EXPLORING THE ROLE OF EARLY TRANSCRIPTION ON THE ZYGOTIC GENOME ACTIVATION IN THE ZEBRAFISH EMBRYO

by

KACPER TOMASZ WOŹNIAK

A thesis submitted to the
University of Birmingham
for the degree of
Doctor of Philosophy

Institute of Cancer and Genomic Sciences
College of Medical and Dental Sciences
University of Birmingham
September 2022

Abstract

Early embryos are transcriptionally quiescent and rely on the signals provided by the mother to guide them through the first hours of development. Through the process of zygotic genome activation, embryos are reprogrammed from a fully differentiated egg cell into totipotent stem cells. This corresponds with a vast transcriptional activation event in the embryo, as transcription of zygotic genes is activated. However, some genes are able to escape the transcriptional silencing before zygotic genome activation, forming an early wave of transcription. In this thesis, I have aimed to utilise a teleost model organism, *Danio rerio* (zebrafish), to investigate the role of one such early wave transcription factor, *mxtx2*, in zygotic genome activation. I have aimed to characterise the role of *mxtx2* by generating a loss of function and null mutants of its gene. Additionally, I have aimed to characterise cell cycle and genome activation dynamics in the first distinct extra-embryonic tissue of the embryo – the yolk syncytial layer – induction of which is mediated by *mxtx2*. In this thesis, I demonstrate the generation of an efficient method of mosaic loss of function of *mxtx2* by Cas9 nuclease. I have shown that *mxtx2* phenotype generated by this loss of function phenocopies previous studies, independently confirming the results. I also demonstrate a previously unreported phenotype caused by Cas13d nuclease, causing unspecific developmental abnormalities and embryonic lethality. The exploration of cell cycle and transcription dynamics in the zebrafish embryo provided evidence showing a specific cell cycle and transcriptional delay in the nuclei of the yolk syncytial layer, which we hypothesise is caused by a delay in zygotic genome activation in the yolk syncytial layer. Aided with further studies, this could provide the first evidence in any organism of distinct zygotic genome activation control in subsets of embryonic cells.

Acknowledgements

Firstly, I would like to thank my supervisor Professor Ferenc Müller for his continued support, patience and resolve over the years, for the guidance and mentoring that I have received and for the trust that I was given to carry this project to its conclusion despite all the difficulties along the way. I would like to especially thank him for having confidence in me and offering me a place in his lab as a co-supervisor and creating an exciting project from scratch for me to join. I also want to thank my co-supervisor Dr Agnieszka Gambus for her support and kind words throughout the process.

I want to thank the BBSRC MIBTP programme for funding the PhD and making this project possible, with special thanks to the MIBTP Directors for their generous support during the project change and COVID disruptions.

Thank you to all the present and past members of the Müller lab for the fruitful discussions, always being there able to help and to advise on experiments, and for creating a fantastic working atmosphere and being overall excellent. I would like to especially thank Yavor Hadzhiev for all his technical knowledge and endless experimental advice, Lucy Wheatley for mentoring me and trying to make me a better scientist, Ada Jimenez-Gonzalez for always being there to offer advice, knowledge or a kind word, and last but not least Haseeb Qureshi for lending me some of his passion for microscopy, all of our scientific and non-scientific discussions, his tremendous help over the years and for being a fantastic friend.

I would like to acknowledge the help and advice from Dr Andreas Zaucker (University of Warwick) who provided plasmids containing a Cas9-nanos expression construct. Thanks to Dr Nev Gilhooly (ICGS, University of Birmingham) for offering his time and expertise to purify Cas13d protein. I would also like to thank the members of the animal facility at the University of Birmingham, especially for their work on maintaining zebrafish during the pandemic.

Finally, I would like to thank my family and friends who gave me an immeasurable amount of support and love and carried me through all those years.

To my Grandmothers Regina and Alinka, Grandfather Stanisław and Aunt Ewa.

Table of contents

Table of figures

List of abbreviations

Abbreviation	Full Description
1k	1000 (in the context of cell stage in zebrafish)
ChIP-seq	chromatin immunoprecipitation with next-generation sequencing
CRISPR	Clustered Regularly Interspaced Short Palindromic Repeats
crRNAs	CRISPR RNAs
dgRNA	duplex guide RNA
DSB	double-stranded break
GFP	green fluorescent protein
gRNA	guide RNA
HDR	homology-mediated DNA repair
hpf	hours post fertilisation
LSFM	light sheet fluorescent microscopy
MBT	mid-blastula transition
MIP	maximum intensity projection
MO	morpholino oligonucleotide
mRNA	messenger RNA
MZT	maternal to zygotic transition
NHEJ	non-homologous end joining
PCR	polymerase chain reaction
RNA Pol II / Pol II / polii	RNA polymerase II
qPCR	quantitative reverse transcription PCR
RNA	ribonucleic acid
RNAseq	RNA sequencing
ROI	region of interest
sgRNA	single guide RNA
TALEN	transcription activator-like effector nuclease
TF	transcription factor
TSS	transcription start site
tracrRNA	trans-activating CRISPR RNA
UTR	untranslated region
WISH	whole mount *in situ* hybridisation
YSL	yolk syncytial layer
YSN/YSNs	yolk syncytial layer nucleus/nuclei
ZGA	zygotic genome activation

Chapter 1: General Introduction

1.1 Zygotic Genome Activation

1.1.1 Overview

After the egg is fertilised by a sperm, the early embryo is initially transcriptionally inactive and undergoes fast cell divisions without gap phases (Newport and Kirschner, 1982b). The first stages of development are entirely controlled by mRNAs and proteins deposited in the oocyte by the mother – in mice, those deposited mRNAs represent 40% of all protein-coding genome (Wang *et al.*, 2004), while in *Drosophila melanogaster* it was estimated at 65% (Lecuyer *et al.*, 2007). These RNAs are required for the egg to convert into totipotent state. The overall transition process between a transcriptionally quiescent and active embryo is referred to as the maternal-to-zygotic transition (MZT), within which the process of activating the zygotic transcription called zygotic genome activation (ZGA) takes place (Tadros and Lipshitz, 2009). This is one of the most dramatic shifts in embryogenesis and perhaps the largest genomic reprogramming event occurring in nature. Through both maternal instructions and zygotic gene expression, fully differentiated cells (egg and sperm) are reprogrammed to become totipotent and reform the whole organism. During MZT, degradation of maternally deposited factors is coordinated with the initiation of zygotic transcription, with thousands of genes becoming transcriptionally active. During ZGA, activation of transcription of a vast number of genes coincides also with lengthening of cell cycles from cleavage to those containing 'gap' phases. Increase in cell cycle length is needed to accommodate increased transcription alongside ongoing fast rounds of DNA replication, as those two

processes can interfere with each other (Schulz and Harrison, 2019). Zygotic genome activation is a period, rather than a singular point, during which the embryo ceases to rely on maternally provided genetic information and starts utilising its own, provided by both mother and father.

However, even though there is a sharp increase of zygotic transcription at ZGA, embryos exhibit low levels of transcriptional activity before ZGA. For clarity purposes, I will continue to refer to this minor transcriptional activity as an early wave of zygotic transcription. This process is primarily regulated by maternally deposited factors and could serve as preparation of the cell environment for the vast change that is ZGA. In a conserved fashion, ZGA and zygotic transcription is required for progression in development, with transcriptional block causing developmental arrest shortly after the approximate time that ZGA would take place (Abe *et al.*, 2018).

As the mechanisms governing ZGA are not fully understood, studying the impact of early transcription on that process can provide some very important answers about the regulation of zygotic transcription, signalling networks involved and the mechanisms by which embryo cells are 'reprogrammed' at ZGA.

1.1.2 Pre-ZGA Development

Following the fertilisation of the egg and fusion of two highly specialised cells – egg and sperm – the resulting genome undergoes large-scale epigenetic reorganisation, with changes to chromatin marks, DNA demethylation and remethylation (Guo *et al.*, 2014; Meehan *et al.*, 2005; Potok *et al.*, 2013; Santos *et al.*, 2005). During the oocyte growth, transcription is dynamically regulated. However, global levels of transcriptional activity gradually decline as oocyte grows, leading to no active

transcription towards the end of that process (De La Fuente *et al.*, 2004; Schultz *et al.*, 2018). The mechanisms underlying this global transcriptional silencing are still elusive, with chromatin compaction, global changes in DNA organisation within the nucleus, and RNA polymerase II localisation being suggested (Abe *et al.*, 2010; Bouniol-Baly *et al.*, 1999; Schultz *et al.*, 2018). An alternative approach suggested that dissociation of chromatin factors (such as transcription factors and other members of the protein complex binding the promoter) away from chromatin on a global scale within cells could be responsible for transcriptional silencing in early embryos, and for removal of maternal transcription program. Global chromatin factor dissociation was observed in late-stage oocytes, followed by an "erase and rebuild" epigenome reprogramming in the zygote (Sun *et al.*, 2007). A similar observation was made in experiments involving somatic cell nuclear transfer and in parthenogenesis (Gao *et al.*, 2007), suggesting a common mechanism and perhaps a requirement for "clearance" of the pre-existing transcriptional setup in major reprogramming events.

Transcriptional quiescence in the zygote is evolutionarily conserved in metazoans. The period of quiescence differs between species, but in most cases coincides with very fast progression through multiple cell cycles that do not contain gap phases (Schulz and Harrison, 2019; Tadros and Lipshitz, 2009; Yuan *et al.*, 2016). In most species those cell cycles also coincide with cleavage cell divisions, as there is very limited time for cell growth. During normal development, ZGA occurs after several rounds of cleavage cell cycles, and how many differs between species (Fig. 1.1). Those cell cycles are primarily directed by the maternally contributed proteins and transcripts, which include products of multiple housekeeping genes (Liu *et al.*, 2014;

Omura and Lott, 2020). Alongside the machinery required for maintaining the zygote and direct the fast cell cycles, maternal contributions also include mRNAs and proteins that would direct the gradual activation of transcription. As the zygote progresses through cleavage cell cycles, the division cycles start to gradually lengthen, which coincides with appearance of first zygotic transcripts and gradual degradation of maternally-provided mRNAs. Maternal to zygotic transition ends the period of cleavage divisions and culminates with ZGA.

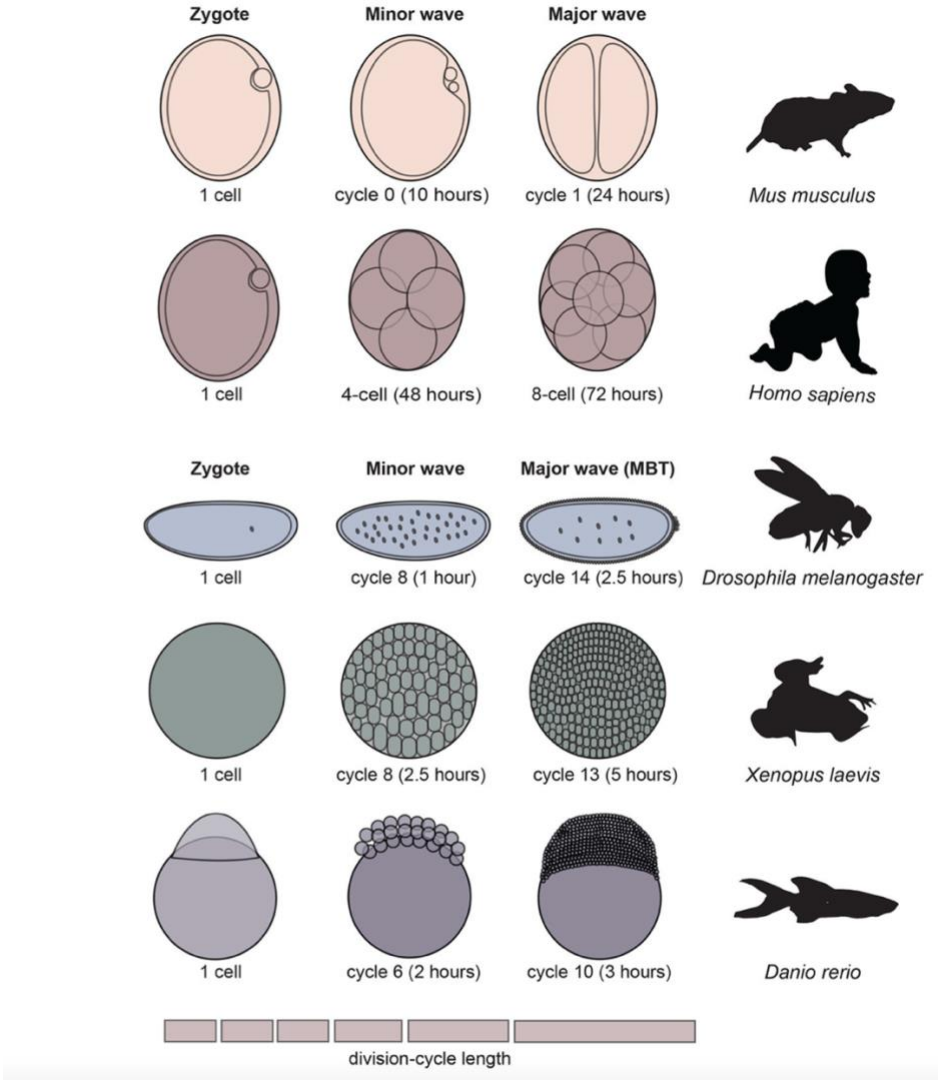


Fig. 1.1. The timing of zygotic genome activation in several model organisms.
The top 2 rows are mammalian model organisms, while the bottom 3 represent a range of other animal models. The first column depicts a diagram of a zygote, second column depicts a developmental stage and time (in hours post fertilisation) at which minor zygotic transcription is first observed, and the third column depicts the developmental stage and time (in hours post fertilisation) when main wave of zygotic genome activation begins. Last column presents a simplified depiction of a corresponding organism. Adapted from (Schulz and Harrison, 2019).

