of 50 embryos became the non-injected fitness control group, >30 embryos were designated for the water/phenol red control injection and the rest were divided into injection groups. Five embryos from a minimum of three different clutches were collected from groups injected with either 165 pg of Slc34a2a RNA, 165 pg of Slc34a2b RNA or 110 pg of Slc34a2a(as) RNA and non-injected controls. RNA extraction and RT-qPCR was completed in triplicate as mentioned in chapter 2. Water and RNA (negative) controls were always run in triplicate alongside the cDNA samples. RNA integrity and concentration were tested for each extraction by agarose gel electrophoresis and OD measurements. The average concentration of RNA from five embryos used in RT-qPCR was ~90 ng/µl. Cycle threshold values and melting peaks were considered to verify specific qPCR products. β-actin was used as reference gene for relative quantification and it maintained a constant Ct value of about 21.

Fertilized eggs that were injected with Slc34a2a, Slc34a2a(as) and Slc34a2b RNA underwent RNA isolation after 10 and 24 hours. Comparable results were found for the two time points (figure 4.8). Only one transcript had a ΔCt difference greater than 2, depicted as red bars in figure 4.8, 165 pg Slc34a2a injection altered Slc34a2b levels greatly between 10 and 24 hpf; transcript levels increased dramatically from a ΔCt 5.40±0.52 to ΔCt 0.35±0.27. Otherwise, transcript levels did not differ greatly between 10 and 24 hours post-injection and as a result, only 24 hour data is presented for analysis below.

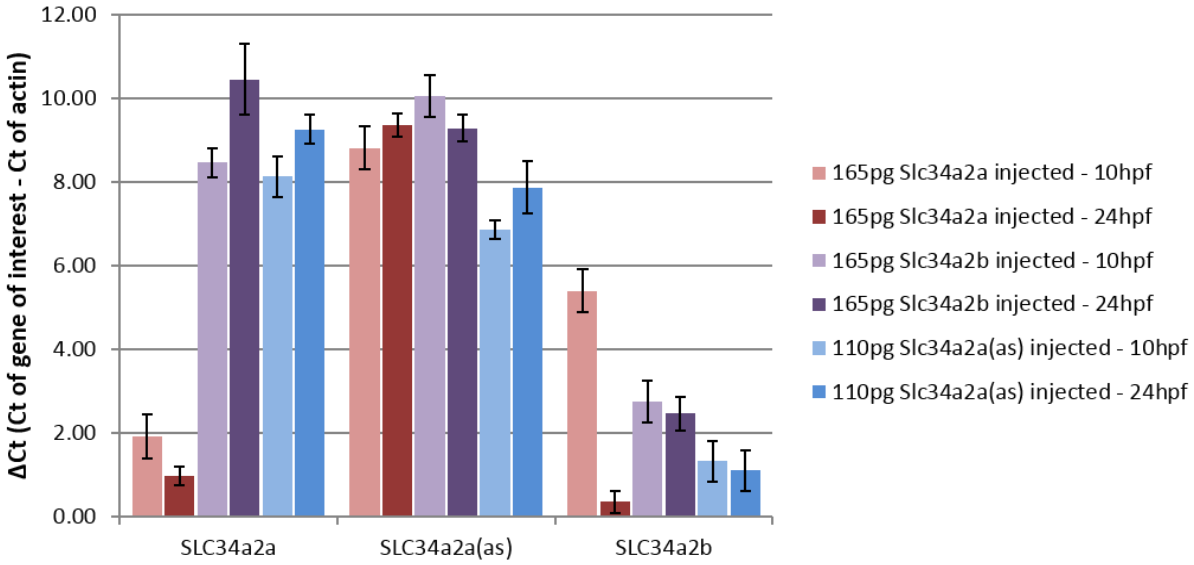


Figure 4.8 RT-qPCR analysis of Slc34a2a, Slc34a2a(as) and Slc34a2b transcript levels on RNA extracted from zebrafish embryos at 10 and 24 hpf. RNA extraction was performed from five embryos for each set of injections. ΔCt is calculated by using β-actin as gene reference. β-actin maintained a consistent Ct throughout all experiments. N=3

To determine the impact of Slc34a2a, Slc34a2a(as) and Slc34a2b RNA injection on the Slc34a2a transcript levels, RT-qPCR was completed on RNA extracted from 24 hpf embryos that underwent each injection. Embryos injected with 165 pg Slc34a2a RNA had considerably higher levels of Slc34a2a transcript when compared to non-injected controls, as well as embryos injected with 165 pg Slc34a2b and 110 pg Slc34a2a(as) RNA (figure 4.9). In embryos injected with 165 pg Slc34a2a, Slc34a2a transcript levels were near β-actin levels (ΔCt = 0.98±0.22). Based on the data provided in figure 4.9, Slc34a2a expression appears to have decreased by a ΔCt of 1-2 in Slc34a2a(as) and Slc34a2b RNA injected samples in comparison to wildtype samples (ΔCt Slc34a2a(as) =9.25 ±0.35, ΔCt Slc34a2b =10.45 ±0.85, ΔCt non-injected control =7.62 ±0.38). However, as established in the previous chapter, the Slc34a2a transcript is not endogenously present at 24 hpf. The discrepancy of the background values for Slc34a2a between this and the previous chapter reflect the fact that two different primer pairs were used.

4. Cerebellar loss

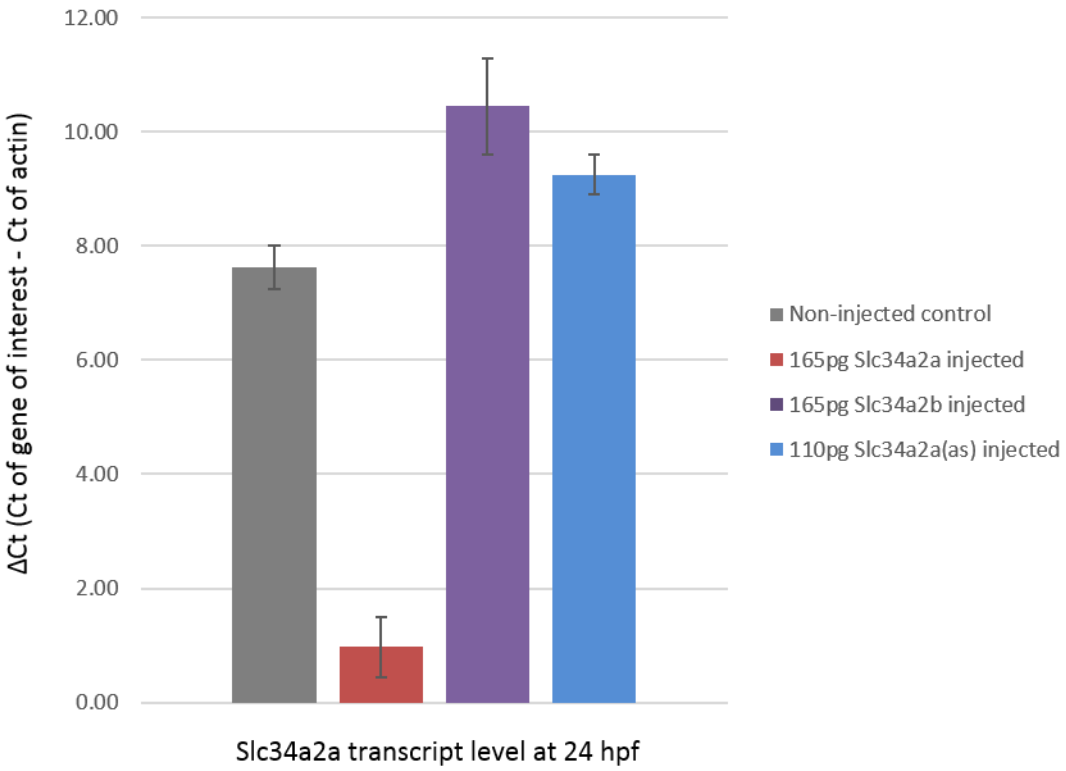


Figure 4.9 Examination of Slc34a2a transcript levels by RT-qPCR after RNA microinjection. RNA was extracted from five 24 hpf zebrafish embryos that underwent microinjection with either Slc34a2a, Slc34a2a(as) or Slc34a2b RNA. ΔCt were calculated by using β-actin as the gene of reference. β-actin maintained a consistent Ct throughout all experiments. N=3

Upon injecting 110 pg Slc34a2a(as) RNA, embryos were found to have decreased Slc34a2a(as) levels since ΔCt increased from 4.61±0.29 in non-injected embryos to 7.87 ±0.62 (figure 4.10, middle). Moreover Slc34a2b levels were found to increase since ΔCt decreased from 3.85±0.99 in non-injected to 1.10±0.49 in injected samples (figure 4.10, right). As mentioned previously, the apparent Slc34a2a expression decrease in Slc34a2a(as) RNA injected samples in comparison to control embryos is irrelevant (ΔCt Slc34a2a of 110 pg Slc34a2a(as) injected =9.25 ±0.35, ΔCt Slc34a2a in non-injected control=7.62 ±0.38) since Slc34a2a is not endogenously expressed at 24 hpf (chapter 3 results).

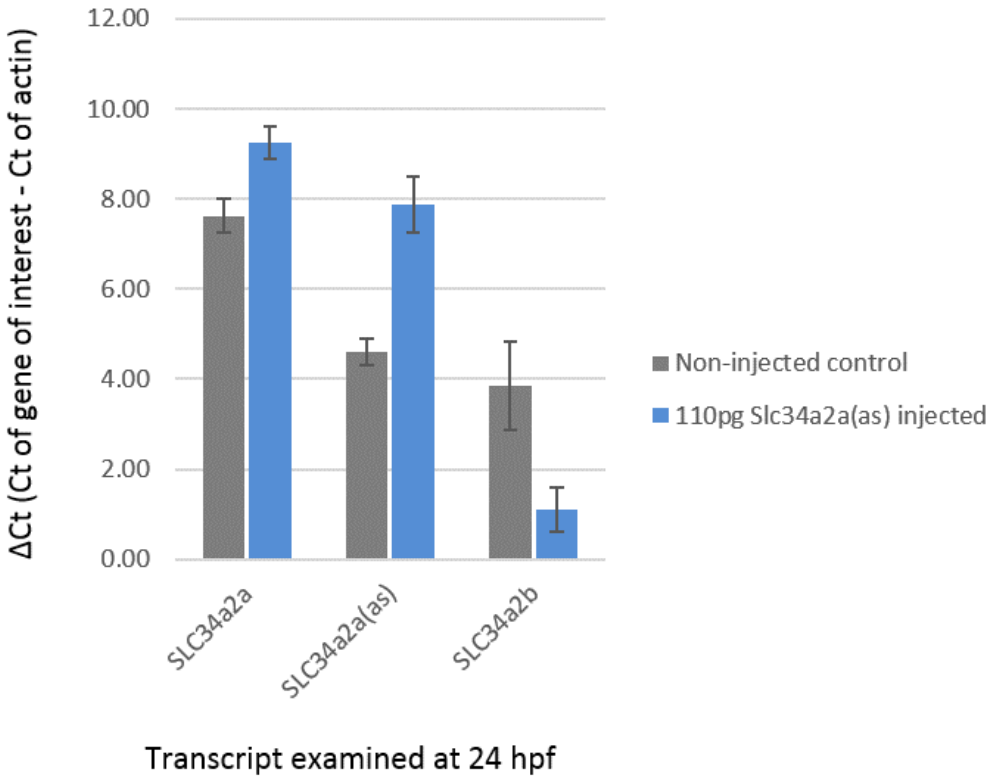


Figure 4.10 Semi quantitative RT-qPCR analysis of RNA extracted from five fertilised zebrafish oocytes 24 hours post injection with 110 pg Slc34a2a(as) RNA (blue) in comparison to non-injected samples (grey). ΔCt is calculated by using β-actin as gene reference. N=3

In a subsequent set of experiments, the consequences of 165 pg Slc34a2b RNA microinjection were examined. Slc34a2a sense transcript levels were found to be below detection level (chapter 3; Slc34a2a is not present at 24 hpf in wildtype embryos). Interestingly, Slc34a2a(as) levels decreased significantly upon Slc34a2b RNA injection with a ΔCt difference of ~4 when compared to the non-injected control group (figure 4.10, ΔCt Slc34a2a(as) in 165 pg Slc34a2b RNA injected=9.28±0.33, ΔCt Slc34a2a(as) in non-injected control=4.61±0.29). Upon injecting Slc34a2b, Slc34a2b transcript levels increased sizeably in 24 hpf embryos as shown in figure 4.11 (right) (ΔCt change of 2.75; ΔCt Slc34a2b in 165 pg Slc34a2b RNA injected=1.10±0.49, ΔCt Slc34a2b in non-injected control= 3.85±0.99).

4. Cerebellar loss

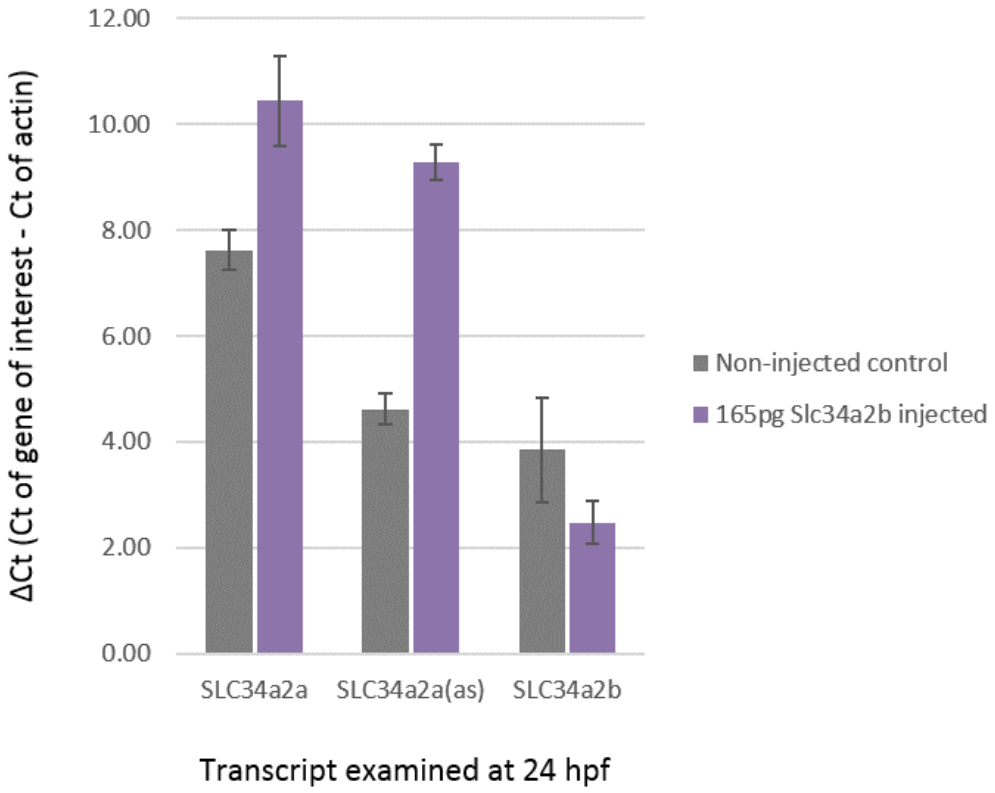


Figure 4.11 Semi quantitative RT-qPCR analysis of RNA extracted from five fertilised zebrafish oocytes 24 hours post injection with 165 pg Slc34a2b RNA (purple) in comparison to non-injected samples (grey). ΔCt is calculated by using β-actin as gene reference. N=3

4.7 Localization of transcripts after microinjection

Endogenous and exogenous levels of Slc34a2a, Slc34a2a(as) and Slc34a2b were monitored by in situ hybridization (ISH) after microinjection of the different transcripts into fertilised embryos. The purpose of these experiments was to monitor the fate of the injected material and investigate possible interactions and cross regulation between the transcripts. Mapping by ISH was completed on microinjected PFA-fixed embryos at 24 hpf and 48 hpf in line with the experiments described in chapter 3. Digoxigenin labelled probes for Slc34a2a, Slc34a2a(as), Slc34a2b and Rbpja were

used. Sonic Hedgehog was used as a control (figures 4.12 and 4.13) and staining located to the central nervous system, floor plate and brain as in published work (Sumanas *et al.*, 2005; Thisse, 2005; Jensen *et al.*, 2012; Chatterjee *et al.*, 2014). These findings included the embryos that were severely developmentally affected by 165 pg Slc34a2a RNA injection (figures 4.12 and 4.13, second column).

Slc34a2a was undetectable in non-injected and all injected 24 hpf embryos as seen in the top row of figure 4.12. The 165 pg Slc34a2a RNA injected embryo, which is developmentally hindered with a lack of eye formation and shortened body length, shows no Slc34a2a purple staining whatsoever. This is contradictory to the above RT-qPCR results which found a significant increase in Slc34a2a transcript levels after Slc34a2a injection. Slc34a2a was undetectable in 24 hpf zebrafish embryos by both ISH and RT-qPCR in chapter 3 (non-injected samples).

Slc34a2a(as) was found to be dispersed in the head region and in the endoderm of 24 hpf non-injected wildtype embryos and 165 pg Slc34a2a2b RNA injected embryos (figure 4.12, second row). The Slc34a2a(as) staining in the Slc34a2b injected embryos appears to be in the same magnitude as that seen in the wildtype embryos. This is contrary to the RT-qPCR analysis as Slc34a2a(as) levels were determined to decrease upon injection of 165 pg Slc34a2b. In the 110 pg Slc34a2a(as) RNA injected embryos, there is reduced staining. This matches the RT-qPCR analysis above which determined that Slc34a2a(as) transcript levels decreased after Slc34a2a(as) injection. In the developmentally affected Slc34a2a injected embryos, minimal staining was visualised behind the eye. This apparent reduction in Slc34a2a(as) levels visualised via ISH is consistent to the qPCR results.

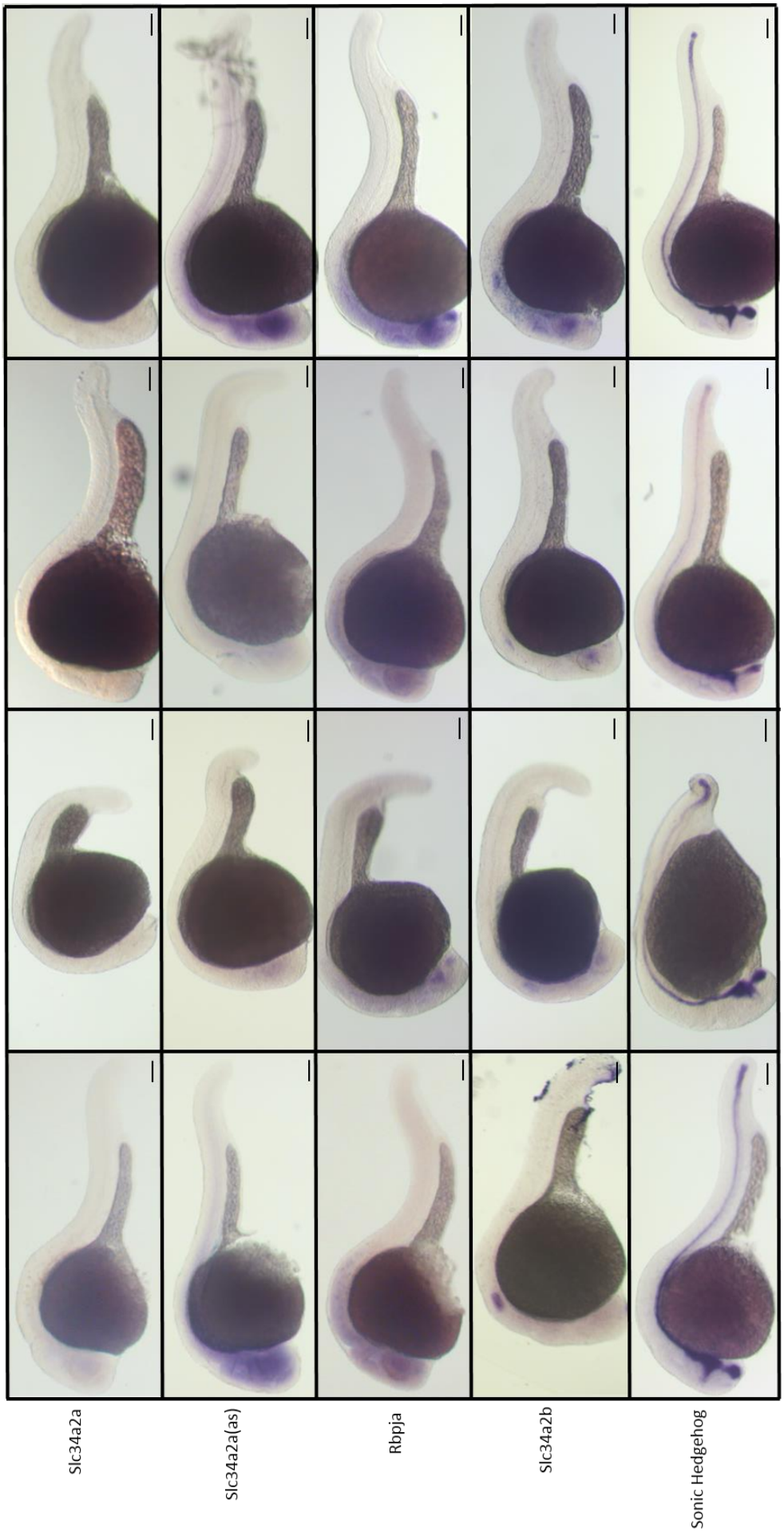


Figure 4.12 Localisation of Slc34a2a related transcripts in non-injected and injected embryos at 24 hpf. Sonic Hedgehog was utilised as positive control. Negative controls not shown. Scale bar at 0.1 mm. N≥ 10

Rbpja, which shares a bidirectional promoter with Slc34a2a(as), was also examined for any change in endogenous levels that could have been influenced by microinjection. Rbpja was found to be dispersed in the head region of the non-injected control, 110 pg Slc34a2a(as) RNA and 165 pg Slc34a2b RNA injected embryos (figure 4.12, third row). The strength and dispersion of the Rbpja staining appears to be equal in these embryos. Differing, in the 165 pg Slc34a2a RNA injected embryos, staining was apparently reduced with only a faint signal found in the head directly behind the eye.

Slc34a2b localised to the otic vesicle (ear) and the eye in the 24 hpf non-injected wildtype embryos as demonstrated in figure 4.12 (fourth row). The same staining pattern and intensity was visible in Slc34a2a and Slc34a2a(as) injected embryos. In the qPCR data above, Slc34a2b levels were found to increase upon Slc34a2a(as) injection; this increase was not apparent in the ISH staining. In Slc34a2b injected embryos, the same staining of the otic vesicle and eye was visualised but there too was extra specific staining located in the midbrain (mesencephalon). This apparent increase is consistent with the Slc34a2b transcript increase measured by RT-qPCR analysis.

In 48 hpf embryos, Slc34a2a localised to the endoderm and the pharynx (future digestive tract and swim bladder) as well as to primordial midbrain and hindbrain channel in wildtype embryos (Figure 4.13, top row). The same staining pattern in alike intensity was visible in 110 pg Slc34a2a(as) RNA and 165 pg Slc34a2b RNA injected embryos. Within the 165 pg Slc34a2a RNA injected embryos, which all had developmental defects, staining seemed more dispersed and differed in localization. In the Slc34a2a injected embryo visualised in figure 4.13, dispersed staining occurred throughout the head region with specific staining in the endoderm, primordial midbrain channel and what appears to the rhombomeres 3-7.

Slc34a2a(as), localised to the pharynx, endoderm and primordial midbrain and hindbrain channel reproducibly in 48 hpf non-injected embryos and those injected with 110 pg Slc34a2a(as) RNA and 165 pg Slc34a2b RNA (figure 4.13, second row). In 48 hpf 165 pg Slc34a2a RNA injected embryos, just as in 24 hpf embryos, there appeared to be reduced staining in the head region.

4. Cerebellar loss

Rbpja was found in the otic vesicle (ear) and the posterior of the mesencephalon (midbrain) in near same intensity in all injected and non-injected samples (figure 4.13, third row). Slc34a2b localised to the otic vesicle and endoderm of non-injected 48 hpf embryos as visualised in figure 4.13 (fourth row). The same staining pattern in a similar intensity was visualised in 110 pg Slc34a2a(as) injected embryos. In 48 hpf 165 pg Slc34a2b injected embryos, staining was visualised in the endoderm and mesencephalon but in apparent higher level than that of the non-injected control and Slc34a2a(as) injected samples. This apparent increase in staining for Slc34a2b was also seen in Slc34a2b injected 24 hpf samples above. In the developmentally malformed 48 hpf embryos injected with 165 pg Slc34a2a RNA, staining was inconsistent. The Slc34a2a injected 48 hpf embryo pictured in figure 4.13 (fourth row), Slc34a2b localised to the otic vesicle and was otherwise dispersed throughout the entire embryo.

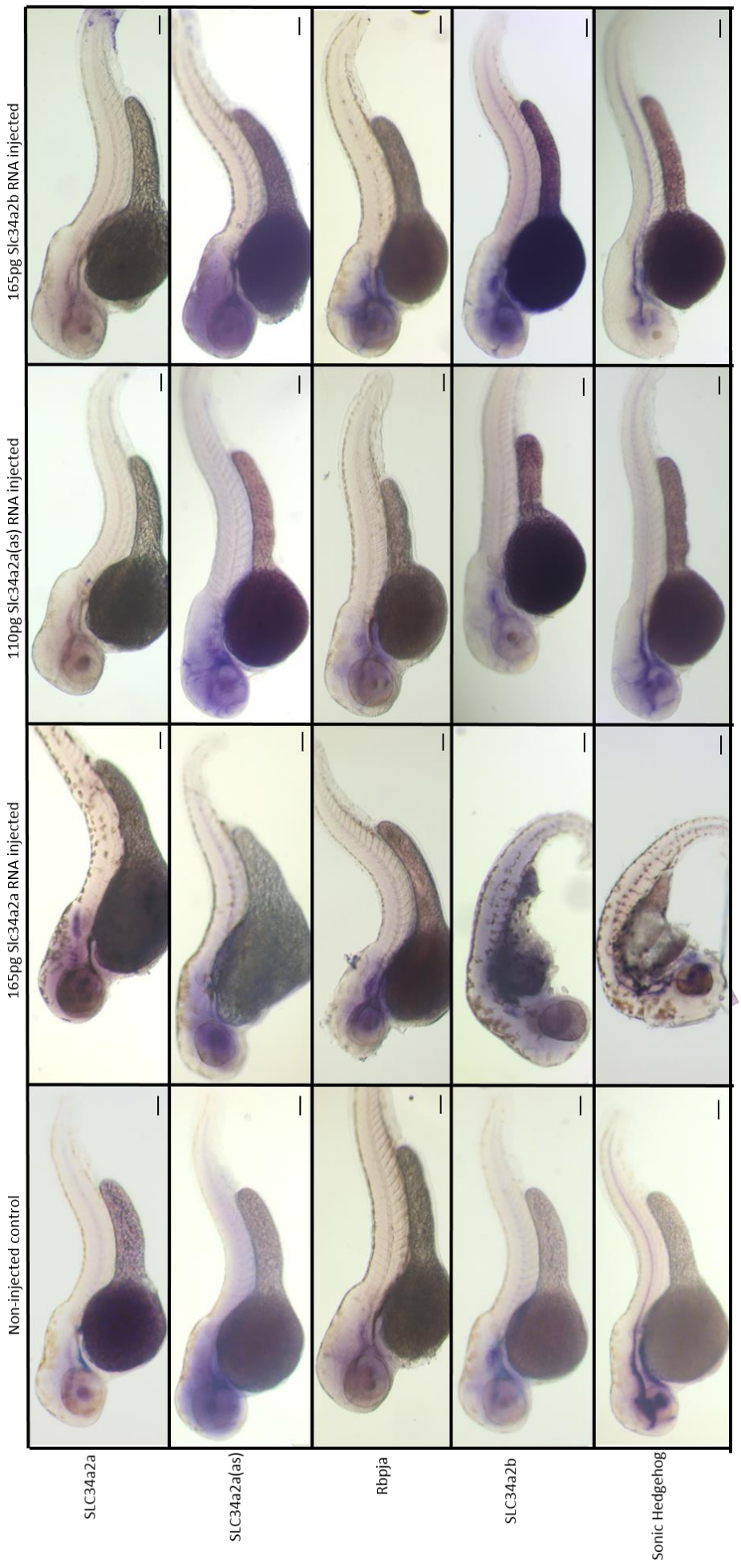


Figure 4.13 Localisation of Slc34a2a related transcripts in non-injected and injected embryos at 48 hpf. Sonic Hedgehog was utilised as positive control. Negative controls not shown. Scale bar at 0.1 mm. N≥ 10

4. Cerebellar loss

4.8 Discussion

Microinjections of Slc34a2a, Slc34a2a(as) and Slc34a2b RNA were completed to determine if and how the three transcripts interact and if they have any individual impact on zebrafish development. Injection of Slc34a2a sense RNA produced a phenotype, which at the lowest injection amount of 82.5 pg, produced specific loss of cerebellum (figure 4.3). Injection of paralog Slc34a2b and Slc34a2a(as) transcripts produced no phenotypic effect.

Slc34a2a RNA injection, which caused the failure of cerebellar development, was confirmed by in situ hybridisation (ISH) using the midbrain-hindbrain marker, Engrailed-2a (Eng2). *Eng2* is essential for vertebrate embryogenesis by coordinating midbrain and hindbrain development. ISH staining for control non-injected embryos was analogous to that seen in current literature (Fjose *et al.*, 1992; Scholpp and Brand, 2001; Thisse, 2005; D'Aniello *et al.*, 2013). The very faint staining in Slc34a2a injected embryos indicates that the cerebellum is completely missing.

Eng2 ISH findings are supported by behavioural phenotypes. The cerebellum is vital for coordinated movement, balance, posture and has been linked to non-motor learning. Therefore any damage to the cerebellum can produce severe effects such as impaired balance, decomposition of movement, involuntary tremors and compromised working memory (Moens and Prince, 2002; Lalonde and Strazielle, 2003; Morton and Bastian, 2004). Accordingly, Slc34a2a RNA injected embryos were noted to have problems swimming upright (loss of balance), did not react immediately to the ‘poke’ (impaired predictive movement), would sometimes have what looked like an epileptic shock (involuntary tremor) and often would either swim in a circle or back to the area of danger (impaired learning). These behavioural deficiencies are in line with the loss of cerebellum that was visually identified under the microscope and confirmed via Eng2 staining.

Injection of Slc34a2a RNA did not have an effect on central nervous system motor neuron development, even though it effected cerebellar development, as seen in figure 4.7. Microinjections were completed on Islet1-GFP transgenic zebrafish as they were available at the Centre for Life, Newcastle University, where all fish work was completed. Alternatively, a transgenic Islet3-GFP fish could have been included

as the Islet3 LIM/homeodomain protein is active during mesencephalon and cerebellar development and thus providing a more specifically localized reference.

In vivo, mRNA is processed to increase stability and translation efficiency. Likewise, *in vitro* synthesized RNA can be stabilized by adding poly-A tails or 5' untranslated regions. Capping is known to provide the RNA with a longer life span and aid in dispersion to the correct subpopulation of cells (Moreno *et al.*, 2009). Capping of synthesized RNA also promotes the translation process (Contreras *et al.*, 1982). Nonetheless, in this study, efficient translation of the Slc34a2a transcript by capping the injected RNA, was found not to be necessary for biological activity (figure 4.5). Embryos injected with capped and non-capped Slc34a2a RNA resulted in the same phenotype, i.e. loss of cerebellum at 48 hpf.

No RNA Slc34a2a/Slc34a2a(as) co-injections were performed since numerous publications reported that injection of dsRNA, even in doses as low as 7.5 pg, leads to non-specific phenotypic effects during early zebrafish embryo development (Li *et al.*, 2000; Oates *et al.*, 2000; Zhao *et al.*, 2001). These reports aimed at reproducing dsRNA induced gene silencing as observed in *C. elegans*. However, all injected dsRNA and co-injected RNA were found to be degraded without sequence specificity thus indicating that dsRNA injection caused a nonspecific effect at the posttranscriptional level (Zhao *et al.*, 2001).

For the injection process, fertilised eggs from a single clutch are separated into multiple groups. A minimum of 50 eggs are used as the non-injected fitness control group and if less than 80% of the embryos survive by 24 hpf, the entire clutch was disregarded. As the membrane and structures protecting the first forming cells are tougher than the yolk membrane, injections were completed so that RNA was injected into the yolk adjacent to the first embryonic cell or through the yolk directly into the first cell, where RNA would be most active. Cytoplasmic flow and diffusion ensure that the RNA injected into the yolk migrates into the cytoplasm of the embryonic cells (Yuan and Sun, 2009; Chang, 2012). If injections are performed at the eight-cell stage or later, a mosaic embryo would be generated in which the injected RNA is only be found in a subset of cells (Carmany-Rampey and Moens, 2006). However, since injections were completed before the four cell stage, the exogenous RNA reached all cells as demonstrated by the in situ hybridization results, figures 4.12 and 4.13.

4. Cerebellar loss

Within the previous chapter, chapter 3, primers purchased from and designed by PrimerDesign were utilised for all qPCR experiments. PCR primers for this and subsequent chapters were designed and optimised in-house and purchased from Sigma Aldrich. Consequently ΔCt values from non-injected controls differed from this chapter to the preceding.

To determine the impact Slc34a2a, Slc34a2a(as) and Slc34a2b RNA injection had on transcript levels, RT-qPCR was completed on RNA extracted from 10 and 24 hpf embryos that underwent each injection. Transcript levels never differed more than a ΔCt of 1.5 between 10 and 24 hpf samples, except Slc34a2b in 165 pg Slc34a2a injected embryos. In Slc34a2a injected embryos, the amount of Slc34a2b changed by ~ΔCt 5 between 10 hpf and 24 hpf. As no major differences in transcript levels were found between 10 and 24 hpf samples that underwent the same injection (figure 4.11), consequently only 24 hpf data was presented.

As described in chapter 3, Slc34a2a is not present from fertilization and only appears at 48 hpf. Logic would dictate that injection of a transcript should increase its abundance. As expected, data from the above qPCR experiments show an increase in Slc34a2a levels at 10 and 24 hpf after 165 pg Slc34a2a RNA injection (figure 4.8 and 4.11). However, this Slc34a2a increase was not visible by ISH; Slc34a2a was undetectable in non-injected and all injected 24 hpf embryos (figure 4.12). ISH is based on the complementary binding of probes to the RNA transcript of interest. If the injected Slc34a2a RNA was binding to the endogenous Slc34a2a(as), ISH probes would be unable to bind and detect the Slc34a2a transcript. Another hypothesis involves the storage of the injected Slc34a2a RNA in either GW-bodies or P-bodies, which are present during *Drosophila melanogaster* and *Danio rerio* embryogenesis. These structures are involved in the removal of non-translated RNAs from the cytoplasm (Liu *et al.*, 2005; Kloosterman *et al.*, 2006; Patel *et al.*, 2016). As Slc34a2a encoded protein is not required till 48 hpf onwards, exogenous Slc34a2a RNA may be compartmentalized in these P-bodies and GW-bodies and thus not available for binding by DIG- labelled probes. For both situations, the phenol chloroform extraction process may have freed up the Slc34a2a RNA for detection via RT-qPCR.

Slc34a2a(as) is endogenously expressed from fertilization as demonstrated in chapter 3. Slc34a2a(as) levels were expected to increase after Slc34a2a(as) injection but the contrary occurred when levels were examined via RT-qPCR and

ISH (figure 4.9 and 4.12). The mechanisms behind how antisense non-coding RNA levels are controlled is largely unknown. In yeast, regulation processes of mRNA and antisense RNA transcripts differ (Marquardt *et al.*, 2011). Nonetheless, there is well documented evidence that endogenous antisense RNA levels are indeed tightly controlled (Lehner *et al.*, 2002; Tufarelli *et al.*, 2003; Nakaya *et al.*, 2007). Moreover, in eukaryotic systems there are a number of degradation pathways to remove extra, unnecessary or faulty transcripts (Houseley and Tollervey, 2009). Therefore, as Slc34a2a(as) is already naturally present in the first 24 hours of development, it is plausible that one or many of the degradation pathways exceeded the amount of Slc34a2a(as) RNA requiring degradation to return to normal levels.

Slc34a2b is consistently expressed throughout the early stages of embryogenesis as demonstrated in the previous chapter. When RNA levels were examined via RT-qPCR, Slc34a2b levels were found to increase after Slc34a2b injection as expected (figure 4.11). This increase was consistent when embryos were examined via ISH. In 24 hpf Slc34a2b injected embryos, the established staining of the otic vesicle and eye was present, but there was additional staining located in the midbrain (figure 4.12). The added transcript increased signal without triggering an intrinsic feedback response to degrade or sequester the exogenous RNA.

Interestingly, in 24 hpf embryos injected with 110 pg Slc34a2a(as), Slc34a2b levels increased. The opposite was seen when Slc34a2a(as) levels decreased upon 165 pg Slc34a2b injection. When Slc34a2b is aligned to the Slc34a2a antisense, the longest stretch of perfect complementarity is less than 11 nucleotides. Nevertheless, as both Slc34a2a(as) and Slc34a2b localise to the same structures (Figure 4.12 and 4.13), it is plausible that the transcripts interact via incomplete sequence complementarity. Therefore, as eukaryotic cells detect dsRNA differently depending on their source (see review: (Gantier and Williams, 2007)), it is possible that Slc34a2a(as) and Slc34a2b are affecting one another either directly or indirectly through a dsRNA pathway. It too is plausible that injection of either Slc34a2a(as) or Slc34a2b is triggering a physiological function to maintain Pi homeostasis. This could be tested, for example, if embryos were exposed to varying levels of Pi and a change in gene expression could be detected.

Whole mount RNA in situ hybridization is a technique used for monitoring the precise localisation of transcripts of interest. As mentioned in the previous chapter, the abundance of a transcript is related to the development of the stain. Nevertheless,

4. Cerebellar loss

ISH is not a quantitative technique. For example, if left in developing solution for 24 hours, any SHH probed embryo would irreversibly stain throughout the entire body no matter if the RNA of interest is present in that structure or not. Consequently, even though the protocol has been optimized for each probe, the results have to be interpreted with care.

One of the limitations is that signal development is terminated based on visual examination. Due to human limitation, some overstaining is possible and could therefore explain the some of the discrepancy seen here between RT-qPCR and ISH data. A second limitation to whole mount ISH in *Danio rerio* embryos is the large yolk. The yolk can hinder examination both by creating a shadow and by making it difficult to examine internal structures. Therefore to confirm endoderm staining, embryos required yolk removal (photos not shown).

4.9 Conclusion

Injection of full length Slc34a2a RNA caused a dose-dependent phenotype in fertilized zebrafish oocytes. A specific failure to develop the cerebellum was noted upon injections of Slc34a2a RNA that was seen with neither Slc34a2a(as) nor Slc34a2b injections. In situ hybridization with Engrailed-2 staining confirmed the complete lack of cerebellar development and indicated a possible widening of the midbrain-hindbrain furrow. However, the mechanism that leads to cerebellar loss is unknown. Full length Slc34a2a transcript was injected which could lead to speculation that ectopic production of Slc34a2a protein might be the cause. Another hypothesis would suggest that the injected sense transcript hybridised to the naturally occurring antisense to create dsRNA. The dsRNA could in turn be inducing a protective response by activing Dicer or another dsRNA-related pathway which consecutively could lead to gene silencing. Another possibility entails the injected Slc34a2a binding to and inhibiting some other unknown sequence that is endogenously present and too activating this dsRNA response. In the subsequent chapter, possible pathways will be examined to determine by which mechanism this phenotype is produced.

Chapter 5. Slc34a2a dsRNA is linked to cerebellar loss

Slc34a2a sense transcript expression commences at 48 hpf, unlike the antisense transcript which is present from fertilization. Full length Slc34a2a RNA microinjection into fertilized zebrafish oocytes was found to cause a specific phenotypic failure in cerebellum development. The mechanism which causes this detrimental phenotype is unknown. This chapter aims to further investigate the relationship of Slc34a2a and its antisense transcript on midbrain/hindbrain development.

5.1 Slc34a2a protein production is not the cause of cerebellar loss

Within the previous chapter, full length *in vitro* synthesized RNA was injected into 1-4 cell embryos. Injection of capped and uncapped Slc34a2a give rise to cerebellar loss, unlike Slc34a2a(as) and Slc34a2b RNA, which resulted in no phenotypic change. To determine if the phenotype observed was due to overexpression of the Slc34a2a encoded phosphate transport protein, a full length capped Slc34a2a RNA which included a frameshift mutation was synthesized *in vitro*. The frameshift was created by removing a single nucleotide 18 bp downstream from the start codon resulting in a non-coding transcript. The mutated Slc34a2a transcript (FS-2a) was injected utilizing the same procedure and technique as all injections completed in chapter 4. Alongside a non-injected fitness group and a negative control water/phenol red injected group, capped FS-2a was injected in conjunction with the same capped Slc34a2a RNA transcript from chapter 4. Due to prolonged storage, synthetic Slc34a2a transcript lost some of its activity and therefore, the concentration required to produce the same phenotype for Slc34a2a increased.

Phenotypic effects caused by microinjections were logged according to the standards described in the aforementioned chapter (Chapter 4, figure 4.1). Of the 364 embryos injected with 165 pg Slc34a2a RNA that were assessed at 24 and 48 hpf, all showed phenotypic abnormalities (figure 5.1). Nearly all embryos (96.2%) had a specific failure in cerebellum development. The remaining embryos, 3.3% at

level 3 and 0.3% at level 4, manifested other effects such as shortened body length, body curvature and improper eye formation in addition to the brain segmentation deformities.

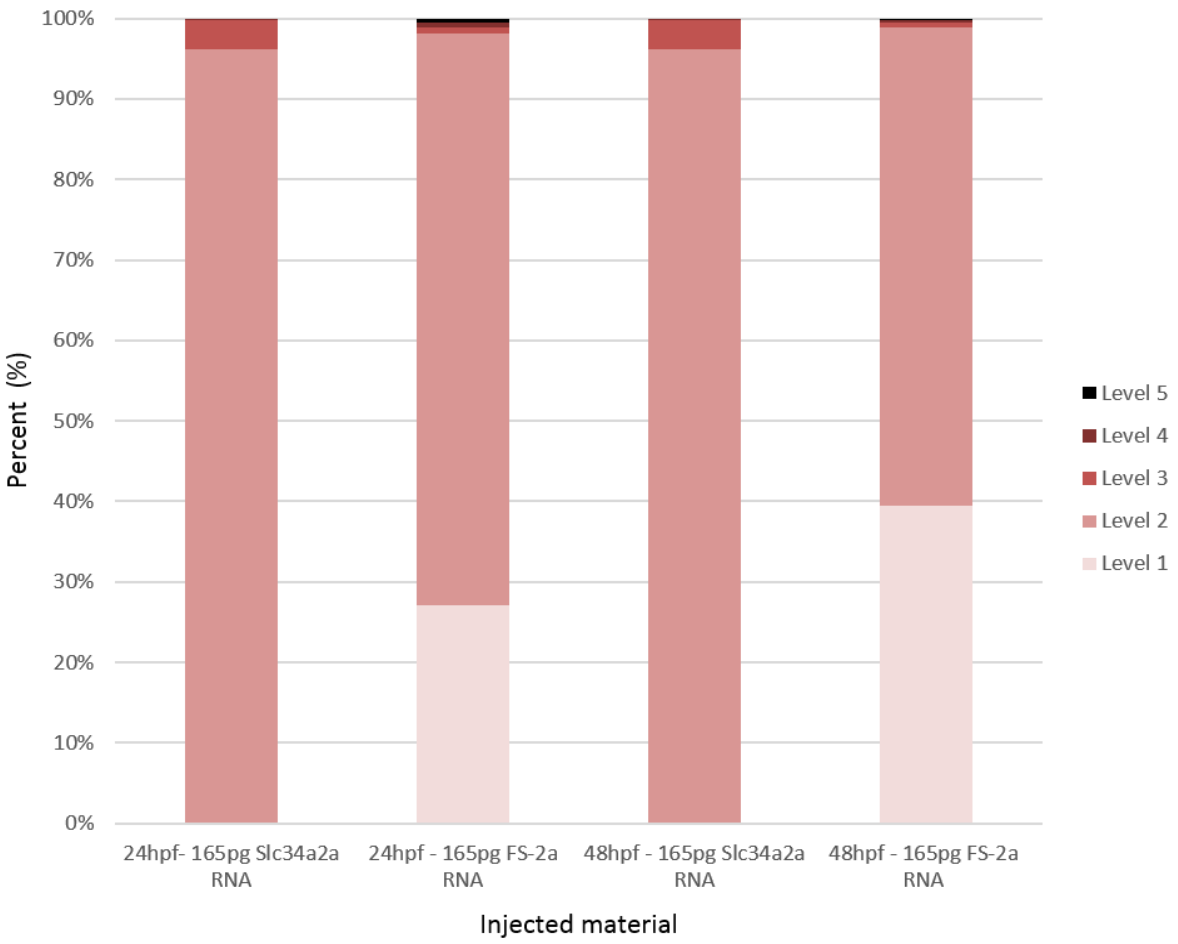


Figure 5.1 Phenotypic effects caused by microinjection of Slc34a2a RNA containing a frameshift mutation (FS-2a) into fertilized 1-4 cell zebrafish oocytes. Minimum 364 embryos were used from at least three clutches for each RNA injected.