1.1.3 ZGA in Model Organisms

1.1.3.1 *Drosophila melanogaster*

In embryos of the invertebrate *Drosophila melanogaster*, the main wave of transcriptional activation occurs during the 14th cell cycle (Darbo *et al.*, 2013) (Fig.

1.1). Prior to that, the fruit fly zygote forms a multicellular syncytium, with nuclear divisions occurring, but not cell divisions. ZGA then coincides with the process of cellularisation of syncytial blastoderm. Transcription of a small subset of early-wave genes involved in, among others, embryo patterning and cellularisation process, starts around the 8th cell cycle (Harrison *et al.*, 2011; Pritchard and Schubiger, 1996). Some studies claim even earlier transcription and competency of *Drosophila* embryos to start transcription as early as cycle 2 (Ali-Murthy *et al.*, 2013). Studies in the fruit fly embryos led to a discovery of the first major regulator of ZGA in any species – Zelda (Zld) (Liang *et al.*, 2008). Zld is a maternally deposited zinc-finger pioneer transcription factor. It was found to bind regulatory elements of nearly all early zygotically transcribed genes in *Drosophila*, but also to mark genes that would be transcribed during main-wave ZGA, which would be activated by already zygotically expressed early wave transcription factors (Harrison *et al.*, 2011). It is therefore understood that Zld binding is priming genes for zygotic transcription.

Additionally, it was shown that Zld, as a pioneer transcription factor, has a role in opening chromatin at the loci where it is bound (Schulz *et al.*, 2015; Sun *et al.*, 2015), an effort in overcoming the inhibitory environment in the nucleus of an early embryo (Dufourt *et al.*, 2018). Zld has been found binding to promoters of some of the earliest zygotic transcripts identified in *Drosophila* embryos (Harrison *et al.*, 2010). Zld is required for early and main-wave gene transcription, and essential for embryo survival through ZGA, with both maternal and zygotic mutants causing the embryo to die. Maternal mutants die at the cellularisation stage, while zygotic mutants die later in embryonic development (Schulz and Harrison, 2019), which shows the necessity of *zld* transcription at the zygotic level despite the maternal deposition. No

orthologues of Zld have been found in species other than insects (Schulz and Harrison, 2019), which spurred the search for factors that could serve the same function, especially in vertebrate embryos.

While Zelda is considered a major activator of transcription in *Drosophila*, there are several other key factors in the MZT process worth mentioning. A maternally deposited RNA-binding factor Smaug was found to be essential for early development, affecting several aspects of MZT. In *Drosophila*, between the 11th and 13th synchronous syncytial divisions the length of cell cycles gradually increases, with an appearance of short gap phases. This coincides not only with the appearance of the first zygotic transcripts, but also with a re-introduction of DNA replication checkpoints into the cell cycle, mediated by an ATM kinase homologue (Sibon *et al.*, 1997). Stricter control of the cell cycle is likely to be directly responsible for slowing down of cell cycles. Mutations in *smaug* have been found to disrupt the cell cycle slowdown and introduction of checkpoints. Smaug is also responsible for destruction of maternally deposited mRNAs by binding them and recruiting the CCR4/POP2/NOT complex, which in turn de-adenylates them and triggers their lysis (Benoit *et al.*, 2009). Smaug is also involved in expression of several micro-RNAs (miRs), including *miR-3, miR-6* and *miR-286* (Bushati *et al.*, 2008). *smaug* mutant was found to stabilise many of the targets of the *miR-309* cluster – a set of miRs involved in destruction of maternal mRNAs, indicating that Smaug plays a role in regulating clearance of maternal material (Bushati *et al.*, 2008; Fu *et al.*, 2014).

Zygotic expression of microRNAs is essential during the time preceding main wave of ZGA, and they are a key component of the early wave of zygotic transcription in blastoderm, conducting a diverse set of processes, including clearance of maternal

mRNAs. Mutants lacking pre-ZGA microRNA expression were found to be embryonic lethal and displayed gastrulation defects (Fu *et al.*, 2014). Expression of *miR-309* cluster is required for destabilisation of over 400 maternal mRNAs during early development (Benoit *et al.*, 2009; Bushati *et al.*, 2008). Both Smaug and Zelda were found to regulate expression of miRs, including the miR-309 cluster (Liang *et al.*, 2008), however, activation of expression by Zelda is more robust and affects many more microRNA loci (Fu *et al.*, 2014).

Taken together, ZGA in *Drosophila* has a master regulator in the zinc finger transcription factor Zelda, which directs the early zygotic transcription wave that coordinates destruction of maternal transcripts by zygotically expressed *Smaug* and *miR-309* and activates transcription of over 100 zygotically expressed genes involved in early embryo development, which in turn leads to the main transcriptional wave of ZGA (Benoit *et al.*, 2009; Liang *et al.*, 2008; Schulz *et al.*, 2015). Zelda therefore connects two of the key events of the maternal-to-zygotic transition – transcriptional activation and clearance of maternal transcripts.

1.1.3.2 *Xenopus*

Amphibian models *Xenopus laevis* and *Xenopus tropicalis* have proved to be very important in understanding of early embryonic processes. *Xenopus laevis* was the first embryo in which the ability of a somatic cell nucleus to form a functioning organism was shown (Gurdon *et al.*, 1958), developing a new field of biology dedicated to cloning and stem cell research and leading to discovery of a number of key pluripotency factors aside from *Drosophila* Zelda. *X. laevis* was also one of the first organisms where a major developmental transition of the early embryo was

discovered – with cell divisions becoming asynchronous, cells becoming motile and widespread activation of transcription (Newport and Kirschner, 1982a; b). This observation coined the term mid-blastula transition (MBT) as it occurs during the blastula stage of the frog embryo development, after 12 rapid cleavage cell divisions (Fig. 1.1).

As in other organisms, pre-ZGA transcription is present and essential. In *Xenopus* it is first observed after 6 cell cycles (Gentsch *et al.*, 2019a; Skirkanich *et al.*, 2011). Regulation of zygotic genome activation in *Xenopus* was recently reviewed (Blitz and Cho, 2021). First cleavage division in *Xenopus* embryos requires a longer period, ending at 1.5h post fertilisation in *X. laevis*, with following cell cycles progressing very fast in a meta-synchronous manner, with slightly larger vegetal blastomeres lagging behind the smaller blastomeres at the animal side of the embryo (Chen *et al.*, 2019), but maintaining a constant cell cycle length between those. In contrast to *Drosophila*, no syncytium is formed and cells forming the blastoderm undergo rounds of DNA replication followed by mitosis and cytokinesis. Following the 12^{th} cell cycle, approximately 5 hours post fertilisation (hpf) in *X. laevis*, cell cycles slow down and become asynchronous. That marks the mid-blastula transition and the beginning of main wave of ZGA, and shortly thereafter the embryo proceeds into gastrulation.

Xenopus was used as a model for several seminal studies in understanding genome activation in embryos, detecting earliest zygotic transcription and incorporation of radiolabelled nucleotides into the nucleus before MBT and observing the global activation of transcription during MBT (Bachvarova *et al.*, 1966; Kimelman *et al.*, 1987; Newport and Kirschner, 1982a). Follow-up studies using new methods of transcript detection, such as incorporation of 5-ethynyl uridine (5-EU) showed that

main wave of genome activation in *Xenopus* embryos does not begin in all cells at the same time, but that smaller animal pole blastomeres begin ZGA earlier, by cell cycle 12, and larger vegetally positioned blastomeres begin ZGA during cell cycle 14 (Fig. 1.2).

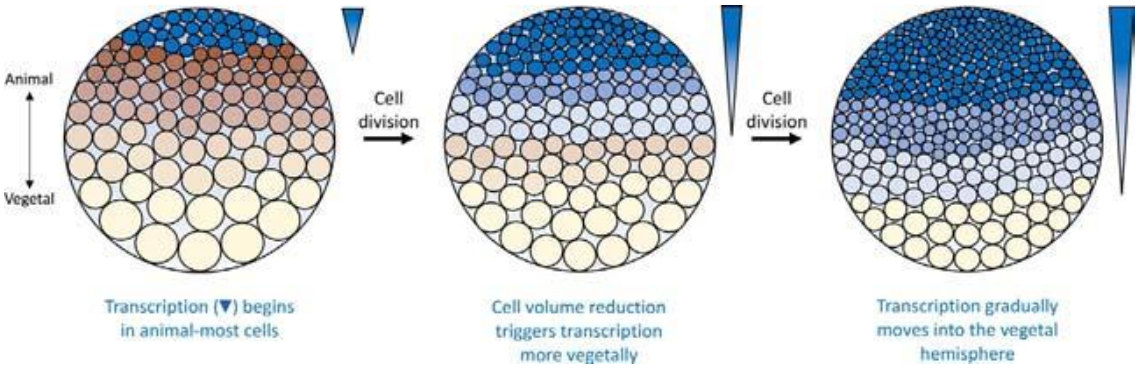


Fig. 1.2. Progression of zygotic genome activation in *Xenopus laevis* embryos.
The simplified diagram shows the timing of major wave of genome activation during cell cycles 12 (left panel), 13 (middle panel) and 14 (right panel). A "wave" of genome activation is observed starting from the animal-most blastomeres and, across the span of 2 cell cycles, progressing to the vegetally positioned, larger blastomeres. From: (Blitz and Cho, 2021).

Xenopus was one of the first models to link blastomere size and onset of genome activation, suggesting that a certain threshold of the ratio of genetic material (that is replicated rapidly in cleavage cell cycles) to the cytoplasm volume (that roughly halves every cell cycle) is required. This helped coin a prevalent model of ZGA timing called nuclear to cytoplasmic (N:C) ratio based on titration of inhibitory factors (Newport and Kirschner, 1982b; Prioleau *et al.*, 1994), which will be discussed in detail later in this thesis.

The earliest transcription in *Xenopus* is observed at around 64-cell stage, with two of the first transcripts detected being *nodal5* and *nodal6* – encoding 2 of 6 identified

Nodal-related proteins in *Xenopus* (Skirkanich *et al.*, 2011; Yang *et al.*, 2002). Those members of TGF-ß superfamily are ligands in the Nodal signalling pathway and are required for induction of mesendoderm during gastrulation, but it was shown that their expression as early as 128-cell-stage (after 8 cell cycles) is regulated by the ß-catenin/Tcf pathway (Wnt signalling pathway) (Blitz and Cho, 2021; Onuma *et al.*, 2002), and more directly by the maternally deposited T-box transcription factor VegT (Skirkanich *et al.*, 2011). Activation of the Nodal signalling before MBT is required for main-wave ZGA transcription of multiple downstream targets, and for the correct progression in embryo development into gastrulation.

Among other early wave transcripts identified in *Xenopus* were several transcription factors such as *bix1.3* (brachyury inducible factor 1) and *sox17a*, several less well-described zinc finger transcription factors, as well as *miR-427*, a microRNA involved in clearance of maternal transcripts, orthologous to zebrafish *miR-430* (Gentsch *et al.*, 2019a; Owens *et al.*, 2016; Skirkanich *et al.*, 2011).

Similarly to *Drosophila*, master regulators of zygotic transcription and ZGA have been identified in *Xenopus*. These are maternally deposited transcription factors Sox3 and Pou5f3 (Gentsch *et al.*, 2019b). With their broad range of binding and the ability to trigger chromatin remodelling at their sites of binding, they are often termed pioneer transcription factors. Sox3 and Pou5f3 belong to a set of pluripotency factors, with their ability to induce pluripotency in transcriptionally quiescent blastomeres, effectively starting the cellular reprogramming during MBT and leading to the formation of three embryonic germ layers (ectoderm, mesoderm, endoderm) in the zygote.

1.1.3.3 Mouse

In mammalian embryos, ZGA happens much earlier from the developmental point of view – in mice genome activation is observed in pre-implantation embryos during the second cell cycle (during 2-cell stage, 24 hpf), with the minor wave of transcription observed during the S/G2 phase of the 1st cell cycle, approximately 10-12 hpf (Aoki *et al.*, 1997; Bouniol *et al.*, 1995; Flach *et al.*, 1982; Xue *et al.*, 2013). The difference in absolute time elapsed after fertilisation between mammals and lower vertebrates is quite noticeable, as it is measured in days, rather than minutes or hours, and occurs after fewer cleavage divisions (Fig. 1.1).

Regulation of the murine ZGA has been relatively well studied, and recently reviewed by (Aoki, 2022). In murine zygotes, there are not enough cell divisions between fertilisation and ZGA to establish a nuclear:cytoplasmic ratio. Experiments with polyspermic fertilisation (increasing the nuclear component) did not affect the transcription before ZGA, ruling it out as a regulatory mechanism. It was suggested that a maternally provided mRNA could be involved in triggering ZGA in mice, as some of the maternal mRNAs are translated shortly after fertilisation. Those mRNAs often require polyadenylation in the zygote but can be translated simultaneously with Poly-A tail elongation (Oh *et al.*, 2000; Temeles and Schultz, 1997). Experiments in 1-cell fertilised eggs have shown that preventing this polyadenylation inhibited transcription and ZGA (Aoki *et al.*, 2003). However, no individual trigger, mRNA or protein, has yet been found to regulate ZGA initiation in mice, but several genes have been proven to be involved.

In order for early wave zygotic transcription to take place during late first cell cycle, it was shown that murine zygotes have a greatly loosened chromatin structure

compared to oocytes, allowing transcription factors to bind to promoters to activate transcription (Ooga *et al.*, 2016). One of the suggested mechanisms for the loosening of chromatin is epigenetics – histone variants that were identified in 1-cell stage murine embryos are usually involved in loosened chromatin structures, such as H3.3/H2A.X and TH2A histone variant compositions (Funaya and Aoki, 2017). The chromatin becomes much more structured and tighter packed with all histone variants present in late 2-cell stage, with major wave ZGA in full swing. Thus, chromatin state and histone composition of nucleosomes are important factors in regulating early wave transcription in mammals. The transition from minor to major wave transcription is also a transition from enhancer-less expression where transcription factors are able to bind promoters in loosened chromatin to one where enhancers regulate the accessibility of the transcriptional machinery to the core promoter, effectively changing from a permissive state at 1-cell stage to a more repressive, regulated state in late 2-cell stage by changes to nucleosome components as well as establishment of chromatin marks such as H3K4me3 (Schultz *et al.*, 2018). As in other previously described models, in mice minor wave transcription is required for major wave ZGA to occur. Inhibiting minor wave transcription with an RNA Pol II inhibitor led to the minor wave pattern of transcription being observed in 2-cell stage embryos, with major wave ZGA not occurring and embryos developmentally arresting in 2nd cell cycle (Abe *et al.*, 2018).

No direct ZGA activators like Zelda or Sox3 and Pou5f3 have been found in mammalian embryos, but several factors were linked with regulatory roles in the process. Oct4, homologue of a key regulator of ZGA in zebrafish (Pou5f3), was found to not have in influence on ZGA in mice, but to be essential in later development (Gao

et al., 2018). Pioneer transcription factor Nfya, transcriptional regulator Yap1 and a homeobox transcription factor Dux have all been shown to influence zygotic transcription in 2-cell mouse embryos (De Iaco *et al.*, 2017; Hendrickson *et al.*, 2017; Lu *et al.*, 2016; Yu *et al.*, 2016). However, these factors might not be required for progression past the 2-cell stage, as several knockdown and knockout studies have revealed (Chen *et al.*, 2021; Chen and Zhang, 2019; De Iaco *et al.*, 2020).

The case of *Dux* (orthologue of human DUX4), encoding a dual homeobox transcription factor is particularly intriguing. It is an intronless, multi-copy gene found to be transcribed in minor ZGA. It was found to regulate expression of ERVL family retrotransposons and have an influence on pluripotency of cells (Hendrickson *et al.*, 2017). Through activation of retrotransposons and their linked genes, Dux is an important factor in cell reprogramming, found to be necessary for transition of mouse embryonic stem cells to the 2-cell-like-cells (2CLCs) *in vitro*. 2CLCs are a small subset of murine embryonic stem cells that, in cell culture, share many characteristics with their 2-cell stage counterparts, such as expression profile and totipotency (De Iaco *et al.*, 2017; Hendrickson *et al.*, 2017). Acting upstream of Dux are maternally provided developmental pluripotency-associated factors Dppa2 and Dppa4, regulating its role in driving ZGA (Eckersley-Maslin *et al.*, 2019). Loss of Dux leads to depletion of several genes found in major wave of ZGA, suggesting Dux as an interesting candidate for a ZGA regulator and making it similar to Zelda in its early wave expression. However, further studies have shown that loss of Dux does not lead to a developmental arrest and has a minor effect on ZGA, with only some of the Dux targets downregulated (Bosnakovski *et al.*, 2021; Chen and Zhang, 2019; De Iaco *et al.*, 2020). *Dux*$^{-/-}$ knockout embryos were able to survive to adulthood, but with

several developmental defects and reduced viability (Chen and Zhang, 2019; De Iaco *et al.*, 2020). It was therefore termed a "ZGA synchroniser", with a search for a key regulator of the murine ZGA continuing.

Perfect candidates for this role are pioneer transcription factors due to their ability to open regions of closed chromatin, allowing transcription in an otherwise repressive environment. As most well-known pioneer factors Nanog and Oct4 are unlikely to be directly involved (with activatory roles in ESCs, but not observed *in vivo*), the search for regulators is expanding. One of the newly suggested regulators of ZGA in mouse, Nr5a2 was found using a *de novo* binding motif search attempting to find potential binding partners of major wave ZGA genes (Gassler *et al.*, 2022). This previously poorly described orphan nuclear receptor has been found to be essential for early embryonic development with deficient embryos not surviving post implantation (Labelle-Dumais *et al.*, 2006). Nr5a2 was found to be required for efficient ZGA, with its knockdown reducing nascent RNA transcription in 2-cell embryos. In combination with it being found enriched near transcription start sites of ZGA genes and promoting chromatin accessibility, Gassler *et al.* claim that Nr5a2 is a pioneer transcription factor and that it regulates murine ZGA, with over 70% of ZGA major wave genes exhibiting reduced expression upon Nr5a2 inhibition (Gassler *et al.*, 2022), however further investigation will be needed to confirm these claims.

1.1.3.4 Human

Understanding of human embryonic development relies on experiments in model organisms, with very few opportunities to test hypotheses *in vivo*, with many studies also using embryonic stem cells or pluripotent stem cells as a model. Studies directly

on human embryos are limited due to the ethical limitations as well as technical difficulties. A major point of progress was generation of induced pluripotent stem cells (iPSCs) from mouse somatic cells (Takahashi and Yamanaka, 2006), making it possible to reprogram somatic cells into a pluripotent state, converting it to a stem cell with self-renewal potential and capacity to form other cell types. This was achieved by inducing ectopic expression of several candidate transcription factors in cultured fibroblasts, transforming them into a pluripotent-like state, which allowed them to express many of embryonic stem cell markers, as well as differentiate into several other cell types both *in vitro* and *in vivo*. Essential combinations of transcription factors responsible for that state were identified, and these factors were termed stem cell pluripotency (or Yamanaka) factors. These key factors were OCT4, SOX2, KLF4 and MYC, with NANOG also playing an important, but non-essential role. These factors are considered crucial in human embryonic development and contribute to establishment of pluripotency and subsequently cell fates during zygotic genome activation. iPSCs allowed another way of modelling human development *in vitro* (Taubenschmid-Stowers *et al.*, 2022).

In contrast to the murine model, early wave of transcription in human embryos has been identified around 2-4 cell stages, with major wave ZGA at 8-cell stage (Fig. 1.1) (Vassena *et al.*, 2011) (Braude *et al.*, 1988; Zhang *et al.*, 2009). There is limited evidence of human embryos showing zygotic transcription as early as 1-cell stage (Xue *et al.*, 2013), but the current consensus is the minor wave between 2- and 4-cell stage and the major wave between 4- and 8-cell stage. Timing of human ZGA is conserved among other mammals like sheep or cow, suggesting that mouse ZGA timing might be an outlier among mammals (Jukam *et al.*, 2017). ZGA also coincides

with the first cell fate decisions in 8-cell stage embryos, with cells later forming the inner cell mass (later responsible for forming the embryonic germ layers) and trophectoderm, later responsible for forming extraembryonic tissues and placenta.

The understanding of key factors driving ZGA in humans is limited. Stem cell pluripotency factors NANOG, OCT4 and SOX2 were all found to regulate a significant portion of transcription in embryonic stem cells, targeting multiple genes essential for correct embryonic development (Boyer *et al.*, 2005; Loh *et al.*, 2006). DUX4 (homologue of mouse Dux) has been found to drive ZGA transcription in cultured 8-cell-like cells, modelling the 8-cell stage human embryo (Taubenschmid-Stowers *et al.*, 2022), but its role in ZGA regulation is contested due to the conflicting results from mouse models. OCT4 has been suggested as a contributor to zygotic transcription in humans, with its binding motif significantly enriched among open chromatin during ZGA, giving rise to a claim that OCT4 targets are transcribed during ZGA (Gao *et al.*, 2018). Very recently, PRD-like homeobox transcription factors TPRXL (maternal) and TPRX1/2 (zygotic, early wave) have been identified as regulators of ZGA and preimplantation development in humans, acting upstream of several DUX-family transcription factors and regulating expression of approximately 20% of major wave genes (Zou *et al.*, 2022).

1.1.4 Zygotic Genome Activation in Zebrafish

1.1.4.1 Overview

In zebrafish (*Danio rerio*), a teleost model organism, ZGA and MZT coincide with mid-blastula transition, similarly to amphibians. Zebrafish is one of the model organisms where ZGA has been most thoroughly studied, however a lot of questions remain

about its regulation. Zebrafish embryos undergo 10 fast synchronous cleavage cell divisions, after which blastomeres lose synchrony, cell cycles slow down with gap phases appearing, and zygotic transcription is activated (Kane and Kimmel, 1993; Kane *et al.*, 1996). After several more rounds of asynchronous divisions, blastula stage embryos proceed into gastrulation to form the embryonic germ layers, all within hours post fertilisation.

Earliest transcription in zebrafish embryos has been detected around 64-cell-stage (during 6th cell cycle) and major wave of ZGA starts at the 1000-cell stage, approximately 3 hours post fertilisation (Fig. 1.1) (Hadzhiev *et al.*, 2019; Heyn *et al.*, 2014; Kane and Kimmel, 1993). Abolition of zygotic transcription causes the embryos to arrest before gastrulation (Kane *et al.*, 1996; Lee *et al.*, 2013), while abolition of translation of maternal mRNAs is able to arrest development even as early as 16-cell stage (Chan *et al.*, 2019).