As shown in figure 5.1 above, of the 372 embryos injected with 165 pg capped FS-2a RNA, 71.0% were classified as level 2 at 24 hpf, indicating they had one developmental malformation. These embryos did not properly develop their

midbrain-hindbrain region as seen by the lack of cerebellum (figure 5.2). The remaining 24 hpf embryos were classified level 3 (0.8%), level 4 (0.5%) and level 5 (0.5%). Those classified level 3 and 4 had shorter body and tail lengths, more-so curved bodies and incorrect eye development alongside improper brain development, comparable to the Slc34a2a RNA injected embryos. Those classified as level 5 did not develop past the blastula stage.

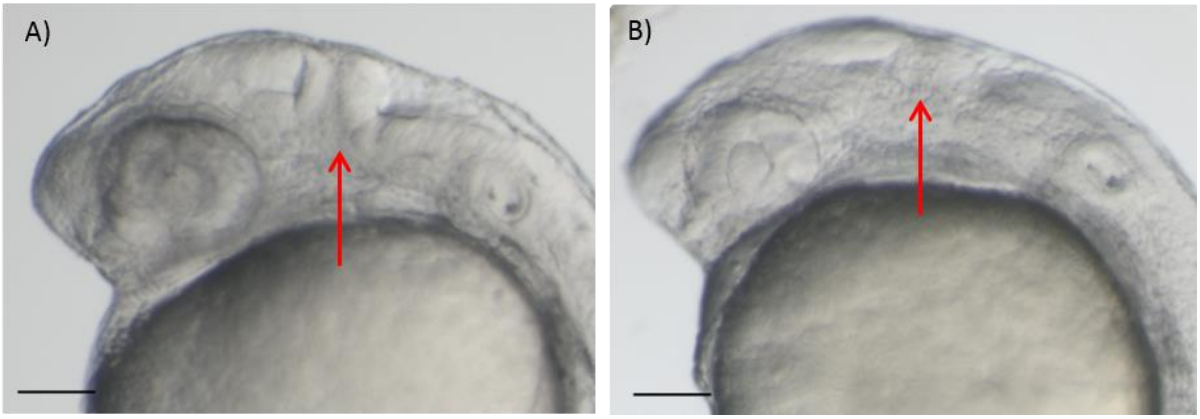


Figure 5.2 Embryos 24 hpf injected with 165 pg capped frameshift Slc34a2a RNA lacked clear midbrain-hindbrain segmentation with an absence of cerebellum (red arrow). (A) Water/phenol red injected controls. (B) 165 pg FS-2a RNA injected embryo. Scale bar at 0.1 mm.

Of the 372 embryos injected with 165 pg capped FS-2a, 370 survived to 48 hpf; 59.5% were classified as level 2, 0.5% were level 3, 0.3% level 4 and 0.3% level 5. Those logged as level 2 showed the same phenotypic loss of cerebellum as those injected with protein-coding Slc34a2a RNA. From a dorsal view, in a representation of a 48 hpf embryo as seen in figure 5.3, A, the cerebellum is denoted as two red semicircles which are clearly visible in non-injected and water injected controls. However in FS-2a injected 48 hpf embryos, the semi circular structures are missing and replaced with a thin line as pointed out by the red arrow in figure 5.3, C. This would indicate that despite the loss of cerebellum the midbrain hindbrain furrow has most likely formed.

5. Mechanism

The phenotypic observations recorded in figure 5.1 show an increase in the number of embryos classified as level 1 (wildtype) from 24 hpf to 48 hpf (27.2% to 39.5%). As the same injected embryos were assessed at these two time points, it appears that some embryos overcame the initial defect.

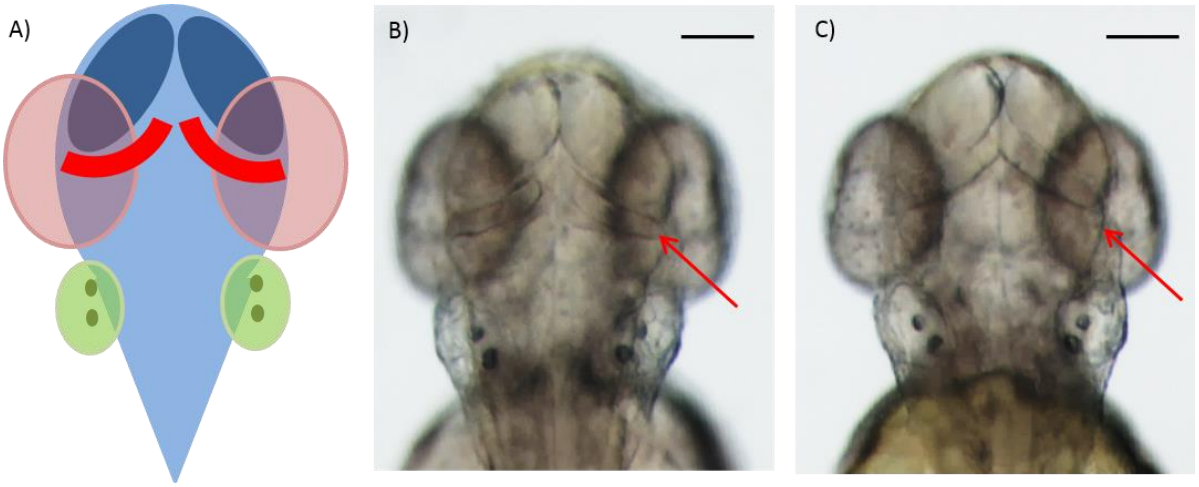


Figure 5.3 Injections of frameshift Slc34a2a RNA causes loss of cerebellum as observed in the dorsal view of 48 hpf embryos. A) Representation of the structures visible from a dorsal view: eyes, pink; otic vesicles (ears), green; mesencephalon, dark blue; cerebellum, red. B) Fitness control embryo with correctly developed brain structures. C) FS-2a injected embryo lacking the cerebellum. Red arrows point to area of cerebellum. Scale bar at 0.1 mm.

As with previous experimental injections, basic behaviour of FS-2a injected embryos were recorded. In line with Slc34a2a RNA injected embryos, FS-2a injected embryos had delayed reactions in the stimuli ‘poke’ test. Many embryos had difficulties swimming in an upright position but those who maintained balance, would either swim away after a delayed period, needed numerous stimuli to react and occasionally only swam in a circle back to the point of danger (videos of reactions in appendix B).

5.2 In situ hybridization confirmation of cerebellar loss due to frameshift Slc34a2a RNA microinjection

To confirm the loss of the cerebellum, as seen in figures 5.2 and 5.3, in situ hybridization (ISH) was completed using a marker that is highly expressed in the cerebellar region of the brain, homeobox protein Engrailed-2 (Eng2). An Eng2 probe was synthesised as mentioned in Chapter 2.

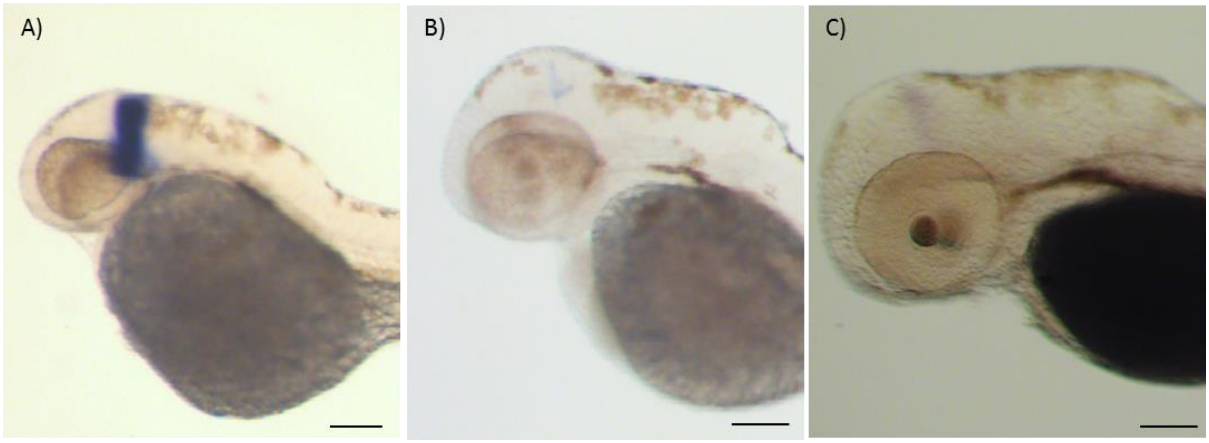


Figure 5.4 Engrailed-2 RNA staining in whole-mount in situ hybridization to show the loss of cerebellum at 48 hpf after the microinjection of full length capped frameshift Slc34a2a RNA. A) non-injected wildtype embryo. B) 110 pg Slc34a2a injected. C) 165 pg FS-2a injected. Scale bar at 0.1 mm.

Eng2, a gene essential for midbrain-hindbrain boundary development, was used to corroborate if cerebellar development was affected in the same fashion by FS-2a RNA injection as it was by Slc34a2a RNA injections. As seen in figure 5.4 A, the entire cerebellum including the midbrain-hindbrain region is stained as a thick purple wedge in the non-injected sample. In the Slc34a2a RNA injected embryo, there is a slight sliver of what would be the back of the midbrain region and a sliver of purple pointing backwards. The cerebellum is missing completely but the midbrain-hindbrain furrow appears developed and widened (figure 5.4, B). In the FS-2a RNA

injected embryo, the staining appears significantly reduced (figure 5.4, C). The thin sliver in FS-2a injected is position like that of the Slc34a2a injected embryos; the thin stained line is above the eye and does not venture further back to create a wedge shape. Therefore, it appears that no cerebellum is present (no thick purple wedge) but the midbrain-hindbrain boundary/furrow must have formed (thin sliver visible).

5.3 Expression profiles of Slc34a2b, Slc34a2a sense and antisense in embryos microinjected with frameshift Slc34a2a RNA

RT-qPCR was completed using in-house designed primers (chapter 2) to monitor the levels of injected RNA, their impact on biologically related transcripts and detect possible cross-regulation. As with other qPCR experiments in this study, β-actin was used as a reference gene and was found to be expressed at constant levels generating Ct values of ~21. The Slc34a2a qPCR primers used do not distinguish between the frameshift Slc34a2a and coding Slc34a2a transcripts. Therefore, as seen in figure 5.5 below, Slc34a2a transcript levels increased at both 10 and 24 hpf time points after frameshift Slc34a2a RNA injection.

Slc34a2a(as) levels were found to decrease at both 10 and 24 hpf in response to FS-2a RNA injection. The same observation was made after full length coding Slc34a2a was injected in chapter 4. The reduction was large at 10 hours post injection with a ΔCt change of 4.61±0.29 in 24 hpf wildtype embryos to 8.12±0.79 in 10 hpf FS-2a injected embryos. Slc34a2a(as) levels seem to decrease closer to wildtype level at 24 hpf (ΔCT 5.70±0.34). FS-2a injection did not have any effect on Slc34a2b levels (wildtype; ΔCT 3.85±0.99, 10 hpf injected; ΔCT 4.21±0.67, 24 hpf injected; ΔCT 3.76±0.42).

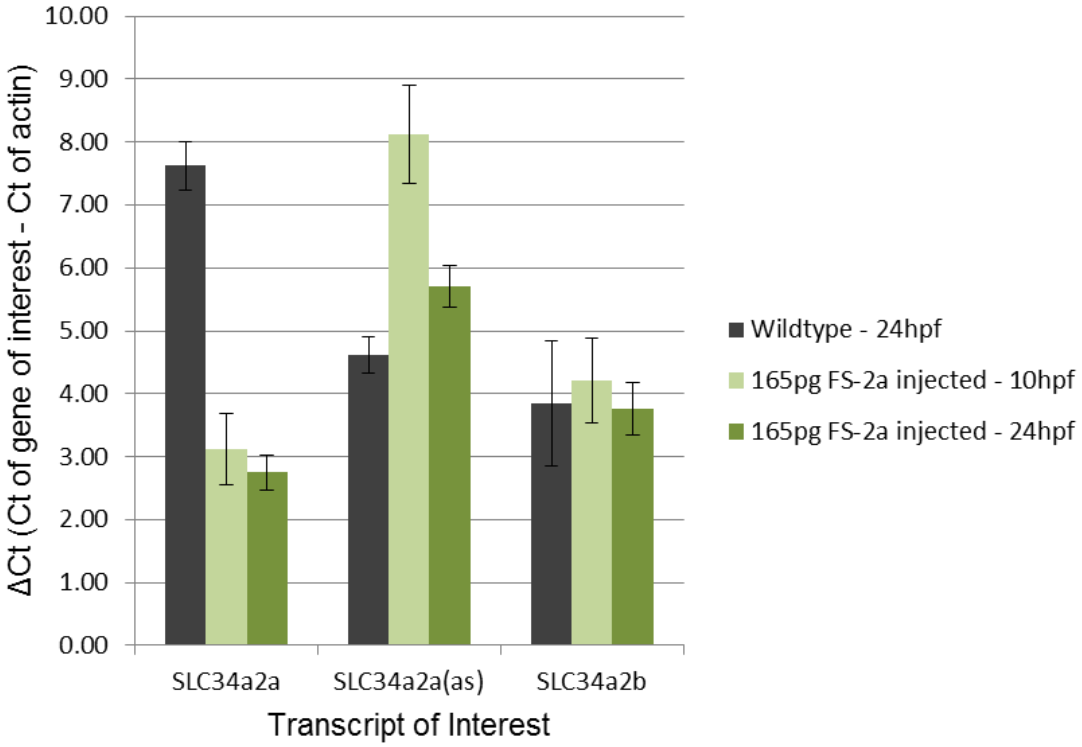


Figure 5.5 Expression changes of Slc34a2a related genes in response to FS-2a RNA injection. Semi quantitative RT-qPCR analysis of RNA extracted from five fertilised zebrafish oocytes 10 and 24 hours post injection with 165 pg frameshift Slc34a2a RNA (green) in comparison to 24 hpf non-injected samples (grey). ΔCt were calculated by using β-actin as gene reference. N=3

5.4 RNA injection of partial Slc34a2a and antisense RNA transcripts

RNA fragments encompassing various regions of the Slc34a2a locus were produced for microinjection into fertilized zebrafish eggs. These *in vitro* synthesized non-capped RNA fragments were created to determine if cerebellar loss was caused by a specific sequence. Six RNA fragments were produced in total as represented in figure 5.6. Three fragments were created from the positive strand (top strand encompassing the Slc34a2a gene) and three from the negative strand (bottom strand encompassing the Slc34a2a(as) sequence).

5. Mechanism

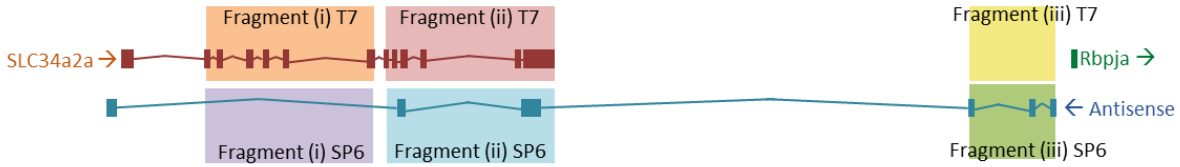


Figure 5.6 Representation of location of synthesized RNA fragments produced for injection. Red connected blocks represent the exons of Slc34a2a. Blue connected blocks represent Slc34a2a(as) exons. Fragments (i) T7 (orange) and (i) Sp6 (purple) are 614 nucleotides in length whilst fragments (ii) T7 (pink) and (ii) SP6 (blue) contain 959 nucleotides. Fragments (iii) T7 (yellow) and SP6 (green) are 263 nucleotides long.

Fragment (i) T7 (highlighted by an orange box in figure 5.6), includes part of the Slc34a2a sequence but would not hybridize to endogenously expressed Slc34a2a(as) as it shares no sequence complementarity at mRNA level. This is represented by the lack of exons (connected blue boxes) on the strand below the (i) T7 fragment. Fragment (i) SP6, purple box in figure 5.6, contains the sequence that is complimentary to the Slc34a2a but these exons are not included in the endogenous, spliced form of the Slc34a2a(as) transcript. Fragment (ii) T7 contains part of the Slc34a2a sequence which would bind to endogenously expressed Slc34a2a(as). Contrary, fragment (ii) SP6 which encodes a section of the Slc34a2a(as) transcript, would bind to Slc34a2a due areas of exon-exon complementarity. The sequence for fragment (iii) T7 (yellow) does not code for exons but it can hybridize to Slc34a2a(as) as it shares complementarity with Slc34a2a(as). The fragment (iii) SP6 (green), which codes for the 5' end of the antisense transcript, is not expected to interact with Slc34a2a related transcripts as it lacks complementarity to expressed transcripts.

Standard microinjections into fertilized eggs were performed as described and phenotypes were recorded as discussed in chapter 4. As shown in figure 5.7 below, the only fragment that interfered with the development of the embryos was the (ii) T7 fragment. This RNA fragment, part of the Slc34a2a coding sequence, was the only one injected that shared exonic sequence complementarity with the Slc34a2a(as) transcript that is present naturally from fertilization. Injections of this RNA produced a

significant proportion of embryos with a single developmental malformation (34.8%, level 2). Almost all these embryos (34.3%) showed evidence of incorrect brain segmentation. A loss of cerebellum was witnessed with the (ii) T7 fragment injected embryos (figure 5.8) comparable to the frameshift containing Slc34a2a RNA injected fish as pointed out with the red arrow in figure 5.8. All other injected RNA fragments produced a normal phenotype.

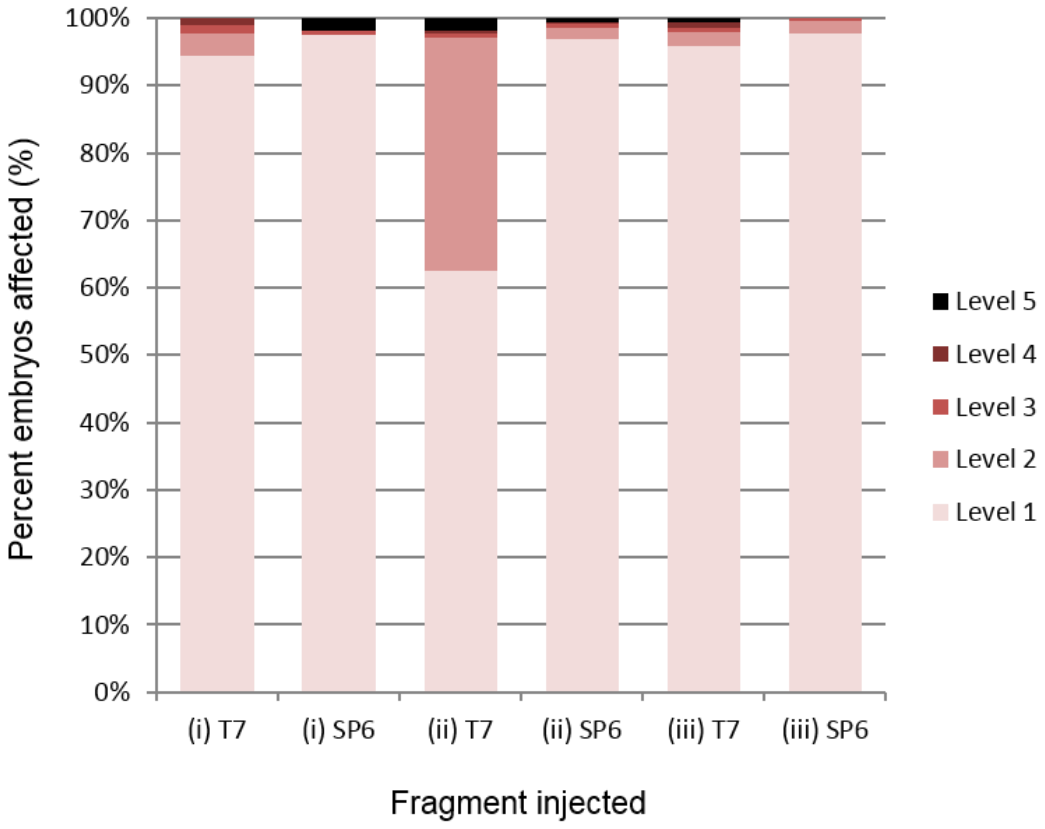


Figure 5.7 Phenotypic effects recorded at 48 hpf caused by microinjection (165pg) of Slc34a2a RNA fragments into fertilised zebrafish oocytes. Results are recorded as percentages based on phenotype severity. A minimum of 90 embryos from three clutches were analyzed for each injected transcript.

Injection results were logged at 24 and 48 hpf. For those embryos injected with fragment (ii) T7, there was a minor reversion to a wildtype phenotype in 48 hpf groups in comparison to 24 hpf embryos (data not shown). Due to their improperly

developed brains, the fragment (ii) T7 RNA injected embryos showed unusual behavioral traits. Many, could not swim upright and showed delayed reaction prior to swimming away in response to a poke. Occasionally the embryos would swim only in a small circle after a delayed reaction to the poke, putting themselves back in the danger zone. Such behavior was never seen in wildtype embryos.

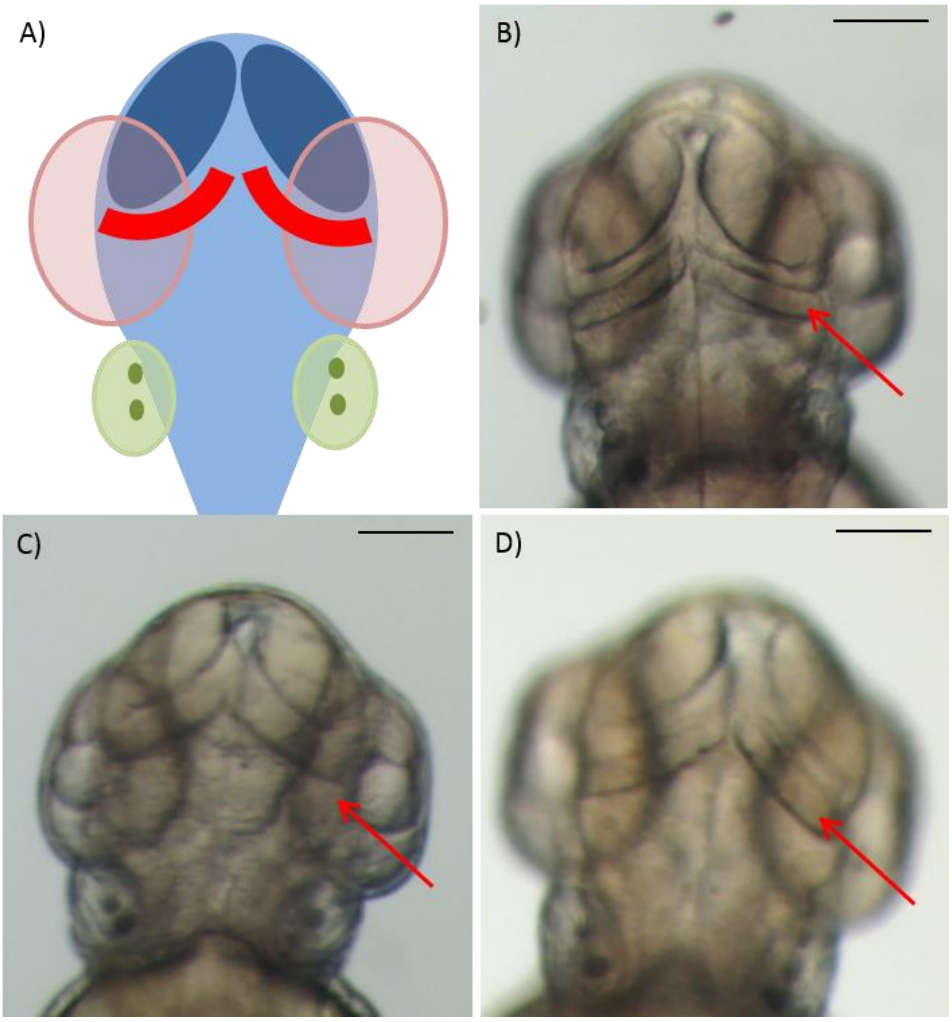


Figure 5.8 Injections of fragment (ii) T7 RNA causes loss of cerebellum. The dorsal view of 48 hpf embryos is shown. A) Schematic representation of a 48 hpf embryo where the cerebellum is denoted as two red semi-circles. B) Non injected control embryo. C) 165 pg fragment (ii) T7 RNA injected embryo. D) 165 pg fragment (ii) SP6 RNA injected embryo. Red arrows point to the cerebellum or the area where the cerebellum should be located. Scale bar at 0.1 mm.

5.5 In situ hybridization confirmation of cerebellar loss due to Slc34a2a RNA fragment

In situ hybridization (ISH) staining was completed on embryos injected with the RNA fragments synthesised from the Slc34a2a locus. The Eng2 probe described in previous ISH experiments was utilised to visualize cerebellum presence or absence. Wildtype embryos show an intensely stained purple wedge that runs vertically throughout the embryo in representation of the midbrain-hindbrain boundary including the cerebellum, (figure 5.9 A). From a lateral view, the purple structure is located posterior to the eye. All embryos injected with RNA fragment (ii) T7 show markedly fainter Eng2 staining as compared to the wild type and the other fragments (figure 5.9 E). This result is comparable to frameshift Slc34a2a RNA injected embryos (figure 5.4). Injection of the (ii) T7 RNA fragment reduced the intensity of the Eng2 staining but not to levels observed after Slc34a2a injection. All other RNA fragments produced an Eng2 staining that was comparable to the wildtype.

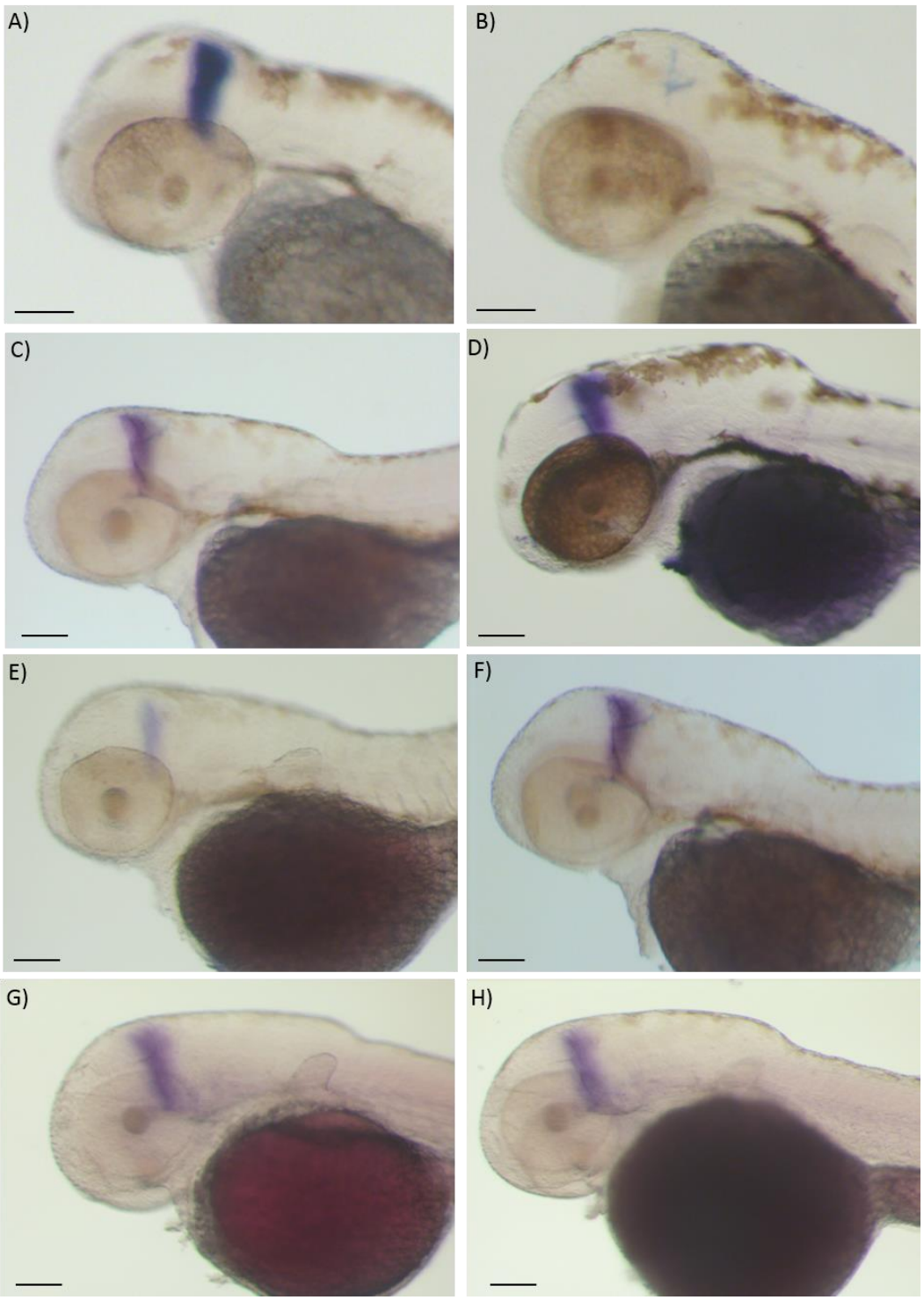


Figure 5.9 Engrailed-2 staining of embryos injected with Slc34a2a fragments. 48 hpf embryos injected with the various Slc34a2a RNA fragments are presented after in situ hybridization using the Eng2 probe. A) Non-injected control embryo; B) Capped full length Slc34a2a RNA injected embryo; C) Fragment (i) T7 injected embryo; D) Fragment (i) SP6 injected embryo; E) Fragment (ii) T7 injected embryo; F) Fragment (ii) SP6 injected embryo; G) Fragment (iii) T7 injected embryo; H) Fragment (iii) SP6 injected embryo. All fragment injections were completed with 165 pg of RNA. Scale bar at 0.1 mm.

5.6 Slc34a2a antisense knockdown by morpholino microinjection

Microinjection of full length antisense RNA did not induce any phenotypic changes in the development of zebrafish embryos as mentioned in chapter 4. In order to investigate a possible function of the antisense transcript directly, a morpholino oligonucleotide (MO) that interferes with the splicing of the transcript, was designed and custom synthesised (Gene Tools, LLC). The morpholino (sequence: 5' – GCCATCTGGTGAAAAGACAGAGTTT – 3') targeted the third exon in the antisense pre-mRNA (chapter 2) in order to cause an exon deletion. The antisense morpholino (AS-MO), alongside a mismatch antisense morpholino control, were injected into 1-4 cell fertilised oocytes. Any developmental changes were logged according to the same phenotypic scale used in the previous chapter at both 24 and 48 hpf.

Injection of 0.4-10 ng of the splice site antisense morpholino caused a severe, nonspecific phenotype at 24 hpf. As seen in figure 5.10 A, only embryos injected with 0.2 ng of the AS-MO had a substantial proportion of wildtype embryos at 24 hpf (67.5%; level 1). Otherwise, the majority of injected embryos at 24 hpf were classified as level 3 when injected with 0.4-10 ng AS-MO. Deformities mainly included shortened body length, curved body, increased area above rhombomeres (engorged fourth ventricle), no clear brain segmentation, and non-circular small eyes. An example of a level 3 antisense morpholino injected embryo is pictured in figure 5.11. Other, less common malformations included improper heart rate, heart or heart sack deformation, non-circular yolks and no tail. Those embryos classified as level 2 had one of the problems mentioned at random.

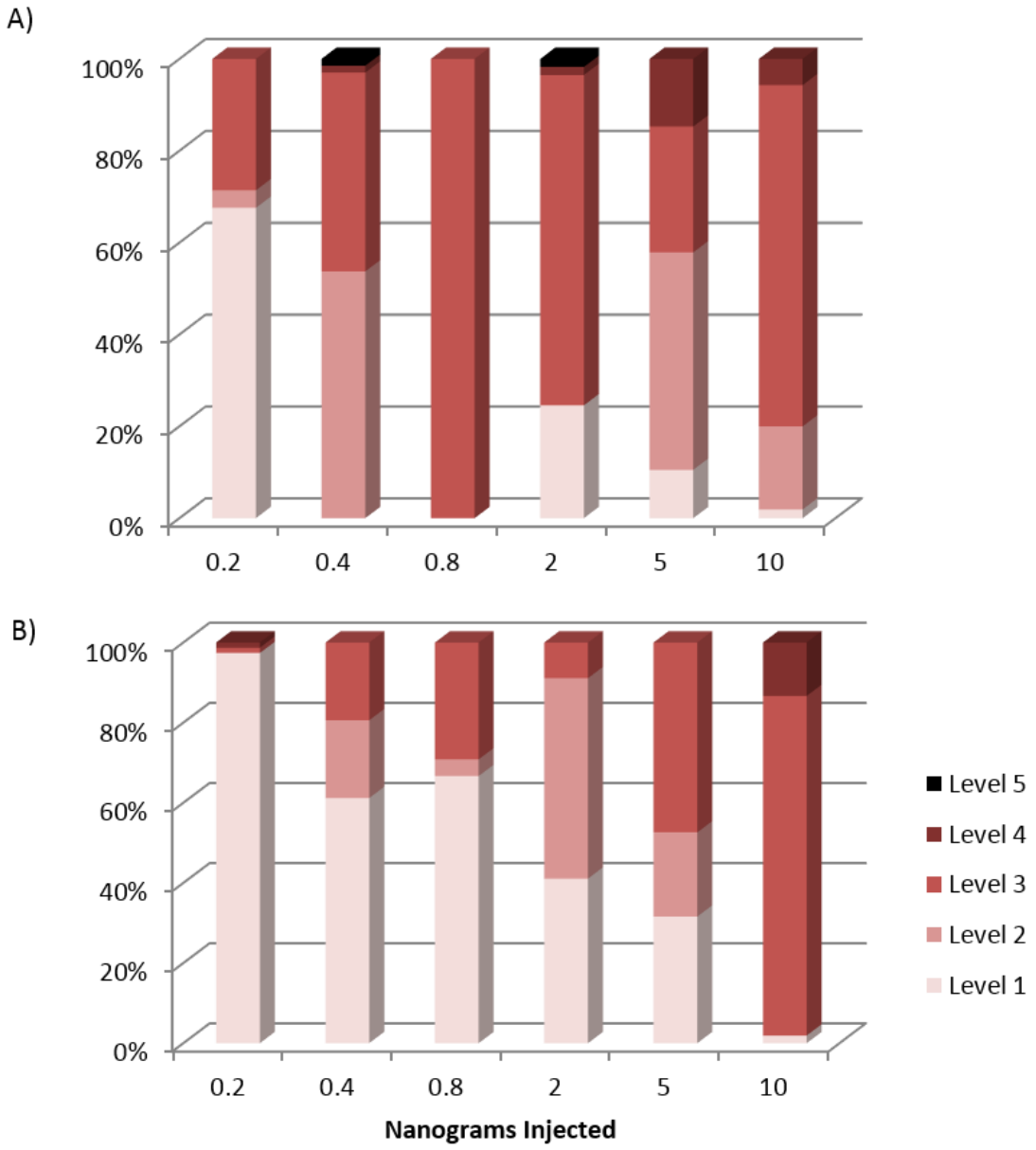


Figure 5.10 Phenotypes observed after Slc34a2a antisense morpholino injection into fertilised zebrafish oocytes. Different dilutions ranging from 0.2 to 10 ng were assessed at developmental time points 24 hpf (A) and 48 hpf (B). A minimum of 90 embryos from three clutches were utilised for each injection set.

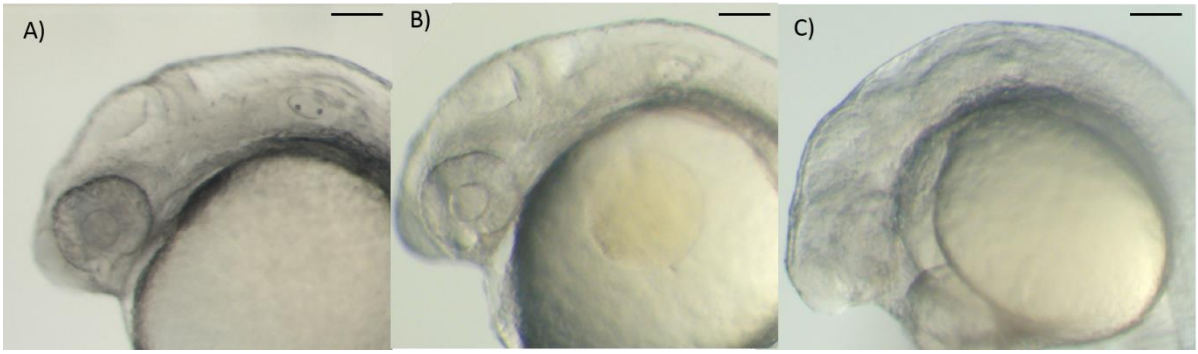


Figure 5.11 Effects of Slc34a2a antisense knockdown by injecting a splice site morpholino (Gene Tools, LLC). Embryos were pictured at 24 hpf. A) Non-injected fitness control. B) 10 ng of mismatch control morpholino. C) 10 ng of antisense morpholino. Scale bar at 0.1 mm.

The injected embryos appeared to look ‘healthier’ after examination at 48 hpf. The percentage of embryos classified as normal (level 1) increased as amount of antisense morpholino injected decreased, indicative of a dilution effect (figure 5.10 B). Embryos injected with 10 ng were still mostly classified as level 3 and had malformations including shorted body length, curved bodies, improper brain segmentation and incorrect eye development (figure 5.12). In comparison to 24 hpf embryos, 48 hpf embryos seemed to recover regardless of the amount injected. At 48 hpf, 0.2 ng injected embryos appeared as wildtype. The mismatch control morpholino caused no phenotypic effect at neither 5 ng nor 10 ng (data not shown).

5. Mechanism

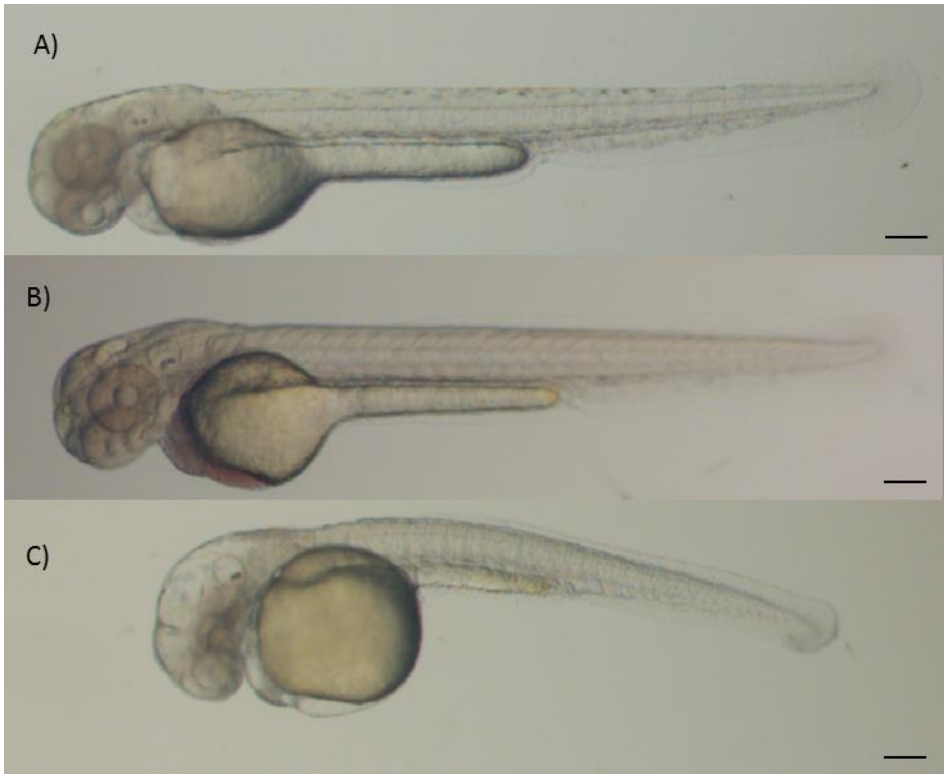


Figure 5.12 Representation of 48 hpf embryos injected with Slc34a2a antisense morpholino. Comparison of length and body shape after Slc34a2a antisense morpholino injection in comparison to control groups. A) Non-injected fitness control. B) 5 ng of mismatch control morpholino. C) 5 ng of Slc34a2a antisense targeting morpholino. Scale bar at 0.1 mm.

5.7 Expression profiles of Slc34a2b, Slc34a2a sense and antisense in embryos microinjected with antisense targeting morpholino

Even though injected Slc34a2a antisense morpholino likely produced a non-specific phenotype, semi quantitative RT-qPCR was completed to determine if the injected Slc34a2a(as) targeting morpholino reduced the concentration of its target transcript or had any effect on Slc34a2a or Slc34a2b transcript levels. Five embryos, chosen at random, from minimum of three different clutches were utilised as mentioned in chapter 2 and assessed by qPCR (figure 5.13).

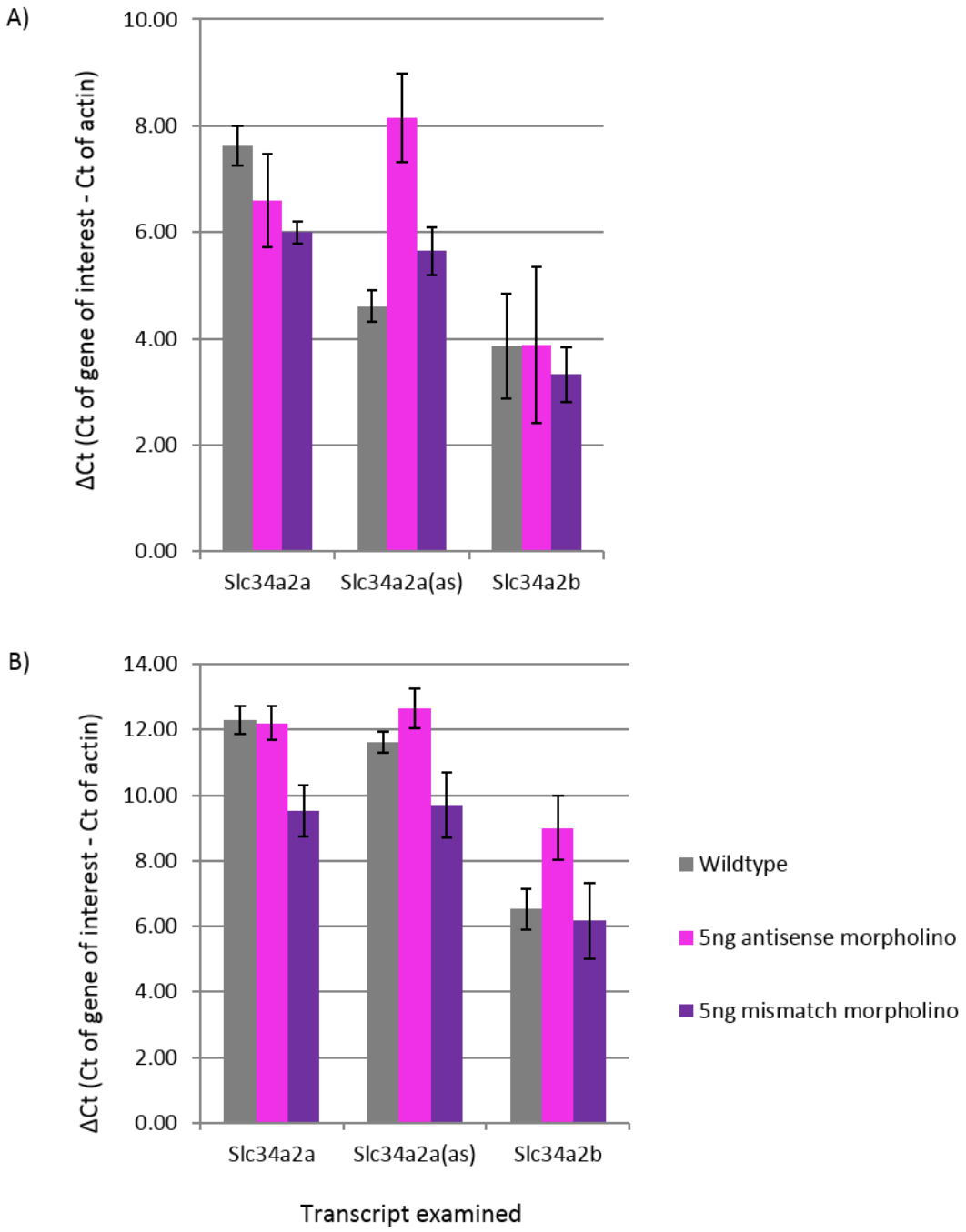


Figure 5.13 Semi quantitative RT-qPCR analysis of Slc34a2a(as) morpholino injected zebrafish embryos. RNA was extracted from five fertilised zebrafish embryos at 24 hpf (A) and 48 hpf (B) after injection with 5 ng Slc34a2a antisense morpholino (pink bars) in comparison to non-injected samples (grey) and mismatch control morpholino (purple bars). ΔCt was calculated by using β-actin as gene reference which maintained a consistent Ct throughout all experiments. N=3

5. Mechanism

In 24 hpf embryos, antisense morpholino injection did affect Slc34a2a(as) levels, as seen in the top graph of figure 5.13. Transcript levels decreased by a ~ΔCt of 4 (ΔCt wildtype; 4.61±0.29, ΔCt AS-MO injected; 8.14±0.83). Antisense morpholino injection did not have an impact on neither Slc34a2a nor Slc34a2b levels. By 48 hpf, Slc34a2a(as) levels of 5 ng antisense morpholino injected samples returned back to near wildtype levels (ΔCt wildtype; 11.62±0.33, ΔCt AS-MO injected; 12.65±0.62, figure 5.13 B). Slc34a2a and Slc34a2b levels remained unaffected.