1.1.4.2. Pioneer factors activating zygotic transcription

Key regulators of zygotic genome activation in zebrafish have been suggested, with Nanog, SoxB1 (family of Sox genes to which SOX2 belongs) and Pou5f3 (homologue of OCT4) stem cell pluripotency factors all required for activation of zygotic transcription, with hundreds of downstream gene targets. All of those factors are provided maternally as mRNAs, and *nanog, pou5f3* and *sox19b* were found to be among the most translated genes in the pre-ZGA embryo (Lee *et al.*, 2013). These transcription factors were found to bind loci of genes activated during ZGA (Leichsenring *et al.*, 2013), and loss of function mutations in these genes led to an extreme reduction of zygotic expression, with more than 80% reduction in gene

expression compared to wild-types by 6hpf, an effect that was shared among all 3 with some synergy and redundancy, but with *nanog* loss of function having the strongest effect (Lee *et al.*, 2013). Nanog, Sox19b and Pou5f3 are also required for increasing chromatin accessibility and remodelling chromatin at their target sites therefore acting as pioneer transcription factors (Miao *et al.*, 2022; Veil *et al.*, 2019). Their binding was found to prime their target genes for expression in pre-ZGA embryos (Palfy *et al.*, 2020). Since these are maternally contributed factors, it places them temporally as acting upstream of any zygotic components in regulation of zygotic transcription. That is indeed the case, with the NPS (Nanog, Pou5f3 and Sox19b) factors suggested to be responsible for activation of the early wave of transcription in zebrafish (Lee *et al.*, 2013). Study with a triple knockout mutant of NPS factors found that these factors are able to overcome the global transcriptional silencing of early embryos by binding to the target sites with high nucleosome occupancy and subsequently inducing a change in chromatin accessibility. This effect is achieved by histone acetylation, and Nanog, Pou5f3 and Sox19b were found to act upstream of histone acetyltransferase p300 and acetylation reader Brd4 (Miao *et al.*, 2022). In parallel, it was found that Pou5f3 and Sox19b are able to modify chromatin and drive expression of genes involved in gastrulation and ventral specification, activating separate sets of genes during major wave of ZGA (Gao *et al.*, 2022). Knockout of these transcription factors seemed to suggest that even those pioneer factors have a localised role within the embryo, as they are not involved in the activation of pathways that are first expressed on the dorsal side of the embryo (especially Nodal pathway). This suggests that specific localisation of those maternal mRNAs into a specific side of the embryo could be an additional aspect determining

which genes are activated during ZGA, and that ZGA is not uniform across the zebrafish embryo.

1.1.4.3 Chromatin remodelling and other factors influencing ZGA

Brd4 and p300-dependent histone acetylation and subsequent opening of chromatin has already been implicated as one of the key elements in ZGA regulation. Study by (Chan *et al.*, 2019) revealed that these proteins, binding to enhancer elements, were required for genome activation – their overexpression was sufficient to prematurely activate zygotic expression, as well as rescue zygotic transcription in embryos that were treated with inhibitors of maternal translation. This study, in combination with one by (Miao *et al.*, 2022) provides a comprehensive model of ZGA regulation in zebrafish, wherein the maternally contributed factors have a supreme role in firstly binding the regions of poorly accessible chromatin to subsequently recruit histone acetyltransferases, leading to opening of those regions, and subsequently allowing transcription machinery to bind and transcription initiation to occur.

This would position translation of maternally deposited transcripts as a key regulatory event upstream of ZGA, as without the pioneer transcription factors and histone modifying proteins the embryo would remain unable to reprogram to the zygotic transcriptional repertoire.

While the maternally provided factors have been shown to play a very important role, they are not the only important pieces of the ZGA regulation puzzle. Consistently with previously mentioned observations, chromatin remodelling and histone modifications such as histone H3K27 acetylation, histone H3K4 and K27 methylation all play a role in the transcription reprogramming of the fertilised egg

into a totipotent zygote and influence zygotic transcription (Sato *et al.*, 2019; Vastenhouw *et al.*, 2019; Zhu *et al.*, 2019).

Availability of the key components of the transcription initiation machinery, such as the TATA-binding protein (Tbp) and its paralogues Tbpl2 (Trf3) and Tbpl1 (Tlf) also has a limiting effect on zygotic genome activation (Ferg *et al.*, 2007). Tbp is not required for all transcription in early embryos (Muller *et al.*, 2001), but depletion of Tbp caused downregulation of a subset of major ZGA genes. Many of the early wave zygotic transcripts originate from promoters containing a TATA box. Interestingly, TBP was also involved in another facet of the maternal-to-zygotic transition, namely degradation of maternal mRNAs. This shows coordination of zygotic transcription activation with maternal mRNA degradation, as several maternal mRNAs were stabilised when *tbp* expression was knocked down using morpholino oligonucleotides(Ferg *et al.*, 2007).

1.1.4.4 Degradation of maternal factors and role of *mir-430*

The process of degradation of maternal mRNAs in zebrafish has one particularly key component, a microRNA *miR-430*. This non-coding RNA is orthologous to the previously mentioned *miR-427* in amphibians, as well as miR-302 in humans. It is encoded by a short, 22nt gene, and is able to regulate gene expression by promoting mRNA deadenylation and decay, as well as more directly by causing reduced translation of its targets (Bazzini *et al.*, 2012; Giraldez *et al.*, 2005; Giraldez *et al.*, 2006). Degradation of mRNAs by miR-430 is highly coordinated with other factors through the RNA-induced silencing complex (RISC), particularly by binding of miRNAs loaded onto Argonaute (AGO) proteins to 3'UTRs of target mRNAs (Kim *et al.*, 2009;

Rosa and Brivanlou, 2009). In zebrafish, miR-430 is required for embryo progression through ZGA, disruption of miR430 locus or *dicer1* RNAse III ribonuclease (involved in microRNA maturation from pre-miRNAs) leads to a range of severe developmental defects with gastrulation, development of body axis, brain morphogenesis and heart development all affected (Giraldez *et al.*, 2005; Giraldez *et al.*, 2006; Liu *et al.*, 2020; Takacs and Giraldez, 2016).

Expression of miR-430 has been identified as early as 64 cell stage (Hadzhiev *et al.*, 2019; Heyn *et al.*, 2014). Due to its extremely unusual locus, *miR-430* is by far the the most expressed transcript in the early embryo (Heyn *et al.*, 2014). This is due to miR-430 locus being organised into triplet repeats of its precursor genes. Those triplet repeats have been observed as duplexes or triplexes in the zebrafish genome, each unit with a single promoter (Hadzhiev *et al.*, 2019). In the current assembly of the zebrafish genome, the *mir-430* locus contains 8 of those transcriptional units. Studies have shown extremely high levels of expression of *miR*-430 in pre-ZGA embryos through both RNA sequencing and microscopy (Hadzhiev *et al.*, 2019; Heyn *et al.*, 2014). The latter showed that transcribed miR-430 organises into very large subnuclear foci or compartments that colocalise with most of the detected pre-ZGA nascent RNA transcripts and transcription machinery, with active or poised RNA polymerase II colocalising in those vast *miR-430* foci (Hadzhiev *et al.*, 2019; Hadzhiev *et al.*, 2021; Hilbert *et al.*, 2021). This suggests a hypothesis that miR-430 locus on chromosome 4 could act as a “transcription organiser” with the local concentration of transcriptional machinery and resources providing a specifically permissive environment for zygotic transcription before ZGA. This hypothesis can be supported well by the previously mentioned activatory function of pioneer factors, which are

required for *miR-430* expression (Lee *et al.*, 2013) and Nanog has been shown to be required for formation of those transcription bodies in pre-ZGA embryos (Hilbert *et al.*, 2021; Kuznetsova *et al.*, 2022). Additionally, the pioneer function could be even more pronounced, as the *miR-430* locus might be significantly larger than previously thought, with a new study using long-read next generation sequencing observing a cluster of not 8, but over 300 promoters and over 750 *mir-430* precursor triplets spanning a region of 0.6 Mb (Hadzhiev *et al.*, 2021). Understanding the mechanisms by which this transcriptional body is formed and whether it truly serves a function in regulating ZGA is yet to be established.

To summarise, ZGA in zebrafish coincides with mid-blastula transition, with dramatic changes in cell cycle, a large-scale genome reorganisation and removal of maternal signals all contributing to this enormous gene reprogramming event.

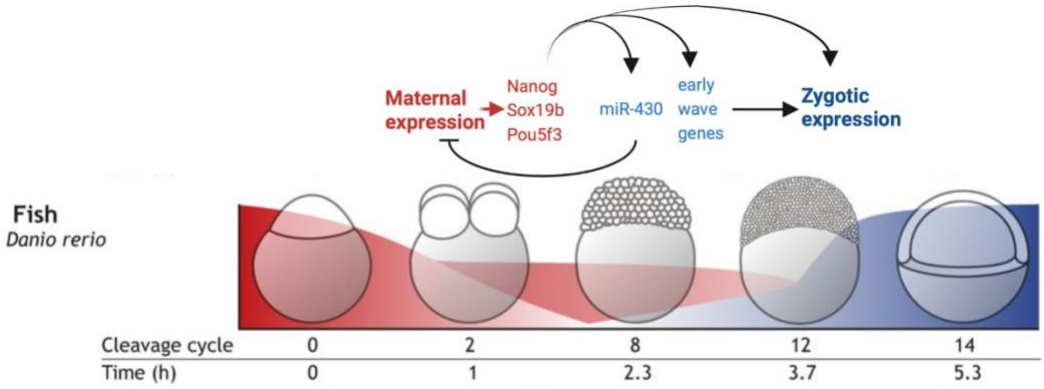


Fig. 1.3. Illustration of the maternal to zygotic transition in a zebrafish embryo.
The cartoon diagram presents simplified drawings of zebrafish embryo progression through developmental stages, from a fertilised egg (left), through blastula up to the beginning of gastrulation (right). In red and blue are simplified and not to scale representations of levels of expression of maternal transcripts and zygotic transcripts respectively. Roles of miR-430 and maternally deposited pioneer factors in regulation of gene expression in the pre-ZGA embryo have been highlighted. Image modified from (Tadros and Lipshitz, 2009).

1.1.5 Timing of Zygotic Genome Activation

ZGA is required to proceed with embryo development, as inhibition of zygotic transcription causes embryonic arrest at a particular stage of development that depends on the species. *Drosophila melanogaster* embryos fail to undergo cellularisation after initial syncytial cell divisions; zebrafish and *Xenopus laevis* embryos fail to gastrulate; while mouse embryos developmentally arrest in the 2nd cell cycle (Abe *et al.*, 2018).

As described previously, the timing of zygotic genome activation differs between organisms (Lee *et al.*, 2014; Schulz and Harrison, 2019). There are several models aiming to explain the mechanism of ZGA timing in embryos, and thereby the mechanism by which genome activation is triggered, however the inter-species divergence remains unanswered.

The existence of ZGA and its delay with respect to egg fertilisation raises a number of important questions. Why is the period of transcriptional quiescence necessary?

Why is zygotic genome activated at a particular point in development, and what is the significance of that point? What triggers the activation and through which mechanisms?

While not all answers are currently known, there are three most prevalent models aiming to explain the timing of ZGA, all of which rely on the significant remodelling of chromatin and an increase in accessibility to explain the activation of gene transcription (reviewed in (Jukam *et al.*, 2017; Schulz and Harrison, 2019; Vastenhouw *et al.*, 2019) among others). Those models are:

a) nucleocytoplasmic (N:C) ratio model in which the trigger to ZGA is passing a threshold of concentration of inhibitory factors in the nuclei of the embryo, below which transcription can take place (Fig. 4a);

b) maternal clock model, dictating that fertilisation starts a 'timer' on ZGA, which is achieved by biochemical reactions taking place in the embryo within the first cell cycles (Fig.4b);

and c) epigenetic remodelling and *de novo* establishment of chromatin states in the embryo.

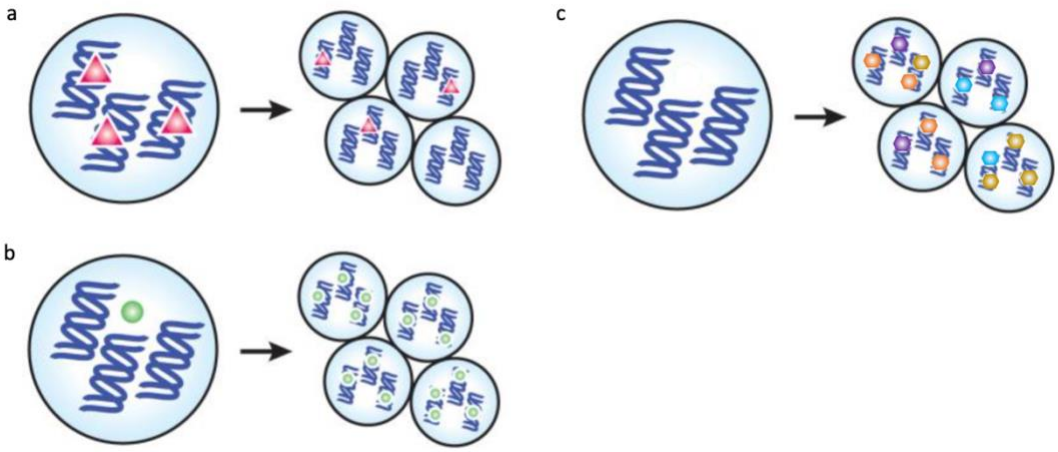


Fig. 1.4. Schematics representing the alternative models of embryonic transcription triggers.
(a) Gradual decrease in concentration of inhibitory factors (red triangles) as the DNA content exponentially grows with each cleavage division, eventually reaching a threshold at which transcription can start. (b) Gradual accumulation of transcription activators, such as pioneer factors, being generated through the cleavage cell cycles from maternal templates, until a particular threshold is reached and transcription can start. (c) *De novo* establishment of various chromatin marks (depicted as orange, blue, purple and gold hexagons), such as histone methylation and acetylation, during cleavage divisions in the embryo. This leads to a gradual remodelling of chromatin and allows transcription to take place in genes where chromatin is open. Image adapted from: (Lee *et al.*, 2014).

Those models are not all mutually exclusive, and most probably, all of these are at least partially correct in a particular species context. In recent years, mixed models have become more prevalent, highlighting the role of epigenetic remodelling and chromatin accessibility as the direct regulators of genome activation, with depletion of inhibitory factors or accumulation of activators and other mechanisms playing a supporting role(Miao *et al.*, 2022; Vastenhouw *et al.*, 2019). Causative relationship of chromatin opening with activation of zygotic transcription is still poorly understood.

1.1.5.1 N:C ratio model and titration of inhibitory factors

Perhaps the most prevalent model is the N:C ratio. In simple terms, during cleavage divisions the total volume of cells does not increase (Lee *et al.*, 2001), but the amount of DNA and thus nuclear material is doubled in every cell cycle. A threshold N:C ratio,

depending on the number of cell cycles undergone by the embryo, would be a key permissive timepoint, after which ZGA can occur. This idea was proposed initially by Newport and Kirschner from their *Xenopus* studies. Polyspermic frog embryos (with higher N/C ratio) were shown to activate their genome earlier than controls (Newport and Kirschner, 1982b). The suggested mechanism was that the fertilised egg contains maternally deposited factors inhibiting transcription that are titrated down, as the cells divide and double the amount of their DNA material each cycle. In support of this hypothesis, levels of core histone proteins – that are maternally provided, can bind the genome ubiquitously, and can keep chromatin compact consequently blocking access to DNA – were found to play a role in negatively regulating zygotic transcription in *Xenopus* and zebrafish (Amodeo *et al.*, 2015; Joseph *et al.*, 2017). Titration of core histone components was found to be one of the events enabling ZGA, but in zebrafish, it was a decrease in levels of unbound histones and not ones sat on DNA, that regulated ZGA progression (Joseph *et al.*, 2017).

Transcription in the early embryo has to be coordinated with utmost temporal precision with DNA replication (Kermi *et al.*, 2017)and key replication factors were also found to be regulated by N:C ratio, influencing cell cycle lengthening and leading to activation of one of the key DNA damage checkpoint kinases Chk1 in *Xenopus* embryos (Collart *et al.*, 2013).

In support of the N:C ratio hypothesis, reduction in cell size (volume) has been found to be sufficient to trigger zygotic transcription in *Xenopus* embryos, with embryos with initially halved or quartered volume (reduced C component of the N:C ratio) activating transcription 1 or 2 cell division cycles earlier, respectively (Chen *et al.*, 2019). This confirms that *Xenopus* follow a cell sizer model of activating transcription,

however these results do not disprove the involvement and importance of other factors in this process. Similar results are yet to be replicated in other species.

The importance of N:C ratio for ZGA was not confirmed in mice, where it influenced morphological changes, but did not affect zygotic transcription (Lee *et al.*, 2001). It is likely that N:C ratio model, involving titration of inhibitors like core histone proteins and a corresponding increase in number of DNA binding sites for transcription activators, may hold more relevance for fast cycling embryos, such as amphibian or fish.

1.1.5.2 Maternal clock and accumulation of activating factors

An alternative mechanism, dubbed the 'maternal clock model' suggests that regulation of ZGA timing is tied to maternally deposited information and independent of cell cycle. A cascade of events is activated at egg fertilisation and regulates the developmental programme through translation (and post-translational modification) of maternally deposited mRNAs, leading to ZGA. This could be achieved through accumulation of transcription-activating or de-repressing factors in the embryo cells during cleavage divisions. There is limited evidence supporting this model. One of main examples is the requirement for accumulation of RNA-binding protein Smaug in *Drosophila* embryos. Smaug is involved in co-ordinating clearance of maternal mRNAs in the embryo through deadenylation and regulating expression of microRNAs, accumulation of Smaug protein is also required in the developmental programme (Benoit *et al.*, 2009; Tadros *et al.*, 2007).

In a similar fashion, accumulation of a general transcription factor – TATA-binding protein (TBP) – is required for transcriptional activation at ZGA, and manipulating

translation of its maternally deposited mRNA affected timing of ZGA (Veenstra, Destrée, and Wolffe 1999) (Veenstra *et al.*, 1999). The precise timing of translation of this general factor does not, however, explain a selective activation of transcription at ZGA (Schulz and Harrison, 2019). A similar case might be made for accumulation of pioneer transcription factors across the early development creating a permissive environment for transcription to initiate (Miao *et al.*, 2022; Veil *et al.*, 2019), however there is not enough evidence to prove that it is a process independent of cell cycle or other factors.

While it is beyond doubt that maternally deposited factors and their translation play a role in ZGA, there is very limited evidence for the maternal clock model. It is therefore likely that timing of this key step in the developmental programme is not controlled independently of the cell cycle.

1.1.5.3 Establishment of chromatin states

The third aforementioned model assumes that chromatin in the zygote does not possess histone marks and other epigenetic markers that prime it for transcription. Upon fertilisation, quite distinctly packaged DNA coming from the egg and sperm is reformed, and histones coming from the mother replace protamines packaging paternal DNA. The model suggests that in early embryo, chromatin states need to be (re)established through DNA and histone modifications, thus changing the epigenetic environment to drive the onset of ZGA. There is a large amount of evidence that chromatin in the embryo is extensively remodelled before and at ZGA, with changes in DNA methylation, histone methylation, acetylation and replacement, and establishment of topologically associated domains. This was reviewed in multiple

publications (Schulz and Harrison, 2019; Tadros and Lipshitz, 2009; Vastenhouw *et al.*, 2019). However, the question whether these are causative for ZGA, a consequence of pre-ZGA developmental programme, or maybe just coinciding events, has still not been resolved.

In mice, altering chromatin structure induced premature transcription (Aoki *et al.*, 1997), suggesting that chromatin state can be causative for genome activation. In zebrafish, an accumulation of chromatin methylation marks – repressive H3K27me3 and activating H3K4me3 – is observed on gene promoters after ZGA. Some genes contain H3K4me3 marks before ZGA, and these are genes poised to be transcriptionally active in the early wave of transcription (Lindeman *et al.*, 2011; Vastenhouw *et al.*, 2010). Many genes are found to contain both activating and repressive marks after ZGA – a bivalency found specifically in embryonic stem cells, poising promoters of key developmental genes for activation with H3K4me3 marks, but keeping them repressed using H3K27me3s mark until a correct stage in the developmental programme (Vastenhouw *et al.*, 2010). While the same mechanism was not found in other model species, it is possible that this way of control of developmental programme is achieved using different histone marks (Schulz and Harrison, 2019).

Histone acetylation was also a subject of interest, as it confers mostly chromatin openness, and has been previously found in early embryos at enhancer domains that can be bound by pluripotency (pioneer) factors (Bogdanovic *et al.*, 2012a; Bogdanovic *et al.*, 2012b; Creyghton *et al.*, 2010). As mentioned in the overview of ZGA in zebrafish, transcriptional activation in this species requires histone H3K27 acetylation, regulated by Brd4 protein (reader) and p300 histone acetyltransferase

(writer). Authors of the Chan *et al.* study argue that neither slowdown of cell cycle nor N:C ratio change are required for genome activation, and it is translation of Brd4 and P300 factors that is necessary and sufficient to activate widespread transcription in embryos (Chan *et al.*, 2019). Additionally H3K27ac was observed to accumulate in the miR-430-associated transcription foci found in zebrafish embryos, and experiments using the small molecule inhibitor of Brd4, JQ-1, led to a loss of this co-localisation with active transcription (Sato *et al.*, 2019). Histone H3K27 acetylation and its associated opening of chromatin has been found to coincide with ZGA in other species than zebrafish, with evidence from mice and *Drosophila* (Dahl *et al.*, 2016; Li *et al.*, 2014).

Another acetylation mark, H3K16 acetylation was found to function similarly to the H3K27 acetylation in both *Drosophila* and murine embryos, with its deposition by the mother in the oocyte being required for correct genome architecture during ZGA and H3K16ac-dependent opening of chromatin was found to be instructive for ZGA genes to be transcribed (Samata *et al.*, 2020).

A mixed model for what triggers ZGA can therefore be proposed, depending on a convoluted interplay between a developmental timer through translation of maternal factors, including pioneer factors, and extensive chromatin reorganisation following the egg fertilisation. It is however unknown how these interplay, and how N:C ratio model can be incorporated into the theory, considering that species like *Xenopus* seem to be heavily dependent on cell volume for triggering ZGA. Early wave zygotic transcription also comes into play, regulating maternal contributions and perhaps further directing the transcriptional reprogramming. It could be hypothesised that different layers of ZGA control exist in different species, having

evolved in parallel and adjusting to the challenges of development faced by each species.

The question of which mechanisms exactly control activation of genes in the early embryo and confer the transition from a quiescent cleavage-dividing cells to totipotent embryo cells remains unanswered. Finding an answer poses a significant challenge for the field and, in order to understand ZGA, it will be important to study all aspects of the intricate interplay of factors during early embryo development and to incorporate multiple models when posing hypotheses of key ZGA regulators.

1.1.6 Role of early zygotic transcripts in ZGA

The pre-ZGA wave of transcription presents a particularly interesting area of research. Most studies focus on maternal contribution, the main source of direction for the developmental programme. However, not all of the pioneer factors are purely maternally provided. Murine Dux and human DUX family pluripotency transcription factors are not only some of the earliest zygotically transcribed genes, but can also drive zygotic transcription (De Iaco *et al.*, 2017; Hendrickson *et al.*, 2017). However, the requirement of these factors for embryo viability and primary regulatory role in ZGA has been disputed (Bosnakovski *et al.*, 2021; De Iaco *et al.*, 2020). There are potentially more pluripotency or pioneer factors hidden within the early wave of transcription. In mice, early activated genes are transcribed from the first cell cycle, and code for proteins involved in basic cellular function, as well as clearance of maternal RNA and protein, while transcription factors and patterning factors activate later. In zebrafish, early transcription appears 4 cell cycles before main wave of ZGA. Genes transcribed in the early wave are mostly short, allowing for rapid transcription

during the extremely fast cleavage cell cycles, and contain genes involved in embryo maintenance as well as transcription factors and chromatin remodellers (Collart *et al.*, 2014; De Renzis *et al.*, 2007; Heyn *et al.*, 2014). 592 genes were found to be expressed early in zebrafish embryos, with 152 of them purely zygotic (compared to over 3000 genes activated at ZGA) (Heyn *et al.*, 2014). During ZGA the proportion is similar, with about 25% of activated genes being purely zygotic, while the rest are genes which mRNAs were also provided by the mother(Lee *et al.*, 2013).