Visualised in both 24 hpf and 48 hpf samples, injected mismatch morpholino did appear to result in a ~ΔCt of 1.5 change in Slc34a2a and Slc34a2a(as) levels. Slc34a2a sense transcript levels changed such that in 24 hpf mismatch samples, ΔCt was 6.00±0.20 and 9.52±0.79 in 48 hpf samples (wt ΔCt at 24 hpf, 7.62±0.38; wt ΔCt at 48 hpf, 12.31±0.42). Slc34a2a(as) transcript levels after mismatch morpholino injection, ΔCt was 5.65±0.79 in 24 hpf samples and 9.69±1.01 in 48 hpf samples (wt ΔCt at 24 hpf, 4.61±0.29; wt ΔCt at 48 hpf, 11.62±0.33). This is unlikely to be specific as the mismatch morpholino does not bind with high affinity to any endogenous targets in zebrafish (Ensembl, BLAST search) but rather a broader systemic effect. Injection of the mismatch morpholino, which is five nucleotides different from the antisense morpholino, did not appear to have any effect on Slc34a2b transcript levels at either 24 nor 48 hpf.

5.8 Co-injection of p53 morpholino reduces morpholino toxicity

Injection of Slc34a2a antisense morpholino decreased Slc34a2a(as) transcript levels in 24 hpf embryos, but not 48 hpf embryos, as shown in the above RT-qPCR. However, upon visual examination, the injected embryos showed a non-specific phenotype with multiple severe deformities including shortened body length, curved body, increased area above rhombomeres (fourth ventricle), lack of brain segmentation, and non-circular small eyes. These disfigurements are consistent with general morpholino toxicity which is caused by the activation of p53 mediated apoptosis (Robu *et al.*, 2007). p53 activation causes cell death in somite tissues and the central nervous system thus producing these non-target specific phenotypes. To inhibit the p53 pathway and mitigate unspecific toxicity, a well characterised p53

morpholino was co-injected alongside the antisense morpholino. The double-knockdown strategy would allow the specific phenotypic effect of antisense RNA knockdown to be determined.

The injection process was completed as described; data was collected from a minimum of three clutches which were divided to include a non-injected fitness control, a negative control (water and phenol red, mismatch morpholino).Two different concentrations of antisense morpholinos were used. All co-injections included 4 ng of Gene Tools LLC p53 morpholino. Co-injection (p53 plus AS-MO injection) data came from a minimum of 129 embryos.

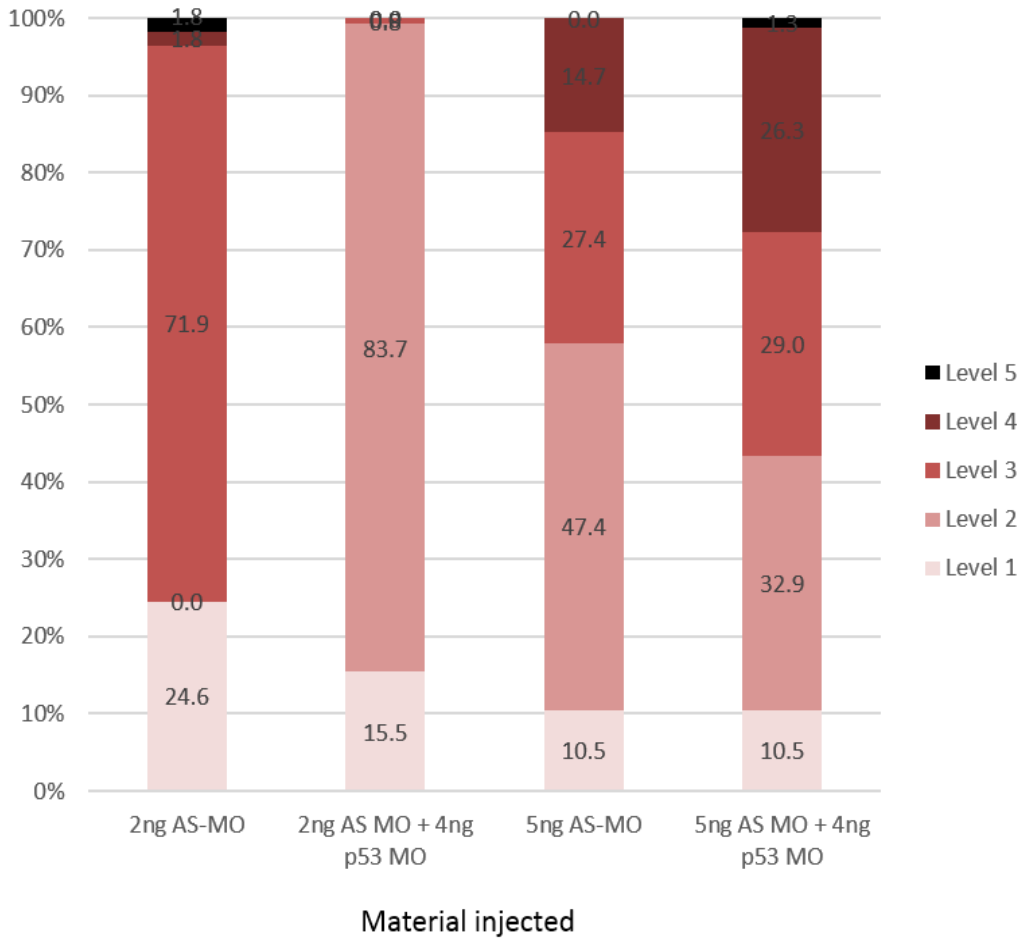


Figure 5.14 Phenotypic classification of 24 hpf embryos after co-injection of Slc34a2a(as) and p53 targeting morpholinos. Embryos were injected with either 2 ng or 5 ng Slc34a2a antisense morpholino with or without 4 ng of p53 morpholino. N=3

5. Mechanism

Co-injected embryos at 24 hpf fared better than their counterparts that were solely injected with 2 ng antisense morpholino (AS-MO) as seen in figure 5.14 (two left bars). 71.9% of embryos injected with 2 ng antisense morpholino were classified as level 3 displaying many non-specific phenotypic deformities mentioned previously. However, the majority of embryos co-injected with 2 ng AS-MO and 4 ng p53 morpholino were classified as level 2 (83.7%), indicating morpholino toxicity was decreased due to p53 knockdown. This significant recovery to near wildtype phenotype was not visible in embryos injected with 5 ng of antisense morpholino. The number of embryos classified as wildtype, level 1, remained unchanged at 10.5% between 5 ng antisense morpholino injected embryos and co-injected embryos (figure 5.14, third and fourth bars). Overall, the co-injected embryos (5 ng AS-MO and 4 ng p53 morpholino) did not fare better than embryos injected with 5 ng AS-MO alone. Overall, no specific phenotype was observed after co-injection.

Injected embryos examined at 48 hpf showed the same trend as seen at 24 hpf. As depicted in figure 5.15, embryos co-injected with 2 ng AS-MO and 4 ng p53 morpholino were nearly all classified as wildtype (level 1, 98.8%) in contrast to those embryos which received solely 2 ng of AS-MO (level 1, 41.1%). When 5 ng of antisense morpholino was injected, co-injection of the p53 morpholino did not diminish the toxic effect as all co-injected embryos were classified as level 2 or 3. In comparison, 5ng AS-MO injected embryos had 31.6% classified as level 1.

Embryos solely injected with 2 ng of antisense morpholino displayed severe developmental malformations at both 24 and 48 hpf; improper body shape and length, incorrect brain segmentation and proportions, non-circular eyes, and tail deformation such as twisted or blunt tail tips. These developmental deformities affected the behaviour so that fish could not swim correctly or not move at all. Therefore none of the fish showed a normal escape reaction in response to a poke. Co-injection of 4 ng of p53 morpholino rescued the severe phenotype so that by 48 hpf, embryos showed wildtype behaviour.

The recovery to wildtype was not seen when 5 ng of antisense morpholino was injected. Embryos injected with solely 5 ng of antisense morpholino displayed the same severe deformities as observed for the 2 ng antisense injected samples and as demonstrated in figures 5.11 and 5.12. Morpholino toxicity was not reduced upon co-injection of p53 morpholino. Malformations in these embryos were not consistent and behaviour continued to be unlike wildtype.

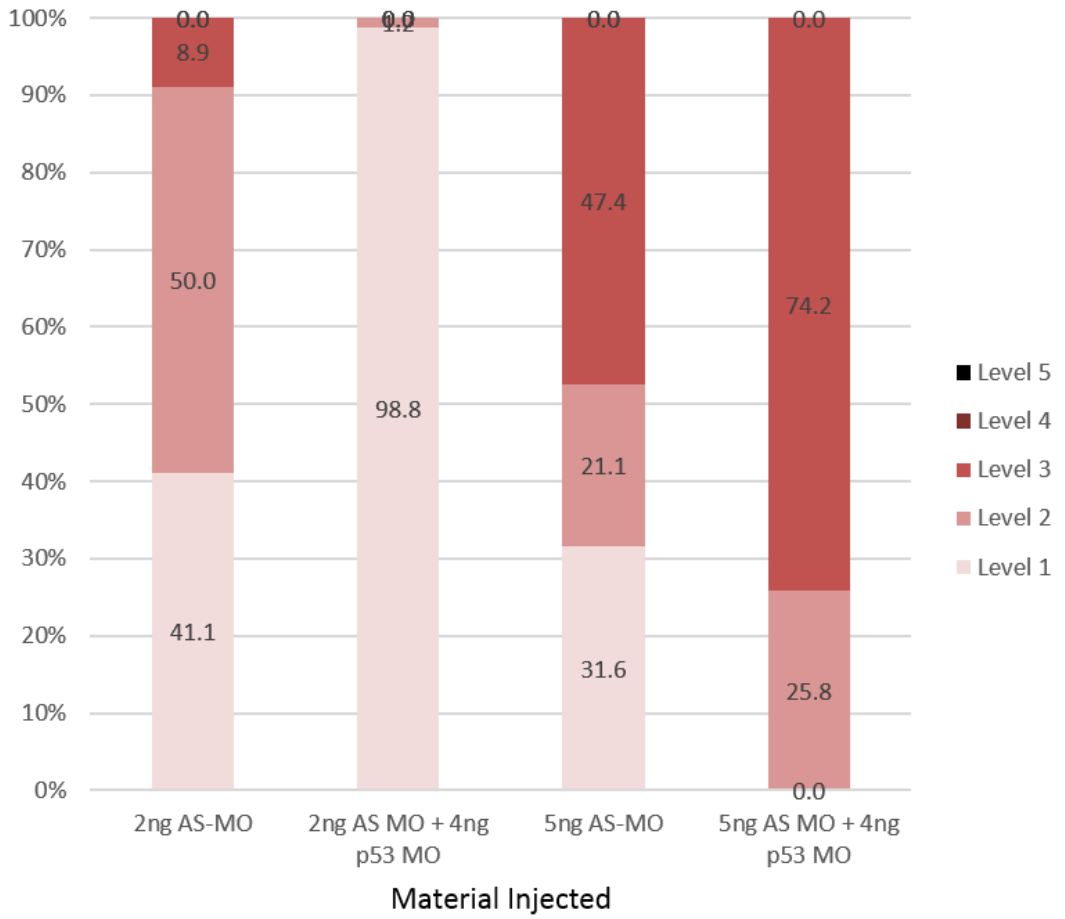


Figure 5.15 Phenotypic classification of 48 hpf embryos after co-injection of Slc34a2a(as) and p53 targeting morpholinos. Embryos were injected with either 2 ng or 5 ng Slc34a2a antisense morpholino with and without 4 ng p53 morpholino. N=3

As shown in figure 5.16 (C), the embryo injected with 5 ng antisense morpholino alone had an engorged area above the rhombomeres (fourth ventricle), smaller non circular eyes, larger nose/face region, and even though the cerebellum was present, it appears smaller in size and pushed forward. (The angle of the pictured embryo looks different, however, this appearance is due to the deformed body and tail. All embryos were pictured laterally with aligned eyes). Knocking down p53 reduced the severity of the deformations but the embryos still had a reduced area above the rhombomeres figure 5.16 (B). This reduction of space above the rhombomeres was

consistent in co-injected embryos whilst the engorged fourth ventricle area was common in 5 ng antisense morpholino injected samples.

Figure 5.16 Injection of Slc34a2a(as) targeting morpholino affected the zebrafish hindbrain. 48 hpf embryos are shown. The size of the area found above the rhombomeres (fourth ventricle) was altered as indicated by the red arrows. Non-injected fitness control embryo (A) in comparison to an embryo co-injected with 5 ng of antisense morpholino and 4 ng of p53 morpholino (B) and an embryo injected with 5 ng of antisense morpholino (C) at 48 hpf. Scale bar at 0.1 mm.

5.9 In situ hybridization on antisense morpholino injected embryos

To address whether the injections of the splice site AS-MO reduced the levels of Slc34a2a(as), in situ hybridisation was performed. ISH staining was completed on 48 hpf embryos injected with the antisense morpholino and mismatch morpholino alongside wildtype non-injected embryos. Antisense and Engrailed-2 (Eng2) probes were utilised as described in Chapters 3 and 4. As shown in figure 5.17 A, the antisense probe stained in the pharynx, endoderm (future digestive tract and swim bladder), primordial midbrain channel and primordial hindbrain channel of the 48 hpf wildtype embryo in agreement with figure 3.3. The mismatch morpholino injected embryos (figure 5.17, C), presented a comparable pattern to the wildtype but showed slightly weaker staining. The 5ng antisense morpholino injected samples

contrasted these findings with no detectable antisense staining in the majority of embryos (B). All these embryos displayed the severe developmental deformities mentioned in the previous subsections.

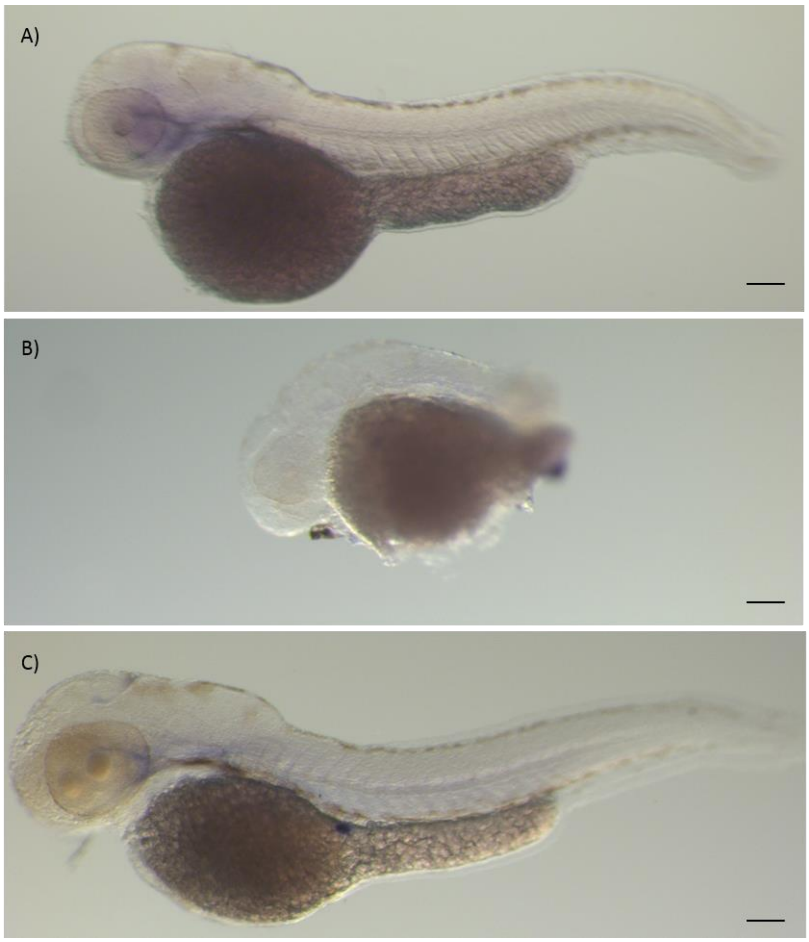


Figure 5.17 48 hpf embryos stained by in situ hybridization with a Slc34a2a(as) probe after antisense morpholino injection. All embryos were stained simultaneously with aliquots of the same probe. A) Wildtype non-injected embryos. B) 5 ng antisense morpholino injected embryos. C) 5 ng mismatch morpholino injected embryos. Scale bar at 0.1 mm.

5. Mechanism

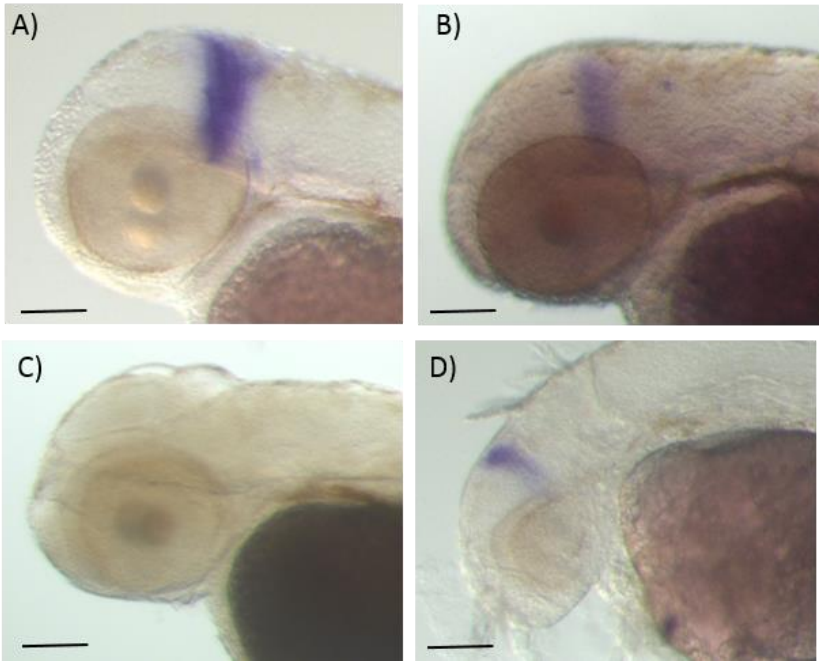


Figure 5.18 48 hpf embryos stained by in situ hybridization with an Engrailed-2 probe after 5 ng antisense morpholino injection. All embryos were stained simultaneously with aliquots of the same probe. A) Wildtype non-injected embryo. B) 5 ng mismatch morpholino injected control. C and D) 5 ng antisense morpholino injected embryos. Scale bar at 0.1 mm.

Embryos from the same sets of injections were stained with Eng2 to determine if the morpholino injections had an effect on cerebellum development. As reported in the previous subsection, antisense morpholino injected 48 hpf embryos displayed correctly segmented brains but the proportions were incorrect. Comparable malformations were detectable after in situ hybridization as shown in figure 5.18. As displayed in both wildtype (A) and mismatch morpholino injected control (B), Eng2 created the purple wedge staining encompassing the midbrain-hindbrain boundary and the cerebellum. Again, the mismatch morpholino injected embryos produced the correct but a slightly weaker staining. Embryos injected with 5 ng antisense morpholino showed neither consistent nor correct staining. About a third of the embryos produced no stain such as in figure 5.18 C, possibly the due to a complete lack of brain segmentation and no brain development outside of frontal lobe. In the remaining embryos, the cerebellum was present and stained with the Eng2 probe but was pushed forward and small (figure 5.18, D). This result is in conjunction with

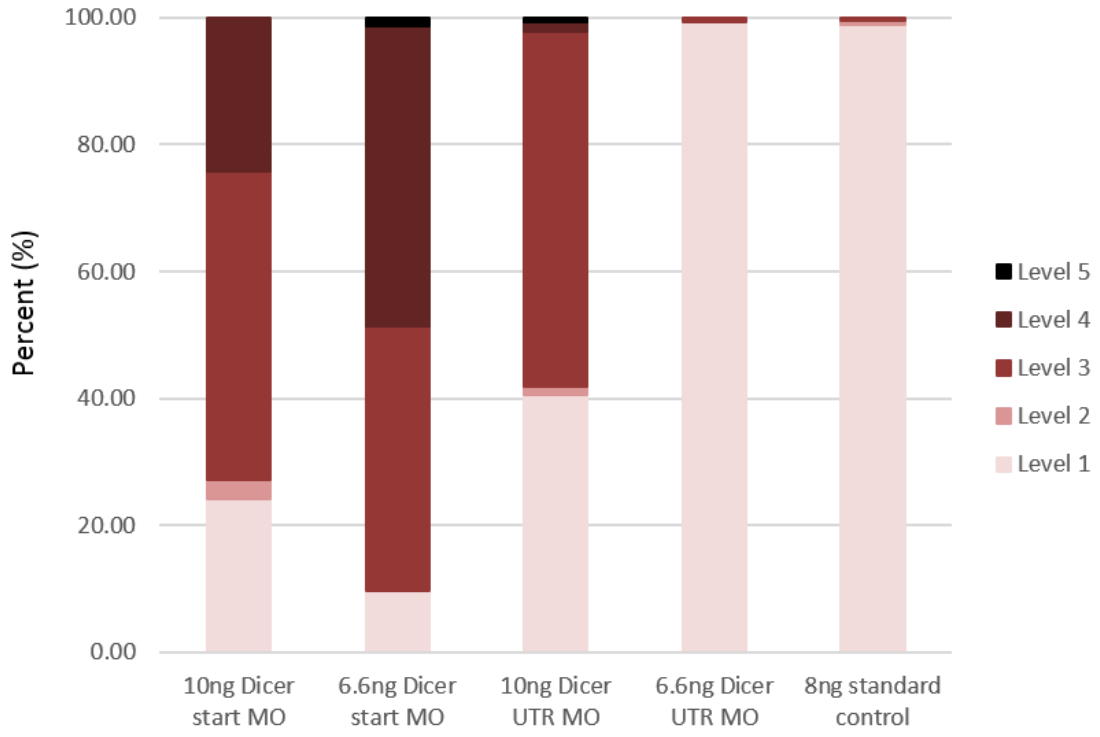
data presented in figure 5.16 that the cerebellum is pinched if the fourth ventricle is engorged. These partly conflicting findings are likely due to the unpredictable effects of morpholino toxicity on zebrafish development rather than the specific consequences of reduced antisense expression.

5.10 Evaluation of Dicer morpholino toxicity

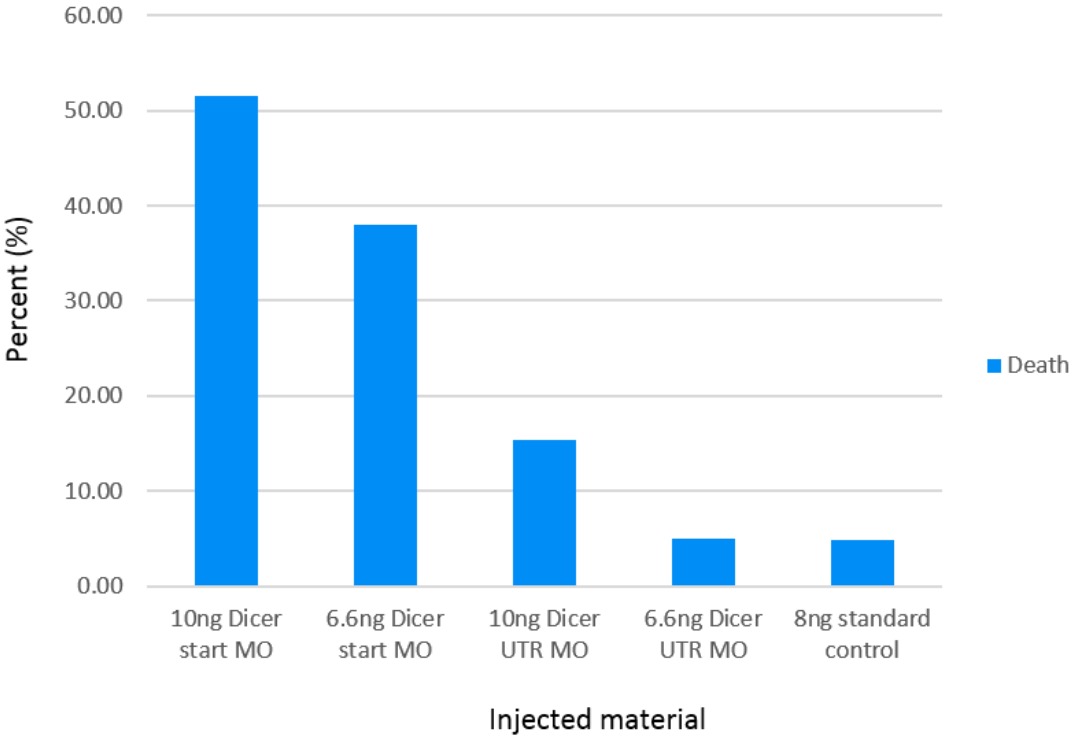
The aforementioned sections established that injection of any RNA containing a complimentary sequence to the endogenously expressed Slc34a2a antisense transcript was found to produce a specific phenotype; failure to develop a cerebellum. Dicer is a key RNAi enzyme which is involved in cutting RNA hybrids (dsRNA) into short interfering RNAs (siRNAs). To determine if siRNAs were involved in the failure to generate a cerebellum, Dicer was knocked down with morpholinos. Aliquots of two morpholinos that have been reported to knockdown Dicer were kindly provided by James Patton (Department of Biological Sciences, Vanderbilt University, Nashville, USA). Both morpholinos were designed by and ordered from Gene Tools LLC. The first antisense morpholino, Dicer start MO, targeted the translational start site whilst the second one, Dicer UTR MO, was designed against the 5' untranslated region of the *Dicer1* mRNA. The sequences can be found in chapter 2, materials and methods. Microinjection procedure was completed as previously described. Both morpholinos were injected at 10 ng and 6.6 ng dilutions. A standard control morpholino with no biological target (Gene Tools LLC) was injected at 8 ng as a negative control and produced no phenotypic effect (figure 5.19).

5. Mechanism

A1)



A2)



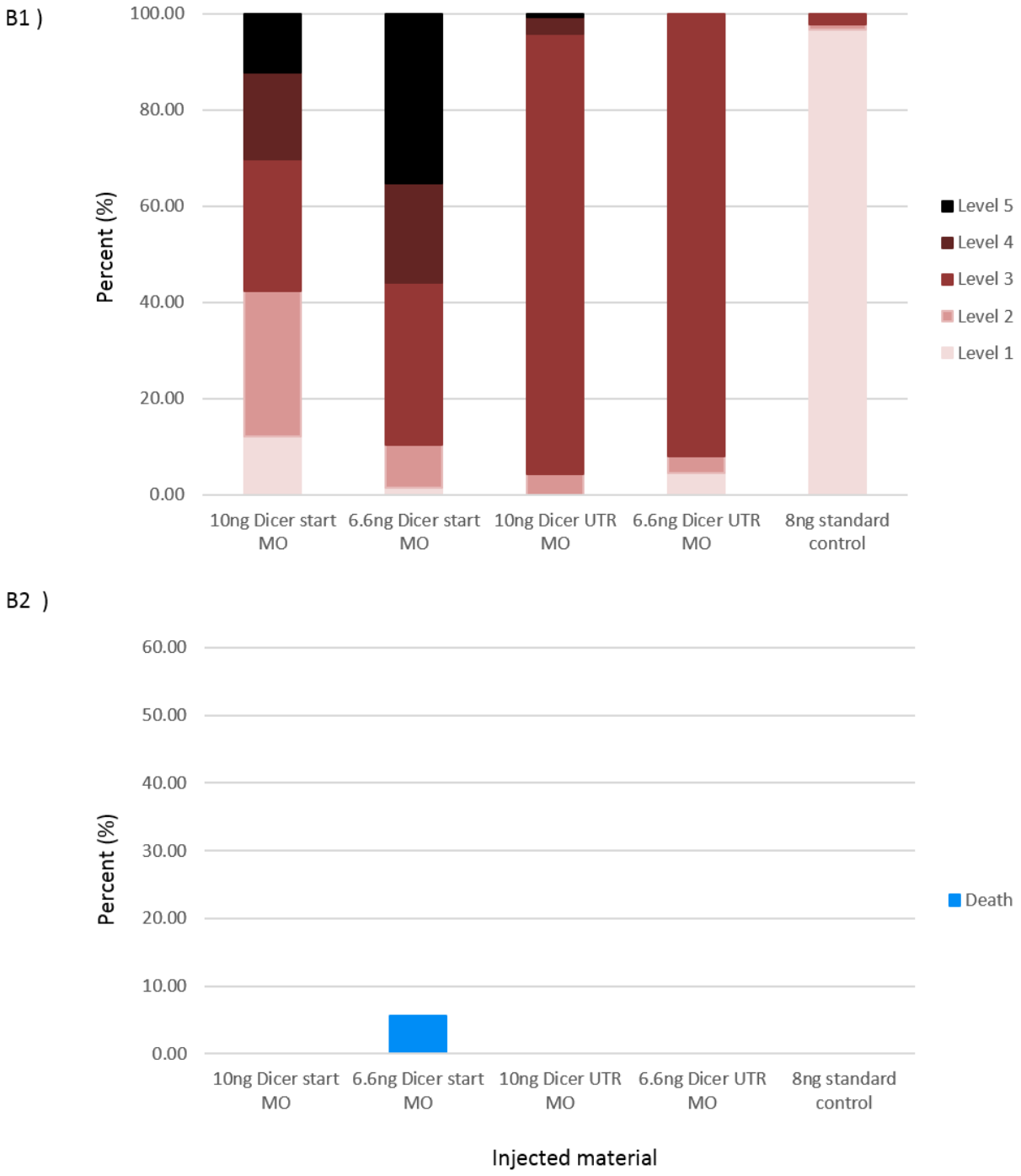


Figure 5.19 Phenotypic classification (A1 and B1) and death rates (A2 and B2) of embryos injected with two different Dicer targeting morpholinos. Embryos were analysed at 24 hpf (A) and 48 hpf (B). Data was collected from a minimum of 60 fertilised oocytes (average 105) from at least three clutches for each injection.

At both, 10 ng and 6.6 ng, the Dicer translation start site morpholino produced a high death rate for embryos examined at 24 hpf; 51.6% of embryos injected with 10 ng and 37.9% of those injected with 6.6 ng died within the first 24 hours as demonstrated in Figure 5.19 A (first two bars, in blue). The majority of surviving embryos were also classified as level 3 or 4 according to their severe phenotypic malformations. Those injected with 10 ng Dicer start morpholino, were 48.5% level 3 whilst 24.2% were level 4. Injection with 6.6 ng did not lessen the severity of malformations as 41.7% of the fish were level 3 and 47.2% were level 4. Embryos were severely developmentally delayed with small heads, no brain formation, extremely short curved or bent tails and occasionally no eye development. By 48 hpf, the embryos only seemed to worsen as demonstrated in Figure 5.19 B, first two bars. The number of embryos classified as level 1, wildtype, dropped from 24.2% to 12.1% in 10 ng injected samples and from 9.7% to 1.5% in 6.6 ng injected samples. Those that were level 1 at 24 hpf, were classified as level 2 or 3 at 48 hpf due to incorrect brain segmentation, improper eye formation or shortened body length. Overall, injection of Dicer start morpholino seemed to severely developmentally hinder the embryos.

Embryos injected with 10 ng of Dicer UTR morpholino were mostly classified as level 1 (40.6%) or level 3 (55.9%) at 24 hpf. When the amount of injected morpholino was lowered to 6.6 ng, nearly all embryos (99.3%) were classified as level 1, as shown in Figure 5.19 A, third and fourth bars. However, embryos worsened by 48 hpf (Figure 5.19 B, third and fourth bar) as seen with Dicer start morpholino injected samples. For both amounts, nearly all embryos were classified now as level 3 (91.6% level 3 at 10 ng, 92.1% at 6.6 ng). Prevalent phenotypic traits were short body/ tail lengths, body curvature and/or small non-circular eyes. Generally, a broader range of phenotypic effects was visible upon injecting Dicer start morpholino in comparison to Dicer UTR morpholino.

A key difference between Dicer start and Dicer UTR morpholino injected samples was the much lower death rate in Dicer UTR injected groups; 15.4% of the 10 ng injected fish died whilst 5.0% died in 6.6 ng injected samples by 24 hours and none passed away afterwards. In comparison, over 30% of injected embryos with Dicer start morpholino passed away in the first 24 hours and a few more in the following

day. The standard control morpholino caused no phenotypic effect with nearly all embryos classified as level 1 (fifth bar in figures 5.19 A and B).

Injection of Dicer start morpholino caused an extreme death rate in embryos prior to 24 hours that was not witnessed with any other injected materials within this study. Alongside the very severe phenotypic deformities Dicer start morpholino caused, it was decided all experiments from this point onwards should utilise the Dicer UTR morpholino.

5.11 Dicer morpholino co-injection with p53 morpholino

Non-target specific phenotypes can be caused by morpholino toxicity through the activation of p53 mediated apoptosis (Robu *et al.*, 2007). Therefore, the previously used p53 morpholino was co-injected with the Dicer UTR morpholino to mitigate any unspecific effects caused by morpholino toxicity. As demonstrated in figure 5.20, co-injection of 4 ng of p53 morpholino caused a partial rescue of 10 ng Dicer UTR injected samples. As established in the previous set of experiments, 91.6% of embryos injected with 10ng of Dicer UTR morpholino were classified as level 3; with co-injection of the p53 morpholino, only 20% of embryos were scored level 3 whilst the remaining majority were either level 1 (46.2%) or level 2 (33.9%). The rescue was only visible at high concentrations of Dicer UTR morpholino injections. When 4 ng p53 morpholino was co-injected with 2.5 ng Dicer UTR morpholino, 98.2% of embryos were classified as level 3. These embryos still had shorter body length and non-circular eyes as shown in figure 5.21.

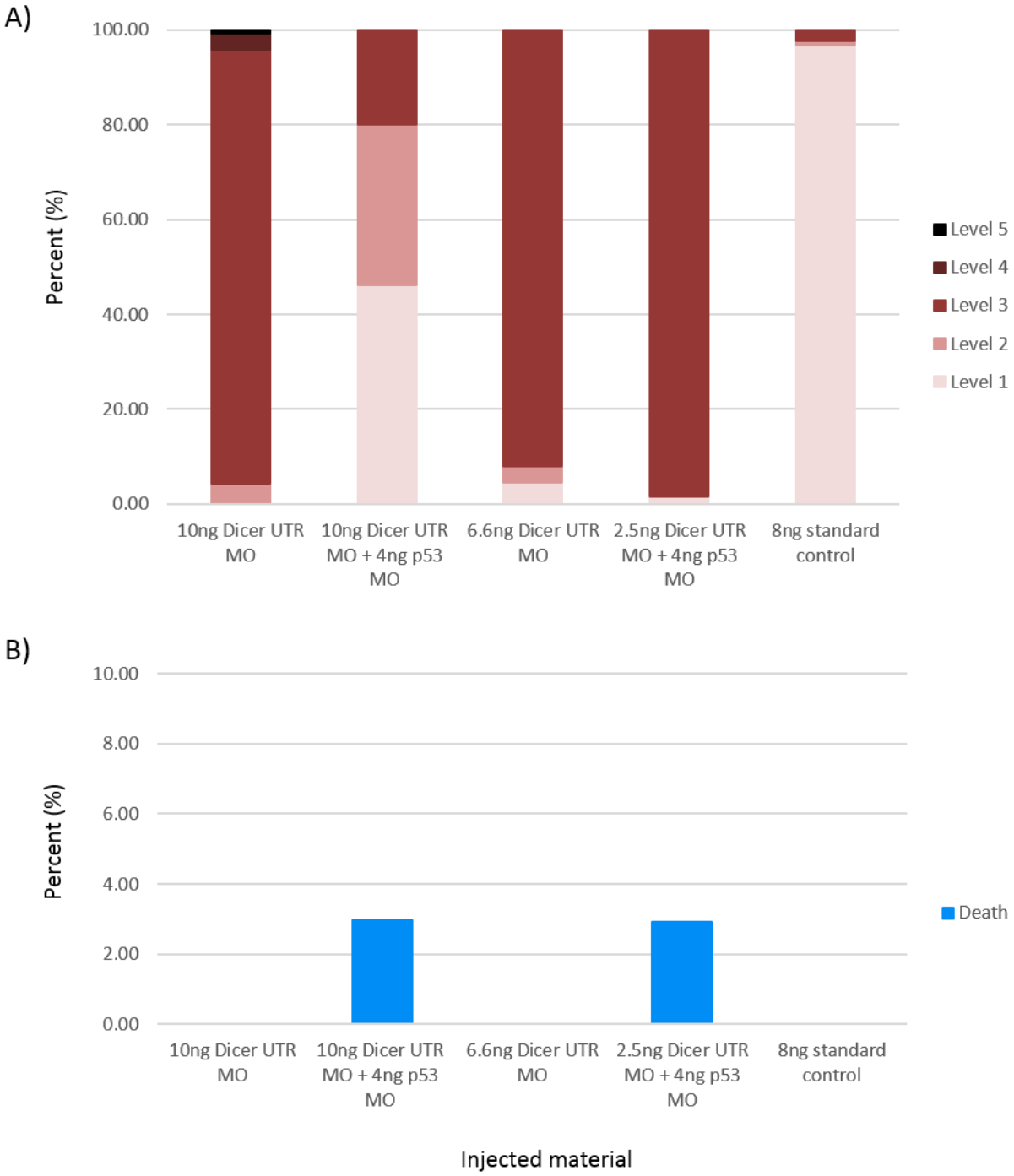


Figure 5.20 Phenotypic classification (A) and death rates (B)) of embryos co-injected with p53 and Dicer UTR morpholinos at 48 hpf. Data collection based on a minimum of 60 embryos from a minimum of three clutches for each injection.

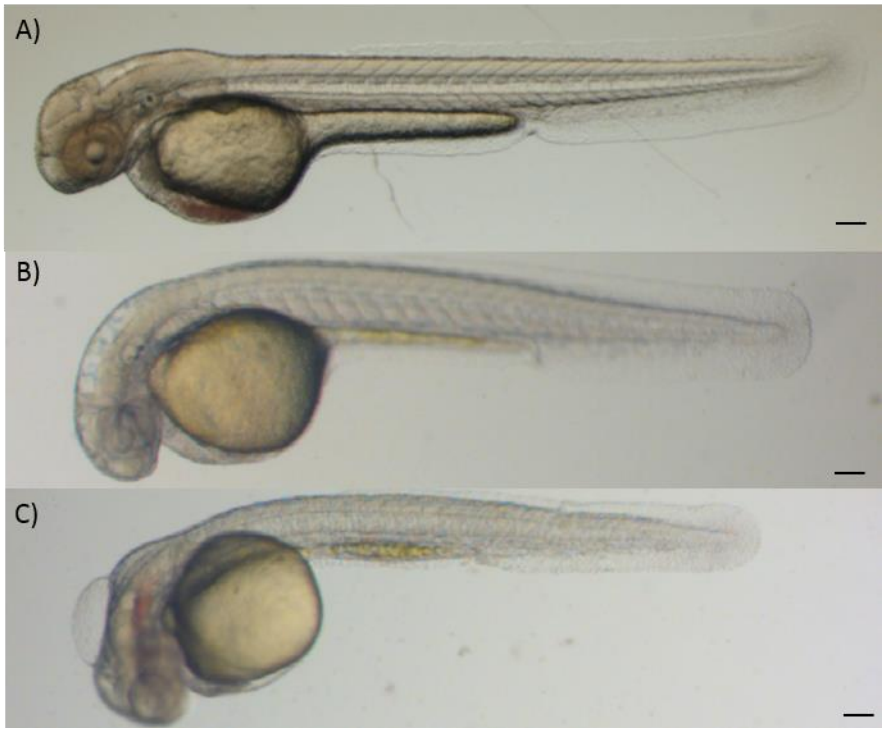


Figure 5.21 Co-injection of p53 morpholino reduced the phenotypic severity witnessed in 10 ng Dicer UTR morpholino injected embryos at 48 hpf. Non-injected fitness control embryo (A) in comparison to a 2.5 ng Dicer UTR morpholino/4 ng p53 morpholino co-injected embryo (B) and an embryo receiving solely 10 ng Dicer UTR morpholino (C). Scale bar at 0.1 mm.

5.12 Dicer knockdown rescues loss of cerebellum caused by Slc34a2a RNA injection

Injection of RNA Slc34a2a fragments that are complementary to the endogenous Slc34a2a(as) RNA transcript produce a consistent, specific phenotype; the failure to develop a cerebellum. As Dicer is the key RNAi enzyme that cuts RNA hybrids (dsRNA) into short interfering RNAs (siRNAs), knocking down Dicer may alleviate the deleterious effects of Slc34a2a RNA injection. It would also determine if Slc34a2a sense/antisense dsRNA formation was required to induce the process that leads to the loss of cerebellum. In this hypothesis, Slc34a2a sense and antisense transcripts would hybridise upon Slc34a2a RNA injection. Without Dicer, the dsRNA

of sense/antisense would not be processed to produce siRNAs. If there is no loss of cerebellum upon co-injection, then a siRNA produced from the complementary region must be involved in the pathway that results in cerebellar development failure.

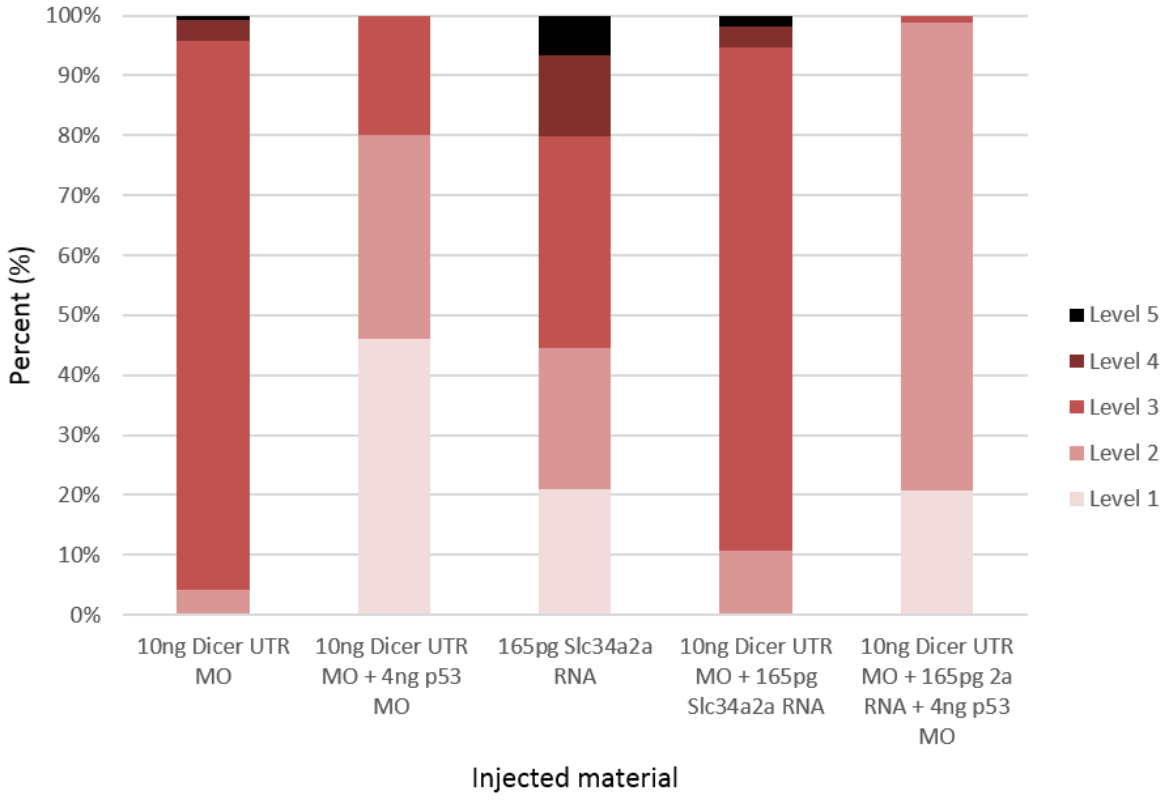


Figure 5.22 Phenotypic classification of fertilized zebrafish embryos injected with combinations of 165 pg capped full length Slc34a2a RNA, 10 ng Dicer UTR morpholino and 4 ng p53 morpholino assessed at 48 hpf. A minimum of 60 embryos from three clutches were used for each data set.

Due to the high death rate caused by Dicer start morpholino, only the Dicer UTR morpholino was used for the following experiments. Co-injection of 10 ng Dicer UTR morpholino and 165 pg capped full length Slc34a2a RNA produced a specific phenotype in all injected embryos (Figure 5.22, fourth bar). All embryos at 24 (data not shown) and 48 hpf examinations had a concoction of the following phenotypes

thus classifying them as level 3-5; small non-circular eyes, no brain formation or improper brain segmentation, shortened body length, curved body, smaller heads, incorrect heart rates, misshapen tails and/or had a clear bump on the head (as seen in figure 5.21 C). All embryos also had behavioural insufficiencies such that they did not react to the poke stimulus.

p53 morpholino co-injection has been found to rescue non-specific phenotypes caused by morpholino toxicity (figure 5.22, bar 1 vs bar 2). A triple injection consisting of Slc34a2a RNA, Dicer UTR morpholino and p53 morpholino was completed to reduce any non-specific phenotypes present in the Dicer UTR/Slc34a2a RNA injected samples. As shown in figure 5.22, last bar, over 98% of triple injected fertilised oocytes were classified as level 1 or 2. Of these embryos, all had a cerebellum and exhibited normal behaviour. The 78.2% that were classified as level 2 all had a specific phenotype that presented as an enlarged hindbrain region. Specifically, there was an enlarged area above the rhombomeres (fourth ventricle) (Figure 5.23, red arrow). To confirm the rescue of cerebellar loss caused by Slc34a2a RNA injection, in situ hybridization was completed utilising the Engrailed-2 probe as in previous in situ experiments. Triple injected embryos were stained alongside non-injected controls and Slc34a2a RNA injected embryos (not included in figure 5.23). As visualised in figure 5.23, triple injected embryos (B and D) developed a cerebellum just as wildtype embryos at 48 hpf (A and C).

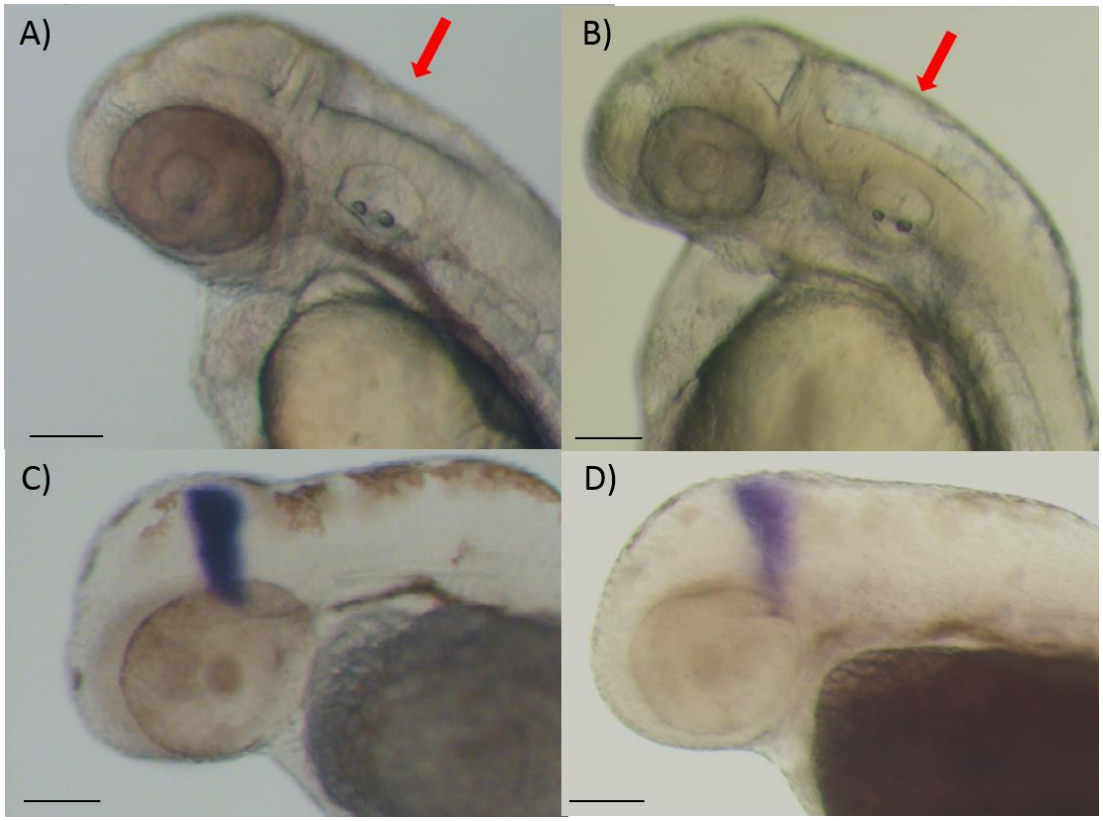


Figure 5.23 Loss of cerebellum caused by Slc34a2a RNA injection was rescued upon co-injection of Dicer UTR and p53 morpholinos. 48 hpf embryos are shown: A) Fitness non-injected control; B) triple injected embryo with an enlarged area above the hindbrain (red arrows); C) fitness non-injected control embryo with Engrailed-2 stain; D) Triple injected embryo with Engrailed-2 staining. Scale bar at 0.1 mm.

5.13 Can endo siRNAs mimic microRNA action?

Sense/antisense generated endo-siRNAs could produce a Dicer and RISC mediated phenotype reminiscent of an off target effect of microRNAs. A BLAST search was performed with the Slc34a2a sense/antisense complementary sequences (exons 10/13). Hits with a minimal length of 15 nucleotides and 95% identity were further investigated. Moreover, matching seed regions were required. Expression patterns and knockdown phenotypes of 76 genes with significant identity were manually researched (Appendix C). Three candidate genes were identified which had a role in

brain development; Presenilin2 (Psen2), Wingless-type MMTV integration site family member 4b (Wnt4b) and RAS-like, family 11, member B (Rasl11b).

Presenilin2 (Psen2), which is transcribed from chromosome 1, is expressed throughout embryogenesis, especially within the neural crest (Thisse, 2004a). This transmembrane protein is found in the membranes of the Golgi apparatus or endoplasmic reticulum (Fraering, 2007). Belonging to the peptidase A22A family, Psen2 forms a homodimer and is part of the gamma secretase complex which is involved in the cleaving of type-1 transmembrane proteins such as Notch (Nornes *et al.*, 2009). Psen2 knockdown studies found a reduction in hindbrain development due to reduced notch signalling (Groth *et al.*, 2002; Campbell *et al.*, 2006; Nornes *et al.*, 2009). Orthologues have been found in various other species including primates, rodents, fish, birds and reptiles (Ensembl, 2016a).

Wingless-type MMTV integration site family member 4b (Wnt4b), which is transcribed from chromosome 16, has three splice variants. They are expressed from fertilization and are located in the central nervous system; floor plate and hindbrain (Liu *et al.*, 2000; Lu *et al.*, 2011). Wnt4b protein is secreted into the extracellular space and has numerous biological functions. It stimulates the Wnt signalling pathway which regulates gene transcription, cytoskeleton formation, calcium homeostasis and most importantly, is necessary in the proper formation of the midbrain-hindbrain boundary (Liu *et al.*, 2000; Verkade and Heath, 2009). It is involved in the regulation of mitotic nuclear division, neuron differentiation and cell fate commitment. Functional orthologues have been noted in mice, rabbits, birds, reptiles, fish and humans (Verkade and Heath, 2009).

RAS-like, family 11, member B, abbreviated as Rasl11b, is transcribed from chromosome 20. Expressed from fertilization through embryogenesis, Rasl11b transcript is located in the central nervous system; specifically in the diencephalon and the anterior neural rod (Rauch, 2003; Pézeron *et al.*, 2008). The protein is a small cytosolic GTPase which is involved in signal transduction and mesendoderm development (Stolle *et al.*, 2007; Ensembl, 2016b).

For each candidate gene, ultramers were designed incorporating T7 and universal M13 primer sites (table 5.1). The hairpin RNAs contain a sequence from Slc34a2a locus which binds to a complimentary sequence in the candidate gene thus allowing the RNA to fold back on itself (dsRNA joined on one end). The designed

oligonucleotides were amplified by end-point PCR using T7/ M13 primers prior to purification. Restriction digest to remove the M13 primer site was completed with *Xba*I prior to synthesising the RNA (section 2.3.4). Microinjection of the hairpin RNA would determine if the three candidate genes could mimic Slc34a2a sense/antisense duplex formation and siRNA production. Each hairpin was injected at two dilutions, 82.5 pg and 165 pg, alongside a water/phenol red negative control. The phenotypic classification of all injected 48 hpf embryos is shown in figure 5.24.

Table 5.1 Areas of identity overlap between Slc34a2a and the three candidate genes. Below is the sequence of the cDNA oligonucleotides utilised to generate RNA hairpins. In red bold is the T7 sequence, purple bold is the M13 universal primer, black is an Xba1 restriction site, blue is the gene overlap sequence whilst linker sequences are in pink and the loop is in green.

Gene	**Overlapping sequence**	**Exon in Slc34a2a**	**Length (bp)**	**Orientation**	**Percent identity**
Psen2	CTGCTGAATTC_ATGCT CAA	10	20	Forward	95
TAATACGACTCACTATAGGGTGAAGCTGCTGAATTCTATGCTCAAGGGACAAAA GTGTCCCTTGAGCATACAATTCAGCAGCTTCATCTAG**ACTGGCCGTCGTTTTAC**					
Wnt4b	GAGCCTTTCGGACCT	10	15	Forward	100
TAATACGACTCACTATAGGGAACACGAGCCTTTCGGACCTGGCGGAAAAGTCC GCCAGGTCCGAAAGGCTCGTGTTTCTAGAC**TGGCCGTCGTTTTAC**					
Rasl11b	TGTTGCTGTTGCTGCA	13	16	Reverse	100
TAATACGACTCACTATAGGGCCCGCTGTTGCTGTTGCTGCAAGTGCAAAAGTG CACTTGCAGCAACAGCAACAGCGGGTCTAGA**CTGGCCGTCGTTTTAC**					

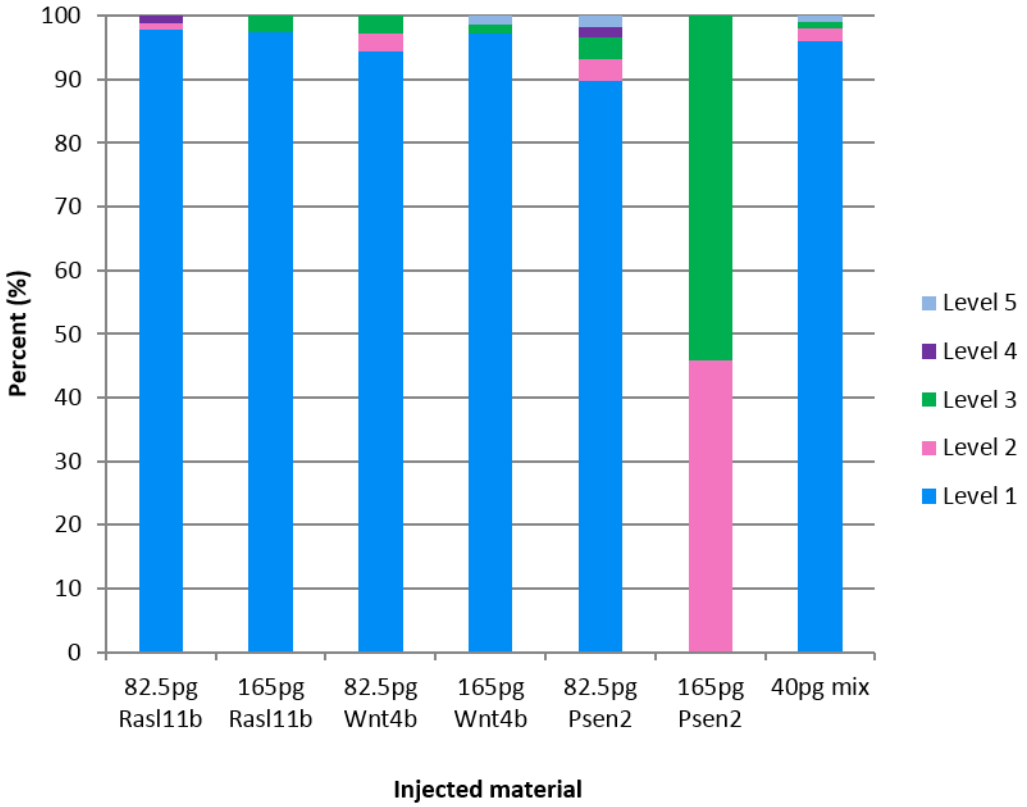


Figure 5.24 Phenotypic classification of embryos injected with RNA hairpins at 48 hpf. Each RNA hairpin was injected in two different amounts. A minimum of 50 embryos from three clutches were used for each data set.

Fertilised oocytes injected with 82.5 pg Rasl11b produced no phenotype when inspected at 24 and 48 hpf. When examined at 24 hpf, embryos injected with 165pg Rasl11b were slightly smaller in size but with correct anatomical features in the accurate proportions (data not shown). However by 48 hpf, embryos were the same size as their non-injected counterparts. No aberrant phenotype was visible when 82.5pg of Rasl11b was injected at both 24 and 48 hpf. Injections of Wnt4b, at both 82.5 and 165pg examination time points, generated no phenotype.

When 82.5 pg Psen2 hairpin RNA injected, no phenotype emerged at both 24 and 48 hpf examinations. However, when 165 pg was injected, a specific phenotype emerged at both examination time points. By 48 hpf, all 165 pg Psen2 hairpin RNA injected embryos were classified as level 2 (45.8%) or level 3 (54.2%) as the result of hindbrain malformation. As shown in figure 5.25 (24 hpf embryos), the seven

rhombomeres that make up the hindbrain were enlarged to the extent that rhombomeres were nearly touching the upper epidermis of the fish (no fourth ventricle present). Those classified as level 3 presented a slightly enlarged/elongated cerebellum in addition to the enlarged rhombomeres. Unlike Slc34a2a RNA injected embryos which did not react the stimuli test, all embryos injected with the RNA hairpins had normal behaviour and correctly responded to any stimuli.

To test for a synergistic effect, a mix of all three hairpin RNAs was injected. However, development was not affected in these embryos (figure 5.24, far right bar).

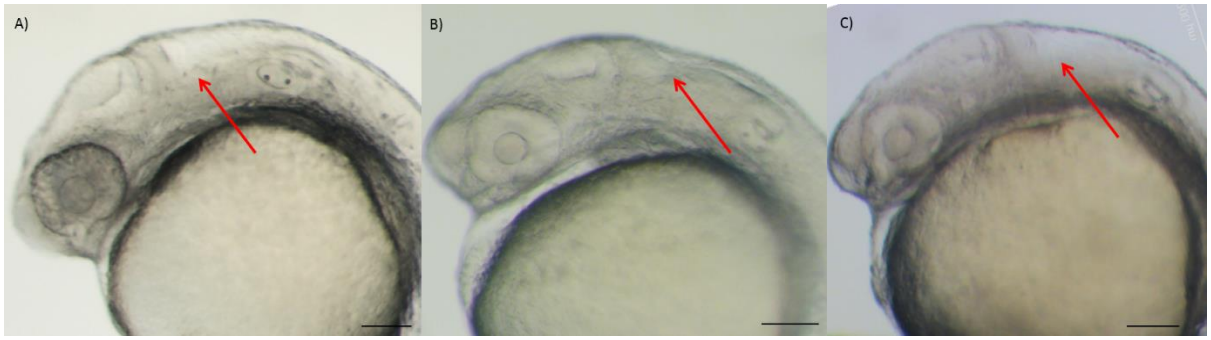


Figure 5.25 The cerebellum is present in all embryos that were injected with RNA hairpins at 24 hpf. The area above the rhombomeres is reduced in 165 pg Psen2 injected embryos as pointed out with the red arrow (B) in comparison to wildtype (A) and 165 pg Ras11b injected embryos (C). Scale bar at 0.1 mm.

5.14 Discussion

5.14.1 Frameshift Slc34a2a RNA and Slc34a2a RNA fragment injections

Injection of capped and uncapped full-length Slc34a2a RNA was found to generate a specific phenotype resulting in the loss of cerebellum in the preceding chapter. To determine if failure to develop a cerebellum was due to overexpression of the sodium-phosphate transporter during early development (first 48 hours), when it is not endogenously present, a non-coding Slc34a2a RNA transcript was generated.

The *in vitro* synthesised non-coding Slc34a2a RNA, which contains a frameshift mutation after the start codon, was found to produce the same specific loss of cerebellum. This failure to develop a cerebellum was confirmed with Engrailed-2 in situ hybridization. Therefore, these results indicate that it is not the Slc34a2a protein causing this developmental malformation, but the introduced RNA molecules.

Not all embryos injected with the frameshift Slc34a2a RNA had a loss of cerebellum; less than 30% were classified as level 1 at 24 hpf which increased to near 40% at 48 hpf. Hence, the severity of the phenotype was slightly less with the frameshift construct as compared to Slc34a2a. Both recovery to wildtype and reduced penetrance of the frameshift Slc34a2a RNA could be attributed to the notation that the injected RNA was not translated. mRNA surveillance mechanisms could be recognizing the frameshift containing Slc34a2a RNA and targeting it for degradation, reduce its halftime and consequently mitigate cerebellar loss. Otherwise, the 100% efficiency seen after injection of coding Slc34a2a RNA, can be attributed to the combination of both RNA sequence and protein production.

Due to the possible cross-reactions between Slc34a2a, Slc34a2a(as) and Slc34a2b RNA, semi-quantitative RT-qPCR was completed to determine the impact of frameshift Slc34a2a RNA injection on each transcript. Slc34a2a transcript levels were elevated at both 10 and 24 hours post injection. As the Slc34a2a qPCR primers utilized do not distinguish between the frameshift Slc34a2a and the coding Slc34a2a transcripts, this increase was expected. The amount of Slc34a2a transcript documented between 10 and 24 hpf did not change, indicating that the injected material was not being degraded in the first 24 hpf. However, as there was significant phenotypic recovery between 24 and 48 hours in the frameshift Slc34a2a RNA injected embryos (figure 5.1) it is plausible that the RNA is either significantly diluted or degraded between these two time points. Alternatively, cells could sequester translated and non-translated RNAs in different compartments which could explain stability but reduced penetrance of the non-translated transcript.

As there is no relationship between Slc34a2b and Slc34a2a, as anticipated, Slc34a2b transcript levels were unaltered after frameshift Slc34a2a RNA injection. Interestingly, Slc34a2a antisense levels were affected by exogenous frameshift Slc34a2a RNA. Slc34a2a(as) levels decreased significantly at 10 hours post injection and returned to near wildtype levels by 24 hours post injection. A similar decrease in Slc34a2a(as) levels was seen after coding Slc34a2a RNA injection (data

not shown). As frameshift containing Slc34a2a and endogenous antisense transcripts share regions of complementarity, it is likely that RNA-RNA interaction occurs. Accordingly, injected frameshift Slc34a2a is potentially generating dsRNA with the endogenously present antisense transcript. This may in turn decrease antisense transcript levels by RNAi or other dsRNA dependent mechanisms. According to this study, Slc34a2a(as) is significantly expressed in the first 48 hpf of embryogenesis. Therefore, it is conceivable that the reduction is triggering a feedback loop resulting in increased transcription of the antisense transcript (Ghosh *et al.*, 2005; Janowski and Corey, 2010). This feedback loop would explain why there is a return to near wildtype levels by 24 hours post injection.

Injection of frameshift Slc34a2a RNA demonstrated that loss of cerebellum is likely attributed to the Slc34a2a RNA and not protein. To define if a specific region was accountable, RNA fragments of both Slc34a2a sense and antisense transcripts were *in vitro* synthesized and injected. Only one of the six RNA fragments produced a phenotype upon injection into fertilized oocytes; fragment (ii) T7. This fragment contained a region of the Slc34a2a sense sequence that is complimentary to the endogenous antisense transcript; Slc34a2a exons 10 and 13 which bind to Slc34a2a(as) exon 4 and 5 (figure 5.6). Just like injection of both protein-coding and non-coding Slc34a2a RNA, injection of fragment (ii) T7 RNA resulted in the failure of cerebellar development (figure 5.8). As Slc34a2a(as) is endogenously expressed from fertilization, such a result would indicate that failure in cerebellar development is due to sense/antisense RNA interaction because none of the other Slc34a2a sense fragments synthesized produced an aberrant phenotype upon injection. Interestingly, injection of fragment (iii) T7 RNA, which is complimentary to the first three exons of Slc34a2a(as) did not produce a phenotype; failure to develop a cerebellum does not solely rely on putative hybridization between sense and antisense. There must be a sequence-specific effect that is specific to the naturally occurring dsRNA hybrid.

All embryos examined that endured cerebellar loss had impaired behavioural traits. The embryos had problems swimming in an upright position, in a straight line and occasionally they could not move in any coordinated way as they exhibited signs that could be described as tremors. In the stimuli test, embryos get poked with the end of a pair of forceps. Wildtype embryos will swim away as quickly as possible to a region of safety even before being touched; either they mix into a large group of embryos or

hide in the shadow of the petri dish lip. Embryos with cerebellar loss could either be poked numerous times without any response or would escape but only swim a circle and return to the area of ‘danger’ (another poke). The cerebellum controls and coordinates fine movement, balance, muscle tone and is involved in non-motor learning (Moens and Prince, 2002; Lalonde and Strazielle, 2003; Morton and Bastian, 2004). Therefore, lack of cerebellum in these young fish is in accordance with the behavioural competencies observed.

Cerebellar loss was confirmed in every case by completing in situ hybridization with an Engrailed-2a probe (Eng2). The *Eng2* gene is essential for midbrain hindbrain development in embryogenesis. The probe marks the midbrain-hindbrain furrow (also known as the midbrain-hindbrain boundary) and the cerebellum, thus giving what appears to be a purple triangular wedge in the brain (Fjose *et al.*, 1992; Thisse, 2005; D'Aniello *et al.*, 2013). Within the embryos that displayed repressed cerebellar development, ISH with Eng2 presented as a very thin, faint purple line. No individual or pair of genes/probes were available for in situ hybridization to determine if both sides of the midbrain-hindbrain furrow have developed. However to confirm cerebellar loss, Eng2 was sufficient and did so in the aforementioned studies.

5.14.2 Slc34a2a(as) knock down

Morpholinos are antisense oligonucleotides that comprise of usually 25 subunits which bind via sequence complementarity to their RNA target. These subunits include a nucleic acid, a morphine ring and a non-ionic phosphorodiamidate intersubunit linkage, which protects the morpholinos from degradation upon injection nor do morpholinos induce the degradation of their RNA target. Morpholinos can knock down the expression of a gene of interest in numerous ways such as blocking translation, modifying pre-mRNA splicing or inhibiting miRNA maturation (Gene Tools, 2016).

To examine a putative function of the Slc34a2a antisense transcript directly, a morpholino that interferes with the splicing of the Slc34a2a(as) RNA transcript was designed and injected. Based on amounts used to induce target-specific phenotypes

in published studies, a range of dilutions from 0.2 to 10 ng were injected. A variety of inconsistent malformations appeared in all Slc34a2a(as) morpholino injected embryos. Generally, those at 48 hpf appeared less affected than their 24 hpf counterparts for all dilutions injected, with the exception of the embryos that received 10 ng (figure 5.10). These observations are in agreement with the semi-quantitative data from RT-qPCR analysis. A decrease in Slc34a2a(as) transcript levels was found in 24 hpf embryos which notions that the knockdown was successful. However, the Slc34a2a(as) transcript levels returned to near wildtype and embryos appeared slightly healthier by 48 hpf indicating that the effect of the antisense morpholino wore off over time. Due to their modified backbones, morpholinos do not get degraded in cells or serum since they are not recognized by nucleases (Hudziak *et al.*, 1996; Youngblood *et al.*, 2007). For that reason, the lack of Slc34a2a(as) knock down and healthier appearance at 48 hpf is likely the result of dilution of the morpholino during cell division. Alternatively, the reduction of Slc34a2a(as) could trigger a feedback loop that increases Slc34a2a(as) levels, such as the one proposed after Slc34a2a RNA injections. A third reason may be due to the functionality and specificity of the morpholino itself (morpholino specificity discussed further below).

In situ hybridization utilising a probe for Slc34a2a(as) suggested that knockdown was successful at 48 hpf since no purple staining was visible in majority of embryos (figure 5.17). However, the 5 ng injected samples utilised were severely deformed due to the many off-target phenotypic effects (most likely due to morpholino toxicity). It is highly plausible that these 5 ng injected embryos were so highly developmental hindered that the organs in which Slc34a2a(as) is normally found, did not develop appropriately.

Various factors need to be considered when using morpholinos; the specificity of a morpholino is not linked to effectiveness. For example, 80% knockdown can be achieved with 5 ng of morpholino but increasing concentration to 9 ng gives over 90% knockdown in addition to numerous off-target effects (Eisen and Smith, 2008; Bill *et al.*, 2009). The 0.4 ng -10 ng of Slc34a2a(as) morpholino injected triggered increasingly severe developmental malformations. However, even though deformity severity increased with quantity injected, it may be speculated that since no specific phenotype was produced, the amount injected was too high. Over 60% of 0.2 ng injected embryos were classified as level 1 at 24 hpf. Perhaps the threshold to

producing unspecific phenotypes/ toxicity line with this specific morpholino lies between 0.2 ng - 0.4 ng.

One shortfall of the Slc34a2a(as) knockdown experiments described in this thesis is that only one morpholino oligonucleotide was used. There is no guarantee for 100% effectiveness as many factors need to be considered when designing the oligonucleotide. Any morpholino designed against a splice site junction will cause either a complete or partial single exon deletion or insertion. However, skipping or retention is dependent on the activation of cryptic donor and acceptor sites which are difficult to predict. Thus, morpholinos targeting the same RNA at different splice sites will have different optimal doses as different insertions/deletions will have incurred. Ultimately, to determine the effect of Slc34a2a(as) knockdown, numerous morpholinos targeting multiple splice site junctions should ideally have been trialled (this was not done for financial considerations). In addition, injecting more than one morpholino against the same target RNA would strengthen the specific effect seen (Bill *et al.*, 2009).

The morpholino used within this study targeted Slc34a2a(as) via interfering with splicing. As mentioned previously, it could be speculated that the natural system of the embryo understood that it requires Slc34a2a(as) for normal development and in turn transcribed more to deal with the lack of functional Slc34a2a(as) present after morpholino knock down. Eventually more Slc34a2a(as) would be transcribed than morpholino injected and consequently a near wildtype level can be restored by 48 hpf. Another option to determine if the Slc34a2a(as) RNA transcript is necessary would have been to inhibit it directly at the transcriptional level with a complete knockdown. This would have involve utilizing CRISPR-Cas9 technology so that either a premature stop codon is inserted in the genomic sequence between the 3' of Slc34a2a and 5' end of Rbpja or insertion of a premature poly-A site. However, this technique is costly and time consuming as it would involve creating a mutant strain of zebrafish. Depending on the technique used for mutation introduction, generations F0-F2 with confirmed changes, could be used in experimental studies to determine if/how Slc34a2a(as) is required (Jao *et al.*, 2013).

Morpholinos do not activate an innate immune response as they do not activate toll-like receptors due to their unnatural backbones (Stessl *et al.*, 2012; Gene Tools, 2016). Nevertheless, they are known to cause many off-target phenotypic effects (Ekker and Larson, 2001). The majority of these effects have been linked to p53

mediated apoptosis. *Danio rerio* p53, which is highly similar in function and structure to the mammalian orthologues, is found in abundance during embryogenesis (Storer and Zon, 2010). The p53 mediated apoptosis has been found to be a direct consequence of losing an endogenous transcript and not of the morpholino itself since using other antisense knockdown technologies produced the same result (Robu *et al.*, 2007). Therefore knockdown of p53 via a p53 targeting morpholino can suppress non-target specific effects.

Co-injection of Slc34a2a(as) morpholino and p53 morpholino were completed to determine if Slc34a2a(as) transcript knockdown produced a specific effect. At 24 hpf, embryos co-injected with 2 ng Slc34a2a(as) morpholino and 4 ng p53 morpholino (majority level 2) fared much better than their counterparts that were only injected with 2 ng Slc34a2a(as) morpholino (majority level 3). Nevertheless, no consistent phenotype emerged. By 48 hpf, nearly 99% of embryos co-injected with 2 ng Slc34a2a(as) morpholino and p53 morpholino returned to wildtype. Even though p53 knockdown rescued the embryos by suppressing non-specific effects, no specific developmental malformation could be attributed to Slc34a2a(as) knockdown.

A phenotypic rescue was not visible in embryos co-injected with 5 ng Slc34a2a(as) morpholino and 4 ng p53 morpholino at neither 24 hpf nor 48 hpf. As mentioned previously, there is a fine line between morpholino efficacy and morpholino toxicity. It is plausible that the extremely high Slc34a2a(as) knockdown efficiency caused by injection of 5 ng of Slc34a2a(as) morpholino has in turn caused Slc34a2a(as) transcription via the previously mentioned feed-back loop. It is also conceivable that the non-specific phenotypes caused by 5 ng of Slc34a2a(as) morpholino are so severe that p53 knock-down could not rescue them since the malformations were irreversible.

5.14.3 RNAi

Dicer, a key enzyme in eukaryotic development, cleaves dsRNA into short interfering RNA (siRNA) and pre-microRNA into mature microRNAs (miRNA) (see reviews (Plasterk, 2006; Carthew and Sontheimer, 2009)). The 20-30 base pair fragments

produced by Dicer are then incorporated into a RNA-induced silencing complex (RISC) which is crucial for RNA interference (RNAi). RISC can capture and silence RNA which shares sequence complementarity to the guide strand of the siRNA or miRNA that is incorporated within the RISC complex.

It has been hypothesised in this study that Slc34a2a and Slc34a2a related transcripts hybridise to Slc34a2a(as). If this is the case, siRNAs may be produced, which could mimic the action of specific miRNAs and affect cerebellar development. To test such possibility, Dicer knock down was completed via morpholino injection. Two Dicer targeting morpholinos were received from James Patton (Vanderbilt University, Nashville, USA). The first morpholino, Dicer start MO, was designed against the translational start site whilst the second, Dicer UTR MO, was designed against the 5' untranslated region of the *dicer1* mRNA.

Injection of Dicer start MO caused severe developmental defects in 24 hpf embryos that progressed into 48 hpf. Over 30% of embryos passed away within the first 24 hpf. No consistent malformations amongst both 24 hpf and 48 hpf embryos were visible. An Ensembl BLAST (EMBL-EBI, 2015) search with this morpholino sequence returned over 100 hits (all with over 94.74% identity). This morpholino could potentially bind to many endogenous transcripts found in zebrafish embryos such as Slc8a (a sodium- calcium exchanger), triadin (a protein necessary for heart and muscle contractions) and hecw1b (protein ubiquitination involved in ubiquitin-dependent protein catabolic process). Due to the phenotypic severity, high death rate and possible non-specificity, Dicer start MO was not utilised in further experiments.

Gene-specific knockdown studies that utilise morpholinos have come under scrutiny recently. Even though it is well established that unspecific morpholino toxicity due to p53 activation exists, it has come to light that many morpholinos appear to produce the same recurring phenotypes. As seen in this study, the engorged area above the rhombomeres ('megamind') and misshapen eyes seem to be a phenotype common to many morpholino injections (Pollard *et al.*, 2006; Nornes *et al.*, 2009; Fischer *et al.*, 2011; Ulitsky *et al.*, 2011; Buhler *et al.*, 2016). Moreover, phenotypes caused by morpholino knockdown have also been found to be unreproducible when mutant zebrafish lines for individual genes were generated (Kok *et al.*, 2015). Therefore, morpholino based studies require concentration curves and thorough examination

and validation of the phenotype, ideally via a non-morpholino based approach (Bedell *et al.*, 2011; Morcos *et al.*, 2015).

Co-injection of 6.6 ng and 2.5 ng Dicer UTR morpholino and p53 morpholino produced a specific phenotype when embryos were visually examined at 48 hpf. Nearly all embryos were classified as level 3 due to having short body/tail lengths, curved bodies and/or small non-circular eyes. This phenotypic result is consistent with that seen in Wienholds *et al.* who injected 5 ng of a Dicer morpholino with the same sequence as well as other morpholino Dicer knockdown studies (Wienholds *et al.*, 2003; Giraldez *et al.*, 2005). As miRNA and siRNAs are vital for development, it was anticipated that knocking down Dicer would significantly slow the development of the embryos, thus producing the specific phenotype described (Wienholds *et al.*, 2003; Giraldez *et al.*, 2005; Thatcher *et al.*, 2008; Rajesh K. Gaur, 2009).

To determine if a siRNA produced by Dicer from Slc34a2a sense/antisense dsRNA was interfering with cerebellum development, a triple injection was completed involving coding Slc34a2a RNA, Dicer UTR morpholino and p53 morpholino. Over 98% of triple injected fertilised oocytes were classified as level 1 or 2. Of these embryos, all had a cerebellum whose existence was confirmed with Eng2 ISH. Those embryos that were classified as level 2 had a specific phenotype that presented as an enlarged hindbrain region (figure 5.23). All triple injected embryos exhibited normal behaviour, including in the ‘poke’ test, affirming cerebellar presence. Therefore, knocking down Dicer rescued the ‘loss of cerebellum’ phenotype produced by Slc34a2a RNA injection. Therefore, knowing Dicers role in siRNA production by cleaving dsRNA, dsRNA formation of Slc34a2a and Slc34a2a(as) is highly likely.

5.14.4 shRNA injection

Endo-siRNAs may mimic microRNA action in silencing endogenous mRNAs which could in turn lead to developmental malformation. A BLAST search with the Slc34a2a sense/antisense complementary sequences, was performed with the aim to identify genes relevant to brain development that would be targeted by

Slc34a2a/Slc34a2a(as) siRNAs. Three candidate genes were identified: Presenilin2 (Psen2), Wingless-type MMTV integration site family member 4b (Wnt4b) and RAS-like, family 11, member B (Rasl11b).

Unlike long dsRNA, injection of short hairpin RNAs (shRNA) into fertilized zebrafish oocytes has been widely validated for inhibiting gene expression by triggering RNAi (Li *et al.*, 2000; Zhao *et al.*, 2001; Su *et al.*, 2008; De Rienzo *et al.*, 2012). Consequently, *in vitro* synthesized hairpin RNAs containing the particular Slc34a2a sequence complimentary to the candidate gene were microinjected. The aim was to determine if knockdown of one of the candidate genes by shRNA injection produced the same 'loss of cerebellum' phenotype seen upon Slc34a2a RNA injection.

Only one of the candidate gene shRNAs resulted in a phenotype. Injection of 165 pg Psen2 shRNA was found to affect hindbrain development in 24 hpf and 48 hpf visually examined embryos. The cerebellum developed as normal but the rhombomeres were so enlarged that they seemed to be touching the upper epidermis of the fish (Figure 5.25). Psen2 is highly expressed within the neural crest during embryogenesis; knock down causes a distortion in hindbrain development due to reduced notch signalling (Groth *et al.*, 2002; Campbell *et al.*, 2006; Nornes *et al.*, 2009). However, no consistent phenotype has been reported in Psen2 knock down studies involving morpholinos. In one study, the opposite phenotype was observed; the area above the rhombomeres increase and produced a 'megamind' like phenotype (Nornes *et al.*, 2009). In another Psen2 morpholino knockdown study, the embryos would here be classified as level 3-4; they are extremely small with curved bodies, no midbrain-hindbrain boundary formation, no eye development and the area above the rhombomeres is non-existent (Campbell *et al.*, 2006). As mentioned previously, morpholino studies can be a valuable tool in gene knockdown analysis but need to be thoroughly controlled and phenotypic results need comprehensive validation (Bedell *et al.*, 2011). Therefore, as this study finds a phenotype of enlarged rhombomeres by shRNA gene silencing, which is different from the published phenotypes by morpholino knockdown (there is no phenotypic validation through a non-morpholino experiment in either paper), the Psen2 knockdown phenotype observed here is difficult to be verified. The Psen2 shRNA phenotype does not match that of 'loss of cerebellum', but this does not mean Psen2 is somehow not involved. The brain specific phenotye suggests that Psen2 may play a role but that additional factors may be required to create the whole phenotype.

5. Mechanism

5.15 Conclusion

The aim of this chapter was to determine the mechanism behind the failure to develop a cerebellum which was caused by Slc34a2a RNA injection. Microinjection of a frameshift containing Slc34a2a RNA indicates that the observed phenotype is not caused by overexpression of the Slc34a2a protein. RNA fragments encompassing different areas of the Slc34a2a locus were generated to identify the biologically active region of the Slc34a2a sequence. One RNA fragment produced a brain phenotype; the fragment complementary to Slc34a2a(as). Triple injection consisting of Slc34a2a RNA, a Dicer morpholino and a p53 morpholino, found that the 'loss of cerebellum' phenotype can be rescued by knocking down Dicer. Therefore, the conclusion was drawn that injected Slc34a2a and Slc34a2a related transcripts were creating dsRNA with the endogenous Slc34a2a(as) transcript. Three target genes for the endo-siRNAs were identified and knocked down with shRNAs: Psen2, Wnt4b and Rasl11b. However, none produced the same 'loss of cerebellum' phenotype. However, this does not mean Psen2, which created a brain phenotype, is not involved through an undetermined pathway. In the subsequent chapter, dsRNA extraction will be attempted for the purpose of eventually sequencing and confirming Slc34a2a/Slc34a2a(as) endo-siRNA existence.

Chapter 6. dsRNA mapping and extraction

Until recently, the existence of endogenous small interfering RNAs (endo-siRNA) was not well established and was believed to be limited to nematodes (Piatek and Werner, 2014; Piatek *et al.*, 2016). Now that endo-siRNAs and their generating enzymes (Dicer/Drosha) have been found to occur in numerous eukaryotes, including fruit flies (Kawamura *et al.*, 2008), mice (Babiarz *et al.*, 2008; Flemr *et al.*, 2013) and humans (Werner *et al.*, 2007; Xia *et al.*, 2013), another layer of complexity has been added to how host gene expression is regulated. In the previous chapter, evidence was provided that a siRNA produced by Dicer from a sense/antisense hybrid (dsRNA) was disrupting zebrafish brain development. The aim of this chapter is to study endo-siRNA precursors; endogenously occurring double stranded RNA (dsRNA). Using the J2 monoclonal antibody, which has an affinity for dsRNA, demonstration of *in vivo* dsRNA existence was attempted by immunofluorescence and end point PCR.

6.1 Immunofluorescent staining of dsRNA using J2 monoclonal antibody

A monoclonal anti-dsRNA antibody, J2, was used in immunofluorescent staining to determine if endogenous dsRNA could be localised in 24, 36 and 48 hpf zebrafish embryos. PFA-fixed embryos, embedded in OCT, were sectioned at 5 µm using a Leica cryostat. The J2 mouse monoclonal IgG2a kappa chain antibody was utilised as primary antibody with both goat anti-mouse alexafluor 488 and TRITC-conjugated secondary antibodies. DAPI nuclear stain was utilised in the mounting process. 5 µm sections of mouse testis (from 6-9 month old mice) were utilised as a positive control; mouse testis are known to have a high abundance of endogenous dsRNA (Werner *et al.*, 2007; Ge *et al.*, 2008; Nejepinska *et al.*, 2012).

Specific dsRNA related fluorescence could not be localised utilising the J2 antibody in any zebrafish embryo sections. Similar fluorescence intensity (as visualised in figure 6.1) was found in comparable embryo structures between anti-dsRNA

sections and control sections where only secondary antibody was applied. Use of two different secondary antibodies with differing emission and excitation spectra made no difference to the results in zebrafish tissues. Within mouse testis sections, specific anti-dsRNA J2 signal was found in elongating spermatids (figure 6.1, arrows in bottom right panel). The connective tissue of the mouse testis stained as well. However, as the connective tissue stained in both anti-dsRNA and the control, it is considered background fluorescence. Nuclei were stained with DAPI.

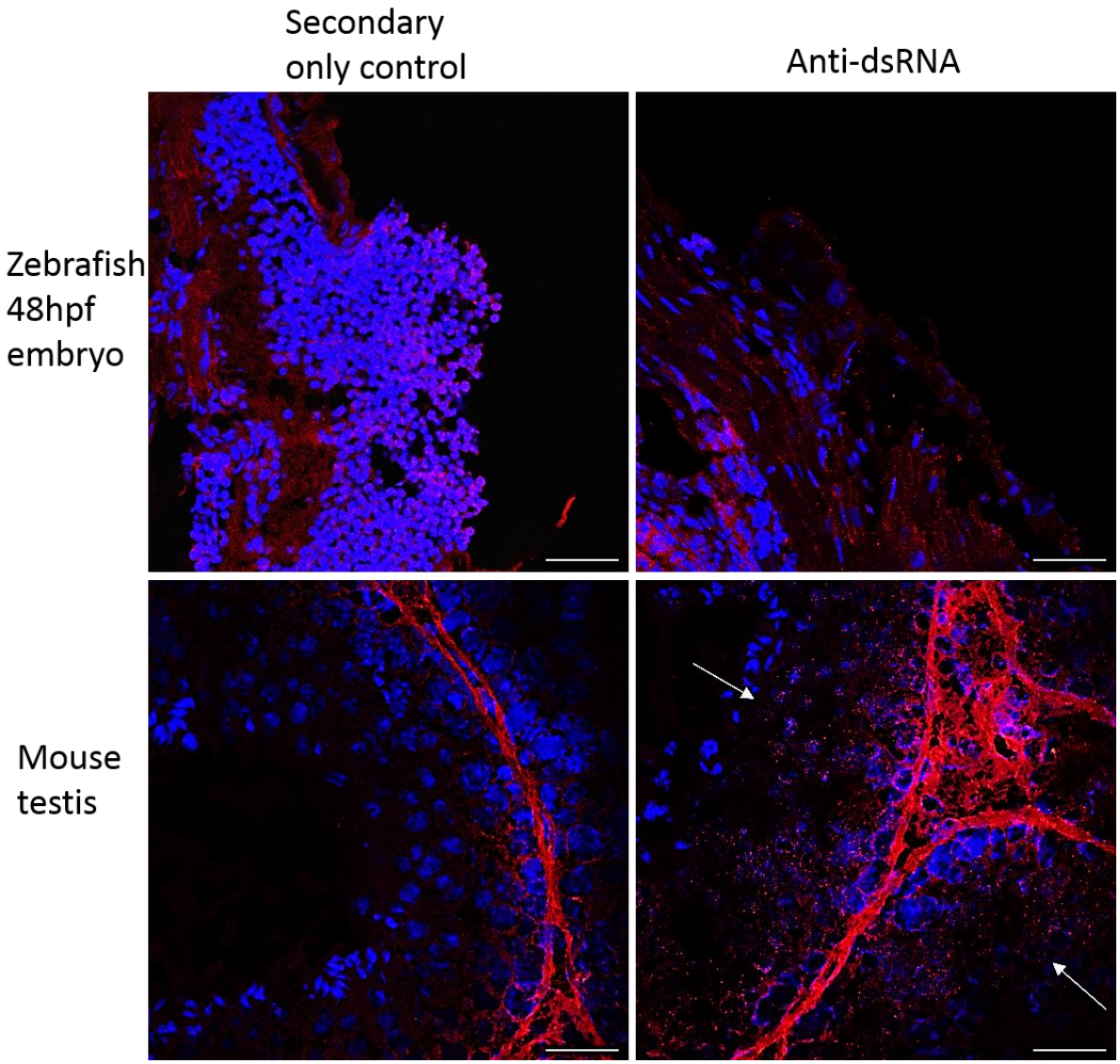


Figure 6.1 Immunofluorescent staining of dsRNA using the J2 IgG monoclonal primary antibody on zebrafish embryos and mouse testis sections. Tissue sections are 5 µm thick. Negative control consisted of staining with secondary antibody only (left panels). Blue stain is DAPI. Scale bar at 30 µm.

6.2 dsRNA immunoprecipitation using J2 monoclonal antibody

With the aim to eventually extract endogenous dsRNA from whole zebrafish embryos, preliminary experiments were completed to determine how *in vitro* synthesized dsRNA responded to different extraction processes. Slc34a2a and Slc34a2a(as) uncapped RNA was *in vitro* synthesized and dsRNA was generated by mixing equal quantities in 1x PBS prior to incubation at 72°C for 10 minutes, followed by cooling on ice (sample called '2a/as dsRNA'). Control samples included single stranded RNA Slc34a2a or Slc34a2a(as) in 1x PBS and a mix of Slc34a2a RNA and Slc34a2a(as) RNA in water without the heating/cooling process, thus non hybridizing conditions. RNA extraction techniques tested included Trizol ® and the Bioline Isolate II miRNA Kit prior to immunoprecipitation with the J2 monoclonal antibody in combination with goat anti-mouse IgG magnetic beads. RNA was eluted from the J2/beads using MACS Separation Columns followed by reverse transcription. End-point PCR was completed on cDNA utilizing primers designed in house (chapter 2, table 2.5).