Perhaps the best example of an early zygotic gene having a large influence on ZGA is the previously described regulator of clearance of maternal transcripts: *miR-430* (Giraldez *et al.*, 2006), which also seems to have a role in spatially organising transcription into massive subnuclear foci, termed transcription bodies (Hadzhiev *et al.*, 2019). Data currently in pre-print also seems to suggest that the transcription bodies formed around miR-430 locus promote transcription of a subset of other genes present on chromosome 4 – most of which are zinc-finger DNA binding factors with unknown function (Hadzhiev *et al.*, 2021). Their role could be in transcription activation during main wave ZGA, as many zinc finger transcription factors are known, such as *Drosophila* Zelda. Dissecting individual roles of these zinc finger family factors will be particularly tough, however, as high sequence homology between them makes them difficult targets of mutagenesis and genome engineering tools.

Analysis of the minor wave genes also found that majority of them have a very particular promoter structure, containing a TATA box and a “sharp” transcription start site (Chen *et al.*, 2013; Hadzhiev *et al.*, 2021; Sandelin *et al.*, 2007). Major wave genes lack this promoter structure (Haberle *et al.*, 2014b). This suggests that there

might be a difference in transcriptional machinery used between minor and major wave genes as their core promoter structure differs, however further studies are needed to confirm this (Haberle *et al.*, 2019; Hadzhiev *et al.*, 2021; Muller *et al.*, 2010).

1.1.7 ZGA and Cell Cycle Interaction

ZGA in embryos coincides with a 'switch' in cell cycle mechanics. In most species, pre-ZGA divisions are fast and contain no gap phases associated with gene expression. At ZGA and beyond, cell cycles become slower. The temporal correlation of ZGA and the introduction of gap phases into the cell cycle, and its consequential slowdown, seemed to indicate that increasing cell cycle length can be causative for ZGA. While that was the case in *Xenopus* embryos, in other species manipulation of cell cycle length did not lead to transcriptional activation, with ZGA being independent of cell cycle manipulations (Kimelman *et al.*, 1987; Zhang *et al.*, 2014).

While cell cycle lengthening is not required to trigger ZGA, the co-incidence of these two processes provides an environment that enables increased levels of gene expression in embryo cells and provides more checkpoint-level control over DNA damage that can result from replication-transcription conflicts. Interactions between DNA replication and transcription in the early embryo were recently reviewed (Kermi *et al.*, 2017). In line with the notion of extensive transcription posing a DNA damage risk in early, fast-dividing embryos, only low levels of transcription are detected before ZGA, and even then the genes transcribed are very short, with few or no introns, and often late replicating (Heyn *et al.*, 2014; Swinburne and Silver, 2008). Experiments using an early gene *knirps (kni)* and its homologue *knirps-related (knrl)*

in *Drosophila* shown that while the 3kb gene encoding *kni* could be transcribed early, the 23kb-long *knrl* could not and resulted in a truncated product (Rothe *et al.*, 1992). Transcription in early embryo associates with S-phase (Hadzhiev *et al.*, 2019) and must therefore coincide with replication.

Interestingly, it was shown that ZGA, and its corresponding lengthening of cell cycles, is required for activation of replication checkpoints and DNA repair in embryos. Studies in *Drosophila* showed mutants of ATR and Chk1 kinases arresting in development and dying due to genome instability (Sibon *et al.*, 1999; Takada *et al.*, 2007), while blocking zygotic transcription led to a reduction in DNA damage – suggesting that transcription-replication conflicts are a major source of DNA damage in the early embryo, but also are to some extent "necessary evil", as transcription is needed to slow down cell cycles and activate repair checkpoints (Blythe and Wieschaus, 2015; Fogarty *et al.*, 1994; Sibon *et al.*, 1997).

As DNA repair and S-phase checkpoints have been found to be inefficient in embryos (Desmarais *et al.*, 2012; Raff and Glover, 1988), it may make them a particularly enticing model to study mechanisms of DNA repair. Fast dividing cells, driven to divide as fast as possible by internal factors, having to co-ordinate very fast replication with transcription, and showing deficiencies in DNA repair mechanisms show many similarities to cancers. Study of the complex relation between transcription and replication in embryos not only unveils the principles governing them, but can also shed light on similar processes relevant in cancer formation and maintenance, and thus give direction to research on cancer treatments.

1.2 Zebrafish as a Model Organism

1.2.1 Overview

Zebrafish (*Danio rerio*) is one of the most prevalent model organisms used in a range of biological areas, valued particularly as a model in studies of embryonic development. They are small freshwater fish from the Cyprinidae family of the teleost class. They are commonly used as aquarium fish due to their small size (<120mm length of adults), easy maintenance and particular bright and dark striped pattern of pigmentation (hence "zebra"-fish). They are not the only commonly used fish model, others include medaka (*Oryzias latipes*), pufferfish (*Takifugu rubripes*) and green spotted pufferfish (*Tetraodon nigroviridis*). Zebrafish remains very popular as a model organism and is used extensively in various branches of biomedical research (Parichy, 2015; Varga *et al.*, 2016).

Adult zebrafish are easy to maintain and breed, with hundreds of embryos being laid in each mating cycle. Embryos develop very rapidly *ex utero*, forming a free-feeding, free-swimming larva in a period of 5 days, and reaching sexual maturity in approximately 3 months (Kimmel *et al.*, 1995; Parichy *et al.*, 2009). Despite their generation time being similar to that of mice, zebrafish lend themselves better to studies of vertebrate development, as they remain transparent through most of the development and organogenesis, allowing better utilisation of microscopy techniques and non-invasive employment of *in vivo* microscopy, a significant advantage over other developmental biology models. Due to the external development and relative sturdiness, zebrafish embryos are also very easy to

manipulate in large numbers, with microinjections into the yolk or cells being a commonly used method of delivery of factors. Zebrafish embryos can also be easily used in high throughput drug or toxicology studies, as many drugs can be dissolved into the water. Zebrafish genome has been fully sequenced and contains over 26000 protein-coding genes on 25 chromosomes. Over 70% of human genes have at least one orthologue in zebrafish (Howe *et al.*, 2013), justifying usage of zebrafish as models of human disorders and suggesting high evolutionary conservation of biological mechanisms between these species.

A repository of zebrafish gene expression patterns, alongside information about genes, mutations and phenotypes, available mutant lines, antibodies and morpholinos, as well as publications has been created, and is actively updated and curated as a database at Zebrafish Information Network (zfin.org) (Bradford *et al.*, 2022), providing an invaluable source of information for researchers.

1.2.2 Zebrafish Embryonic Development Overview

Zebrafish embryonic development has been extensively studied and well described (Kimmel *et al.*, 1995). A brief overview of the early stages of the zebrafish development, pertinent to the scope of this thesis, will be given below. Zebrafish eggs are laid by the mother, protected by a thin layer of chorion, and fertilised externally by the father's sperm. After fertilisation, the egg becomes activated in an extended first cell cycle, lasting around 20-25 minutes, before moving through multiple cycles of rapid (10-15 mins) and synchronous cell divisions. During the early cleavage stages, cells sit on top of the yolk, positioned at the animal pole of the embryo (Fig. 1.5 A-D). The formed blastula progresses through synchronous cleavage

cell cycles until reaching 1000-cell stage (referred to as 1k cell stage, with approximately 1024 cells) at approximately 3 hours post fertilisation (Fig. 1.5 E). At that point, as mentioned before, main wave of ZGA is observed, and blastomeres lose synchrony while individual cell cycles increase in length. Blastomeres continue to divide at the animal pole (Fig. 1.5 F, dark grey) and cell migration starts occurring after around 4 hours of development, marking the beginning of gastrulation and a process called epiboly, in which the epiblast (dark grey in 1.5 G) forms and begins the directed cell migration to envelop the yolk cell. Fig. 1.5 G and H show that cell migration, as deep blastomeres migrate towards the surface and start forming a ring structure centred directly around the animal pole (germ ring), while the epiblast continues to envelop the yolk. Some of the leading cells of the epiblast start to involute and extend, forming the embryonic shield (Fig. 1.5 I, the thickening on the right side), while the epiboly continues as the epiblast envelops the yolk (Fig. 1.5 J-K).

The involution of cells continues from the embryonic shield, forming the anterior-posterior axis in the process of gastrulation, as the 3 embryonic germ layers of ectoderm, mesoderm and endoderm are formed (Kimmel *et al.*, 1995; Schier, 2001). At around 10 hpf (Fig. 1.5K) epiboly completes and a period of segmentation starts in zebrafish embryos (Fig. 1.5 L-M), during which first organogenesis and neurogenesis takes place, as well as formation of somites, while a tail starts forming and extending (Fig. 5N). Around 22 hours post fertilisation most somites have formed and the embryo enters the pharyngula stage (Fig. 5O), a phylotypic stage in zebrafish development. At this point embryos are still confined to the chorion and are relatively immobile, however as organogenesis continues, the heart starts to beat,

fins start to appear and pigmentation develops at the surface of the embryo, embryos hatch from their chorions around or before 48 hours post fertilisation (Kimmel *et al.*, 1995; Westerfield, 2000). At around 72 hours zebrafish embryos reach a larva stage (Fig. 1.5 P), at which point the swim bladder inflates and around 120 hpf (5 days post fertilisation) they become free-feeding organisms.

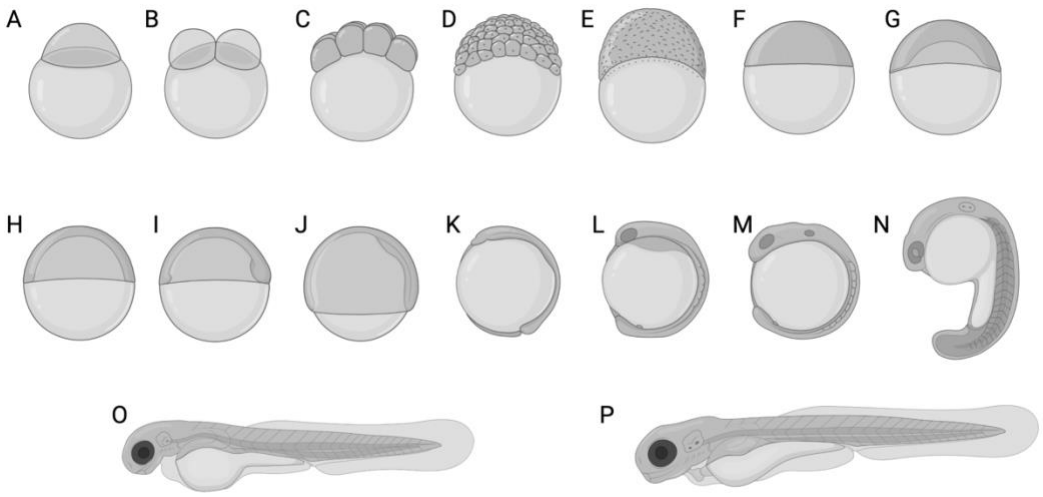


Fig. 1.5. Zebrafish embryonic development stages.
A panel of not-to-scale cartoon sketches of zebrafish embryo development, starting from the (A) 1-cell-stage zygote (fertilised egg) and progressing to (P) protruding mouth embryo 72 hours post fertilisation. Lighter grey is used to depict the yolk cell. Stages and their approximate time post fertilisation when kept at 28.5°C as described by (Kimmel *et al.*, 1995) are given below:
(A) 1 cell stage embryo (20-25 mins post fertilisation), (B) 2-cell stage (0.5 hpf), (C) 8-cell stage (1.25 hpf), (D) 128-cell stage (2.25 hpf), (E) 1k-cell stage (3 hpf), (F) sphere stage (4 hpf), (G) dome stage (4.33 hpf), (H) 50% epiboly stage (5.25hpf), (I) shield stage (6 hpf), (J) 75% epiboly stage (8 hpf), (K) bud stage (10 hpf), (L) 5 somites stage (11.66 hpf), (M) 10 somites stage (14 hpf), (N) 21 somites stage (19.5 hpf), (O) prim-5 stage (24 hpf), (P) protruding mouth stage (72 hpf). Figure created with BioRender.com.