RNA extraction utilising the Trizol® method was found to favour single stranded RNA in comparison to dsRNA after immunoprecipitation since higher intensity Slc34a2a and Slc34a2a(as) bands were visible in the Slc34a2a plus Slc34a2a(as) mix (2a+as RNA mix) than in the 2a/as dsRNA sample as seen in figure 6.2. Immunoprecipitation with J2 without first undergoing RNA extraction confirmed that the J2 antibody favours dsRNA over ssRNA since band intensity was stronger in the 2a/as dsRNA sample in comparison to the 2a+as RNA sample. In both the Slc34a2a RNA and Slc34a2a(as) RNA control samples, only the individual respective product was found. Bands for both Slc34a2a and Slc34a2a(as) were of higher intensity in the samples that did not undergo Trizol® extraction (J2 immunoprecipitation only). As only primer bands were visible when PCR was completed with actin primers, no contamination occurred (figure 6.2, bottom panel). Negative controls utilising respective RNA and water in the PCR reaction produced no result (data not shown).

6. dsRNA

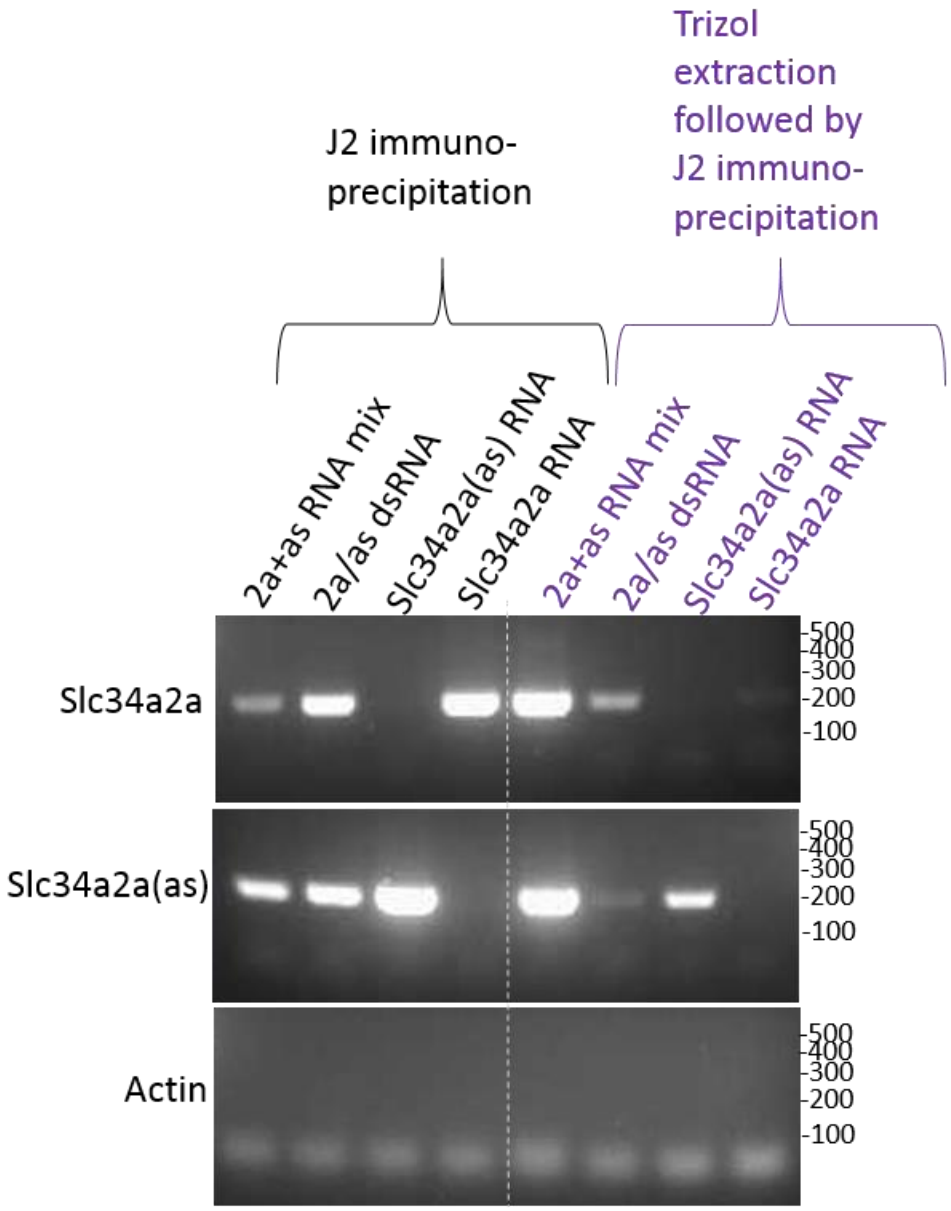


Figure 6.2 Double stranded RNA immunoprecipitation utilizing the J2 monoclonal antibody on goat anti-mouse IgG magnetic beads with and without prior RNA extraction with Trizol®. Expected product sizes were 174 bp for Slc34a2a, 198 bp for Slc34a2a(as) and 149 bp for actin primers. PCR with actin was completed to control for contamination. N= 2

In contrast, RNA extraction utilising the Bioline Isolate II miRNA Kit prior to immunoprecipitation with the J2 antibody was found to favour 2a/as dsRNA over the mix of Slc34a2a and Slc34a2a(as) RNA as visualised in figure 6.3. Bands for Slc34a2a and Slc34a2a(as) had the same intensity in samples 2a/as dsRNA,

Slc34a2a RNA and Slc34a2a(as) RNA which were more intense in comparison to the Slc34a2a+Slc34a2a(as) RNA mix. When no extraction was performed prior to immunoprecipitation, Slc34a2a and Slc34a2a(as) band intensity did not differ between single stranded RNA or dsRNA samples. No contamination was found in no-cDNA PCR controls and PCR completed with actin primers (RNA in PCR reaction data not shown).

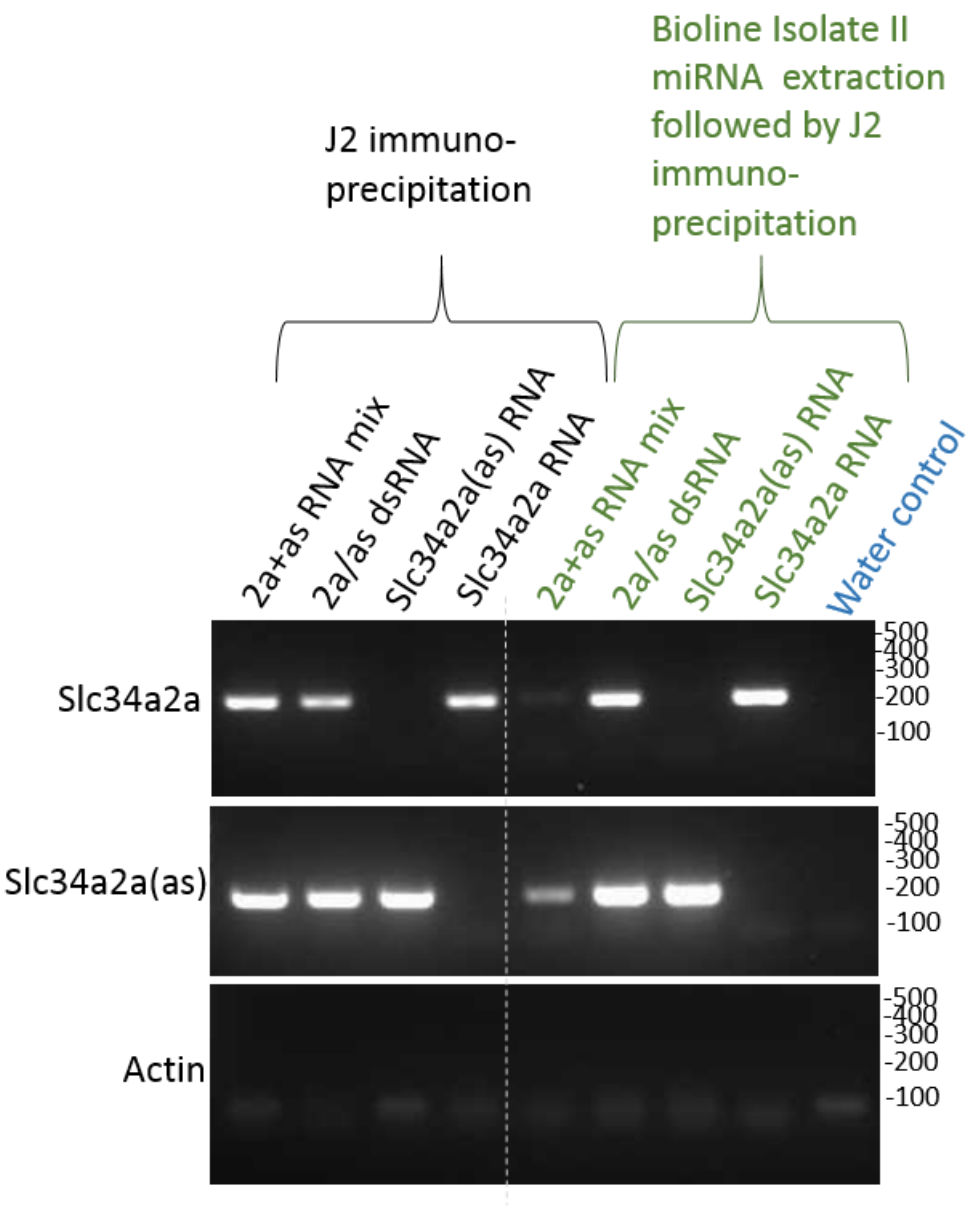


Figure 6.3 Double stranded RNA immunoprecipitation utilizing the J2 monoclonal antibody with and without first undergoing RNA extraction with the Bioline Isolate II miRNA Kit. Expected product sizes were 174 bp for Slc34a2a, 198bp for Slc34a2a(as) and 149 bp for actin primers. PCR with actin was completed to control for contaminations. N=2

The J2 monoclonal antibody has been used in various immunocapture methods in both plant extracts and in cell lines but there is limited literature involving eukaryotic tissues (Lybecker *et al.*, 2014; Nejepinska *et al.*, 2014; Blouin *et al.*, 2016). It has been hinted that J2 antibody favours naturally occurring dsRNA over synthetic dsRNA (Schonborn *et al.*, 1991). Even though it remained undermined how the above conditions or J2 affected dsRNA stability, dsRNA immunoprecipitation with J2 antibody was attempted utilizing whole zebrafish embryos. Transcript levels were examined via semi-quantitative RT-qPCR. Five embryos were flash frozen without liqid on dry ice prior to being crushed with a pestle. Total RNA was examined by 10% of the tissue slurry undergoing RNA extraction with Trizol® prior to reverse transcription as mentioned in the preceding chapters (input control). The remaining 90% of the sample was divided with one half receiving J2 antibody (with J2) and the other half had nothing added (no J2). Five embryos from a minimum of three clutches were used for the data presented in figure 6.4.

The content of Slc34a2a, Slc34a2a(as) and Slc34a2b in the flash frozen embryos was examined and did not differ significantly from previous experiments (figure 6.4, yellow and blue bars, input control). When immunoprecipitation with the J2 antibody was completed using the tissue slurry without prior RNA extraction, levels for all transcripts of interest increased (figure 6.4, red bars). When immunoprecipitation was completed without J2 antibody, all transcript levels decreased (figure 6.4, grey bars). Error bars for no J2 added immunoprecipitation controls, are much larger than all others; Ct values within triplicates for each clutch had larger standard deviation than the triplicates completed for Trizol® extracted RNA and dsRNA precipitated with J2 antibody. The fact that the presence or absence of the J2 antibody had indiscriminate effects on all transcripts, whether or not putative dsRNA could be present, suggests a lack of specificity for the entire process.

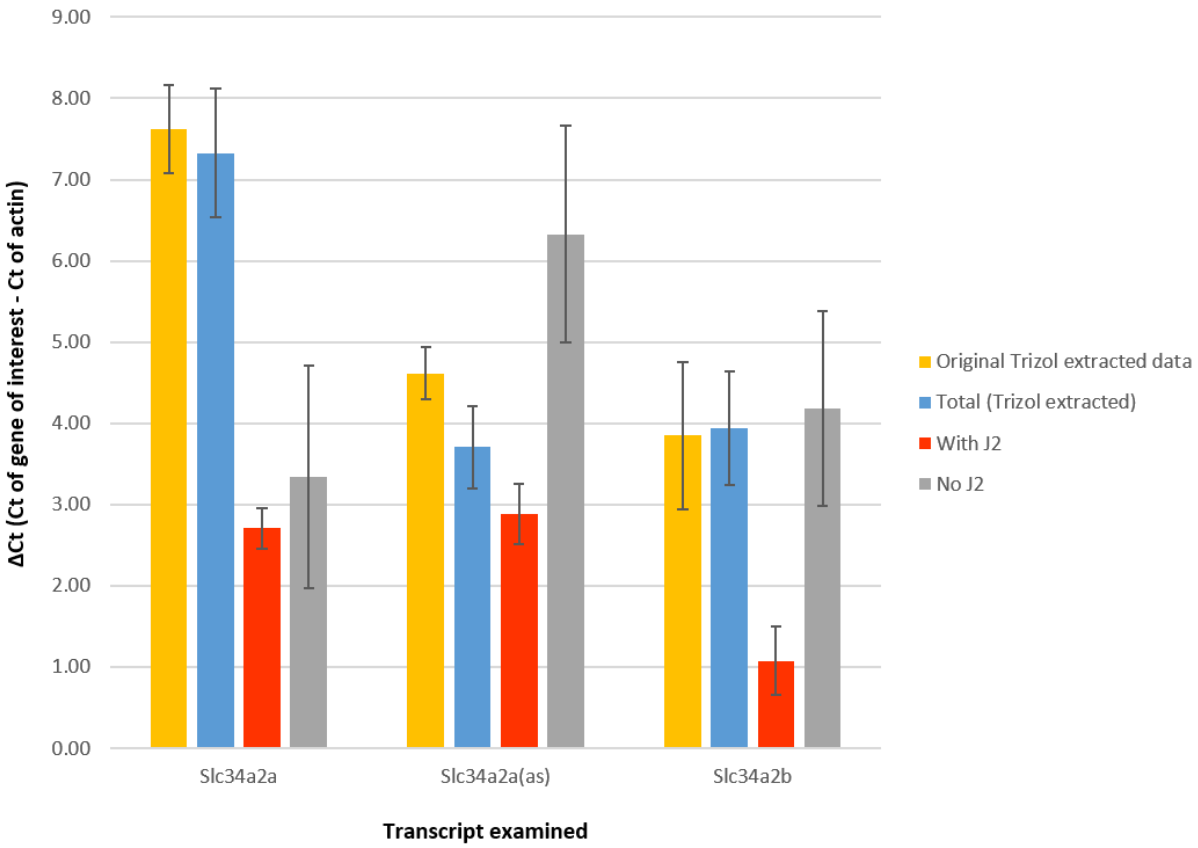


Figure 6.4 Semi quantitative RT-qPCR analysis of Trizol® extracted total RNA (blue) and dsRNA precipitated with J2 antibody (red) from five flash frozen fertilised zebrafish oocytes 48 hours post fertilization. A no J2 antibody control was included (grey). RNA extracted from embryos using the previously described Trizol protocol (chapter 4) is presented in yellow. ΔCt was calculated by using β-actin as gene reference. N≥3

6.3 Discussion

The J2 antibody has been successfully used in dsRNA detection caused by RNA and/or DNA viruses in cell lines and parasites (Weber *et al.*, 2006; Targett-Adams *et al.*, 2008; Zangger *et al.*, 2013). Recently, the J2 antibody was used to demonstrate dsRNA accumulation after Dicer-knockdown in HEK293 cells (White *et al.*, 2014). Thus, it was attempted to determine if endogenous dsRNA could be localised in sectioned zebrafish embryos. Attempts were made with two different secondary

antibodies but in both cases, the search proved fruitless. As seen in figure 6.1, no specific staining was found in 24-48 hpf zebrafish embryos.

Accumulation of dsRNA within cells activates the interferon-response pathway which consequently leads to cell death (Sledz and Williams, 2004; Reynolds *et al.*, 2006; White *et al.*, 2014). The function of mammalian cytoplasmic Dicer is well documented; as a key enzyme in the RNA interference (RNAi) pathway, Dicer cleaves dsRNA into short RNAs. It has been suggested that human nuclear Dicer, which interacts with RNA polymerase II, has a specific function in cleaving endogenous dsRNA formed from overlapping long non-coding RNA (Gagnon *et al.*, 2014; White *et al.*, 2014). Knocking down nuclear Dicer, which results in dsRNA accumulation, has a catastrophic effect on human cell-line function by activating the interferon-response pathway subsequently leading to cell death (White *et al.*, 2014).

Zebrafish, just like mammals, have an interferon-response pathway and express Dicer, which functions in RNAi. The detection of dsRNA could have failed in the above experiments due to wildtype embryos being utilized; Dicer would have cleaved all dsRNA present to prevent any interferon-response. Thus, if dsRNA exists transiently, the J2 antibody may not be able to bind due Dicer masking or quickly cleaving the > 40 bp dsRNA structures required for J2 binding. Rapid degradation/masking is important as dsRNA presence may trigger an immune response since dsRNA is normally foreign and due to viral infection. To date, only hyperedited dsRNAs have been found not to trigger the interferon-response pathway and apoptosis (Vitali and Scadden, 2010). Therefore, it is conceivable that only minimal quantities of dsRNA are formed endogenously; not enough to produce a detectable signal. Accordingly, no publications report quantification of endogenous dsRNA levels.

dsRNA detection was unsuccessful in zebrafish embryos. However, many different suggestions can be made for further experiments. White *et al.* detected high levels of dsRNA with the J2 antibody after knocking down Dicer in human cells (White *et al.*, 2014). Attempts for dsRNA detection could be made utilising the Dicer UTR morpholino injected embryos from the preceding chapter. Slc34a2a sense and antisense transcripts are both expressed at 48 hpf; this however does not mean 48 hpf has the highest levels of endogenous dsRNA. Attempts to detect dsRNA could be made using juvenile and adult fish to determine if dsRNA presence/existence can be detected. Other options include manipulations to the immunofluorescence

protocol. In this study, embryos were fixed with PFA prior to freezing in OTC which could have affected tissue integrity. White *et al*. used an acetone/methanol fixation prior to monitoring immunofluorescence (White *et al.*, 2014). Another route may be to permeabilize the cells with Triton x-100 as done in in situ hybridization prior to adding the primary antibody.

Mouse testis, which are potentially high in dsRNA, were utilised as a positive control for anti-dsRNA J2 signal (Werner *et al.*, 2007; Ge *et al.*, 2008; Nejepinska *et al.*, 2012). Within the mouse testis sections, a J2 specific signal appeared to be present in elongating spermatids. Dicer-dependent miRNAs and endo-siRNAs play a vital role in gene regulation, chromatid condensation and nuclear polarization in developing spermatids (Gou *et al.*, 2014; Yadav and Kotaja, 2014). Therefore, endogenous dsRNA was expected to be mapped to areas containing developing spermatids as visualised in figure 6.1. This finding corresponds with the anti-dsRNA granular staining within spermatocyte nuclei that was found by Noora Kotaja and group (personal communication).

Trial dsRNA immunoprecipitation experiments with the J2 antibody were performed to confirm the specificity of the J2 monoclonal antibody. In all experiments where synthesized dsRNA immunoprecipitation was completed without prior RNA extraction, the J2 antibody favoured Slc34a2a/Slc34a2a(as) dsRNA over the mix of Slc34a2a and Slc34a2a(as) RNA in water (figure 6.2). However, control samples that contained only one RNA strand (either Slc34a2a or Slc34a2a(as) at the same concentration used to make the dsRNA or mix) produced bands of similar intensity when compared to the dsRNA sample. Slc34a2a and Slc34a2a(as) bands were also produced from the sample that contained both the sense and antisense strand under non hybridizing conditions. These results indicate that the J2 antibody used under these conditions for immunoprecipitation, was not highly specific for dsRNA (figures 6.2 and 6.3). Alternatively, single and double-stranded RNA could feature differently at various steps throughout the test protocol (see below).

Two different RNA extraction methods were tested prior to dsRNA precipitation with J2 to arbitrate if an extraction method could ameliorate the precipitation process: the standard Trizol® phenol/chloroform technique and the Bioline Isolate II miRNA Kit which can be adjusted to extract both long (>200bp) and short RNA. Phenol chloroform extraction using Trizol® prior to dsRNA precipitation was found to favour single stranded RNA since Slc34a2a and Slc34a2a(as) bands were of higher

intensity in the Slc34a2a plus Slc34a2a(as) mixed sample rather than the synthesized dsRNA sample (figure 6.2). Contrary, immunoprecipitation after using the Bioline Kit for extraction, favoured the synthesized dsRNA in comparison to the Slc34a2a and Slc34a2a(as) mix. However, in both experiments, bands were visible for Slc34a2a and Slc34a2a(as) in control samples containing each individually. Therefore even though the two different RNA extraction process seem to favour different forms of RNA (single versus double stranded), both are extracted nevertheless. An additional complication represents the detection by RT-PCR. Experience in our lab have confirmed that hairpin RNA in particular and dsRNA in general, is not efficiently reverse transcribed. This would favour the detection of the individual transcripts. The applied strategy using the J2 antibody, which aimed to be dsRNA specific, does not appear to be precise enough to only extract the synthesized Slc34a2a/Slc34a2a(as) dsRNA under the conditions tested above.

In vitro synthesized Slc34a2a and Slc34a2a(as) RNA were hybridized to form dsRNA as recommend by numerous groups under low salt conditions with heating followed by a slow cooling on ice (DRSC, 2015; BrownUniversity, 2016). Phenol chloroform was utilised for RNA extraction as it is the most commonly utilized method resulting in high quantities of RNA. Apart from trialling other RNA extraction kits produced by differing manufacturers, another option would have been to attempt RNA extraction using lithium salt buffers. Specific double stranded RNA extraction from plants using lithium salts has determined that the quantity is adequate and the quality pure enough for numerous downstream applications including RT-qPCR (Tzanetakis and Martin, 2008). Perhaps this extraction process that claims to be better for dsRNA extraction could have ameliorated the J2 antibody immunoprecipitation method by providing more dsRNA than ssRNA. Other options to optimise J2 antibody dsRNA precipitation include manipulating the salt concentrations (differing salts), binding time and trialling differing ratios of J2 antibody to magnetic beads to dsRNA.

One publication has hinted that J2 antibody favours naturally occurring dsRNA over synthetic dsRNA (Schonborn *et al.*, 1991). Therefore dsRNA immunoprecipitation was attempted utilizing crushed flash frozen whole zebrafish embryos prior to examining transcript levels via semi-quantitative RT-qPCR. Flash freezing embryos on dry ice did not have an impact on total RNA content since ΔCt values for all

transcripts examined were consistent with the data from embryos frozen and stored in Trizol® (figure 6.4, yellow and blue bars).

Interestingly, after anti-dsRNA J2 immunoprecipitation, levels for all transcripts examined (Slc34a2a, Slc34a2a(as) and Slc34a2b) increased. Phenol chloroform with Trizol® extracts both single and double stranded RNA (Flegr, 1987; Castillo and Cifuentes, 1994; Sambrook and Russell, 2006) but the above data suggests it may partially favour single stranded (similar findings have been mentioned in Biological online forums). If phenol chloroform extraction does partially favour ssRNA, then J2 precipitation from whole embryos could have extracted a higher level of dsRNA than Trizol®. Alongside the ssRNA that is eluded through the J2 immunoprecipitation process (figure 6.2 and 6.3), this would explain the higher levels of Slc34a2a and Slc34a2a(as) RNA detected (figure 6.4, red bars). However, there is a significant increase in Slc34a2b RNA levels as well and Slc34a2b does not have a cis-NAT. The manufacturer claims that J2 antibody recognition of dsRNA is independent of sequence and nucleotide composition as long as the dsRNA is over 40 bp (SCICIONS, 2014). However, through scanning force microscopy, some preferential binding sites have been determined and verified by site-directed mutagenesis (Bonin *et al.*, 2000). Therefore increase in detectable levels may be due to either Slc34a2b transiently hybridizing to an unknown RNA partner or J2 antibody having an undetermined sequence specify that Slc34a2b binds.

In the control samples without J2, Slc34a2a RNA levels were found to be less than that precipitated with J2 antibody but still higher than that found in the total RNA (figure 6.4). The expected result was that the minus J2 control would have less of each transcript since no RNA was extracted and all RNA should have been in the flow-through after passing through the MACS columns. This was the case for Slc34a2a(as) and Slc34a2b RNA levels, which were found to be less than in total RNA extracted with Trizol® and J2 precipitated RNA levels. All Ct values for transcripts of interest in the minus J2 samples differed substantially within clutch triplicates and amongst clutches thus resulting in higher standard deviations for each clutch. This in turn gave rise to the large standard error bars presented (figure 6.4, error bars on grey bars). Both single and double stranded RNA should flow through the columns without J2 added and any signal is due to unspecific binding, consistent with the inconsistent Ct values.

6. dsRNA

The above preliminary experiments suggest that the anti-dsRNA J2 antibody in the tested set-up is unable to solely extract synthesized Slc34a2a/Slc34a2a(as) dsRNA; both single and double stranded RNA is present after immunoprecipitation. The semi-quantitative RT-qPCR data would suggest that J2 or the magnetic beads bind more RNA than Trizol ® extracts (possibly due to both dsRNA and ssRNA being present in eluted samples). This highlights essential problems in the experimental setup. Neither of these preliminary studies demonstrate dsRNA *in vivo* existence in zebrafish, nor is the stability of dsRNA determined. Questions such as if the extraction process allows for the dsRNA to remain intact or if it disrupted and re-hybridises further downstream, still need to be answered. It appears that many steps of the experimental protocol have a bias toward single or double stranded RNA molecules that, at the moment are not fully understood. Secondly, the amount of dsRNA in the zebrafish embryos may simply be too low to warrant reliable detection by J2 precipitation and RT-PCR. Many further experiments need to be completed prior to determining if endogenously occurring dsRNA can be extracted from whole zebrafish tissues using J2 antibody.

6.4 Conclusion

The J2 antibody has been successfully used in dsRNA mapping and extraction caused by RNA/DNA viruses or Dicer knockdown in cell lines and parasites (Weber *et al.*, 2006; Targett-Adams *et al.*, 2008; Zangger *et al.*, 2013; White *et al.*, 2014). The anti-dsRNA J2 antibody, however, was unable to detect any Slc34a2a/ Slc34a2a(as) dsRNA in 24, 36 and 48 hpf section zebrafish embryos. However, dsRNA was localised to elongating spermatids in mouse testis validating the protocol. The protocol used was found to be unspecific for dsRNA immunoprecipitation conducted with both synthesised dsRNA and in whole tissues. Eventually, an improved immunoprecipitation protocol would be useful to determine whether dsRNA is formed in vivo and how this avoids triggering an interferon response throughout development. The sequence of the siRNA generated by Slc34a2a/ Slc34a2a(as) dsRNA as seen in the previous chapters could further be determined by size selected RNAseq experiments from staged zebrafish embryos.

Chapter 7. Discussion

7.1 Summary of major findings

Quantity and temporal expression patterns of a protein or transcript can be a key indicator as to function. Within this study, the interplay of four RNA transcripts from three genes were examined during *Danio rerio* embryogenesis; Slc34a2a and its naturally occurring antisense transcript (Slc34a2a(as)), Slc34a2b and Rbpja. RT-qPCR and in situ hybridization data determined that the Slc34a2a sense and antisense transcripts demonstrate a reciprocal relationship were both transcripts are expressed at a near equal amounts at 48 hpf. Slc34a2a(as) expression was present from fertilisation and decreased gradually over time whilst Slc34a2a expression commenced at 48 hpf and progressively increased from there on. The two transcripts located to the same tissues during the examined 72 hours of development: the pharynx, endoderm (future digestive tract and swim bladder), primordial midbrain channel and primordial hindbrain channel. These findings support the possibility of antisense-mediated Slc34a2a gene regulation.

Injection of full length *in vitro* synthesised Slc34a2a RNA caused a dose-dependent phenotype in fertilized zebrafish oocytes. A specific loss of cerebellar development was noted upon injections of Slc34a2a RNA that was neither seen with Slc34a2a(as) nor Slc34a2b RNA injections. Engrailed-2 in situ hybridization confirmed the loss of cerebellum and malformation of the midbrain/hindbrain boundary. Microinjection of a frameshift containing Slc34a2a RNA produced the same phenotypic failure to develop a cerebellum, indicating that it was not overexpression of the Slc34a2a protein that caused the observed phenotype. RNA fragments encompassing different areas of the Slc34a2a locus were generated for microinjection to determine the specific region of the Slc34a2a sequence that caused the phenotype. One RNA fragment produced a brain phenotype; the fragment encompassing the Slc34a2a sequences (exons 10 and 13) that are complementarity to Slc34a2a(as) caused failure to develop a cerebellum. Triple injection consisting of Slc34a2a RNA, a Dicer morpholino and a p53 morpholino, found that the ‘loss of cerebellum’ phenotype could be rescued by Dicer knockdown. Therefore, the conclusion was drawn that injected Slc34a2a and Slc34a2a related RNA transcripts were creating dsRNA with

the endogenous Slc34a2a(as) transcript to form a Dicer substrate. The Slc34a2a/Slc34a2a(as) derived endo-siRNAs were hypothesised to interfere with a transcript necessary for correct brain development. With the aim of defining target genes, identified by BLAST, short hairpin RNAs were produced to knockdown three candidate genes: Psen2, Wnt4b and Rasl11b. None produced the same 'loss of cerebellum' phenotype demonstrating that they are not the specific target for the Slc34a2a/Slc34a2a(as) siRNA. However, it is plausible that the Slc34a2a/Slc34a2a(as) siRNA works through an intricate pathway which Psen2 may be part of, but other unknown factors are involved to create the 'loss of cerebellum' phenotype.

The J2 monoclonal antibody has been successfully used in previous studies to map and extract dsRNA caused by RNA/DNA viruses or Dicer knockdown in cell lines and parasites. However, the anti-dsRNA J2 antibody was unable to clearly detect any endogenous dsRNA in 24, 36 and 48 hpf sectioned zebrafish embryos within this project. Nevertheless, dsRNA was localised to elongating spermatids in positive control mouse testis sections. Thus, either dsRNA exist briefly and transiently in zebrafish embryos or the dsRNA is bound by other molecules preventing J2 antibody binding. Additionally, immunoprecipitation using the J2 antibody followed by RT-PCR detection gave inconclusive results likely due to technical difficulties and/ or low expression of dsRNA in zebrafish whole tissues.

7.2 Putative mechanisms and the biological function of Slc34a2a(as)

In recent years, bioinformatics and transcriptome studies have proposed that thousands if not hundreds of thousands of non-coding RNAs (ncRNA) exist in higher eukaryotic species (Bertone *et al.*, 2004; Washietl *et al.*, 2005; Jeggari *et al.*, 2012). Non-coding RNAs have been linked to numerous pathways involved in RNA splicing, DNA replication, chromosome structure and gene regulation (Wilusz *et al.*, 2009; Baguma-Nibasheka *et al.*, 2012; Modarresi *et al.*, 2012). Therefore it is no surprise that some ncRNAs have been linked directly to diseases such as Alzheimer's (review:(Tan *et al.*, 2013)), autism (Velmeshev *et al.*, 2013), Prader-Willi syndrome (Galiveti *et al.*, 2014) and cancer (Matouk *et al.*, 2007). Many ncRNAs have yet to

have their function validated but it is likely that many are non-functional products of spurious or pervasive transcription.

A form of non-coding RNA, natural antisense transcripts (NATs), are fully processed mRNA-like transcripts originating from the opposite strand of protein coding genes. These antisense transcripts can contain sequences of perfect complementarity with other endogenous transcripts and thus, can act in *cis* or *trans* via hybridization. There are a few possible mechanisms as to how NATs regulate cognate sense transcripts; Transcriptional inference like in the case of initiation factor 2a (eIF2a) (Silverman *et al.*, 1992), RNA masking like thyroid hormone receptor ErbAa2 (Munroe and Lazar, 1991), dsRNA-dependent mechanisms (such as RNA interference) like SINE B2 (Fan and Papadopoulos, 2012), antisense-induced methylation such as hemoglobin α-2 in α-thalassemia (Tufarelli *et al.*, 2003) or chromatin remodelling such as tumour suppressor gene *p15* in leukaemia (Yu *et al.*, 2008).

Even though the mechanisms through which NATs function have been theorised, few validated and thoroughly understood examples exist. Within this study, an attempt was made to understand the role of Slc34a2a(as) during embryogenesis. As transcriptional interference has only been validated in yeast, and both Slc34a2a sense and antisense transcripts are present in the same tissues at 48-72 hpf (qPCR and in situ hybridization data), it is unlikely that Slc34a2a(as) is preventing Slc34a2a transcription through TI. However, since knock down studies of Slc34a2a(as) by morpholino were unsuccessful, it is difficult to pinpoint what other mechanism through which this specific NAT functions, and more so, what is its designated target(s).

NATs can work in *cis* on their cognate sense partner or in *trans* – affecting a distant gene, transcript or even protein. RNA masking of Slc34a2a by Slc34a2a(as) seems unlikely as there are no known alternate splice forms to Slc34a2a. Nevertheless, this does not mean RNA masking can be ruled out completely. As no successful knock out of Slc34a2a(as) was completed, it is hard to determine if endogenous Slc34a2a(as) is involved in regulating polyadenylation, translation or Slc34a2a stability. However, this hypothesis too seems fairly unlikely. Otherwise injected Slc34a2a would have been stabilized and/or translation would have been aided by Slc34a2a(as) and signs of increased sodium-phosphate transport would be exhibited. Experiments observing phosphate level uptake would have been valuable.

It too is plausible that Slc34a2a(as) works in trans and this study has been unable to determine the specific target by absence of Slc34a2a(as) knock out experiment. If Slc34a2a(as) works by masking or unveiling critical factors in a *trans* transcript, then injection of Slc34a2a could be inhibiting Slc34a2a(as) normal function, thus producing the loss of cerebellum phenotype. However, knock down of Dicer would cleave all Slc34a2a/Slc34a2a(as) duplexes, and as Slc34a2a(as) is transcribed within the first 72 hpf, it could resume its normal function in cerebellar development after the ratio of Slc34a2a to Slc34a2a(as) switches in its favor.

To date, there are only a few well established examples of naturally occurring endo-siRNAs that function in higher eukaryotes as humans and mice. As Dicer knock down rescued the 'loss of cerebellum' phenotype observed upon Slc34a2a RNA injection, it suggests that a siRNA from Slc34a2a/Slc34a2a(as) dsRNA might be a possible mediator for the phenotype observed. However, as endo-siRNA presence in the first 48 hours is due to exogenously added sense transcript, the loss of cerebellum phenotype may simply be an artefact. It is doubtful that endogenous Slc34a2a(as) regulates Slc34a2a through RNAi prior to 48 hpf, otherwise lack of cerebellar development would be common in zebrafish. Additionally, as Slc34a2a(as) knockdown studies were unsuccessful by morpholino injection, it remains unknown if Slc34a2a(as) regulates another RNA in *trans* through a dsRNA-dependent mechanism.

If Slc34a2a(as) RNA does regulate Slc34a2a, it might do so through the least understood mechanism, antisense-induced methylation or chromatin remodelling. As Slc34a2a and Slc34a2a(as) show a temporal reciprocal pattern, it may be that the reduction in Slc34a2a(as) levels affects an unknown transcriptional enhancer or repressor of Slc34a2a. It too is possible that Slc34a2a(as) exists to regulate a distant gene by chromatin remodelling.

The work completed within this study highlights the biological importance of maintaining a sense/antisense balance and provides a legitimate example as to how NATs can affect development through ectopic dsRNA formation. However, many questions still remain. Slc34a2a RNA presence in the first 48 hours of development caused a detrimental phenotype involving the failure to develop a cerebellum due to siRNA production by Dicer from Slc34a2a/Slc34a2a(as) dsRNA. Therefore, if Slc34a2a presence is harmful during the first 48 hpf, why does Slc34a2a(as) exist? Does Slc34a2a(as) suppress Slc34a2a during the first 48 hours through a

mechanism that is not transcriptional interference or RNAi? Would the repercussion of Slc34a2a expression during the first 48 hpf without Slc34a2a(as) be worse than Slc34a2a/Slc34a2a(as) dsRNA formation? Does Slc34a2a(as) regulate any other genes or proteins *in trans*? Or is Slc34a2a(as) simply the result of transcriptional noise?

Additional questions arise due to Slc34a2a sense and antisense RNA co-existence. Both Slc34a2a and Slc34a2a(as) are endogenously present at 48 hpf and it is fair to assume that they hybridize to form dsRNA at this time point as both localise to the same tissues (these findings are supported by (Carlile *et al.*, 2009)). Double stranded RNA can elicit an immune response due to its resemblance to viruses or an intermediate formed in genome defence strategies against transposable elements. Therefore, it remains to be determined how the endogenous Slc34a2a sense/antisense duplex at 48 hpf does not trigger an immune response.

An antisense RNA transcript for Slc34a2a (and orthologs) is found within humans, mice, zebrafish and other higher organisms (Piatek *et al.*, 2016). It seems logical that evolution would not have kept a disadvantageous RNA transcript. Therefore, even though the exact function of Slc34a2a(as) remains undetermined within this project, many possibilities have been eliminated such as TI and RNAi. Many questions still remain to be answered about the zebrafish Slc34a2a antisense RNA transcript, but this study has proved that maintaining a natural balance is vital for normal development.

7.3 Impact on the medical field

Within the last decade, the biological importance of regulatory RNA has been recognized with major advances in understanding their individual roles in health and disease. With a surprisingly large number of human genes being transcribed in both directions, both sense and antisense RNA transcripts play a role in gene expression regulation and thus make promising targets for therapeutics. Due to emerging and rapidly changing challenges to human health, modern medicine requires new forms of treatment for both common and unusual ailments. Antisense oligonucleotides

designed to bind to target RNAs via well-characterized Watson-Crick base pairing, can modulate the target RNA's function through an assortment of post-hybidizing events such as cleaving, degradation or translation prevention (Boisguérin *et al.*, 2015).

Many factors however, need to be considered prior to trialling and use. Not only must the sequence and function of the target RNA be known, but also the nature of related transcripts to avoid unwanted hybridization and off target effects. The work within this study highlights the importance of sense/ antisense transcript balance. By knocking-down the NAT by microinjecting the Slc34a2a sense transcript, a negative phenotype emerged that lead to behavioural impairment and severe physical malformation dependent upon dose. Thus, alongside dose, not only is target specificity a necessity, knowing no other targets are present is critical.

Cell-specific delivery of antisense oligonucleotides is also vital as knocking down the target RNA in normal healthy tissues might result in negative outcomes. To date, numerous delivery strategies have been tested, each having its own advantages and disadvantages, including nanocarriers, conjugates of oligonucleotides and various ligands (see reviews;. (Juliano *et al.*, 2008; Boisguérin *et al.*, 2015)).

There is a natural balance between cellular components and regulatory pathways within a host, as seen in this study. The Slc34a2a antisense transcript examined within this project also naturally occurs in humans and other mammalian species (Piatek *et al.*, 2016). Hopefully this study provides insight into how the human equivalent functions thus adding to the breadth of knowledge required for designing future therapeutic antisense oligonucleotides. Heterozygous loss-of-function mutation in SCN1A, a voltage-gated sodium channel, causes a brain disorder called Dravet syndrome. Antisense oligonucleotide interference of SCN1ANAT, a NAT which controls SCN1A, has been determined to relieve the symptoms of Dravet syndrome in both mouse and non-human primate models (Hsiao *et al.*, 2016). In the future, a comparable strategy could be used to increase Pi reabsorption by SLC34A1 in patients with Pi wasting and kidney stones. Antisense and siRNA oligonucleotide clinical trials are already underway with RNA targets involved in cancer progression and metastasis (Chi *et al.*, 2010; Cresce and Koropatnick, 2010; Bedikian *et al.*, 2014; Natale *et al.*, 2014), neurodegenerative disease such as Duchenne's muscular dystrophy (Takeshima *et al.*, 2006; Koo and Wood, 2013),

allergen induced inflammation control (Gauvreau *et al.*, 2008) and Crohn's disease (Monteleone *et al.*, 2015).

7.4 Future work

Alongside mapping and more extensively quantifying Slc34a2a sense and antisense transcripts, this study has increased the knowledge on how zebrafish Slc34a2a/Slc34a2a(as) RNA transcripts interact and affect one another. However, much more research needs to be accomplished to complete our understanding of Slc34a2a(as) existence. Slc34a2a(as) knockout studies, such as CRISPR-Cas9 insertion of a premature stop codon within the first 100 bp of the antisense genome sequence, would provide insight into the role of Slc34a2a(as). It is plausible that Slc34a2a(as) may regulate an unknown transcript in *trans* through a dsRNA-dependent mechanism. Antisense-induced methylation is the least understood mechanism by which NATs can regulate gene expression. Bisulfite treatment of DNA (cytosine converts to uracil unless methylated) from wildtype fish versus Slc34a2a(as) knockout fish, followed by next-generation sequencing, would provide insight in the methylation status and changes coinciding with presence of Slc34a2a(as) RNA.

Preliminary experiments were completed with anti-dsRNA J2 monoclonal antibody. Further experiments need to be conducted to determine the specific conditions under which J2 antibody would be able to extract dsRNA from whole tissues. RNAseq on all dsRNA extracted would provide novel insight into the quantity of endogenously occurring dsRNA at different developmental stages. Completing RNAseq on dsRNA extracted from Dicer knock-down fish would be interesting to determine the amount of dsRNA that gets processed through RNAi. Perhaps this would provide insights as to why some endogenous dsRNAs do not activate the interferon response pathway. Additionally, genomic areas with higher levels of dsRNA could be determined after alignment of compiled RNAseq data, thus providing supporting information for the number of NATs theorised to exist. If dsRNA extraction was found to be successful from zebrafish whole tissues, then it would be possible to perform the extraction on human tissues. All sequencing

information from human dsRNA could add to the knowledge base to reduce off target effects for future antisense and siRNA oligonucleotide therapeutics.

This thesis will provide a solid foundation to inform future work either in the field of antisense-mediated gene regulation or in studies focussing on dsRNA using the J2 antibody.

References

Alexander, S.P.H., Kelly, E., Marrion, N., Peters, J.A., Benson, H.E., Faccenda, E., Pawson, A.J., Sharman, J.L., Southan, C., Davies, J.A. and Collaborators, C. (2015) 'SLC superfamily of solute carriers' *Concise Guide to PHARMACOLOGY* 10/08/2016.

Altman, N.R., Naidich, T.P. and Braffman, B.H. (1992) 'Posterior fossa malformations', *AJNR Am J Neuroradiol*, 13(2), pp. 691-724.

Amaral, P.P., Dinger, M.E. and Mattick, J.S. (2013) 'Non-coding RNAs in homeostasis, disease and stress responses: an evolutionary perspective', *Briefings in Functional Genomics*, 12(3), pp. 254-278.

Aruga, J., Minowa, O., Yaginuma, H., Kuno, J., Nagai, T., Noda, T. and Mikoshiba, K. (1998) 'Mouse Zic1 Is Involved in Cerebellar Development', *The Journal of Neuroscience*, 18(1), pp. 284-293.

Augustine-Rauch, K., Zhang, C.X. and Panzica-Kelly, J.M. (2010) 'In vitro developmental toxicology assays: A review of the state of the science of rodent and zebrafish whole embryo culture and embryonic stem cell assays', *Birth Defects Research Part C: Embryo Today: Reviews*, 90(2), pp. 87-98.

Babiarz, J.E., Ruby, J.G., Wang, Y., Bartel, D.P. and Blelloch, R. (2008) 'Mouse ES cells express endogenous shRNAs, siRNAs, and other Microprocessor-independent, Dicer-dependent small RNAs', *Genes & Development*, 22(20), pp. 2773-2785.

Baguma-Nibasheka, M., Macfarlane, L.A. and Murphy, P.R. (2012) 'Regulation of fibroblast growth factor-2 expression and cell cycle progression by an endogenous antisense RNA', *Genes (Basel)*, 3(3), pp. 505-20.

Ballarati, L., Rossi, E., Bonati, M.T., Gimelli, S., Maraschio, P., Finelli, P., Giglio, S., Lapi, E., Bedeschi, M.F., Guerneri, S., Arrigo, G., Patricelli, M.G., Mattina, T., Guzzardi, O., Pecile, V., Police, A., Scarano, G., Larizza, L., Zuffardi, O. and Giardino, D. (2007) '13q Deletion and central nervous system anomalies: further insights from karyotype–phenotype analyses of 14 patients', *Journal of Medical Genetics*, 44(1), p. e60.

Barbazuk, W.B., Korf, I., Kadavi, C., Heyen, J., Tate, S., Wun, E., Bedell, J.A., McPherson, J.D. and Johnson, S.L. (2000) 'The Syntenic Relationship of the Zebrafish and Human Genomes', *Genome Research*, 10(9), pp. 1351-1358.