1.2.3 Transgenesis in Zebrafish

Zebrafish have also proven relatively easy to genetically manipulate with a vast selection of previous overexpression, knockdown and knockout studies, which contributed to understanding of gene function and disease mechanisms (Meyers, 2018). Forward genetic screens were common in early days of zebrafish research, identifying numerous mutations and phenotypes, including many developmental defects (Amsterdam *et al.*, 1999; Grunwald *et al.*, 1988; Kimmel *et al.*, 1989).

Reverse genetics adopt an inverse approach, wherein a known genetic target is disrupted in order to assess its function in the organism. As is the case with most model organisms, common methods of targeted gene disruption and mutagenesis have been adopted to use with zebrafish (reviewed in (Sertori *et al.*, 2016)). Gene

knockouts can be achieved through use of modified nucleases that can be targeted to the target gene by sequence homology. These tools include zinc finger nucleases (ZFNs) (Ekker, 2008), and transcription activator-like effector nucleases (TALENs) (Hwang *et al.*, 2014; Sander *et al.*, 2011; Sertori *et al.*, 2016).These tools have been successfully utilised to create knockouts of multiple genes, including a knockout of *nanog* (Gagnon *et al.*, 2018), however the costs and laboriousness of their use prevented their widespread popularity. That was mostly due to the rise in popularity of CRISPR-Cas9 methods. CRISPR stands for clustered regularly interspaced short palindromic repeats, and it is a bacterial and archaeal system of adaptive immunity against viruses, especially bacteriophages. In brief, small fragments of phage genome become incorporated into a cluster of palindromic repeats, which can be transcribed to produce short RNAs. These short, 20-22nt RNAs, termed guide RNAs, form a complex with a DNA endonuclease Cas9, that uses the guide RNA sequence as a template to bind to complementary DNA, at which point a double strand break is formed. This form of bacterial defence has been engineered and adapted for use in eukaryotes to target virtually any genomic sequence, as long as a valid guide RNA can be generated (Jinek *et al.*, 2012). Due to inefficient DNA repair of double strand breaks by non-homologous end joining (NHEJ) or homologous recombination pathways, targeted use of Cas9 often results in mutations in the targeted locus. Most common Cas9-induced mutations are insertions or deletions of single nucleotides, often leading to frameshift mutations(Allen *et al.*, 2018; Naert *et al.*, 2020). The method has quickly been adapted for use in zebrafish and has been commonly used since, thanks to relative easiness of generating guide RNAs and rapidly decreasing costs (Hruscha and Schmid, 2015; Hwang *et al.*, 2013; Jao *et al.*, 2013; Liu *et al.*, 2019).

Constant advancements in the CRISPR technology and its high adaptability provides zebrafish researchers with many new tools of gene disruption, as Cas9 can also be used to e.g. disrupt transcription by sterically interfering with the gene locus, displacing transcriptional machinery (CRISPRi).

CRISPRi is a particularly exciting development for the zebrafish community. Commonly used approaches of RNA interference (RNAi) using short interfering RNAs (siRNAs) or short hairpin RNAs (shRNAs) offered mixed results, leading to efficient gene knockdown (Liu *et al.*, 2005), but also causing unspecific developmental defects in embryos, later attributed to the loss of a key contributor to maternal-to-zygotic transition *miR-430* (Zhao *et al.*, 2008). Due to this effect, RNAi is rarely used in studies of zebrafish embryonic development, with alternative tools for gene knockdown favoured (Giacomotto *et al.*, 2015).

Morpholino oligonucleotides (referred to as morpholinos) are commonly used in zebrafish research as such an alternative. Those short, chemically modified antisense oligonucleotides (20-25 nt) bind to complementary mRNA and are able to block its translation, either by hindering correct splicing (if bound to the splice sites in pre-mRNA) or by hindering ribosome assembly on the mRNA (if bound to the 5' untranslated region) (Nasevicius and Ekker, 2000). Morpholino-bound mRNAs can then be targets of RNAse H – as they are recognised as DNA-RNA hybrids and degraded. This leads to a knockdown of expression without disruption to the genome (Bill *et al.*, 2009; Nasevicius and Ekker, 2000). Morpholinos have been commonly used *in vivo*, injected into 1-cell stage zebrafish embryos, and they were successfully able to phenocopy many of the known gene depletion phenotypes, leading to its widespread use by zebrafish community (Corey and Abrams, 2001; Stainier *et al.*,

2017). Morpholinos do, however, have some limitations. Morpholinos are transient and only produce a fully penetrant phenotype in the first 48 hours after injection. There were also common occurrences of non-specific phenotypes resulting from off-target or unspecific morpholino binding, most notably causing activation of the p53 pathway and subsequent apoptosis (Bedell *et al.*, 2011; Lai *et al.*, 2019; Robu *et al.*, 2007). More recently, it was shown that mutants generated with site-specific nucleases (such as TALENs or CRISPR-Cas9) did not produce phenotypes that were shown by multiple morpholinos, especially those targeting genes involved in early embryonic development (Kok *et al.*, 2015). These morphant phenotypes are therefore likely to be misattributed to the target genes due to an off-target or unspecific effect. It is therefore important to approach morphant results with caution and seek verification using alternative tools.

In the context of morpholino efficiency, it is worth mentioning genetic compensation (El-Brolosy and Stainier, 2017). In short, genetic compensation refers to a phenomenon where a loss of function mutation in one gene results in an increase in expression of related genes to restore normal cellular function (Rossi *et al.*, 2015). In zebrafish embryos, studies have shown that genetic compensation can occur in response to mutations in genes involved in developmental processes, such as axis patterning and segmentation (Rossi *et al.*, 2015). The mechanism of compensation involves activation of gene networks that functionally replace the lost gene, and can result in the preservation of embryonic development. This mechanism occurs only when mutant mRNA is decaying or degraded (El-Brolosy *et al.*, 2019). Genetic compensation can serve as a potential explanation why morpholino-directed RNA knockdowns could exhibit phenotypes that were different or more severe than the

targeted gene knockouts. Knockdowns cannot trigger the genetic compensation pathways, while a subset of knockout mutants (producing mRNA product of its gene) can. This opens more questions regarding what can be considered a "true" phenotype of a mutant – the one with or without genetic compensation – and how would transcriptional adaptation of some mutants affect inferring function from mutated genes. An argument could be made that the most precise method of generating loss-of-function mutations is a full knockout of a gene, preventing any transcription taking place, thus making it impossible for genetic compensation pathways to activate.

1.2.4 Zebrafish and the Toolset For ZGA Studies

Zebrafish have been used extensively as a model for studies of embryonic development thanks to their significant advantages over other model organisms. As the time between fertilisation and ZGA is only 3 hours, and zebrafish embryos are laid in hundreds, this gives a good opportunity for large-scale studies, especially when compared to rodents. In addition, lack of pigmentation and embryo transparency allows easy detection of phenotypes as well as use of microscopy to observe and study development, lending zebrafish well to use of fluorescence markers in particular (Ko *et al.*, 2011), which can and have been used in tracking transcription in zebrafish. Zebrafish embryos can easily be used in *in vivo* studies using live microscopy without any hindrance to their development (Keller, 2013). An example of that is Light-Sheet Microscopy, where zebrafish embryos can be mounted in chorions and imaged in 3 dimensions for multiple hours (Abu-Siniyeh and Al-Zyoud, 2020), or even days (Wragg, unpublished data).

Early development in zebrafish is not identical to that seen in mammals, with much faster cell cycles and ZGA occurring after multiple rounds of cell divisions. However, zebrafish genome contains a significant number of genes homologous to their mammalian counterparts, and can be a powerful model in discerning the basic mechanisms governing embryogenesis and transcriptional activation. Studies like (Lee *et al.*, 2013) showing homologues of human stem cell pluripotency factors (Nanog, Pou5f3, Sox1b) playing a key role in activating transcription and reprogramming of embryo cells into pluripotency in zebrafish embryos are encouraging examples of how zebrafish can be used to model ZGA despite the differences to mammalian ZGA.

The rise of next generation sequencing offered new insights into studies of ZGA with possibilities to study chromatin openness (ATAC-seq) (Palfy *et al.*, 2020) and conformation (3C, 4C, Hi-C) (Kaaij *et al.*, 2018), binding of proteins to chromatin and enrichment of histone marks (ChIP-seq) (Vastenhouw *et al.*, 2010; Zhu *et al.*, 2019), analyse the transcriptome using RNA-seq and its multiple developments (Heyn *et al.*, 2014; Vejnar *et al.*, 2019; White *et al.*, 2017), study mechanisms of promoter usage and architecture using CAGE-seq (Haberle *et al.*, 2014a; Nepal *et al.*, 2013), as well as many other genomic and epigenomic tools emerging in recent years (Akdogan-Ozdilek *et al.*, 2020). The significant limitation of sequencing studies in embryos is the small number of available cells, difficult especially when subsets of cells are studied (Wragg and Müller, 2016). Zebrafish embryos are much better suited for that application than, for example, mouse embryo due to larger number of sibling embryos in each lay, as well as the easiness of manipulation of multiple embryos. However, those relatively small cell numbers, as well as the intrinsic variabilities

coming with use of *in vivo* models still must be taken into account and can hinder studies in comparison to use of cultured cells.

On the cellular level, zebrafish development might not be as tractable as that of *C. elegans*, but new studies using single-cell genome sequencing are able to discern cell populations with particular cell fates and expression profiles, aiming to create lineage maps of those populations (Jiang *et al.*, 2021; Tatarakis *et al.*, 2021; Wagner *et al.*, 2018). Some studies even managed to visualise and track individual mRNA localisation across space and time by combining single-cell RNA sequencing with embryo cryosectioning (Holler *et al.*, 2021; Liu *et al.*, 2022a), providing even more information on the arising cell fates and their relation to maternal mRNA distribution in the embryo.

New methods have also enabled studies into the nascent transcription of mRNAs in embryos, identifying which genes, when, and (in case of single cell studies) in which cells genes are first transcribed. Various methods of labelling the nascent RNAs produced in the embryo can be used to distinguish zygotically produced transcripts in sequencing or in microscopy (Schulz and Harrison, 2019). The methods include exposing embryos to modified ribonucleotides such as 4-thio-UTP (4-sUTP) (Heyn *et al.*, 2014) or 5-ethynyluridine (EU) (Chen and Good, 2020) that get incorporated only into the zygotically transcribed mRNAs and which can later be isolated for sequencing (Fig. 1.6 A) or specifically imaged. Another method allows detection of nascent transcription by use of MS2-based reporters, RNA of which can be bound by MS2 coat protein fused with a fluorescent marker (Fig. 1.6 B) (Campbell *et al.*, 2015). Advent of CRISPR/Cas9 technologies has also allowed use of catalytically dead Cas9 nuclease (dCas9) as a fluorescent fusion protein to bind directly to mRNAs of interest

and fluorescently mark them (Fig. 1.6 C) (Chan *et al.*, 2019; Nelles *et al.*, 2016). Specifically modified morpholino oligonucleotides bound with FITC fluorophore have also been used to detect transcripts of specific target genes during early development in zebrafish and allowing to visualise nascent transcription of *miR-430* as early as 64 cell stage (Hadzhiev *et al.*, 2019).