Barraud, P. and Allain, F.H.T. (2012) 'ADAR Proteins: Double-stranded RNA and Z-DNA Binding Domains', in Samuel, E.C. (ed.) *Adenosine Deaminases Acting on RNA (ADARs) and A-to-I Editing*. Berlin, Heidelberg: Springer Berlin Heidelberg, pp. 35-60.

Barry, G. and Mattick, J.S. (2012) 'The role of regulatory RNA in cognitive evolution', *Trends in Cognitive Sciences*, 16(10), pp. 497-503.

Bass, B.L. (2002) 'RNA editing by adenosine deaminases that act on RNA', *Annual Review Biochemistry*, 71, pp. 817-46.

Bedell, V.M., Westcot, S.E. and Ekker, S.C. (2011) 'Lessons from morpholino-based screening in zebrafish', *Briefings in Functional Genomics*, 10(4), pp. 181-188.

Bedikian, A.Y., Garbe, C., Conry, R., Lebbe, C., Grob, J.J. and the Genasense Melanoma Study, G. (2014) 'Dacarbazine with or without oblimersen (a Bcl-2 antisense oligonucleotide) in chemotherapy-naive patients with advanced melanoma and low–normal serum lactate dehydrogenase: ‘The AGENDA trial’', *Melanoma Research*, 24(3), pp. 237-243.

Beltran, M., Puig, I., Pena, C., Garcia, J.M., Alvarez, A.B., Pena, R., Bonilla, F. and de Herreros, A.G. (2008) 'A natural antisense transcript regulates Zeb2/Sip1 gene expression during Snail1-induced epithelial-mesenchymal transition', *Genes & Development* 22(6), pp. 756-69.

Berezhna, S.Y., Supekova, L., Supek, F., Schultz, P.G. and Deniz, A.A. (2006) 'siRNA in human cells selectively localizes to target RNA sites', *Proceedings of the National Academy of Sciences*, 103(20), pp. 7682-7.

Bertone, P., Stolc, V., Royce, T.E., Rozowsky, J.S., Urban, A.E., Zhu, X., Rinn, J.L., Tongprasit, W., Samanta, M., Weissman, S., Gerstein, M. and Snyder, M. (2004) 'Global Identification of Human Transcribed Sequences with Genome Tiling Arrays', *Science*, 306(5705), pp. 2242-2246.

Bill, B.R., Petzold, A.M., Clark, K.J., Schimmenti, L.A. and Ekker, S.C. (2009) 'A primer for morpholino use in zebrafish', *Zebrafish*, 6(1), pp. 69-77.

Billy, E., Brondani, V., Zhang, H., Müller, U. and Filipowicz, W. (2001) 'Specific interference with gene expression induced by long, double-stranded RNA in mouse embryonal teratocarcinoma cell lines', *Proceedings of the National Academy of Sciences*, 98(25), pp. 14428-14433.

Blouin, A.G., Ross, H.A., Hobson-Peters, J., O'Brien, C.A., Warren, B. and MacDiarmid, R. (2016) 'A new virus discovered by immunocapture of double-stranded RNA, a rapid method for virus enrichment in metagenomic studies', *Molecular Ecology Resources*, pp. n/a-n/a.

Boisguérin, P., Deshayes, S., Gait, M.J., O'Donovan, L., Godfrey, C., Betts, C.A., Wood, M.J.A. and Lebleu, B. (2015) 'Delivery of therapeutic oligonucleotides with cell penetrating peptides', *Advanced Drug Delivery Reviews*, 87, pp. 52-67.

Boland, E., Clayton-Smith, J., Woo, V.G., McKee, S., Manson, F.D.C., Medne, L., Zackai, E., Swanson, E.A., Fitzpatrick, D., Millen, K.J., Sherr, E.H., Dobyns, W.B. and Black, G.C.M. (2007) 'Mapping of Deletion and Translocation Breakpoints in 1q44 Implicates the Serine/Threonine Kinase AKT3 in Postnatal Microcephaly and Agenesis of the Corpus Callosum', *The American Journal of Human Genetics*, 81(2), pp. 292-303.

Bonin, M., Oberstrass, J., Lukacs, N., Ewert, K., Oesterschulze, E., Kassing, R. and Nellen, W. (2000) 'Determination of preferential binding sites for anti-dsRNA antibodies on double-stranded RNA by scanning force microscopy', *RNA*, 6(4), pp. 563-570.

Brand, M., Beuchle, D., Endres, F., Hafter, P., Hammerschmidt, M., Mullins, M., Schulte-Merker, S., Nusslein-Volhard, C., Lucl, R., Jurgen, K. and Schwarz, S. (1995) 'Keeping and raising zebrafish (Danio rerio) in Tubingen', *The Zebrafish Science Monitor*, 3, pp. 2-7.

BrownUniversity (2016) *Preparation of dsRNA*.

Buhler, A., Kustermann, M., Bummer, T., Rottbauer, W., Sandri, M. and Just, S. (2016) 'Atrogin-1 Deficiency Leads to Myopathy and Heart Failure in Zebrafish', *International Journal of Molecular Sciences* 17(2).

Burgess, D.J. (2013) 'Small RNAs: antiviral RNAi in mammals', *Nature Reviews Genetics* 14(12), p. 821.

Bustin, S.A., Benes, V., Garson, J.A., Hellemans, J., Huggett, J., Kubista, M., Mueller, R., Nolan, T., Pfaffl, M.W., Shipley, G.L., Vandesompele, J. and Wittwer, C.T. (2009) 'The MIQE Guidelines: Minimum Information for Publication of Quantitative Real-Time PCR Experiments', *Clinical Chemistry*, 55(4), pp. 611-622.

Cachat, J., Kyzar, E.J., Collins, C., Gaikwad, S., Green, J., Roth, A., El-Ounsi, M., Davis, A., Pham, M., Landsman, S., Stewart, A.M. and Kalueff, A.V. (2013) 'Unique and potent effects of acute ibogaine on zebrafish: The developing utility of novel aquatic models for hallucinogenic drug research', *Behavioural Brain Research*, 236, pp. 258-269.

Campbell, W.A., Yang, H., Zetterberg, H., Baulac, S., Sears, J.A., Liu, T., Wong, S.T.C., Zhong, T.P. and Xia, W. (2006) 'Zebrafish lacking Alzheimer presenilin enhancer 2 (Pen-2) demonstrate excessive p53-dependent apoptosis and neuronal loss', *Journal of Neurochemistry*, 96(5), pp. 1423-1440.

Carlile, M., Nalbant, P., Preston-Fayers, K., McHaffie, G.S. and Werner, A. (2008) 'Processing of naturally occurring sense/antisense transcripts of the vertebrate Slc34a gene into short RNAs', *Physiological Genomics*, 34(1), pp. 95-100.

Carlile, M., Swan, D., Jackson, K., Preston-Fayers, K., Ballester, B., Flicek, P. and Werner, A. (2009) 'Strand selective generation of endo-siRNAs from the Na/phosphate transporter gene Slc34a1 in murine tissues', *Nucleic Acids Research*, 37(7), pp. 2274-2282.

Carmany-Rampey, A. and Moens, C.B. (2006) 'Modern mosaic analysis in the zebrafish', *Methods*, 39(3), pp. 228-238.

Carrieri, C., Cimatti, L., Biagioli, M., Beugnet, A., Zucchelli, S., Fedele, S., Pesce, E., Ferrer, I., Collavin, L., Santoro, C., Forrest, A.R.R., Carninci, P., Biffo, S., Stupka, E. and Gustincich, S. (2012) 'Long non-coding antisense RNA controls Uchl1 translation through an embedded SINEB2 repeat', *Nature*, 491(7424), pp. 454-457.

Carthew, R.W. and Sontheimer, E.J. (2009) 'Origins and Mechanisms of miRNAs and siRNAs', *Cell*, 136(4), pp. 642-655.

Casadei, R., Pelleri, M.C., Vitale, L., Facchin, F., Lenzi, L., Canaider, S., Strippoli, P. and Frabetti, F. (2011) 'Identification of housekeeping genes suitable for gene

expression analysis in the zebrafish', *Gene Expression Patterns*, 11(3–4), pp. 271-276.

Castillo, A. and Cifuentes, V. (1994) 'Presence of double-stranded RNA and virus-like particles in Phaffia rhodozyma', *Current Genetics*, 26(4), pp. 364-8.

Chan, J., Atianand, M., Jiang, Z., Carpenter, S., Aiello, D., Elling, R., Fitzgerald, K.A. and Caffrey, D.R. (2015) 'Cutting Edge: A Natural Antisense Transcript, AS-IL1alpha, Controls Inducible Transcription of the Proinflammatory Cytokine IL-1alpha', *Journal of Immunology*, 195(4), pp. 1359-63.

Chang, C.C., Zhang, B., Li, C.Y., Chang, H.C (2012) 'Exploring cytoplasmic dynamics in zebrafish yolk cells by single particle tracking of fluorescent nanodiamonds', *Proceedings of SPIE - The International Society for Optical Engineering*, 8272(1).

Chatterjee, M., Guo, Q., Weber, S., Scholpp, S. and Li, J.Y.H. (2014) 'Pax6 regulates the formation of the habenular nuclei by controlling the temporospatial expression of Shhin the diencephalon in vertebrates', *BMC Biology*, 12(1), pp. 1-16.

Chen, C.-P. and Shih, J.-C. (2005) 'Association of partial trisomy 9p and the Dandy-Walker malformation', *American Journal of Medical Genetics Part A*, 132A(1), pp. 111-112.

Chen, J., Sun, M., Kent, W.J., Huang, X., Xie, H., Wang, W., Zhou, G., Shi, R.Z. and Rowley, J.D. (2004) 'Over 20% of human transcripts might form sense-antisense pairs', *Nucleic Acids Res*, 32(16), pp. 4812-20.

Cheng, J., Kapranov, P., Drenkow, J., Dike, S., Brubaker, S., Patel, S., Long, J., Stern, D., Tammana, H., Helt, G., Sementchenko, V., Piccolboni, A., Bekiranov, S., Bailey, D.K., Ganesh, M., Ghosh, S., Bell, I., Gerhard, D.S. and Gingeras, T.R. (2005) 'Transcriptional maps of 10 human chromosomes at 5-nucleotide resolution', *Science*, 308(5725), pp. 1149-54.

Chi, K.N., Hotte, S.J., Yu, E.Y., Tu, D., Eigl, B.J., Tannock, I., Saad, F., North, S., Powers, J., Gleave, M.E. and Eisenhauer, E.A. (2010) 'Randomized Phase II Study of Docetaxel and Prednisone With or Without OGX-011 in Patients With Metastatic Castration-Resistant Prostate Cancer', *Journal of Clinical Oncology*, 28(27), pp. 4247-4254.

Chu, C.-y. and Rana, T.M. (2006) 'Translation Repression in Human Cells by MicroRNA-Induced Gene Silencing Requires RCK/p54', *PLoS Biology*, 4(7), p. e210.

Chugh, P., Tamburro, K. and Dittmer, D.P. (2010) 'Profiling of pre-micro RNAs and microRNAs using quantitative real-time PCR (qPCR) arrays', *Journal of Visualized Experiments* (46).

Clancy, S. (2008) 'RNA Transcription by RNA Polymerase: Prokaryotes vs Eukaryotes', *Nature Education*, 1(1), p. 125.

Clancy, S. and Brown, W. (2008) 'Translation: DNA to mRNA to Protein', *Nature Education*, 1(1), p. 101.

Cogoni, C. and Macino, G. (2000) 'Post-transcriptional gene silencing across kingdoms', *Current Opinion in Genetics & Development*, 10(6), pp. 638-643.

Consortium, T.E.P. (2007) 'Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project', *Nature*, 447(7146), pp. 799-816.

Consortium, T.E.P. (2012) 'An Integrated Encyclopedia of DNA Elements in the Human Genome', *Nature*, 489(7414), pp. 57-74.

Contreras, R., Cheroutre, H., Degrave, W. and Fiers, W. (1982) 'Simple, efficient in vitro synthesis of capped RNA useful for direct expression of cloned eukaryoti genes', *Nucleic Acids Research*, 10(20), pp. 6353-6362.

Cresce, C.D. and Koropatnick, J. (2010) 'Antisense Treatment in Human Prostate Cancer and Melanoma', *Current Cancer Drug Targets*, 10(6), pp. 555-565.

Croen, K.D., Ostrove, J.M., Dragovic, L.J., Smialek, J.E. and Straus, S.E. (1987) 'Latent Herpes Simplex Virus in Human Trigeminal Ganglia', *New England Journal of Medicine*, 317(23), pp. 1427-1432.

D'Aniello, E., Rydeen, A.B., Anderson, J.L., Mandal, A. and Waxman, J.S. (2013) 'Depletion of retinoic acid receptors initiates a novel positive feedback mechanism that promotes teratogenic increases in retinoic acid', *PLoS Genetics*, 9(8), p. e1003689.

Dahlem, T.J., Hoshijima, K., Jurynec, M.J., Gunther, D., Starker, C.G., Locke, A.S., Weis, A.M., Voytas, D.F. and Grunwald, D.J. (2012) 'Simple Methods for Generating

and Detecting Locus-Specific Mutations Induced with TALENs in the Zebrafish Genome', *PLoS Genetics*, 8(8), p. e1002861.

Dash, S.N., Lehtonen, E., Wasik, A.A., Schepis, A., Paavola, J., Panula, P., Nelson, W.J. and Lehtonen, S. (2014) 'sept7b is essential for pronephric function and development of left–right asymmetry in zebrafish embryogenesis', *Journal of Cell Science*, 127(7), pp. 1476-1486.

Day, D.A. and Tuite, M.F. (1998) 'Post-transcriptional gene regulatory mechanisms in eukaryotes: an overview', *Journal of Endocrinology*, 157(3), pp. 361-371.

De Luca, E., Zaccaria, G.M., Hadhoud, M., Rizzo, G., Ponzini, R., Morbiducci, U. and Santoro, M.M. (2014) 'ZebraBeat: a flexible platform for the analysis of the cardiac rate in zebrafish embryos', *Scientific Reports*, 4, p. 4898.

De Rienzo, G., Gutzman, J.H. and Sive, H. (2012) 'Efficient shRNA-Mediated Inhibition of Gene Expression in Zebrafish', *Zebrafish*, 9(3), pp. 97-107.

Deng, W. and Roberts, S.G. (2006) 'Core promoter elements recognized by transcription factor IIB', *Biochemical Society Transactions*, 34(Pt 6), pp. 1051-3.

Dragan, A.I., Pavlovic, R., McGivney, J.B., Casas-Finet, J.R., Bishop, E.S., Strouse, R.J., Schenerman, M.A. and Geddes, C.D. (2012) 'SYBR Green I: fluorescence properties and interaction with DNA', *Journal of Fluorescence* 22(4), pp. 1189-99.

DRSC, H.M.S. (2015) *Protocol for dsRNA synthesis*. Available at: http://www.flyrnai.org/DRSC-PRS.html (Accessed: 23 July).

Echeverri, K. and Oates, A.C. (2007) 'Coordination of symmetric cyclic gene expression during somitogenesis by Suppressor of Hairless involves regulation of retinoic acid catabolism', *Developmental Biology*, 301(2), pp. 388-403.

Eisen, J.S. and Smith, J.C. (2008) 'Controlling morpholino experiments: don't stop making antisense', *Development*, 135(10), pp. 1735-43.

Ekker, S.C. and Larson, J.D. (2001) 'Morphant technology in model developmental systems', *Genesis (New York, N.Y. : 2000)*, 30(3), pp. 89-93.

EMBL-EBI (2016) *Kalign Multiple Sequence Alignment* [Computer program]. EMBL-EBI. Available at: http://www.ebi.ac.uk/Tools/msa/ (Accessed: 7 March 2016).

EMBL-EBI, W. (2015) *e!Ensembl BLAST/BLAT* [Computer program]. EMBL-EBI. Available at: http://www.ensembl.org/index.html (Accessed: 7 March 2016).

Ensembl 84 (2016a) 'Gene: psen2 ' *Ensembl*. Available at: http://www.ensembl.org/Danio_rerio/Share/ab73d43ea13c8f8b0147089157b251ad249192570.

Ensembl 84 (2016b) 'Gene: rasl11b ' *Ensembl*. Available at: http://www.ensembl.org/Danio_rerio/Gene/Summary?db=core;g=ENSDARG00000015611;r=20:23123885-23127123;t=ENSDART00000015755.

Esteller, M. (2011) 'Non-coding RNAs in human disease', *Nature Reviews Genetics* 12(12), pp. 861-74.

Eszterhas, S.K., Bouhassira, E.E., Martin, D.I.K. and Fiering, S. (2002) 'Transcriptional Interference by Independently Regulated Genes Occurs in Any Relative Arrangement of the Genes and Is Influenced by Chromosomal Integration Position', *Molecular and Cellular Biology*, 22(2), pp. 469-479.

Faghihi, M.A., Modarresi, F., Khalil, A.M., Wood, D.E., Sahagan, B.G., Morgan, T.E., Finch, C.E., St Laurent, G., Kenny, P.J. and Wahlestedt, C. (2008) 'Expression of a noncoding RNA is elevated in Alzheimer's disease and drives rapid feed-forward regulation of beta-secretase', *Nature Medicine*, 14, pp. 723 - 730.

Faghihi, M.A. and Wahlestedt, C. (2009) 'Regulatory roles of natural antisense transcripts', *Nat Rev Mol Cell Biol*, 10(9), pp. 637-43.

Faghihi, M.A., Zhang, M., Huang, J., Modarresi, F., Van der Brug, M.P., Nalls, M.A., Cookson, M.R., St-Laurent, G. and Wahlestedt, C. (2010) 'Evidence for natural antisense transcript-mediated inhibition of microRNA function', *Genome Biology*, 11.

Fan, J. and Papadopoulos, V. (2012) 'Transcriptional regulation of translocator protein (Tspo) via a SINE B2-mediated natural antisense transcript in MA-10 Leydig cells', *Biology of Reproduction*, 86(5), pp. 147, 1-15.

Farrell, C.M. and Lukens, L.N. (1995) 'Naturally Occurring Antisense Transcripts Are Present in Chick Embryo Chondrocytes Simultaneously with the Down-regulation of the 1(I) Collagen Gene', *Journal of Biological Chemistry*, 270(7), pp. 3400-3408.

Fire, A., Xu, S., Montgomery, M.K., Kostas, S.A., Driver, S.E. and Mello, C.C. (1998) 'Potent and specific genetic interference by double-stranded RNA in Caenorhabditis elegans', *Nature*, 391(6669), pp. 806-811.

Fischer, S., Filipek-Gorniok, B. and Ledin, J. (2011) 'Zebrafish Ext2 is necessary for Fgf and Wnt signaling, but not for Hh signaling', *BMC Developmental Biology*, 11, pp. 53-53.

Fjose, A., Njølstad, P.R., Nornes, S., Molven, A. and Krauss, S. (1992) 'Structure and early embryonic expression of the zebrafish engrailed-2 gene', *Mechanisms of Development*, 39(1), pp. 51-62.

Flegr, J. (1987) 'A Rapid Method for Isolation of Double Stranded RNA', *Preparative Biochemistry*, 17(4).

Flemr, M., Malik, R., Franke, V., Nejepinska, J., Sedlacek, R., Vlahovicek, K. and Svoboda, P. (2013) 'A Retrotransposon-Driven Dicer Isoform Directs Endogenous Small Interfering RNA Production in Mouse Oocytes', *Cell*, 155(4), pp. 807-816.

Fraering, P.C. (2007) 'Structural and Functional Determinants of Secretase, an Intramembrane Protease Implicated in Alzheimer's Disease', *Current Genomics*, 8(8), pp. 531-549.

Fuertes, M.B., Woo, S.-R., Burnett, B., Fu, Y.-X. and Gajewski, T.F. (2013) 'Type I interferon response and innate immune sensing of cancer', *Trends in Immunology*, 34(2), pp. 67-73.

Gagnon, Keith T., Li, L., Chu, Y., Janowski, Bethany A. and Corey, David R. (2014) 'RNAi Factors Are Present and Active in Human Cell Nuclei', *Cell Reports*, 6(1), pp. 211-221.

Galiveti, C.R., Raabe, C.A., Konthur, Z. and Rozhdestvensky, T.S. (2014) 'Differential regulation of non-protein coding RNAs from Prader-Willi Syndrome locus', *Scientific Reports*, 4, p. 6445.

Gantier, M.P. and Williams, B.R.G. (2007) 'The response of mammalian cells to double-stranded RNA', *Cytokine & growth factor reviews*, 18(5-6), pp. 363-371.

Gauvreau, G.M., Boulet, L.P., Cockcroft, D.W., Baatjes, A., Cote, J., Deschesnes, F., Davis, B., Strinich, T., Howie, K., Duong, M., Watson, R.M., Renzi, P.M. and

O'Byrne, P.M. (2008) 'Antisense Therapy against CCR3 and the Common Beta Chain Attenuates Allergen-induced Eosinophilic Responses', *American Journal of Respiratory and Critical Care Medicine*, 177(9), pp. 952-958.

Ge, X., Rubinstein, W.S., Jung, Y.C. and Wu, Q. (2008) 'Genome-wide analysis of antisense transcription with Affymetrix exon array', *BMC Genomics*, 9, p. 27.

Gene Tools, L. (2016) *Morpholinos*. Available at: http://www.gene-tools.com/.

Ghosh, M.K., Katyal, A., Chandra, R. and Brahmachari, V. (2005) 'Targeted activation of transcription in vivo through hairpin-triplex forming oligonucleotide in Saccharomyces cerevisiae', *Molecular and Cellular Biochemistry*, 278(1), pp. 147-155.

Giraldez, A.J., Cinalli, R.M., Glasner, M.E., Enright, A.J., Thomson, J.M., Baskerville, S., Hammond, S.M., Bartel, D.P. and Schier, A.F. (2005) 'MicroRNAs Regulate Brain Morphogenesis in Zebrafish', *Science*, 308(5723), pp. 833-838.

Gou, L.-T., Dai, P. and Liu, M.-F. (2014) 'Small noncoding RNAs and male infertility', *Wiley Interdisciplinary Reviews: RNA*, 5(6), pp. 733-745.

Graham, C., Nalbant, P., Schölermann, B., Hentschel, H., Kinne, R.K.H. and Werner, A. (2003) 'Characterization of a type IIb sodium-phosphate cotransporter from zebrafish (Danio rerio) kidney', *American Journal of Physiology - Renal Physiology*, 284(4), pp. F727-F736.

Grinberg, I. and Millen, K.J. (2005) 'The ZIC gene family in development and disease', *Clinical Genetics*, 67(4), pp. 290-296.

Groth, C., Nornes, S., McCarty, R., Tamme, R. and Lardelli, M. (2002) 'Identification of a second presenilin gene in zebrafish with similarity to the human Alzheimer's disease gene presenilin2', *Development Genes and Evolution*, 212(10), pp. 486-490.

Guo, J.H., Cheng, H.P., Yu, L. and Zhao, S. (2006) 'Natural antisense transcripts of Alzheimer's disease associated genes', *DNA Seq*, 17(2), pp. 170-3.

Han, J., Lee, Y., Yeom, K.-H., Nam, J.-W., Heo, I., Rhee, J.-K., Sohn, S.Y., Cho, Y., Zhang, B.-T. and Kim, V.N. (2006) 'Molecular Basis for the Recognition of Primary microRNAs by the Drosha-DGCR8 Complex', *Cell*, 125(5), pp. 887-901.

Hanks, M., Wurst, W., Anson-Cartwright, L., Auerbach, A.B. and Joyner, A.L. (1995) 'Rescue of the En-1 mutant phenotype by replacement of En-1 with En-2', *Science*, 269(5224), pp. 679-682.

Hanks, M.C., Loomis, C.A., Harris, E., Tong, C.X., Anson-Cartwright, L., Auerbach, A. and Joyner, A. (1998) 'Drosophila engrailed can substitute for mouse Engrailed1 function in mid-hindbrain, but not limb development', *Development*, 125(22), pp. 4521-4530.

Hastings, M.L., Milcarek, C., Martincic, K., Peterson, M.L. and Munroe, S.H. (1997) 'Expression of the thyroid hormone receptor gene, erbAalpha, in B lymphocytes: alternative mRNA processing is independent of differentiation but correlates with antisense RNA levels', *Nucleic Acids Research*, 25(21), pp. 4296-4300.

Higashijima, S.-i. (2008) 'Transgenic zebrafish expressing fluorescent proteins in central nervous system neurons', *Development, Growth & Differentiation*, 50(6), pp. 407-413.

Hoffmann, H.-H., Schneider, W.M. and Rice, C.M. (2015) 'Interferons and viruses: an evolutionary arms race of molecular interactions', *Trends in Immunology*, 36(3), pp. 124-138.

Hornung, V., Guenthner-Biller, M., Bourquin, C., Ablasser, A., Schlee, M., Uematsu, S., Noronha, A., Manoharan, M., Akira, S., de Fougerolles, A., Endres, S. and Hartmann, G. (2005) 'Sequence-specific potent induction of IFN-[alpha] by short interfering RNA in plasmacytoid dendritic cells through TLR7', *Nature Medicine*, 11(3), pp. 263-270.

Houseley, J. and Tollervey, D. (2009) 'The Many Pathways of RNA Degradation', *Cell*, 136(4), pp. 763-776.

Hsiao, J., Yuan, T.Y., Tsai, M.S., Lu, C.Y., Lin, Y.C., Lee, M.L., Lin, S.W., Chang, F.C., Liu Pimentel, H., Olive, C., Coito, C., Shen, G., Young, M., Thorne, T., Lawrence, M., Magistri, M., Faghihi, M.A., Khorkova, O. and Wahlestedt, C. (2016) 'Upregulation of Haploinsufficient Gene Expression in the Brain by Targeting a Long Non-coding RNA Improves Seizure Phenotype in a Model of Dravet Syndrome', *EBioMedicine*, 9, pp. 257-77.

Hudziak, R.M., Barofsky, E., Barofsky, D.F., Weller, D.L., Huang, S.B. and Weller, D.D. (1996) 'Resistance of morpholino phosphorodiamidate oligomers to enzymatic degradation', *Antisense Nucleic Acid Drug Dev*, 6(4), pp. 267-72.

Imamura, T., Yamamoto, S., Ohgane, J., Hattori, N., Tanaka, S. and Shiota, K. (2004) 'Non-coding RNA directed DNA demethylation of Sphk1 CpG island', *Biochemical and Biophysical Research Communications*, 322(2), pp. 593-600.

'Initial sequencing and comparative analysis of the mouse genome', (2002) *Nature*, 420(6915), pp. 520-562.

Isogai, S., Horiguchi, M. and Weinstein, B.M. (2001) 'The Vascular Anatomy of the Developing Zebrafish: An Atlas of Embryonic and Early Larval Development', *Developmental Biology*, 230(2), pp. 278-301.

Janowski, B.A. and Corey, D.R. (2010) 'Minireview: Switching on Progesterone Receptor Expression with Duplex RNA', *Molecular Endocrinology*, 24(12), pp. 2243-2252.

Jao, L.-E., Wente, S.R. and Chen, W. (2013) 'Efficient multiplex biallelic zebrafish genome editing using a CRISPR nuclease system', *Proceedings of the National Academy of Sciences of the United States of America*, 110(34), pp. 13904-13909.

Jeggari, A., Marks, D.S. and Larsson, E. (2012) 'miRcode: a map of putative microRNA target sites in the long non-coding transcriptome', *Bioinformatics*, 28(15), pp. 2062-2063.

Jensen, Lasse D., Cao, Z., Nakamura, M., Yang, Y., Bräutigam, L., Andersson, P., Zhang, Y., Wahlberg, E., Länne, T., Hosaka, K. and Cao, Y. (2012) 'Opposing Effects of Circadian Clock Genes Bmal1 and Period2 in Regulation of VEGF-Dependent Angiogenesis in Developing Zebrafish', *Cell Reports*, 2(2), pp. 231-241.

Jiang, W., Wang, P.-z., Yu, H.-t., Zhang, Y., Zhao, K., Du, H. and Bai, X.-f. (2014) 'Development of a SYBR Green I based one-step real-time PCR assay for the detection of Hantaan virus', *Journal of Virological Methods*, 196, pp. 145-151.

Jing, Q., Huang, S., Guth, S., Zarubin, T., Motoyama, A., Chen, J., Di Padova, F., Lin, S.-C., Gram, H. and Han, J. (2005) 'Involvement of MicroRNA in AU-Rich Element-Mediated mRNA Instability', *Cell*, 120(5), pp. 623-634.

Judge, A.D., Sood, V., Shaw, J.R., Fang, D., McClintock, K. and MacLachlan, I. (2005) 'Sequence-dependent stimulation of the mammalian innate immune response by synthetic siRNA', *Nature Biotechnology* 23(4), pp. 457-462.

Juliano, R., Alam, M.R., Dixit, V. and Kang, H. (2008) 'Mechanisms and strategies for effective delivery of antisense and siRNA oligonucleotides', *Nucleic Acids Research*, 36(12), pp. 4158-4171.

Karikó, K., Bhuyan, P., Capodici, J., Ni, H., Lubinski, J., Friedman, H. and Weissman, D. (2004) 'Exogenous siRNA Mediates Sequence-Independent Gene Suppression by Signaling through Toll-Like Receptor 3', *Cells Tissues Organs*, 177(3), pp. 132-138.

Katayama, S., Tomaru, Y., Kasukawa, T., Waki, K., Nakanishi, M., Nakamura, M., Nishida, H., Yap, C.C., Suzuki, M., Kawai, J., Suzuki, H., Carninci, P., Hayashizaki, Y., Wells, C., Frith, M., Ravasi, T., Pang, K.C., Hallinan, J., Mattick, J., Hume, D.A., Lipovich, L., Batalov, S., Engström, P.G., Mizuno, Y., Faghihi, M.A., Sandelin, A., Chalk, A.M., Mottagui-Tabar, S., Liang, Z., Lenhard, B. and Wahlestedt, C. (2005a) 'Antisense Transcription in the Mammalian Transcriptome', *Science*, 309(5740), pp. 1564-1566.

Katayama, S., Tomaru, Y., Kasukawa, T., Waki, K., Nakanishi, M., Nakamura, M., Nishida, H., Yap, C.C., Suzuki, M., Kawai, J., Suzuki, H., Carninci, P., Hayashizaki, Y., Wells, C., Frith, M., Ravasi, T., Pang, K.C., Hallinan, J., Mattick, J., Hume, D.A., Lipovich, L., Batalov, S., Engstrom, P.G., Mizuno, Y., Faghihi, M.A., Sandelin, A., Chalk, A.M., Mottagui-Tabar, S., Liang, Z., Lenhard, B., Wahlestedt, C., Group, R.G.E.R., Genome Science, G. and Consortium, F. (2005b) 'Antisense transcription in the mammalian transcriptome', *Science*, 309(5740), pp. 1564-6.

Kawamura, Y., Saito, K., Kin, T., Ono, Y., Asai, K., Sunohara, T., Okada, T.N., Siomi, M.C. and Siomi, H. (2008) 'Drosophila endogenous small RNAs bind to Argonaute[thinsp]2 in somatic cells', *Nature*, 453(7196), pp. 793-797.

Kawashima, T., Kosaka, A., Yan, H., Guo, Z., Uchiyama, R., Fukui, R., Kaneko, D., Kumagai, Y., You, D.-J., Carreras, J., Uematsu, S., Jang, Myoung H., Takeuchi, O., Kaisho, T., Akira, S., Miyake, K., Tsutsui, H., Saito, T., Nishimura, I. and Tsuji, Noriko M. (2013) 'Double-Stranded RNA of Intestinal Commensal but Not

Pathogenic Bacteria Triggers Production of Protective Interferon-β', *Immunity*, 38(6), pp. 1187-1197.

Kim, J.H., Lee, S.-R., Li, L.-H., Park, H.-J., Park, J.-H., Lee, K.Y., Kim, M.-K., Shin, B.A. and Choi, S.-Y. (2011) 'High Cleavage Efficiency of a 2A Peptide Derived from Porcine Teschovirus-1 in Human Cell Lines, Zebrafish and Mice', *PLoS ONE*, 6(4), p. e18556.

Kimmel, C.B., Ballard, W.W., Kimmel, S.R., Ullmann, B. and Schilling, T.F. (1995) 'Stages of embryonic development of the zebrafish', *Developmental Dynamics*, 203(3), pp. 253-310.

Kloosterman, W.P., Steiner, F.A., Berezikov, E., de Bruijn, E., van de Belt, J., Verheul, M., Cuppen, E. and Plasterk, R.H.A. (2006) 'Cloning and expression of new microRNAs from zebrafish', *Nucleic Acids Research*, 34(9), pp. 2558-2569.

Knee, R.S., Pitcher, S.E. and Murphy, P.R. (1994) 'Basic Fibroblast Growth Factor Sense (FGF) and Antisense (GFG) RNA Transcripts Are Expressed in Unfertilized Human Oocytes and in Differentiated Adult Tissues', *Biochemical and Biophysical Research Communications*, 205(1), pp. 577-583.

Knudsen, S. (2013) 'Promoter 2.0 Prediction Server'. 23 March 2016. Available at: http://www.cbs.dtu.dk/services/Promoter/.

Kok, Fatma O., Shin, M., Ni, C.-W., Gupta, A., Grosse, Ann S., van Impel, A., Kirchmaier, Bettina C., Peterson-Maduro, J., Kourkoulis, G., Male, I., DeSantis, Dana F., Sheppard-Tindell, S., Ebarasi, L., Betsholtz, C., Schulte-Merker, S., Wolfe, Scot A. and Lawson, Nathan D. (2015) 'Reverse Genetic Screening Reveals Poor Correlation between Morpholino-Induced and Mutant Phenotypes in Zebrafish', *Developmental Cell*, 32(1), pp. 97-108.

Koo, T. and Wood, M.J. (2013) 'Clinical trials using antisense oligonucleotides in duchenne muscular dystrophy', *Human Gene Therapy*, 24(5), pp. 479-88.

Kraus, P., Sivakamasundari, V., Lim, S.L., Xing, X., Lipovich, L. and Lufkin, T. (2013) 'Making sense of Dlx1 antisense RNA', *Developmental Biology*, 376(2), pp. 224-235.

Kumar, M. and Carmichael, G.G. (1997) 'Nuclear antisense RNA induces extensive adenosine modifications and nuclear retention of target transcripts', *Proceedings of*

the National Academy of Sciences of the United States of America, 94(8), pp. 3542-3547.

Lalonde, R. and Strazielle, C. (2003) 'The effects of cerebellar damage on maze learning in animals', *The Cerebellum*, 2(4), pp. 300-309.

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

Lavorgna, G., Dahary, D., Lehner, B., Sorek, R., Sanderson, C.M. and Casari, G. (2004) 'In search of antisense', *Trends in Biochemical Sciences*, 29(2), pp. 88-94.

Lee, Y., Ahn, C., Han, J., Choi, H., Kim, J., Yim, J., Lee, J., Provost, P., Radmark, O., Kim, S. and Kim, V.N. (2003) 'The nuclear RNase III Drosha initiates microRNA processing', *Nature*, 425(6956), pp. 415-419.

Lee, Y., Kim, M., Han, J., Yeom, K.-H., Lee, S., Baek, S.H. and Kim, V.N. (2004) 'MicroRNA genes are transcribed by RNA polymerase II', *The EMBO Journal*, 23(20), pp. 4051-4060.

Lehner, B., Williams, G., Campbell, R.D. and Sanderson, C.M. (2002) 'Antisense transcripts in the human genome', *Trends in Genetics*, 18(2), pp. 63-65.

Lepilina, A., Coon, A.N., Kikuchi, K., Holdway, J.E., Roberts, R.W., Burns, C.G. and Poss, K.D. (2006) 'A Dynamic Epicardial Injury Response Supports Progenitor Cell Activity during Zebrafish Heart Regeneration', *Cell*, 127(3), pp. 607-619.

Li, D., Li, X.P., Wang, H.X., Shen, Q.Y., Li, X.P., Wen, L., Qin, X.J., Jia, Q.L., Kung, H.F. and Peng, Y. (2012) 'VEGF induces angiogenesis in a zebrafish embryo glioma model established by transplantation of human glioma cells', *Oncology Reports* 28(3), pp. 937-42.

Li, Y.-X., Farrell, M.J., Liu, R., Mohanty, N. and Kirby, M.L. (2000) 'Double-Stranded RNA Injection Produces Null Phenotypes in Zebrafish', *Developmental Biology*, 217(2), pp. 394-405.

Lieschke, G.J. and Currie, P.D. (2007) 'Animal models of human disease: zebrafish swim into view', *Nature Reviews Genetics* 8(5), pp. 353-367.

Liu, A., Majumdar, A., Schauerte, H.E., Haffter, P. and Drummond, I.A. (2000) 'Zebrafish wnt4b expression in the floor plate is altered in sonic hedgehog and gli-2 mutants', *Mechanisms of Development*, 91(1–2), pp. 409-413.

Liu, J., Valencia-Sanchez, M.A., Hannon, G.J. and Parker, R. (2005) 'MicroRNA-dependent localization of targeted mRNAs to mammalian P-bodies', *Nature Cell Biology*, 7(7), pp. 719-723.

Lopes, S.S., Lourenço, R., Pacheco, L., Moreno, N., Kreiling, J. and Saúde, L. (2010) 'Notch signalling regulates left-right asymmetry through ciliary length control', *Development*, 137(21), pp. 3625-3632.

Lu, F.-I., Thisse, C. and Thisse, B. (2011) 'Identification and mechanism of regulation of the zebrafish dorsal determinant', *Proceedings of the National Academy of Sciences*, 108(38), pp. 15876-15880.

Lybecker, M., Zimmermann, B., Bilusic, I., Tukhtubaeva, N. and Schroeder, R. (2014) 'The double-stranded transcriptome of Escherichia coli', *Proceedings of the National Academy of Sciences*, 111(8), pp. 3134-3139.

Marquardt, S., Hazelbaker, D.Z. and Buratowski, S. (2011) 'Distinct RNA degradation pathways and 3' extensions of yeast non-coding RNA species', *Transcription*, 2(3), pp. 145-154.

Martin, H.C., Wani, S., Steptoe, A.L., Krishnan, K., Nones, K., Nourbakhsh, E., Vlassov, A., Grimmond, S.M. and Cloonan, N. (2014) 'Imperfect centered miRNA binding sites are common and can mediate repression of target mRNAs', *Genome Biology*, 15(3), p. R51.

Matouk, I.J., DeGroot, N., Mezan, S., Ayesh, S., Abu-lail, R., Hochberg, A. and Galun, E. (2007) 'The H19 Non-Coding RNA Is Essential for Human Tumor Growth', *PLoS ONE*, 2(9), p. e845.

Matranga, C., Tomari, Y., Shin, C., Bartel, D.P. and Zamore, P.D. (2005) 'Passenger-Strand Cleavage Facilitates Assembly of siRNA into Ago2-Containing RNAi Enzyme Complexes', *Cell*, 123(4), pp. 607-620.

Matsui, K., Nishizawa, M., Ozaki, T., Kimura, T., Hashimoto, I., Yamada, M., Kaibori, M., Kamiyama, Y., Ito, S. and Okumura, T. (2008) 'Natural antisense transcript

stabilizes inducible nitric oxide synthase messenger RNA in rat hepatocytes', *Hepatology*, 47(2), pp. 686-697.

Mattick, J.S. (2001) 'Non-coding RNAs: the architects of eukaryotic complexity', *EMBO reports*, 2(11), pp. 986-991.

Mencia, A., Modamio-Hoybjor, S., Redshaw, N., Morin, M., Mayo-Merino, F., Olavarrieta, L., Aguirre, L.A., del Castillo, I., Steel, K.P., Dalmay, T., Moreno, F. and Moreno-Pelayo, M.A. (2009) 'Mutations in the seed region of human miR-96 are responsible for nonsyndromic progressive hearing loss', *Nature Genetics*, 41(5), pp. 609-613.

Modarresi, F., Faghihi, M.A., Lopez-Toledano, M.A., Fatemi, R.P., Magistri, M., Brothers, S.P., van der Brug, M.P. and Wahlestedt, C. (2012) 'Inhibition of natural antisense transcripts in vivo results in gene-specific transcriptional upregulation', *Nature Biotechnology* 30(5), pp. 453-459.

Moens, C.B. and Prince, V.E. (2002) 'Constructing the hindbrain: Insights from the zebrafish', *Developmental Dynamics*, 224(1), pp. 1-17.

Monteleone, G., Neurath, M.F., Ardizzone, S., Di Sabatino, A., Fantini, M.C., Castiglione, F., Scribano, M.L., Armuzzi, A., Caprioli, F., Sturniolo, G.C., Rogai, F., Vecchi, M., Atreya, R., Bossa, F., Onali, S., Fichera, M., Corazza, G.R., Biancone, L., Savarino, V., Pica, R., Orlando, A. and Pallone, F. (2015) 'Mongersen, an Oral SMAD7 Antisense Oligonucleotide, and Crohn's Disease', *New England Journal of Medicine*, 372(12), pp. 1104-1113.

Morcos, P.A., Vincent, A.C. and Moulton, J.D. (2015) 'Gene Editing Versus Morphants', *Zebrafish*, 12(5), p. 319.

Moreno, P.M.D., Wenska, M., Lundin, K.E., Wrange, Ö., Strömberg, R. and Smith, C.I.E. (2009) 'A synthetic snRNA m3G-CAP enhances nuclear delivery of exogenous proteins and nucleic acids', *Nucleic Acids Research*, 37(6), pp. 1925-1935.

Morton, S.M. and Bastian, A.J. (2004) 'Cerebellar Control of Balance and Locomotion', *The Neuroscientist*, 10(3), pp. 247-259.

Much, C., Auchynnikava, T., Pavlinic, D., Buness, A., Rappsilber, J., Benes, V., Allshire, R. and O'Carroll, D. (2016) 'Endogenous Mouse Dicer Is an Exclusively Cytoplasmic Protein', *PLoS Genetics*, 12(6), p. e1006095.

Munroe, S.H. and Lazar, M.A. (1991) 'Inhibition of c-erbA mRNA splicing by a naturally occurring antisense RNA', *Journal of Biological Chemistry* 266(33), pp. 22083-6.

Murer, H., Forster, I. and Biber, J. (2004) 'The sodium phosphate cotransporter family SLC34', *Pflügers Archiv*, 447(5), pp. 763-767.

Najm, J., Horn, D., Wimplinger, I., Golden, J.A., Chizhikov, V.V., Sudi, J., Christian, S.L., Ullmann, R., Kuechler, A., Haas, C.A., Flubacher, A., Charnas, L.R., Uyanik, G., Frank, U., Klopocki, E., Dobyns, W.B. and Kutsche, K. (2008) 'Mutations of CASK cause an X-linked brain malformation phenotype with microcephaly and hypoplasia of the brainstem and cerebellum', *Nature Genetics*, 40(9), pp. 1065-1067.

Nakaya, H.I., Amaral, P.P., Louro, R., Lopes, A., Fachel, A.A., Moreira, Y.B., El-Jundi, T.A., da Silva, A.M., Reis, E.M. and Verjovski-Almeida, S. (2007) 'Genome mapping and expression analyses of human intronic noncoding RNAs reveal tissue-specific patterns and enrichment in genes related to regulation of transcription', *Genome Biology*, 8(3), pp. 1-25.

Nalbant, P., Boehmer, C., Dehmelt, L., Wehner, F. and Werner, A. (1999) 'Functional characterization of a Na+-phosphate cotransporter (NaPi-II) from zebrafish and identification of related transcripts', *Journal of Physiology*, 520 Pt 1, pp. 79-89.

Nasevicius, A. and Ekker, S.C. (2000) 'Effective targeted gene 'knockdown' in zebrafish', *Nature Genetics*, 26(2), pp. 216-20.

Natale, R., Blackhall, F., Kowalski, D., Ramlau, R., Bepler, G., Grossi, F., Lerchenmüller, C., Pinder-Schenck, M., Mezger, J., Danson, S., Gadgeel, S.M., Summers, Y., Callies, S., André, V., Das, M., Lahn, M. and Talbot, D. (2014) 'Evaluation of Antitumor Activity Using Change in Tumor Size of the Survivin Antisense Oligonucleotide LY2181308 in Combination with Docetaxel for Second-Line Treatment of Patients with Non–Small-Cell Lung Cancer: A Randomized Open-Label Phase II Study', *Journal of Thoracic Oncology*, 9(11), pp. 1704-1708.

NatureEducation (2014) 'Scitable'. 22 March 2016. Available at: http://www.nature.com/scitable/about.

Neeman, Y., Dahary, D., Levanon, E.Y., Sorek, R. and Eisenberg, E. (2005) 'Is there any sense in antisense editing?', *Trends in Genetics*, 21(10), pp. 544-7.

Nejepinska, J., Malik, R., Filkowski, J., Flemr, M., Filipowicz, W. and Svoboda, P. (2012) 'dsRNA expression in the mouse elicits RNAi in oocytes and low adenosine deamination in somatic cells', *Nucleic Acids Research*, 40(1), pp. 399-413.

Nejepinska, J., Malik, R., Wagner, S. and Svoboda, P. (2014) 'Reporters Transiently Transfected into Mammalian Cells Are Highly Sensitive to Translational Repression Induced by dsRNA Expression', *PLoS ONE*, 9(1), p. e87517.

Nornes, S., Newman, M., Wells, S., Verdile, G., Martins, R.N. and Lardelli, M. (2009) 'Independent and cooperative action of Psen2 with Psen1 in zebrafish embryos', *Experimental Cell Research*, 315(16), pp. 2791-2801.

Nusslein-Volhard, C. and Dahm, R. (2002) *Zebrafish*. Great Clarendon Street, Oxford: Oxford University Press.

O'Brien, L.L., Grimaldi, M., Kostun, Z., Wingert, R.A., Selleck, R. and Davidson, A.J. (2011) 'Wt1a, Foxc1a, and the Notch mediator Rbpj physically interact and regulate the formation of podocytes in zebrafish', *Developmental Biology*, 358(2), pp. 318-330.

Oates, A.C., Bruce, A.E.E. and Ho, R.K. (2000) 'Too Much Interference: Injection of Double-Stranded RNA Has Nonspecific Effects in the Zebrafish Embryo', *Developmental Biology*, 224(1), pp. 20-28.

Ohhata, T., Hoki, Y., Sasaki, H. and Sado, T. (2008) 'Crucial role of antisense transcription across the Xist promoter in Tsix-mediated Xist chromatin modification', *Development*, 135(2), pp. 227-235.

Orekhova, A.S. and Rubtsov, P.M. (2013) 'Bidirectional promoters in the transcription of mammalian genomes', *Biochemistry (Moscow)*, 78(4), pp. 335-341.

Pandey, R.R., Mondal, T., Mohammad, F., Enroth, S., Redrup, L., Komorowski, J., Nagano, T., Mancini-DiNardo, D. and Kanduri, C. (2008) 'Kcnq1ot1 Antisense

Noncoding RNA Mediates Lineage-Specific Transcriptional Silencing through Chromatin-Level Regulation', *Molecular Cell*, 32(2), pp. 232-246.

Panula, P., Chen, Y.C., Priyadarshini, M., Kudo, H., Semenova, S., Sundvik, M. and Sallinen, V. (2010) 'The comparative neuroanatomy and neurochemistry of zebrafish CNS systems of relevance to human neuropsychiatric diseases', *Neurobiology of Disease*, 40(1), pp. 46-57.

Parameswaran, P., Sklan, E., Wilkins, C., Burgon, T., Samuel, M.A., Lu, R., Ansel, K.M., Heissmeyer, V., Einav, S., Jackson, W., Doukas, T., Paranjape, S., Polacek, C., dos Santos, F.B., Jalili, R., Babrzadeh, F., Gharizadeh, B., Grimm, D., Kay, M., Koike, S., Sarnow, P., Ronaghi, M., Ding, S.-W., Harris, E., Chow, M., Diamond, M.S., Kirkegaard, K., Glenn, J.S. and Fire, A.Z. (2010) 'Six RNA Viruses and Forty-One Hosts: Viral Small RNAs and Modulation of Small RNA Repertoires in Vertebrate and Invertebrate Systems', *PLoS Pathology*, 6(2), p. e1000764.

Patel, P.H., Barbee, S.A. and Blankenship, J.T. (2016) 'GW-Bodies and P-Bodies Constitute Two Separate Pools of Sequestered Non-Translating RNAs', *PLoS ONE*, 11(3), p. e0150291.

Pauli, A., Valen, E., Lin, M.F., Garber, M., Vastenhouw, N.L., Levin, J.Z., Fan, L., Sandelin, A., Rinn, J.L., Regev, A. and Schier, A.F. (2012) 'Systematic identification of long noncoding RNAs expressed during zebrafish embryogenesis', *Genome Research*, 22(3), pp. 577-591.

Perry, A.K., Chen, G., Zheng, D., Tang, H. and Cheng, G. (2005) 'The host type I interferon response to viral and bacterial infections', *Cell Research*, 15(6), pp. 407-422.

Person, A.D., Beiraghi, S., Sieben, C.M., Hermanson, S., Neumann, A.N., Robu, M.E., Schleiffarth, J.R., Billington, C.J., Jr., van Bokhoven, H., Hoogeboom, J.M., Mazzeu, J.F., Petryk, A., Schimmenti, L.A., Brunner, H.G., Ekker, S.C. and Lohr, J.L. (2010) 'WNT5A mutations in patients with autosomal dominant Robinow syndrome', *Developmental Dynamics*, 239(1), pp. 327-37.

Pézeron, G., Lambert, G., Dickmeis, T., Strähle, U., Rosa, F.M. and Mourrain, P. (2008) 'Rasl11b Knock Down in Zebrafish Suppresses <italic>One-Eyed-Pinhead</italic> Mutant Phenotype', *PLoS ONE*, 3(1), p. e1434.

Piatek, M.J., Henderson, V., Zynad, H.S. and Werner, A. (2016) 'Natural antisense transcription from a comparative perspective', *Genomics*.

Piatek, Monica J. and Werner, A. (2014) 'Endogenous siRNAs: regulators of internal affairs', *Biochemical Society Transactions*, 42(4), pp. 1174-1179.

Plasterk, R.H.A. (2006) 'Micro RNAs in Animal Development', *Cell*, 124(5), pp. 877-881.

Pollard, S.M., Parsons, M.J., Kamei, M., Kettleborough, R.N.W., Thomas, K.A., Pham, V.N., Bae, M.-K., Scott, A., Weinstein, B.M. and Stemple, D.L. (2006) 'Essential and overlapping roles for laminin α chains in notochord and blood vessel formation', *Developmental Biology*, 289(1), pp. 64-76.

Poss, K.D., Wilson, L.G. and Keating, M.T. (2002) 'Heart Regeneration in Zebrafish', *Science*, 298(5601), pp. 2188-2190.

Postlethwait, J., Amores, A., Cresko, W., Singer, A. and Yan, Y.-L. (2004) 'Subfunction partitioning, the teleost radiation and the annotation of the human genome', *Trends in Genetics*, 20(10), pp. 481-490.

Prescott, E.M. and Proudfoot, N.J. (2002) 'Transcriptional collision between convergent genes in budding yeast', *Proceedings of the National Academy of Sciences*, 99(13), pp. 8796-8801.

Promega (2016) *Promega Restriction Enzyme Tool*. Available at: http://www.promega.com/a/apps/reTool/.

Proudfoot, N.J. (1986) 'Transcriptional interference and termination between duplicated [alpha]-globin gene constructs suggests a novel mechanism for gene regulation', *Nature*, 322(6079), pp. 562-565.

Rajan, K.S. and Ramasamy, S. (2014) 'Retrotransposons and piRNA: The missing link in central nervous system', *Neurochemistry International*, 77, pp. 94-102.

Rajesh K. Gaur, J.J.R. (2009) *Regulation of Gene Expression by Small RNAs*. CRC Press.

Rauch, G.J., Lyons, D.A., Middendorf, I., Friedlander, B., Arana, N., Reyes, T., and Talbot, W.S. (2003) 'Submission and Curation of Gene Expression Data - Rasl11b' *ZFIN Direct Data Submission* Available at: http://zfin.org.

Reese, M.G. (2015) 'Neural Network Promoter Prediction'. 23 March 2016. Available at: http://www.fruitfly.org/seq_tools/promoter.html.

Reynolds, A., Anderson, E.M., Vermeulen, A., Fedorov, Y., Robinson, K., Leake, D., Karpilow, J., Marshall, W.S. and Khvorova, A. (2006) 'Induction of the interferon response by siRNA is cell type– and duplex length–dependent', *RNA*, 12(6), pp. 988-993.

Robu, M.E., Larson, J.D., Nasevicius, A., Beiraghi, S., Brenner, C., Farber, S.A. and Ekker, S.C. (2007) 'p53 activation by knockdown technologies', *PLoS Genetics*, 3(5), p. e78.

Romaniello, R. and Borgatti, R. (2013) 'Cerebellar Agenesis', in Manto, M., Schmahmann, J.D., Rossi, F., Gruol, D.L. and Koibuchi, N. (eds.) *Handbook of the Cerebellum and Cerebellar Disorders*. Dordrecht: Springer Netherlands, pp. 1855-1872.

Saayman, S., Ackley, A., Turner, A.-M.W., Famiglietti, M., Bosque, A., Clemson, M., Planelles, V. and Morris, K.V. (2014) 'An HIV-Encoded Antisense Long Noncoding RNA Epigenetically Regulates Viral Transcription', *Molecular Therapy*, 22(6), pp. 1164-1175.

Sacilotto, N., Monteiro, R., Fritzsche, M., Becker, P.W., Sanchez-Del-Campo, L., Liu, K., Pinheiro, P., Ratnayaka, I., Davies, B., Goding, C.R., Patient, R., Bou-Gharios, G. and De Val, S. (2013) 'Analysis of Dll4 regulation reveals a combinatorial role for Sox and Notch in arterial development', *Proceedings of the National Academy of Sciences*, 110(29), pp. 11893-8.

Saitoh, K., Sado, T., Mayden, R.L., Hanzawa, N., Nakamura, K., Nishida, M. and Miya, M. (2006) 'Mitogenomic Evolution and Interrelationships of the Cypriniformes (Actinopterygii: Ostariophysi): The First Evidence Toward Resolution of Higher-Level Relationships of the World's Largest Freshwater Fish Clade Based on 59 Whole Mitogenome Sequences', *Journal of Molecular Evolution*, 63(6), pp. 826-841.

Salta, E. and De Strooper, B. (2012) 'Non-coding RNAs with essential roles in neurodegenerative disorders', *The Lancet Neurology*, 11(2), pp. 189-200.

Sambrook, J. and Russell, D.W. (2006) 'Purification of Nucleic Acids by Extraction with Phenol:Chloroform', *Cold Spring Harbor Protocols*, 2006(1), p. pdb.prot4455.

Scholpp, S. and Brand, M. (2001) 'Morpholino-induced knockdown of zebrafish engrailed genes eng2 and eng3 reveals redundant and unique functions in midbrain–hindbrain boundary development', *Genesis*, 30(3), pp. 129-133.

Schonborn, J., Oberstraβ, J., Breyel, E., Tittgen, J., Schumacher, J. and Lukacs, N. (1991) 'Monoclonal antibodies to double-stranded RNA as probes of RNA structure in crude nucleic acid extracts', *Nucleic Acids Research*, 19(11), pp. 2993-3000.

Schwarz, D.S., Hutvagner, G., Du, T., Xu, Z., Aronin, N. and Zamore, P.D. (2003) 'Asymmetry in the assembly of the RNAi enzyme complex', *Cell*, 115(2), pp. 199-208.

Schwarz, D.S., Hutvágner, G., Haley, B. and Zamore, P.D. (2002) 'Evidence that siRNAs Function as Guides, Not Primers, in the Drosophila and Human RNAi Pathways', *Molecular Cell*, 10(3), pp. 537-548.

Schyth, B.D. (2008) 'RNAi-mediated gene silencing in fishes?', *Journal of Fish Biology*, 72(8), pp. 1890-1906.

SCICIONS (2014) *Antibodies - J2*. Available at: https://scicons.eu/en/ (Accessed: 24 June).

Scott, M.S. and Ono, M. (2011) 'From snoRNA to miRNA: Dual function regulatory non-coding RNAs', *Biochimie*, 93(11), pp. 1987-1992.

Sellick, G.S., Barker, K.T., Stolte-Dijkstra, I., Fleischmann, C., J Coleman, R., Garrett, C., Gloyn, A.L., Edghill, E.L., Hattersley, A.T., Wellauer, P.K., Goodwin, G. and Houlston, R.S. (2004) 'Mutations in PTF1A cause pancreatic and cerebellar agenesis', *Nature Genetics*, 36(12), pp. 1301-1305.

Shearwin, K.E., Callen, B.P. and Egan, J.B. (2005) 'Transcriptional interference – a crash course', *Trends in Genetics*, 21(6), pp. 339-345.

Shu, J., Jelinek, J., Chang, H., Shen, L., Qin, T., Chung, W., Oki, Y. and Issa, J.P. (2006) 'Silencing of bidirectional promoters by DNA methylation in tumorigenesis', *Cancer Research*, 66(10), pp. 5077-84.

Sieger, D., Tautz, D. and Gajewski, M. (2003) 'The role of Suppressor of Hairless in Notch mediated signalling during zebrafish somitogenesis', *Mechanisms of Development*, 120(9), pp. 1083-1094.

Siekmann, A.F. and Lawson, N.D. (2007) 'Notch signalling limits angiogenic cell behaviour in developing zebrafish arteries', *Nature*, 445(7129), pp. 781-784.

Sigova, A.A., Mullen, A.C., Molinie, B., Gupta, S., Orlando, D.A., Guenther, M.G., Almada, A.E., Lin, C., Sharp, P.A., Giallourakis, C.C. and Young, R.A. (2013) 'Divergent transcription of long noncoding RNA/mRNA gene pairs in embryonic stem cells', *Proceedings of the National Academy of Sciences*, 110(8), pp. 2876-2881.

Silverman, T.A., Noguchi, M. and Safer, B. (1992) 'Role of sequences within the first intron in the regulation of expression of eukaryotic initiation factor 2 alpha', *Journal of Biological Chemistry*, 267(14), pp. 9738-9742.

Sioud, M. (2005) 'Induction of Inflammatory Cytokines and Interferon Responses by Double-stranded and Single-stranded siRNAs is Sequence-dependent and Requires Endosomal Localization', *Journal of Molecular Biology*, 348(5), pp. 1079-1090.

Sledz, C.A. and Williams, B.R.G. (2004) 'RNA interference and double-stranded-RNA-activated pathways', *Biochemical Society Transactions*, 32(6), pp. 952-956.

Sleutels, F., Zwart, R. and Barlow, D.P. (2002) 'The non-coding Air RNA is required for silencing autosomal imprinted genes', *Nature*, 415(6873), pp. 810-813.

Sneppen, K., Dodd, I.B., Shearwin, K.E., Palmer, A.C., Schubert, R.A., Callen, B.P. and Egan, J.B. (2005) 'A Mathematical Model for Transcriptional Interference by RNA Polymerase Traffic in Escherichia coli', *Journal of Molecular Biology*, 346(2), pp. 399-409.

SoftBerry, I. (2016) 'Softberry'. 23 March 2016. Available at: http://www.softberry.com/berry.phtml.

Stessl, M., Noe, C.R. and Winkler, J. (2012) 'Off-Target Effects and Safety Aspects of Phosphorothioate Oligonucleotides', in Erdmann, A.V. and Barciszewski, J. (eds.)

From Nucleic Acids Sequences to Molecular Medicine. Berlin, Heidelberg: Springer Berlin Heidelberg, pp. 67-83.

Stolle, K., Schnoor, M., Fuellen, G., Spitzer, M., Cullen, P. and Lorkowski, S. (2007) 'Cloning, genomic organization, and tissue-specific expression of the RASL11B gene', *Biochimica et Biophysica Acta (BBA) - Gene Structure and Expression*, 1769(7–8), pp. 514-524.

Storer, N.Y. and Zon, L.I. (2010) 'Zebrafish Models of p53 Functions', *Cold Spring Harbor Perspectives in Biology*, 2(8).

Su, J., Zhu, Z., Wang, Y., Xiong, F. and Zou, J. (2008) 'The Cytomegalovirus Promoter-Driven Short Hairpin RNA Constructs Mediate Effective RNA Interference in Zebrafish In Vivo', *Marine Biotechnology*, 10(3), pp. 262-269.

Sumanas, S., Zhang, B., Dai, R. and Lin, S. (2005) '15-Zinc finger protein Bloody Fingers is required for zebrafish morphogenetic movements during neurulation', *Developmental Biology*, 283(1), pp. 85-96.

Takeshima, Y., Yagi, M., Wada, H., Ishibashi, K., Nishiyama, A., Kakumoto, M., Sakaeda, T., Saura, R., Okumura, K. and Matsuo, M. (2006) 'Intravenous Infusion of an Antisense Oligonucleotide Results in Exon Skipping in Muscle Dystrophin mRNA of Duchenne Muscular Dystrophy', *Pediatric Research* 59(5), pp. 690-694.

Tan, L., Yu, J.-T., Hu, N. and Tan, L. (2013) 'Non-coding RNAs in Alzheimer's Disease', *Molecular Neurobiology*, 47(1), pp. 382-393.

Targett-Adams, P., Boulant, S. and McLauchlan, J. (2008) 'Visualization of Double-Stranded RNA in Cells Supporting Hepatitis C Virus RNA Replication', *Journal of Virology*, 82(5), pp. 2182-2195.

Taylor, J.A., Klemfuss, N.M. and Ivry, R.B. (2010) 'An Explicit Strategy Prevails When the Cerebellum Fails to Compute Movement Errors', *The Cerebellum*, 9(4), pp. 580-586.

Thatcher, E.J., Paydar, I., Anderson, K.K. and Patton, J.G. (2008) 'Regulation of zebrafish fin regeneration by microRNAs', *Proceedings of the National Academy of Sciences of the United States of America*, 105(47), pp. 18384-18389.

Thisse, B., Thisse, C. (2004a) 'Fast Release Clones: A High Throughput Expression Analysis' *ZFIN Direct Data Submission*. Available at: http://zfin.org.

Thisse, B., Thisse, C. (2004b) 'Fast Release Clones: A High Throughput Expression Analysis.' *ZFIN Direct Data Submission*. 29 March 2016. ZFIN. Available at: http://zfin.org.

Thisse, C., and Thisse, B. (2005) 'High Throughput Expression Analysis of ZF-Models Consortium Clones' *ZFIN Direct Data Submission* 29 March 2016. Available at: http://zfin.org.

Townend, J. (2002) *Practical Statistics for Environmental and Biological Scientists*. 1 edn. Chichester, United Kingdom John Wiley and Sons Ltd

Tran, S., Nowicki, M., Muraleetharan, A., Chatterjee, D. and Gerlai, R. (2016) 'Neurochemical factors underlying individual differences in locomotor activity and anxiety-like behavioral responses in zebrafish', *Progress in Neuro-Psychopharmacology and Biological Psychiatry*, 65, pp. 25-33.

Trinklein, N.D., Aldred, S.F., Hartman, S.J., Schroeder, D.I., Otillar, R.P. and Myers, R.M. (2004) 'An Abundance of Bidirectional Promoters in the Human Genome', *Genome Research*, 14(1), pp. 62-66.

Tufarelli, C., Frischauf, A.-M., Hardison, R., Flint, J. and Higgs, D.R. (2001) 'Characterization of a Widely Expressed Gene (LUC7-LIKE; LUC7L) Defining the Centromeric Boundary of the Human α-Globin Domain', *Genomics*, 71(3), pp. 307-314.

Tufarelli, C., Stanley, J.A., Garrick, D., Sharpe, J.A., Ayyub, H., Wood, W.G. and Higgs, D.R. (2003) 'Transcription of antisense RNA leading to gene silencing and methylation as a novel cause of human genetic disease', *Nature Genetics*, 34(2), pp. 157-65.

Tzanetakis, I.E. and Martin, R.R. (2008) 'A new method for extraction of double-stranded RNA from plants', *Journal of Virological Methods*, 149(1), pp. 167-170.

Uchida, T., Rossignol, F., Matthay, M.A., Mounier, R., Couette, S., Clottes, E. and Clerici, C. (2004) 'Prolonged hypoxia differentially regulates hypoxia-inducible factor (HIF)-1alpha and HIF-2alpha expression in lung epithelial cells: implication of natural antisense HIF-1alpha', *Journal of Biological Chemistry* 279(15), pp. 14871-8.

Uesaka, M., Nishimura, O., Go, Y., Nakashima, K., Agata, K. and Imamura, T. (2014) 'Bidirectional promoters are the major source of gene activation-associated non-coding RNAs in mammals', *BMC Genomics*, 15, p. 35.

Ulitsky, I., Shkumatava, A., Jan, Calvin H., Sive, H. and Bartel, David P. (2011) 'Conserved Function of lincRNAs in Vertebrate Embryonic Development despite Rapid Sequence Evolution', *Cell*, 147(7), pp. 1537-1550.

Vance, K.W., Sansom, S.N., Lee, S., Chalei, V., Kong, L., Cooper, S.E., Oliver, P.L. and Ponting, C.P. (2014) 'The long non-coding RNA Paupar regulates the expression of both local and distal genes', *EMBO Journal*, 33(4), pp. 296-311.

Vanhée-Brossollet, C., Thoreau, H., Serpente, N., D'Auriol, L., Lévy, J.-P. and Vaquero, C. (1995) 'A natural antisense RNA derived from the HIV-1 env gene encodes a protein which is recognized by circulating antibodies of HIV+ individuals', *Virology*, 206(1), pp. 196-202.

Velioğlu, S.K., Kuzeyli, K. and Özmenoğlu, M. (1998) 'Cerebellar agenesis: a case report with clinical and MR imaging findings and a review of the literature', *European Journal of Neurology*, 5(5), pp. 503-506.

Velmeshev, D., Magistri, M. and Faghihi, M.A. (2013) 'Expression of non-protein-coding antisense RNAs in genomic regions related to autism spectrum disorders', *Molecular Autism*, 4(1), pp. 1-12.

Venter, J.C., Adams, M.D., Myers, E.W., Li, P.W., Mural, R.J., Sutton, G.G., Smith, H.O., Yandell, M., Evans, C.A., Holt, R.A., Gocayne, J.D., Amanatides, P., Ballew, R.M., Huson, D.H., Wortman, J.R., Zhang, Q., Kodira, C.D., Zheng, X.H., Chen, L., Skupski, M., Subramanian, G., Thomas, P.D., Zhang, J., Gabor Miklos, G.L., Nelson, C., Broder, S., Clark, A.G., Nadeau, J., McKusick, V.A., Zinder, N., Levine, A.J., Roberts, R.J., Simon, M., Slayman, C., Hunkapiller, M., Bolanos, R., Delcher, A., Dew, I., Fasulo, D., Flanigan, M., Florea, L., Halpern, A., Hannenhalli, S., Kravitz, S., Levy, S., Mobarry, C., Reinert, K., Remington, K., Abu-Threideh, J., Beasley, E., Biddick, K., Bonazzi, V., Brandon, R., Cargill, M., Chandramouliswaran, I., Charlab, R., Chaturvedi, K., Deng, Z., Francesco, V.D., Dunn, P., Eilbeck, K., Evangelista, C., Gabrielian, A.E., Gan, W., Ge, W., Gong, F., Gu, Z., Guan, P., Heiman, T.J., Higgins, M.E., Ji, R.-R., Ke, Z., Ketchum, K.A., Lai, Z., Lei, Y., Li, Z., Li, J., Liang, Y., Lin, X., Lu, F., Merkulov, G.V., Milshina, N., Moore, H.M., Naik, A.K., Narayan, V.A.,

Neelam, B., Nusskern, D., Rusch, D.B., Salzberg, S., Shao, W., Shue, B., Sun, J., Wang, Z.Y., Wang, A., Wang, X., Wang, J., Wei, M.-H., Wides, R., Xiao, C., Yan, C., et al. (2001) 'The Sequence of the Human Genome', *Science*, 291(5507), pp. 1304-1351.

Verkade, H. and Heath, J.K. (2009) 'Wnt Signaling Mediates Diverse Developmental Processes in Zebrafish', in Vincan, E. (ed.) *Wnt Signaling*. Totowa, NJ: Humana Press, pp. 225-251.

Vitali, P. and Scadden, A.D.J. (2010) 'Double-stranded RNAs containing multiple IU pairs are sufficient to suppress interferon induction and apoptosis', *Nature Structural & Molecular Biology* 17(99), pp. 1043-1050.

Wang, L., Jiang, N., Wang, L., Fang, O., Leach, L.J., Hu, X. and Luo, Z. (2014) '3′ Untranslated Regions Mediate Transcriptional Interference between Convergent Genes Both Locally and Ectopically in <italic>Saccharomyces cerevisiae</italic>', *PLoS Genetics*, 10(1), p. e1004021.

Wang, Y., Joh, K., Masuko, S., Yatsuki, H., Soejima, H., Nabetani, A., Beechey, C.V., Okinami, S. and Mukai, T. (2004) 'The Mouse Murr1 Gene Is Imprinted in the Adult Brain, Presumably Due to Transcriptional Interference by the Antisense-Oriented U2af1-rs1 Gene', *Molecular and Cellular Biology*, 24(1), pp. 270-279.

Washietl, S., Hofacker, I.L., Lukasser, M., Huttenhofer, A. and Stadler, P.F. (2005) 'Mapping of conserved RNA secondary structures predicts thousands of functional noncoding RNAs in the human genome', *Nature Biotechnology* 23(11), pp. 1383-1390.

Watanabe, T., Totoki, Y., Toyoda, A., Kaneda, M., Kuramochi-Miyagawa, S., Obata, Y., Chiba, H., Kohara, Y., Kono, T., Nakano, T., Surani, M.A., Sakaki, Y. and Sasaki, H. (2008) 'Endogenous siRNAs from naturally formed dsRNAs regulate transcripts in mouse oocytes', *Nature*, 453(7194), pp. 539-543.

Waterhouse, P.M., Wang, M.-B. and Finnegan, E.J. (2001) 'Role of short RNAs in gene silencing', *Trends in Plant Science*, 6(7), pp. 297-301.

Weber, F., Wagner, V., Rasmussen, S.B., Hartmann, R. and Paludan, S.R. (2006) 'Double-Stranded RNA Is Produced by Positive-Strand RNA Viruses and DNA

Viruses but Not in Detectable Amounts by Negative-Strand RNA Viruses', *Journal of Virology*, 80(10), pp. 5059-5064.

Wei, W., Pelechano, V., Jarvelin, A.I. and Steinmetz, L.M. (2011) 'Functional consequences of bidirectional promoters', *Trends in Genetics*, 27(7), pp. 267-76.

Werner, A. and Kinne, R.K. (2001) 'Evolution of the Na-P(i) cotransport systems', *American Journal Physiology: Regulatory Integrative Comparative Physiology* 280(2), pp. R301-12.

Werner, A., Schmutzler, G., Carlile, M., Miles, C.G. and Peters, H. (2007) 'Expression profiling of antisense transcripts on DNA arrays', *Physiological Genomics*, 28(3), pp. 294-300.

Westerfield, M. (1993) *The Zebrafish Book*. Eugene, OR: Oregon Press.

White, E., Schlackow, M., Kamieniarz-Gdula, K., Proudfoot, N.J. and Gullerova, M. (2014) 'Human nuclear Dicer restricts the deleterious accumulation of endogenous double-stranded RNA', *Nature Structural & Molecular Biology* 21(6), pp. 552-559.

Wienholds, E., Koudijs, M.J., van Eeden, F.J., Cuppen, E. and Plasterk, R.H. (2003) 'The microRNA-producing enzyme Dicer1 is essential for zebrafish development', *Nature Genetics*, 35(3), pp. 217-8.

Wight, M. and Werner, A. (2013) 'The functions of natural antisense transcripts', *Essays Biochem*, 54, pp. 91-101.

Wilusz, J.E., Sunwoo, H. and Spector, D.L. (2009) 'Long noncoding RNAs: functional surprises from the RNA world', *Genes & Development*, 23(13), pp. 1494-1504.

Xia, J., Joyce, C.E., Bowcock, A.M. and Zhang, W. (2013) 'Noncanonical microRNAs and endogenous siRNAs in normal and psoriatic human skin', *Human Molecular Genetics*, 22(4), pp. 737-748.

Xu, C., Chen, J. and Shen, B. (2012) 'The preservation of bidirectional promoter architecture in eukaryotes: what is the driving force?', *BMC Systems Biology*, 6 Suppl 1, p. S21.

Xue, Z., Ye, Q., Anson, S.R., Yang, J., Xiao, G., Kowbel, D., Glass, N.L., Crosthwaite, S.K. and Liu, Y. (2014) 'Transcriptional interference by antisense RNA is required for circadian clock function', *Nature*, 514(7524), pp. 650-3.

Yadav, R.P. and Kotaja, N. (2014) 'Small RNAs in spermatogenesis', *Molecular and Cellular Endocrinology*, 382(1), pp. 498-508.

Yamanaka, Y., Faghihi, Mohammad A., Magistri, M., Alvarez-Garcia, O., Lotz, M. and Wahlestedt, C. (2015) 'Antisense RNA Controls LRP1 Sense Transcript Expression through Interaction with a Chromatin-Associated Protein, HMGB2', *Cell Reports*, 11(6), pp. 967-976.

Yan, X., Liang, H., Deng, T., Zhu, K., Zhang, S., Wang, N., Jiang, X., Wang, X., Liu, R., Zen, K., Zhang, C.-Y., Ba, Y. and Chen, X. (2013) 'The identification of novel targets of miR-16 and characterization of their biological functions in cancer cells', *Molecular Cancer*, 12, pp. 92-92.

Yen, J., White, R.M. and Stemple, D.L. (2014) 'Zebrafish models of cancer: progress and future challenges', *Current Opinion in Genetics & Development*, 24, pp. 38-45.

Youngblood, D.S., Hatlevig, S.A., Hassinger, J.N., Iversen, P.L. and Moulton, H.M. (2007) 'Stability of cell-penetrating peptide-morpholino oligomer conjugates in human serum and in cells', *Bioconjug Chem*, 18(1), pp. 50-60.

Yu, F., Jiang, Q.-j., Sun, X.-y. and Zhang, R.-w. (2014) 'A new case of complete primary cerebellar agenesis: clinical and imaging findings in a living patient', *Brain*.

Yu, R.M.K., Lin, C.C., Chan, P.K., Chow, E.S.H., Murphy, M.B., Chan, B.P., Müller, F., Strähle, U. and Cheng, S.H. (2006) 'Four-Dimensional Imaging and Quantification of Gene Expression in Early Developing Zebrafish (Danio rerio) Embryos', *Toxicological Sciences*, 90(2), pp. 529-538.

Yu, W., Gius, D., Onyango, P., Muldoon-Jacobs, K., Karp, J., Feinberg, A.P. and Cui, H. (2008) 'Epigenetic silencing of tumour suppressor gene p15 by its antisense RNA', *Nature*, 451(7175), pp. 202-206.

Yuan, S., Schuster, A., Tang, C., Yu, T., Ortogero, N., Bao, J., Zheng, H. and Yan, W. (2016) 'Sperm-borne miRNAs and endo-siRNAs are important for fertilization and preimplantation embryonic development', *Development*, 143(4), pp. 635-47.

Yuan, S. and Sun, Z. (2009) 'Microinjection of mRNA and morpholino antisense oligonucleotides in zebrafish embryos', *Journal of Visualized Experiments* (27).

Zangger, H., Ronet, C., Desponds, C., Kuhlmann, F.M., Robinson, J., Hartley, M.-A., Prevel, F., Castiglioni, P., Pratlong, F., Bastien, P., Müller, N., Parmentier, L., Saravia, N.G., Beverley, S.M. and Fasel, N. (2013) 'Detection of Leishmania RNA Virus in Leishmania Parasites', *PLOS Neglected Tropical Diseases* 7(1), p. e2006.

ZFIN (2002) 'Gene Ontology Annotation Through Association of InterPro Records with GO Terms. Automated Data Submission '.

Zhang, Z. and Carmichael, G.G. (2001) 'The Fate of dsRNA in the Nucleus: A p54nrb-Containing Complex Mediates the Nuclear Retention of Promiscuously A-to-I Edited RNAs', *Cell*, 106(4), pp. 465-476.

Zhao, Z., Cao, Y., Li, M. and Meng, A. (2001) 'Double-Stranded RNA Injection Produces Nonspecific Defects in Zebrafish', *Developmental Biology*, 229(1), pp. 215-223.

Zong, X., Nakagawa, S., Freier, S.M., Fei, J., Ha, T., Prasanth, S.G. and Prasanth, K.V. (2016) 'Natural antisense RNA promotes 3′ end processing and maturation of MALAT1 lncRNA', *Nucleic Acids Research*, 44(6), pp. 2898-2908.

Appendix A

Sequence similarity and alignments of Danio rerio Slc34a2a, Slc34a2a(as) and Slc34a2b.

1.0 EMBOSS matcher: pairwise nucleotide sequence alignment of Slc34a2a and paralog Slc34a2b (Waterman-Eggert local alignment)

68.4% nucleotide sequence similarity (1169/1710)

```
Slc34a2b        146 GAGGTGGACCCGTGGGAGCTGCCGGAGCTCCTGGACACCGGGGTGAAGTG    195
                    ||||..||.||.||||||..||..|||||||.|.|||||||||||||||.||
Slc34a2a        366 GAGGATGATCCCTGGGAAATGATGGAGCTGCAGGACACCGGGGTGAAATG    415

Slc34a2b        196 GTCAGAGTTGGACCGGCGCGGAAAGGTCCTGCGGGTCTGCACCTCTGTCC    245
                    |.|||||.|||||.........|||||||||.|.||.|..||..|.|...
Slc34a2a        416 GGCAGATCTGGACACTAAAAAGAAGGTCCTGAGAGTTTTTACTACAGCAG    465

Slc34a2b        246 TGAAGTTGCTGCTGCTGCTCGGCCTGCTCTACATGTTCGTCTGCTCCCTG    295
                    ..||..||.|..|||||||.||..|||||||||||||.||.||||||.|||
Slc34a2a        466 CAAAACTGATCATGCTGCTTGGTTTGCTCTACATGTTTGTTTGCTCTCTG    515

Slc34a2b        296 GACATCCTCAGCTCTGCTTTCCAGCTGGTCGGAGGGAAGGCAGCGGGGGA    345
                    |||.||||.||||||||||||||.||.||||||.||.||||||.|||||
Slc34a2a        516 GACGTCCTAAGCTCAGCTTTCCAGCTTGTTGGAGGAAAAGCAGCAGGGGA    565

Slc34a2b        346 CATCTTTAAAGATAATGCGGTGCTGTCCAACCCTGTGGCCGGGCTGGTGA    395
                    |||.||..|.||.||||..||||||.||.||.|||.||||.||.||||||.|
Slc34a2a        566 CATTTTCCAGGAGAATAAGGTGCTCTCAAATCCTTTGGCAGGCCTGGTTA    615

Slc34a2b        396 TCGGGGTGCTGGTCACGGTTCTAGTCCAGAGCTCCAGCACCTCCTCCTCC    445
                    |||||.||||||||||..|..||||||||||.||||||||.||||||.||.
Slc34a2a        616 TCGGGATGCTGGTCACTCTGTTAGTCCAGAGTTCCAGCACATCCTCATCT    665

Slc34a2b        446 ATCGTGGTCAGCATGGTATCGTCTGGATTGCTGGAGGTGAAGTCAGCAGT    495
                    ||.||||||||||||||.||.|||||||.||||||||||.....|.||.||
Slc34a2a        666 ATTGTGGTCAGCATGGTCTCTTCTGGAATGCTGGAGGTCGCAACTGCTGT    715

Slc34a2b        496 GCCTATCATCATGGGTGCAAATATCGGCACTTCAGTAACAAACACTATTG    545
                    .||.||||||.||||||..||||.||||||||.||.||.||.|||||||.|||
Slc34a2a        716 TCCCATCATTATGGGCACAAACATCGGCACATCCGTCACCAACACTCTTG    765

Slc34a2b        546 TGGCCGTGATGCAAGCTGGAGACCGCAATGAGTTTCGCAGGGCTTTTGCG    595
                    |.|||.|....||||.|||.|||.|.||..||||.||.||||||.||||||.
Slc34a2a        766 TAGCCATTGCACAAGTTGGGGACAGGAACAAGTTCCGTAGGGCATTTGCA    815

Slc34a2b        596 GGGGCGACGGTGCATGACTTCTTCAACTGGCTGTCGGTGCTGGTGCTCCT    645
                    ||.||.||.||||||||.||||||.||||||||.||.||.||.||||||.||
Slc34a2a        816 GGAGCCACTGTGCATGATTTCTTTAACTGGCTCTCAGTCCTAGTGCTACT    865

Slc34a2b        646 GCCTCTGGAAGTGGCCTCAGGGATGCTCTACAGACTCACCAAACTTGTGA    695
                    ||||||||||||||||||.||....||..|.|...|.||.|..||..|..
Slc34a2a        866 GCCTCTGGAAGTGGCCTCGGGTTACCTGGAGAAGGTTACGAGCCTCATCG    915
```

```
Slc34a2b         696 TCGATTCCTTCAACATCCAGACGGGCGCAGACGCTCCAGACCTGCTGAAA    745
                     |....||.||.||||||.|||..||.|.|.|.||.||||..||.||.||.
Slc34a2a         916 TGAGATCGTTTAACATCGAGAGTGGAGAAAAAGCACCAGCTCTCCTAAAT    965

Slc34a2b         746 GTCATCACAGAGCCGCTCACCAAGAACATCATTGAGCTGGACACCTCTGT    795
                     ||.||||||.||.||||||.|||.|....||||||.||||||||...|||||
Slc34a2a         966 GTTATCACCGATCCGCTTACCCACTCAATCATTCAGCTGGACGAGTCTGT   1015

Slc34a2b         796 GATTCGTGACATCGCCACTGGAGACCCTGCTGCCCGGAACAAAAGTTTGA    845
                     |||..|.|.|||.||...|||.||.||.|..|||.|.||.|||..|.|.|
Slc34a2a        1016 GATGAGCGGCATTGCGGTTGGCGATCCCGAAGCCAGAAATAAATCTCTTA   1065

Slc34a2b         846 TCAAGATCTGGTGTAAAACGGAAAAAATTACGAACCTGGTGAACGTGACC    895
                     |||||.|.||||||..|.||.|.....|..||||...|...|||.||||||
Slc34a2a        1066 TCAAGGTTTGGTGCCACACAGCCTCGAACACGACTGTCCAGAATGTGACC   1115

Slc34a2b         896 GTCCCGGGAATCGCAAACTGCACTCCGGACGCCCTGTGCTGGGTGGATGG    945
                             |...||||||||||      .|..||.||.||||
Slc34a2a        1116 ---------ACAACAAACTGCAC------AGATCTTTGTTGGG-------   1143

Slc34a2b         946 AGACTTGATCTGGACCCAAAAGAACCAAACCGATACCATCTACTTGAAGA    995
                       |.||||            ||||....||.||.|..||..|..|.||.|
Slc34a2a        1144 --AATTGA------------AGAATGTTACAGAAATTATAAATATTAAAA   1179

Slc34a2b         996 AATGTACGCACATGTTCGTGTTTGCGGATCTGCCTGATCTGGCGGTGGGT   1045
                     ||||.|..|||||.||||||....||...||..|.||.||||||||||||
Slc34a2a        1180 AATGCAGTCACATCTTCGTGAACACGAGCCTTTCGGACCTGGCGGTGGGT   1229

Slc34a2b        1046 TTGATCCTGCTGGCTCTGTCTCTGCTGGCTCTCTGCACATGCCTGATCCT   1095
                     .||||.||||||||....||||||||...|||.|||||.|||||.||...
Slc34a2a        1230 CTGATTCTGCTGGCGGGTTCTCTGCTCATTCTGTGCACTTGCCTCATTTG   1279

Slc34a2b        1096 CATAGTTAAACTCCTGAACTCCATGCTGAAGGGTCAGGTGGCCGTCGTCA   1145
                     .||.||.||.||.|||||.||.||||||.|||||.|||||.||.||.||.|
Slc34a2a        1280 TATTGTGAAGCTGCTGAATTCGATGCTCAAGGGACAGGTTGCTGTTGTGA   1329

Slc34a2b        1146 TCAAAAAAGTGCTCAACACAGATTTCCCCTTCCCCTTCGCCTGGGTTACA   1195
                     ||||.||..|..|.||||||||||||||.||.||.||.||.|||.|.||.
Slc34a2a        1330 TCAAGAAGATAGTGAACACAGATTTCCCTTTTCCATTTGCTTGGCTGACT   1379

Slc34a2b        1196 GGATACCTGGCCATCCTCGTGGGAGCCGGAATGACCTTCATTGTTCAGAG   1245
                     ||||||..|.||.||.||.||.||.||.||||||||.||.|||||.|||||
Slc34a2a        1380 GGATATATTGCAATTCTGGTTGGTGCGGGAATGACTTTTATTGTCCAGAG   1429

Slc34a2b        1246 CAGCTCCGTCTTCACCTCCGCCATCACTCCTCTAGTCGGTATTGGTGTGA   1295
                     |||.||||||||||||.||.||.||.||||||||.||.|||||.|||||.|
Slc34a2a        1430 CAGTTCCGTCTTCACATCGGCTATAACTCCTCTTGTTGGTATCGGTGTTA   1479

Slc34a2b        1296 TCAGTTTGGAGAGAGCATATCCACTCACACTGGGCTCCAACATCGGCACA   1345
                     |.|||.|.||.||.||||||||.||..|.|||||.|||||.||.||.|||
Slc34a2a        1480 TAAGTATTGAAAGGGCATATCCTCTATCCCTGGGATCCAATATTGGAACA   1529

Slc34a2b        1346 ACCACCACTGCTATACTGGCGGCCATGGCCAGTCCTGGAGAAACACTGGC   1395
                     ||.||.||||||.|||.|.||.||||||||.||||||||.||||||||.|.
Slc34a2a        1530 ACTACTACTGCAATATTAGCCGCCATGGCTAGTCCTGGTGAAACACTTGG   1579

Slc34a2b        1396 CAACTCGCTACAGATTTCTCTGTGTCACTTCTTTTTCAACATCGCTGGTA   1445
                     .||||||.||||||||.|.||..||.||||||||||.||||||.|..|.||.|
Slc34a2a        1580 AAACTCATTACAGATTGCATTGGTTCACTTCTTCTTCAACTTATCCGGGA   1629

Slc34a2b        1446 TCCTGCTGTGGTACCCCATCCCCTTCACACGTGTGCCCATCCGACTGGCC   1495
                     |..||.||||||||||.||||||.||||||||.|.|||||||||||||||.
Slc34a2a        1630 TATTGTTGTGGTACCCAATCCCCATCACACGTATCCCCATCCGACTGGCA   1679

Slc34a2b        1496 AAAGCGCTGGGCAACCGTACAGCTAAATACCGCTGGTTCGCGGGCCTGTA   1545
                     ||.|..||.||..|....||||||..|.||||||.||||.||.|.|.|.||
Slc34a2a        1680 AAGGGTCTTGGTGAAACAACAGCCCAGTACCGTTGGTTTGCAGCCTTTTA   1729

Slc34a2b        1546 TCTCATTCTCTGCTTTCTGCTGTTTCCGTTGACCG-TCCTGGGACTCTCC   1594
                     ..||||.||.|||||  ||.|.||||.|.|..||.|  ||.|.||.|||||.
Slc34a2a        1730 CATCATCCTTTGCTT-CTTCGGTTTGCCTCTACTGGTCTTTGGTCTCTCT   1778
```

```
Slc34a2b        1595 ATTGCAGGATGGCAGGTCCTGGTGGGCATCGGGGTGCCAGTGTTGGTGCT   1644
                     ||.||.||.||||||||.||..||||..|....||.||..|...|||..|
Slc34a2a        1779 ATGGCCGGCTGGCAGGTACTCATGGGAGTTCTTGTCCCCATAGCGGTTAT   1828

Slc34a2b        1645 CGCCATCTTTGTGATTGTGGTGAACGTCATGCAGAAACGATGCCCTCGCT   1694
                     ....|||||.|..||..|.||.|||.|..|||||||.|.....||||..|
Slc34a2a        1829 TCTGATCTTCGCCATCATTGTCAACATTCTGCAGAAGCACAAACCTCAAT   1878

Slc34a2b        1695 TTCTGCCCTCGTTCATCCGCAGTTGGGAGTTTCTGCCCAAACCGCTGCAC   1744
                     ..||.||.||.....||||...||||||.||.||.||...|..|...|||
Slc34a2a        1879 GGCTTCCTTCTGCTCTCCGGTCTTGGGATTTCCTTCCTCTATGGGCTCAC   1928

Slc34a2b        1745 TCTCTGAAGCCCTGGGACCGAGTGGTGACTGCGGGAATGAGCTTCTGCAG   1794
                     |||||..|.||.||||||.||||.||.||||.         |.| |||.|
Slc34a2a        1929 TCTCTTGATCCATGGGACAGAGTAGTCACTGT---------CAT-TGCCG   1968

Slc34a2b        1795 GACTCGCTGCTGCTGCTGCTGTAAATGCTGCAGAAATGAAGAGAAAAACC   1844
                       |.|||||.|||||.|||||.||.|||||||....|.|.||..||.|..
Slc34a2a        1969 --CCCGCTGTTGCTGTTGCTGCAAGTGCTGCAATTCTAATGAAGAAGATG   2016

Slc34a2b        1845 ACATGGAGAA   1854
                     |.|.||.|||
Slc34a2a         2017 AGAAGGCGAA   2026
```

2.0 CLUSTAL O(1.2.1) multiple sequence alignment

1: Slc34a2b	100.00	67.45	55.85
2: Slc34a2a	67.45	100.00	72.41
3: Antisense	55.85	72.41	100.00

```
Slc34a2b        ------------------------------------------------------------
Slc34a2a        agccactctgaagaacagcacagctgagtctggatctgaacgtggtctccatcagcacca
Antisense       ------------------------------------------------------------


Slc34a2b        ------------------------------------------------------------
Slc34a2a        gtgcttctgctcaggccaggactgaaggaaacagaagtgcatagagttcaggaaagactg
Antisense       ------------------------------------------------------------


Slc34a2b        ------------------------------------------------------------
Slc34a2a        ttggaggagagatacgcttcagactgtagataagtcactgagctcaaaagcaggtgaaac
Antisense       ------------------------------------------------------------
```

```
Slc34a2b      ------------------------------------caggc----------tgatggctc
Slc34a2a      gtagagtgaacagaagtgttgcacgtgaaattcagaagggcagaagcaacttgattacat
Antisense     ------------------------------------------------------------


Slc34a2b      ctctggctgagactcatcccgcctcgc-cgccaccggaactcggagcggacgcggacaaa
Slc34a2a      ctcggacaatggcacc-------acgtccaaagcatgagcatgaatctgacgaaaaacaa
Antisense     ------------------------------------------------------------


Slc34a2b      catgagccccaaacc----ccgacgcgcaccctcagtccctggcccgcgcagcagtggga
Slc34a2a      ccagaaacacttgacggtgccagaaagaagtctctgtctatggctccggctgtttctaca
Antisense     ------------------------------------------------------------


Slc34a2b      gccggagcaggaggaagaggtggacccgtgggagctgccggagctcctggacaccggggt
Slc34a2a      -gcggct---ctgattgaggatgatccctgggaaatgatggagctgcaggacaccggggt
Antisense     ------------------------------------------------------------


Slc34a2b      gaagtggtcagagttggaccggcgcggaaaggtcctgcgggtctgcacctctgtcctgaa
Slc34a2a      gaaatgggcagatctggacactaaaaagaaggtcctgagagtttttactacagcagcaaa
Antisense     ------------------------------------------------------------


Slc34a2b      gttgctgctgctgctcggcctgctctacatgttcgtctgctccctggacatcctcagctc
Slc34a2a      actgatcatgctgcttggtttgctctacatgtttgtttgctctctggacgtcctaagctc
Antisense     ------------------------------------------------------------


Slc34a2b      tgctttccagctggtcggagggaaggcagcgggggacatctttaaagataatgcggtgct
Slc34a2a      agctttccagcttgttggaggaaaagcagcaggggacattttccaggagaataaggtgct
Antisense     ------------------------------------------------------------


Slc34a2b      gtccaaccctgtggccgggctggtgatcggggtgctggtcacggttctagtccagagctc
Slc34a2a      ctcaaatcctttggcaggcctggttatcgggatgctggtcactctgttagtccagagttc
Antisense     ------------------------------------------------------------


Slc34a2b      cagcacctcctcctccatcgtggtcagcatggtatcgtctggattgctggaggtgaagtc
Slc34a2a      cagcacatcctcatctattgtggtcagcatggtctcttctggaatgctggaggtcgcaac
Antisense     ------------------------------------------------------------


Slc34a2b      agcagtgcctatcatcatgggtgcaaatatcggcacttcagtaacaaacactatt-----
Slc34a2a      tgctgttcccatcattatgggcacaaacatcggcacatccgtcaccaaca--ctctt---
Antisense     -------------------------------------------acatacacacacatttt
                                                          **.:***  .:


Slc34a2b      ---gtggccgtgatgc--aagctggagaccgcaatg--agttt----cgcagggcttttg
Slc34a2a      ---gtagccat--tgcacaagttggggacaggaacaa--gttccgta----gggcatttg
Antisense     aataaagccacattaaacagtatttgatctaaaacagaaatacagtatatatgtaactta
                 .:.***.   *..  *.  *  ..:* . ** .   .*:          * .: **.


Slc34a2b      cggg------ggc--gacggtgcatg---------acttcttcaactggctgtcggtgct
Slc34a2a      cag------gagcc--actgtgcatg---------atttctttaactggctctcagtcct
Antisense     cagtaataaaggccgtactgtgtattaccatcataactactaataaggacttttata--t
              *.*       .**   ** *** **          * *:**: :*. *.** * . :  *


Slc34a2b      ggtgctcctgcctctggaagtggcctcagggatgctctaca---------gactcaccaa
Slc34a2a      agtgctactgcctctggaagtggcctcgggttacctggagaaggttacgagcctca----
Antisense     atatttttacctttta--aaatccact--ttatctg-----------atgccaca----
              . :  *  *.*** * .  .:. **:*    :: **             *.*:**
```

```
Slc34a2b      acttgtgatcgattccttcaacatccagacgggcgcagacgctccagacctgctgaaagt
Slc34a2a      tcgtgagatc-----gtttaacatcgagagtggagaaaaagcaccagctctcctaaatgt
Antisense     gcgga--cac-----tttgaac--catg--------cacagcacacacacac--acatgc
               *  .  .:*      ** ***  * :*        ....**:*.... *:   ..*:*

Slc34a2b      catcac--agagccg------ctcaccaagaacatcattgagctggacacct--------
Slc34a2a      tat--caccgatc------cgcttacccactcaatcattcagctggacgag--------t
Antisense     acacacgcagacacacacacacacacacacacagacaaacttcctgtcgcagagagttat
               .:  *  .** .        *: **..* :...:**:: : *  *:*..

Slc34a2b      -----ctgtgat-------------------tcgt---gacatcgccactggagaccctg
Slc34a2a      -----ctgtgatga-----------g-----------cggcattgcggttggcgatcccg
Antisense     cagtcccgtggtaaaccttggaacaggggaatcctggatgatctgcggagggctata--a
                   * ***.*                           ..:  ** .  **. * .  .

Slc34a2b      ctgcccggaacaaaagtttgatcaagatctggtgtaaaacggaaaaaa-ttac----gaa
Slc34a2a      aagccagaaataaatctcttatcaaggtttggtgccacacagcctcg-aacacgact---
Antisense     atgcctgtgacccaacagag---gccgaactgagccactctgaagaacagcacagctgag
              .:*** * .* ..*: : :    .. .:   *:* .*.:* *.. ..    **

Slc34a2b      cctgg---tgaacgtgaccgtcccgggaatcgcaaactgcact---ccggacgccctgtg
Slc34a2a      ----gtccagaatgtgacc------------acaaca-----------aactgcacagat
Antisense     tctggatctgaacgtggtc------------tccatcagcaccagtgcttctgctcaggc
                  *   :*** ***. *            *.* .            . ** *:*

Slc34a2b      ctgggtggatggagacttgatctggacccaaaagaaccaaaccgataccatctacttgaa
Slc34a2a      ctttgttg----ggaattg------------aagaatgttacagaaattataaatattaa
Antisense     cag---------gatctt----------------------ttatg--ttgttctctattca
              *:          ..:.**                     ::. *  :  :*.:. :* .*

Slc34a2b      gaaatgtacgcacatgttcgtgtttgcggatctgcctgatctggcggtgggtttgatcct
Slc34a2a      aaaatgcagtcacatcttcgtgaacacgagcctttcggacctggcggtgggtctgattct
Antisense     cacaggcagtcacatcttcgtgaacacgagcctttcggacctggcggtgggtctgattct
               *.* * *  ***** ******:: .**.. **  * ** ************ **** **

Slc34a2b      gctggctctgtctctgctggctctctgcacatgcctgatcctcatagttaaactcctgaa
Slc34a2a      gctggcgggttctctgctcattctgtgcacttgcctcatttgtattgtgaagctgctgaa
Antisense     gctggcgggttctctgctcattctgtgcacttgcctcatttgtattgtgaagctgctgaa
              ******    ******** . *** *****:***** **    **:** **.** *****

Slc34a2b      ctccatgctgaagggtcaggtggccgtcgtcatcaaaaaagtgctcaacacagatttccc
Slc34a2a      ttcgatgctcaagggacaggttgctgttgtgatcaagaagatagtgaacacagatttccc
Antisense     ttcgatgctcaagggacaggttgctatag------------------------------
               ** ***** *****:***** ** .* *

Slc34a2b      cttccccttcgcctgggttacaggatacctggccatcctcgtgggagccggaatgacctt
Slc34a2a      ttttccatttgcttggctgactggatatattgcaattctggttggtgcgggaatgacttt
Antisense     ------------------------------------------------------------


Slc34a2b      cattgttcagagcagctccgtcttcacctccgccatcactcctctagtcggtattggtgt
Slc34a2a      tattgtccagagcagttccgtcttcacatcggctataactcctcttgttggtatcggtgt
Antisense     ------------------------------------------------------------


Slc34a2b      gatcagtttggagagagcatatccactcacactgggctccaacatcggcacaaccaccac
Slc34a2a      tataagtattgaaagggcatatcctctatccctgggatccaatattggaacaactactac
Antisense     ------------------------------------------------------------

Slc34a2b      tgctatactggcggccatggccagtcctggagaaacactggccaactcgctacagatttc
Slc34a2a      tgcaatattagccgccatggctagtcctggtgaaacacttggaaactcattacagattgc
Antisense     ------------------------------------------------------attgc
                                                                      *** *
```

```
Slc34a2b       tctgtgtcacttctttttcaacatcgctggtatcctgctgtggtaccccatccccttcac
Slc34a2a       attggttcacttcttcttcaacttatccgggatattgttgtggtacccaatccccatcac
Antisense      attggttcacctcttcttcaacttatccgggatattgttgtggtacccaatccccatcac
               : **  **** **** ******:*. * ** **. ** **********.******:****

Slc34a2b       acgtgtgcccatccgactggccaaagcgctgggcaaccgtacagctaaataccgctggtt
Slc34a2a       acgtatccccatccgactggcaaagggtcttggtgaaacaacagcccagtaccgttggtt
Antisense      acgtatccccatccgactggcaaagggtcttggtgaaacaacagcccagtaccgttggtt
               ****.* **************.**.*  ** ** .*.. :***** .*.***** *****

Slc34a2b       cgcgggcctgtatctcattctctgctttctgctgtt-tccgttgaccgtcctgggactct
Slc34a2a       tgcagccttttacatcatcctttgcttc-ttcggtttgcctctactggtctttggtctct
Antisense      tgcagccttttacatcatcctttgcttc-ttcggtttgcctctactggtctttggtctct
                **.* * * ** .**** ** *****  * * ***  **  *..  *** * **:****

Slc34a2b       ccattgcaggatggcaggtcctggtgggcatcggggtgccagtg---------ttggtgc
Slc34a2a       ctatggccggctggcaggtactcatgggagttcttgtccccatagcggttattctgatct
Antisense      ctatggccggctggcaggtactcatgggagttcttgtccccatagcggttattctgatct
               * ** **.**.********.** .****..*    ** **..*.          **.*

Slc34a2b       tcgccatctttgtgattgtggtgaacgtcatgcagaaacgatgccctcgctttctgccct
Slc34a2a       tcgccatcattgt---------caacattctgcagaagcacaaacctcaatggcttcctt
Antisense      tcgccatcattgt---------caacattctgcagaagcacaaacctcaatggcttcctt
               ********:****          ***.* .*******.*..:..****..*  ** ** *

Slc34a2b       cgttcatccgcagttgggagtttctgcccaaaccgctgcactctctgaagccctgggacc
Slc34a2a       ctgctctccggtcttgggatttccttcctctatgggctcactctcttgatccatgggaca
Antisense      ctgctctccggtcttgggatttccttcctctatgggctcactctcttgatccatgggaca
               *    .**** : ****** ** ** ** .:*  *   ******** .* **.******.

Slc34a2b       gagtggtgactgcgggaatgagcttctgcaggactcgctgctgctgctgctgtaaatgct
Slc34a2a       gagtagtcactgtcat----------tgc--cgcccgctgttgctgttgctgcaagtgct
Antisense      gagtagtcactgtcat----------tgc--cgcccgctgttgctgttgctgcaagtgct
               ****.** ****  .           ***   .* ***** ***** ***** **.****

Slc34a2b       gcagaaatgaagagaaaaa--------c-cacatggagaacaacgac--aggagcctgga
Slc34a2a       gcaattctaatgaagaagatgagaaggcgaaacttgaaaatttggccaatggaatcgaga
Antisense      gcaattcttaaccatcttcttggt---cctcccaggagtcccagcacactgc----agaa
               ***.::.* *: .. .: .        *  ...: **.:.  :  .*  :*      ..*

Slc34a2b       gatgt---acgacaatccag-------------------cactcgg-tatagaggacg-a
Slc34a2a       taaatgataatacgatgacg-accgttgaaatcatagagcccaagaaaacagtggacagt
Antisense      taaatatttcctcagtgaagcaacactgccatcttaa----------aaca---------
                *:.*   :. :*..* ..*                           :* *

Slc34a2b       ggctaaa--gtgactg---ccacacatttataatgctacatggactccattttaatttat
Slc34a2a       tgcgaaatcctgaaggcaacatctt-tat----agatgcatacaaagcatt------at-
Antisense      ggctttatactgaaggaatggtcgcatat----ggatgatttctctgagat-----tga-
                ** ::*   ***. *     :*   *:*     *.*..:*  :.: ..:*       :

Slc34a2b       gcaagtcgtgttttgatgatcgcagggtgaagc------accatgacttttttaagctgct
Slc34a2a       ---a---ggttattattaaaca-------tatctaaatcaaactggctttga------tt
Antisense      ---aggagggtctatataagaatatgtagtaatatcatcactctagctcatt------cc
                  *   *  * *: :*.* ..       :*        *. .*..** : :

Slc34a2b       attggtttttagtttggttataatgc-----------------------------------
Slc34a2a       tttggtacccagagttgacagtatgcaata-tcagtggactttaaagagtttatgtcact
Antisense      --tga---atggaggagtcttgttgcaaccgtccgtc-----------------cccact
                 **.     .*:   *: :  :***

Slc34a2b       -------------------------------tgcat-------------------------
Slc34a2a       ctgtatttcttaaattaatttttttatacagcaatgctatttttataggacattgtaaa
Antisense      ctccctacct-------------tttgatctgctcc-----------------------
                                      :**:
```

```
Slc34a2b      ------------------------------------------------------------
Slc34a2a      tgttgcaatatatttaaatatacaatcttatgttacacactgtgagatct--ggtgattg
Antisense     ----------------------------------------acgtcagatctccatttcctg


Slc34a2b      ------------------------------------------------------------
Slc34a2a      ttgtcgtacttgttccagtgtgaaaaggatacttgttt----atttttaataccattcgg
Antisense     gtgtcggt------cctcagtgctcgggtctccttccttccagaaatgtttttctcacgg


Slc34a2b      ------------------------------------------------------------
Slc34a2a      aggatataaaagccaaagacaaatatatattcagatttaaataagacgtcaaccctgatc
Antisense     ------------------------------------------------------------


Slc34a2b      ------------------------------------------------------------
Slc34a2a      tgtgcattgttctttattcatgataatgttcttgtatttatgtcaaatattttaacatgt
Antisense     ------------------------------------------------------------


Slc34a2b      --------------------------------------------------
Slc34a2a      gcctttctatactgtaaaatattattatgatattaaacaataatttcctgc
Antisense     --------------------------------------------------
```

3.0 EMBOSS matcher: pairwise nucleotide sequence alignment of Slc34a2a sense and antisense (Waterman-Eggert local alignment)

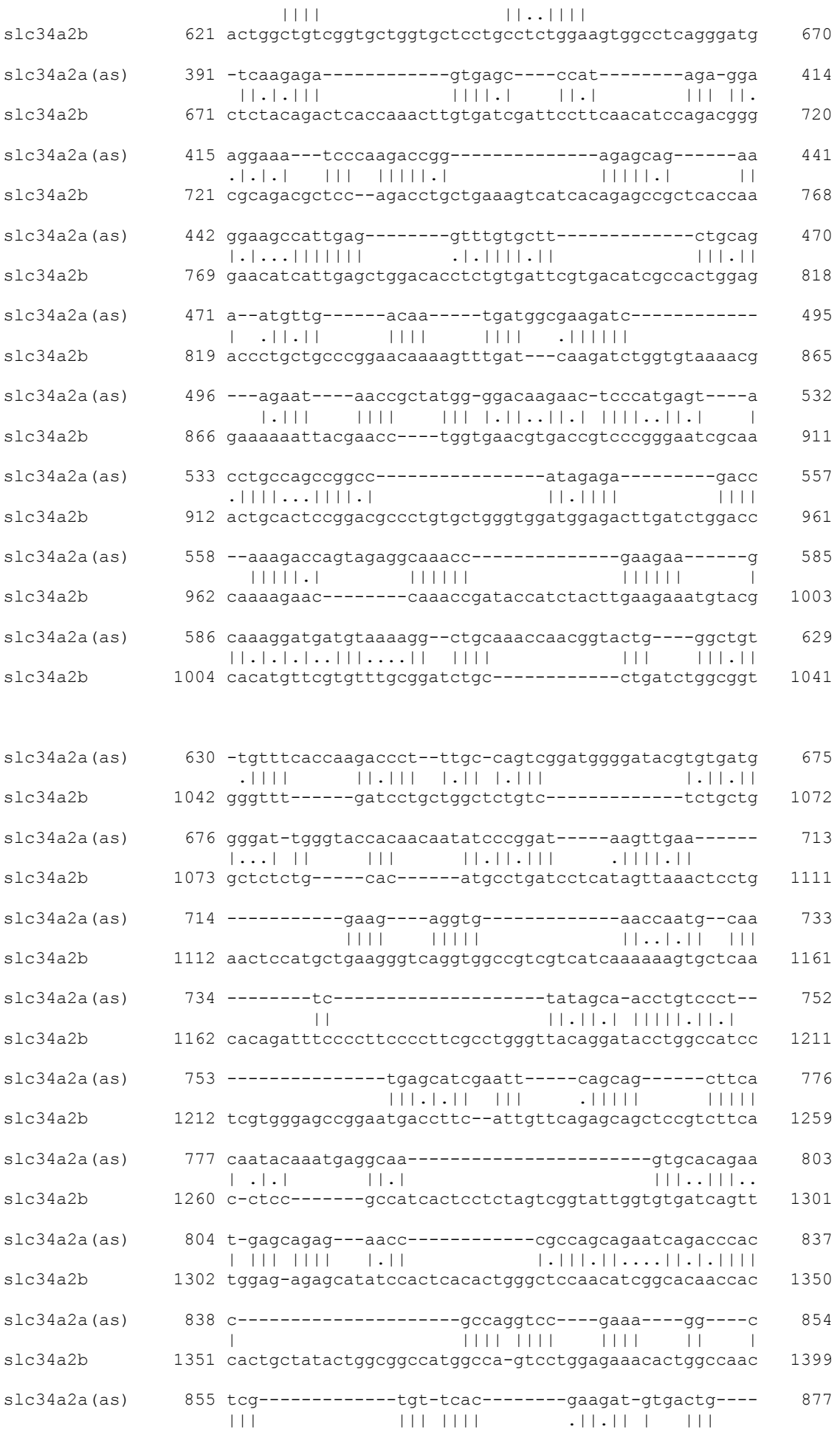
58.7 % nucleotide sequence similarity (88/150)

```
Slc34a2a        2436 GAGGATATAAAAGCCAAAGACAAATATATA----TTCAGATT--TAAATA   2479
                     |.||||.|.|||....||.|.|.|||||||    |.|..|||  ||..||
Slc34a2a(as)    1192 GTGGATTTTAAAAGGTAAAAAATATATATAAAAGTCCTTATTAGTAGTTA   1241

Slc34a2a        2480 AGACGTCAACCCTGA-TCTGTGCATTGTT-CTTTA---TTCATG-ATAAT   2523
                     .||.|..||..|..| |..|..|.||.|| ||.||   |.|||. |||.|
Slc34a2a(as)    1242 TGATGGTAATACACAGTACGGCCTTTATTACTGTAAGTTACATATATACT   1291

Slc34a2a        2524 GTTCTTGTATTTATG-TCAAATATTTTAACATGTGCCTTTCT-ATACTGT   2571
                     ||..||.|.|||..| |||||||.|.|...|||||.||||.| |.|.|||
Slc34a2a(as)    1292 GTATTTCTGTTTTAGATCAAATACTGTTTAATGTGGCTTTATTAAAATGT   1341
```

4.0 EMBOSS needle: pairwise nucleotide sequence alignment of Slc34a2b and Slc34a2a(as) (Needleman-Wunsch global alignment)

37.7% nucleotide sequence similarity

```
slc34a2a(as)       1 --------------------------------------------------      0

slc34a2b           1 caggctgatggctcctctggctgagactcatcccgcctcgccgccaccgg     50

slc34a2a(as)       1 --------------------------------------------------      0

slc34a2b          51 aactcggagcggacgcggacaaacatgagccccaaaccccgacgcgcacc    100

slc34a2a(as)       1 ----------------------ccgtgagaaaaacatttctgga--aggaa     27
                                          |.|||.||        .|.|||  ||| |
slc34a2b         101 ctcagtccctggcccgcgcagcagtggga--------gccggagcagg-a    141

slc34a2a(as)      28 gga-------gaccc----gagc--actgaggacc--gacaccag---ga     59
                     |||       |||||    ||||  .|.|||..||  ||||||.|   ||
slc34a2b         142 ggaagaggtggacccgtgggagctgccggagctcctggacaccggggtga    191

slc34a2a(as)      60 aatggagatctgacgtgga-----gcagatcaaaaggtagggagagtggg    104
                     |.|||   ||.||..||||     ||.|    ||||||
slc34a2b         192 agtgg---tcagagttggaccggcgcgg----aaaggt------------    222

slc34a2a(as)     105 gacggacggt-tgcaacaagactcctccattcaggaatgagctagag-tg    152
                      .|.|..||| ||||        ||||          ||..||..|| ||
slc34a2b         223 -cctgcgggtctgca--------cctc----------tgtcctgaagttg    253

slc34a2a(as)     153 atgatatt-----------actacatattcttat-----atagac--cct    184
                     .||.|..|           .||||||.|||.|.|     .|.|||  |||
slc34a2b         254 ctgctgctgctcggcctgctctacatgttcgtctgctccctggacatcct    303

slc34a2a(as)     185 --------ccttcaa-----tctcagagaaatcatc------cat-----    210
                             |.|||.|     ||..||.|||..||.|      |||
slc34a2b         304 cagctctgctttccagctggtcggagggaaggcagcgggggacatcttta    353

slc34a2a(as)     211 ------atgc------gaccattccttcagtataaagc----ctgtttta    244
                           ||||      |.|||..|||   ||    .||    |||.|
slc34a2b         354 aagataatgcggtgctgtccaaccct---gt----ggccgggctggt---    393

slc34a2a(as)     245 agatggcagtgttgcttcactgaggaaatatttattct------------    282
                      |||.|..|||.|| .||||.|            ||||
slc34a2b         394 -gatcggggtgctg-gtcacgg------------ttctagtccagagctc    429

slc34a2a(as)     283 --gca--------------gtg-------tg-------ctgggactcctg    302
                       |||              |||       ||       || |||.|.||
slc34a2b         430 cagcacctcctcctccatcgtggtcagcatggtatcgtct-ggattgct-    477

slc34a2a(as)     303 ggaggaccaagaag------------------atggttaagaattgcagc    334
                     |||||    .||||                  ||||.|...|||..|.||
slc34a2b         478 ggagg----tgaagtcagcagtgcctatcatcatgggtgcaaatatcggc    523

slc34a2a(as)     335 acttgcagcaacagcaac--------------------agcggg---cgg    361
                     |||| |||.||||  |||                    |||.||   |.|
slc34a2b         524 actt-cagtaaca--aacactattgtggccgtgatgcaagctggagaccg    570

slc34a2a(as)     362 caatga------cag------------------------tgactact---    378
                     ||||||      |||                        |||||.||
slc34a2b         571 caatgagtttcgcagggcttttgcgggggcgacggtgcatgacttcttca    620

slc34a2a(as)     379 -----ctgt-----------------cccatgga----------------    390
```


slc34a2b 621 actggctgtcggtgctggtgctcctgcctctggaagtggcctcagggatg 670
slc34a2a(as) 391 -tcaagaga----------gtgagc----ccat--------aga-gga 414
slc34a2b 671 ctctacagactcaccaaacttgtgatcgattccttcaacatccagacggg 720
slc34a2a(as) 415 aggaaa---tcccaagaccgg--------------agagcag------aa 441
slc34a2b 721 cgcagacgctcc--agacctgctgaaagtcatcacagagccgctcaccaa 768
slc34a2a(as) 442 ggaagccattgag--------gtttgtgctt-------------ctgcag 470
slc34a2b 769 gaacatcattgagctggacacctctgtgattcgtgacatcgccactggag 818
slc34a2a(as) 471 a--atgttg------acaa-----tgatggcgaagatc------------ 495
slc34a2b 819 accctgctgcccggaacaaaagtttgat---caagatctggtgtaaaacg 865
slc34a2a(as) 496 ---agaat----aaccgctatgg-ggacaagaac-tcccatgagt----a 532
slc34a2b 866 gaaaaaattacgaacc----tggtgaacgtgaccgtcccgggaatcgcaa 911
slc34a2a(as) 533 cctgccagccggcc-----------------atagaga---------gacc 557
slc34a2b 912 actgcactccggacgccctgtgctgggtggatggagacttgatctggacc 961
slc34a2a(as) 558 --aaagaccagtagaggcaaacc--------------gaagaa------g 585
slc34a2b 962 caaaagaac--------caaaccgataccatctacttgaagaaatgtacg 1003
slc34a2a(as) 586 caaaggatgatgtaaaagg--ctgcaaaccaacggtactg----ggctgt 629
slc34a2b 1004 cacatgttcgtgtttgcggatctgc------------ctgatctggcggt 1041
slc34a2a(as) 630 -tgtttcaccaagaccct--ttgc-cagtcggatggggatacgtgtgatg 675
slc34a2b 1042 gggttt------gatcctgctggctctgtc-------------tctgctg 1072
slc34a2a(as) 676 gggat-tgggtaccacaacaatatcccggat-----aagttgaa------ 713
slc34a2b 1073 gctctctg-----cac------atgcctgatcctcatagttaaactcctg 1111
slc34a2a(as) 714 -----------gaag----aggtg-------------aaccaatg--caa 733
slc34a2b 1112 aactccatgctgaagggtcaggtggccgtcgtcatcaaaaaagtgctcaa 1161
slc34a2a(as) 734 --------tc--------------------tatagca-acctgtccct-- 752
slc34a2b 1162 cacagatttccccttcccttcgcctgggttacaggatacctggccatcc 1211
slc34a2a(as) 753 ---------------tgagcatcgaatt-----cagcag------cttca 776
slc34a2b 1212 tcgtgggagccggaatgaccttc--attgttcagagcagctccgtcttca 1259
slc34a2a(as) 777 caatacaaatgaggcaa-----------------------gtgcacagaa 803
slc34a2b 1260 c-ctcc-------gccatcactcctctagtcggtattggtgtgatcagtt 1301
slc34a2a(as) 804 t-gagcagag---aacc------------cgccagcagaatcagacccac 837
slc34a2b 1302 tggag-agagcatatccactcacactgggctccaacatcggcacaaccac 1350
slc34a2a(as) 838 c---------------------gccaggtcc----gaaa----gg----c 854
slc34a2b 1351 cactgctatactggcggccatggcca-gtcctggagaaacactggccaac 1399
slc34a2a(as) 855 tcg-------------tgt-tcac--------gaagat-gtgactg---- 877

```
slc34a2b        1400 tcgctacagatttctctgtgtcacttctttttcaacatcg---ctggtat   1446

slc34a2a(as)     878 cctg-tgtgaatagagaacaacat-------aaaag-------atcctgg    912
                     |||| ||||      |.||..|||       |.|.|       ||||  |
slc34a2b        1447 cctgctgtg------gtaccccatccccttcacacgtgtgcccatcc--g   1488

slc34a2a(as)     913 cctgagcagaagcactg-------gtgctgatggagaccac--gttcag-    952
                     .||| ||..||||.|||       ||.|.|.|..|.|||.|  |||| |
slc34a2b        1489 actg-gccaaagcgctgggcaaccgtacagctaaataccgctggttc-gc   1536

slc34a2a(as)     953 --------atc-cagactcagc--tgtgctgt------------------    973
                             ||| ||..|||.||  |.||||||
slc34a2b        1537 gggcctgtatctcattctctgctttctgctgtttccgttgaccgtcctgg   1586

slc34a2a(as)     974 ---tcttca------gagtggctcagttc------ggcctc---------    999
                        |||.||      || |||  |||.||      |||.||
slc34a2b        1587 gactctccattgcagga-tgg--caggtcctggtgggcatcggggtgcca   1633

slc34a2a(as)    1000 -tgttgg------------------------------gtca--cag-----   1012
                      ||||||                             ||||  |||
slc34a2b        1634 gtgttggtgctcgccatctttgtgattgtggtgaacgtcatgcagaaacg   1683

slc34a2a(as)    1013 --------gcatttatagccctccgcagatcatcc------aggattccc   1048
                             || |||.| ||||||    |.||||||      .|||
slc34a2b        1684 atgccctcgc-tttct-gccctc----gttcatccgcagttggga-----   1722

slc34a2a(as)    1049 ctgtt-----ccaaggtttaccacgggactgataactct------ctgcg   1087
                       |||     ||||     |||.|.|.|||     ||||      |||.|
slc34a2b        1723 --gtttctgcccaa-----accgctgcact-----ctctgaagccctggg   1760

slc34a2a(as)    1088 acaggaagtttgtctgtgtgtgtgtgtgtgtgtgtctgcgtg------   1131
                     ||.|  |||       .|||..||.|.|..||.|..||||||.|
slc34a2b        1761 accg--agt-------ggtgactgcgggaatgagcttctgcaggactcgc   1801

slc34a2a(as)    1132 tg-tgcatgtgtgtgtgtg-tgctgtgcatggttca-aagtgtccgctgt   1178
                     || |||           || ||||||..|||.|.|| ||.||
slc34a2b        1802 tgctgc-----------tgctgctgtaaatgctgcagaaatg--------   1832

slc34a2a(as)    1179 ggcatcagataaa----gtggattttaaaaggtaaaaa------------   1212
                        |..|||.|||    .||||          .||.||
slc34a2b        1833 ---aagagaaaaaccacatgga----------gaacaacgacaggagcct   1869

slc34a2a(as)    1213 --atatatataaaagtccttattagtagttat-------gatggtaatac   1253
                       |.||.||..|.|.|||  |..|.|.|.|||       ||.|.|||
slc34a2b        1870 ggagatgtacgacaatcc--agcactcggtatagaggacgaggctaa---   1914

slc34a2a(as)    1254 acagt-acggcc-----tttattactgtaa-gttacat------------   1284
                       ||| ||.|||     ||||      ||| |.|||||
slc34a2b        1915 --agtgactgccacacattta------taatgctacatggactccatttt   1956

slc34a2a(as)    1285 -atatactgtatttc-tgtttt--agatc--------aa-----at-act   1316
                      ||.|| ||.|..|| ||||||  .||||        ||     || |||
slc34a2b        1957 aattta-tgcaagtcgtgttttgatgatcgcagggtgaagcaccatgact   2005

slc34a2a(as)    1317 gtttaa--tg----tggcttt-------attaaaatgtgtgtgtatgt   1351
                     .|||||  ||    |||.|||       .|||.||||   .||.||
slc34a2b        2006 ttttaagctgctattggtttttagtttggttataatg---ctgcat--   2048
```

Appendix B

Video recordings demonstrating reactions of wildtype and injected embryos to the ‘poke’ test are located on CD-ROM.

Appendix C

The complimentary sequences of Slc34a2a sense/antisense from exons 10 and 13 were compared to coding and non-coding zebrafish transcripts using BLAST. Hits with a minimal length of 15 nucleotides and 95% identity were further investigated. Moreover, matching seed regions were required. Below are the raw hits of this search:

A) Non-coding RNA genes matching to exon 10
B) Non-coding RNA genes matching to exon 13
C) cDNA and splice variants

A)

Name	Gene	Length (bp)	Start bp of exon	Start bp of query	End bp of exon	End bp of Query	Exon Orientation	Query Orinetatiom	Biotype	Gene Transcripts	Location
											-
ENSDART00000153628	si:ch211-237h22.2	15	92	1123	106	1137	Forward	Reverse	Antisense	2 -si:ch211-237h22.2-001, si:ch211-237h22.2-002	Chromosome 14 41,625,292-41,628,341 (f)
ENSDART00000147961	si:ch211-147m20.3	14	100	453	113	466	Forward	Forward	Processed Transcripted	3-si:ch211-147m20.3-001, 002, 003	Chromosome 18: 24,171,780-24,260,570 r
ENSDART00000121853	si:dkey-11o18.5	13	101	123	113	135	Forward	Forward	Processed Transcripted	2- si:dkey-11o18.5-001, 201	Chromosome 21: 12,987,419-12,997,088 r
ENSDART00000145742	si:dkey-11o18.5	13	101	13	113	25	Forward	Forward	Processed Transcripted	2-si:dkey-11o18.5-001, 201	Chromosome 21: 12,987,049-12,996,669 r
ENSDART00000150747	si:dkey-115f9.3	13	65	276	77	288	Forward	Forward	LNC RNA	3- si:dkey-115f9.3-001, 002, 003	Chromosome 4: 10,832,785-10,836,764 (f)
ENSDART00000156950	si:dkey-115f9.3	13	65	348	77	260	Forward	Forward	LNC RNA	3- si:dkey-115f9.3-001,002,003	Chromosome 4: 10,826,621-10,836,508 (f)
ENSDART00000153714	si:dkey-115f9.3	13	65	252	77	264	Forward	Forward	LNC RNA	3-si:dkey-115f9.3-001,002,003	Chromosome 4: 10,826,621-10,836,508 (f)
ENSDART00000115828	CABZ01078988.1	14	101	78	114	91	Forward	Reverse	miRNA	1- CABZ01078988.1-201	Chromosome 19: 1,133,823-1,133,956 (f)
ENSDART00000131247	si:dkey-182g1.5	13	70	1110	82	1122	Forward	Reverse	Processed Transcripted	si:dkey-182g1.5-001	Chromosome 22: 8,719,772-8,723,909 (f)
ENSDART00000139571	si:dkey-222f2.3	13	70	1105	82	1117	Forward	Reverse	Processed Transcripted	3- si:dkey-222f2.3-001, 002, 003	Chromosome 22: 5,670,292-5,673,535 r
ENSDART00000115007	si:ch73-338o16.3	13	33	880	45	892	Forward	Reverse	LNC RNA	2- si:ch73-338o16.3-001, 201	Chromosome 3: 5,677,205-5,678,356 r
ENSDART00000128481	U3	13	67	200	79	212	Forward	Reverse	sno RNA	U3.49-201	Chromosome 14: 11,839,256-11,839,470 r
ENSDART00000129695	U3	13	67	200	79	212	Forward	Reverse	sno RNA	U3.7-201	Chromosome 14: 11,898,840-11,899,054 r

B)

Name	Gene	Length (bp)	Start bp of exon	Start bp of query	End bp of exon	End bp of Query	Exon Orientation	Query Orinetation	Biotype	Gene Transcripts	Location
ENSDART00000144470	si:dkey-111e8.3	15	382	55	396	69	Forward	Forward	Processed Transcript	si:dkey-111e8.3-001	Chromosome 5: 39,623,034-39,628,086 (f)
ENSDART00000139427	si:dkey-111e8.2	15	382	66	396	80	Forward	Forward	Processed Transcript	si:dkey-111e8.2-001	Chromosome 5: 39,610,081-39,614,664 (f)
ENSDART00000152348	si:dkey-56e5.2	15	279	606	143	620	Forward	Forward	LNC RNA	si:dkey-56e5.2-001	Chromosome 12: 11,938,019-12,113,718 (f)
ENSDART00000112696	si:ch73-353p21.5	14	12	3499	25	3512	Forward	Forward	Processed Transcript	2- si:ch73-353p21.5-001, 201	Chromosome 22: 5,625,347-5,637,849 r
ENSDART00000124591	si:ch211-127b16.4	14	97	2353	110	2366	Forward	Forward	Processed Transcript	2- si:ch211-127b16.4-001, 201	Chromosome 16: 1,690,053-1,702,545 (F)
ENSDART00000138645	si:dkey-27b3.3	17	317	131	333	147	Forward	Reverse	Processed Transcript	si:dkey-27b3.3-001	Chromosome 10: 8,766,375-8,768,313 r
ENSDART00000150373	si:ch211-162i8.1	16	288	592	303	607	Forward	Reverse	Antisense	si:ch211-162i8.1-001	Chromosome 4: 50,774,775-50,828,776 (f)
ENSDART00000134572	si:ch211-237a4.1	16	288	592	303	607	Forward	Reverse	Processed Transcript	si:ch211-237a4.1-001	Chromosome 4: 57,501,986-57,502,942 r
ENSDART00000135377	si:dkey-254e13.7	16	288	592	303	607	Forward	Reverse	Processed Transcript	si:dkey-254e13.7-001	Chromosome 4: 56,247,584-56,248,549 (f)
ENSDART00000149672	si:ch211-205a14.12	16	288	592	303	607	Forward	Reverse	Processed Transcript	si:ch211-205a14.12-001	Chromosome 1: 37,829,562-37,830,518 r
ENSDART00000149164	si:ch211-76m11.16	16	288	592	303	607	Forward	Reverse	Processed Transcript	si:ch211-76m11.16-001	Chromosome 1: 37,728,211-37,729,168 (f)
ENSDART00000121339	SNORD27	15	250	5	264	19	Forward	Reverse	snoRNA	SNORD27.15-201	Chromosome 9: 10,869,245-10,869,311 (f)
ENSDART00000097863	si:ch73-373m9.1	14	143	22	156	35	Forward	Reverse	Processed Transcript	2- si:ch73-373m9.1-001, 201	Chromosome 16: 55,703,200-55,824,084 (f)
ENSDART00000150025	si:ch73-373m9.1	14	143	73	156	86	Forward	Reverse	Processed Transcript	2- si:ch73-373m9.1-001, 201	Chromosome 16: 55,703,149-55,728,633 (f)
ENSDART00000155839	cab39	14	259	874	272	887	Forward	Reverse	LNC RNA	cab39-001	Chromosome 15: 38,641,893-38,643,424 (f)
ENSDART00000119872	SNORD27	14	251	5	264	18	Forward	Reverse	snoRNA	SNORD27.3-201	Chromosome 17: 34,796,528-34,796,593 r

C)

Name	Gene	Length (bp)	Start bp of exon	Start bp of query	End bp of exon	End bp of Query	Exon Orientation	Query Orinetation	Biotype	Gene Transcripts	Location
ENSDART00000024328	slc34a2a	140	1	1187	140	1326	Forward	Forward	Protein Coding	ENSDART00000024328	Chromosome 1: 13,317,179-13,327,647 f
ENSDART00000156072	hmgb1b	18	81	630	98	647	Forward	Forward	Protein Coding		Chromosome 15: 30,811,097-30,814,116 f
ENSDART00000053527	slc34a2b	26	36	1053	61	1078	Forward	Forward	Protein Coding	slc34a2b-201, slc34a2b-202	Chromosome 23: 1,084,886-1,114,746 f
ENSDART00000115189	slc34a2b	26	36	1053	61	1078	Forward	Forward	Protein Coding	slc34a2b-201, slc34a2b-202	Chromosome 23: 1,084,886-1,115,886 f
ENSDART00000146451	sqrdl	16	87	1471	102	1486	Forward	Forward	Protein Coding	sqrdl-001, 002, 003	Chromosome 25: 33,451,918-33,474,004 f
ENSDART00000009484	cct6a	16	68	2127	83	2142	Forward	Forward	Protein Coding	cct6a-201, 001, 002	Chromosome 21: 40,509,533-40,528,854 r
ENSDART00000137630	cct6a	16	68	2128	83	2143	Forward	Forward	Protein Coding	cct6a-201, 001, 002	Chromosome 21: 40,509,530-40,528,855 r
ENSDART00000006381	psen2	20	108	544	127	563	Forward	Forward	Protein Coding	psen2-001	Chromosome 1: 51,558,297-51,571,727 r
ENSDART00000042074	sema4ba	16	91	6806	106	6821	Forward	Forward	Protein Coding	sema4ba-201, 001	Chromosome 18: 25,561,623-25,638,958 r
ENSDART00000153922	si:ch211-202f3.3	16	104	339	119	354	Forward	Forward	Protein Coding	si:ch211-202f3.3-001, 002, 003	Chromosome 17: 45,744,842-45,766,706 f
ENSDART00000074854	si:ch211-202f3.3	16	104	370	119	385	Forward	Forward	Protein Coding	c1qtnf6a-201, 001	Chromosome 3: 25,322,121-25,323,822 f
ENSDART00000055440	c1qtnf6a	16	7	413	22	428	Forward	Forward	Protein Coding	c1qtnf6a-201, 001	Chromosome 3: 25,318,057-25,323,929 f
ENSDART00000156544	c1qtnf6a	16	7	431	22	446	Forward	Forward	Protein Coding	slc34a1a-201, 001, 002, 003	Chromosome 14: 836,172-849,972 r
ENSDART00000041988	slc34a1a	36	36	805	71	840	Forward	Forward	Protein Coding	slc34a1a-201, 001, 002, 003	Chromosome 14: 835,780-858,118 r
ENSDART00000148687	slc34a1a	36	36	1060	71	1095	Forward	Forward	Protein Coding	wnt4b-201, 001	Chromosome 16: 34,024,131-34,034,943 f
ENSDART00000058739	wnt4b	15	23	412	37	426	Forward	Forward	Protein Coding	aspg-001, 201, 002, 003, 004	Chromosome 17: 45,905,614-45,942,116 f
ENSDART00000110834	wnt4b	15	23	412	37	426	Forward	Forward	Protein Coding	aspg-001, 201, 002, 003, 004	Chromosome 17: 45,905,508-45,942,108 f
ENSDART00000109525	aspg	15	100	1664	114	1678	Forward	Forward	Protein Coding	cers2b-201, 001	Chromosome 16: 1,410,909-1,425,912 f
ENSDART00000135073	aspg	15	100	1770	114	1784	Forward	Forward	Protein Coding	cftr-001	Chromosome 1: 13,317,179-13,327,647 f
ENSDART00000081980	cers2b	15	12	3695	26	3709	Forward	Forward	Protein Coding	soat1-201, 001	Chromosome 8: 15,403,403-15,427,174 f
ENSDART00000060243	cftr	15	76	767	90	781	Forward	Forward	Protein Coding	si:ch73-177h5.2-201, 001	Chromosome 4: 1,784,931-1,790,555 r
ENSDART00000146589	soat1	15	67	501	81	515	Forward	Forward	Protein Coding	A2ML1 (12 of 12)-201, 001	Chromosome 15: 19,953,600-19,975,663 r
ENSDART00000082242	si:ch73-177h5.2	15	66	1088	80	1102	Forward	Forward	Protein Coding	A2ML1 (12 of 12)-201, 001	Chromosome 15: 19,953,463-19,975,912 r
ENSDART00000144991	A2ML1	20	85	4206	104	4225	Forward	Reverse	Protein Coding	narf-001, 002, 003	Chromosome 12: 34,955,702-34,968,937 f
ENSDART00000154019	A2ML1	20	85	4349	104	4368	Forward	Reverse	Protein Coding	ENSDART00000108818, ENSDART00000152428	Chromosome 15: 17,938,567-18,022,230 r
ENSDART00000007053	narf	15	72	1363	86	1377	Forward	Reverse	Protein Coding	ENSDART00000108818, ENSDART00000152428	Chromosome 15: 17,936,831-18,029,159 r
ENSDART00000108818	ENSDARG00000075334	15	63	4689	77	4703	Forward	Reverse	Protein Coding	ENSDART00000130435	Scaffold Zv9_NA129: 4,183-8,616 forward
ENSDART00000152428	ENSDARG00000075334	15	63	4970	77	4984	Forward	Reverse	Protein Coding	elovl5-001	Chromosome 13: 2,174,483-2,197,358 f
ENSDART00000130435	hmgxb4b	15	46	1008	60	1022	Forward	Reverse	Protein Coding	BX649490.1-201	Chromosome 22: 33,282,815-33,290,185 r
ENSDART00000006990	elovl5	15	101	40	115	54	Forward	Reverse			
ENSDART00000131138	BX649490.1	15	46	390	60	404	Forward	Reverse			
ENSDART00000133038	uqcrc1	15	65	10	79	24	Forward	Reverse			
ENSDART00000105374	narf	15	72	1083	86	1097	Forward	Reverse			
ENSDART00000123368	ek1	15	44	1072	58	1086	Forward	Reverse			
ENSDART00000105774	ek1	15	44	1082	58	1096	Forward	Reverse			

List of publications and abstracts

1.0 Publications

Piatek, M.J., Henderson, V., Zynad, H.S. and Werner, A. (2016) 'Natural antisense transcription from a comparative perspective', *Genomics*.

Werner, A., Piatek, M.J. and Mattick, J.S. (2015) 'Transpositional shuffling and quality control in male germ cells to enhance evolution of complex organisms', *Annals of the New York Academy of Sciences*, 1341(1), pp. 156-163.

Piatek, Monica J. and Werner, A. (2014) 'Endogenous siRNAs: regulators of internal affairs', *Biochemical Society Transactions*, 42(4), pp. 1174-1179.

2.0 Conferences and courses attended

Oct 2015	RNP and Disease	Marrakech, MAR
Sept 2014	ESF-EMBO Long Regulatory RNAs	Pultusk, PL
Feb 2014	Non-Coding Genome- International course	Institut Curie, Paris, FR
Oct 2013	North East Postgraduate Conference	Newcastle, UK