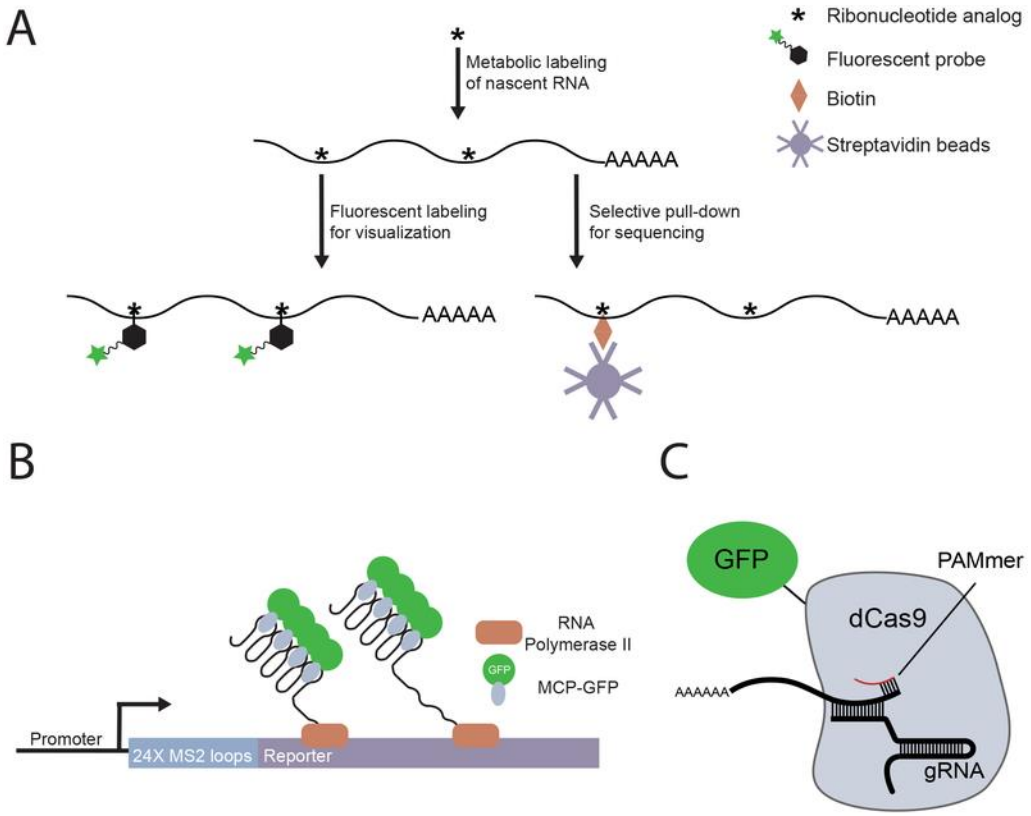


Fig. 1.6. A schematic of methods used to visualise nascent transcription in zebrafish.
(A) Detection of nascent transcription by metabolic labelling using ribonucleotide analogs such as 5-EU or 4sUTP, showing steps of labelling followed by fluorescence labelling or pull-down for sequencing. (B) Detection of a nascent transcript by MS2 based reporters. Through transgenesis, a construct containing a reporter gene of interest and sequence coding for multiple MS2 motifs is introduced into the genome. MS2 sequences form stem loops when transcribed, and they are transcribed alongside mRNA of interest. MS2 stem loops are bound by a MS2 capture protein (MCP) fused with GFP. (C) Detection of transcripts using dCas9-GFP fusion. dCas9-GFP is able to bind the target mRNA through complementary binding of a specifically designed guide RNA (gRNA) and an additionally provided DNA oligo that contains the PAM motif (PAMmer), which is required for Cas9 binding to mRNA of interest, as Cas9 traditionally binds to double-stranded DNA. The fusion of dCas9 and GFP is utilised for imaging. Adapted from (Schulz and Harrison, 2019).

These methods of detection of nascent RNA complement well with established methods of detection of mRNAs in zebrafish embryos, such as fluorescent in situ hybridisation (FISH) in which an antisense modified probe is generated to target an mRNA of interest, and later visualised through colourimetric visualisation (Thisse and Thisse, 2008) or by fluorescence detection (Welten *et al.*, 2006). This method has been particularly useful in establishing anatomical patterns of gene expression throughout zebrafish development. A modification of this method, termed single-molecule FISH (smFISH) is used for quantitative detection of mRNAs in individual cells (Oka and Sato, 2015), as opposed to more qualitative detection using classic FISH.

As in other models, tools like quantitative reverse transcription polymerase chain reaction (qPCR) are at the disposal of researchers to measure relative abundance of mRNAs of interest.

1.3 Mxtx2 in Zebrafish Development

Most of the known regulators of ZGA are maternally provided and often act as chromatin regulators as well as transcription activators – pioneer factors. However, the role of early expressed genes in ZGA has not been as extensively studied. In zebrafish embryos, a number of those genes were identified and their early expression quantified across early development stages (Heyn *et al.*, 2014). miR-430 the most expressed zygotic transcript and its role in MZT has been well described, but some other interesting candidates were identified. One of these was a transcription factor Mxtx2, or mix-type homeobox gene 2. It is a purely zygotic (i.e. not maternally provided) factor, expressed first in the 256 cell-stage (2.5 hpf) embryo

(Heyn *et al.*, 2014) – 2 cell cycles before main wave of ZGA. Mxtx2 is encoded by a short (1537 bp) gene located on chromosome 12 and containing few introns (Fig. 1.7 A), matching the previous description of early-expressed zebrafish genes. Indeed, *mxtx2* is expressed early, with a peak around dome stage (4.33 hpf), and finishes at approximately shield stage (6 hpf), during epiboly (Fig. 1.7 B). Previous studies have implicated *mxtx2* to be involved in induction of embryonic germ layers of mesoderm and endoderm through an interaction with the Nodal signalling pathway (Hong *et al.*, 2011; Xu *et al.*, 2012), and also as the gene responsible for induction of first morphogenetic cell movements in the embryo – epiboly (Wilkins *et al.*, 2008).

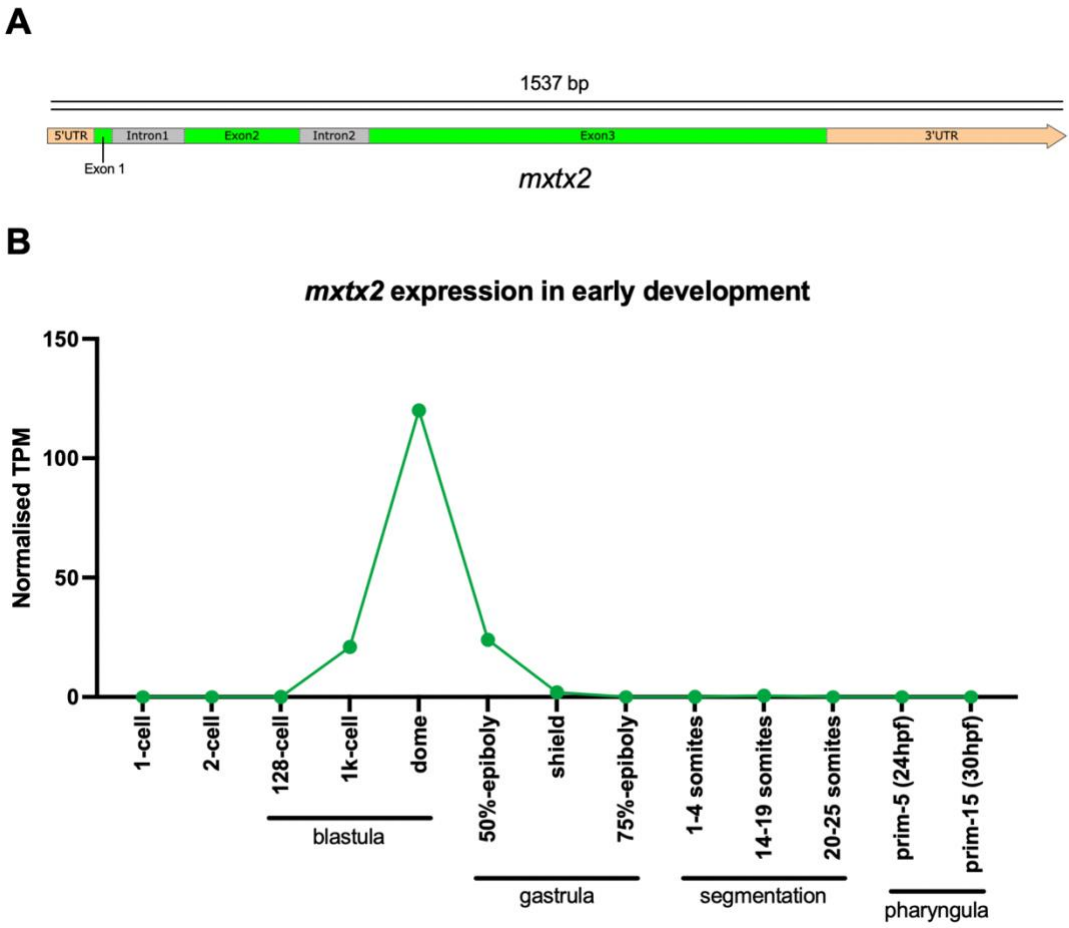


Fig. 1.7. An overview of mxtx2 gene structure and expression

(A) Schematic representation of the structure of the *mxtx2* gene, including UTR's (orange), exons (green) and introns (grey) to scale. Arrow direction indicates the gene orientation on the coding strand of the chromosome 12. Image obtained using SnapGene software. (B) Timing of *mxtx2* expression in early development, across multiple stages ranging from 1-cell to the pharyngula stage of prim-15 (30 hpf), indicated by RNA-seq data (White *et al.*, 2017) from 5 biological replicates. Gene expression is measured by transcripts per million (TPM), normalised by authors. Original datasets available from EMBL Expression Atlas.

1.3.1 Mxtx2 as a Homeobox Protein

Mxtx2 contains an N-terminal homeobox domain (helix-turn-helix DNA binding domain), allowing it to bind to sequence-specific sites on DNA, and it is able to act as a transcriptional activator. Genes containing mix-type homeodomains have been previously implicated in several roles in embryonic development, such as coordination of axis formation and determination of particular cell fates, including that of mesendoderm – later forming the germ layers of mesoderm and endoderm (Pereira *et al.*, 2012). There has been no previous research indicating that Mxtx2 could possess a chromatin remodelling capacity, and it is not known whether it could function as a pioneer factor.

Many of the key transcriptional activators such as Nanog, Pou5f3 (Oct4), murine DUX and human DUX4 are also members of the homeobox family of DNA-binding proteins. However, this is a very broad family of proteins sharing a conserved domain of 60 amino acids, allowing DNA binding and regulation of gene expression. Nevertheless, homeobox proteins function in a variety of processes. Presence of a homeobox domain in Mxtx2 does not therefore reveal much detail about its function. BLAST alignment of Mxtx2 amino acid sequence to proteins of other model organisms does, however, show an interesting pattern. Among the genes showing the highest percentage identity to Mxtx2 are human DUX3, DUX4 and DUX5, murine homeobox proteins SEBOX and MIXL1, *Xenopus laevis* pituitary homeobox 3 protein (Pitx3), and zebrafish Mxtx1 – paralogue of Mxtx2 (Fig. 1.8 A). As most of the identity between those proteins stemmed from the homeobox domain, I have decided to align and compare the 60 amino acid homeobox domains of zebrafish Mxtx2, Mxtx1 and Nanog, as well as homeobox domains present in murine Dux and human DUX4

(each of these protein contains 2 homeobox domains). Dux, DUX4 and Nanog were selected to compare functional domains of known (or candidate) homeobox stem cell pluripotency factors to that of Mxtx2. When aligning the sequence of homeobox domain of zebrafish Mxtx2 to functional domains of murine Dux and human DUX4 (each containing 2 homeobox domains), we can observe 30-56% identity between these (Fig. 1.8 B). There are, however, several stretches of conserved residues shared between Mxtx2, Dux and DUX4 that could suggest a degree of shared function. It would require further structural analysis (e.g. using Alphafold software to visualise protein folding) to discern the role of these residues in DNA binding and whether these serve as key residues for homeobox function. The consensus DNA binding motif of Mxtx2 has been proposed based on ChIPseq analysis in Mxtx2 overexpression experiment (Xu *et al.*, 2012) (Fig 1.8 C).

Mxtx1 has been mentioned as an Mxtx2 paralogue in the sequence alignment analyses(Fig. 1.8 A-B). Likely formed as a result of a genome duplication during teleost ancestral evolution (Howe *et al.*, 2013), Mxtx1 is another mix-type homeobox transcription factor, and is involved in cell migration and heart morphogenesis (Langenbacher *et al.*, 2012; Pereira *et al.*, 2012; Sakaguchi *et al.*, 2006). It is only transcribed after ZGA, suggesting its lack of involvement in ZGA regulation.

Zebrafish Mxtx2 functional domain also shares around 30% identity with that of zebrafish Nanog (Fig. 1.8 B), and additionally has been annotated on ZFIN (https://zfin.org/ZDB-GENE-000710-6, accessed 09/08/2022) as orthologous to human DUX4L5, DUX4L6 and DUX4L7, members of the DUX4 cluster.

All of this suggests *mxtx2* as an interesting target to study further, with the relationship with contested genome activators Dux and DUX4 suggesting a broader

role for Mxtx2 in the context of embryonic development, potentially even one affecting the zygotic transcription.

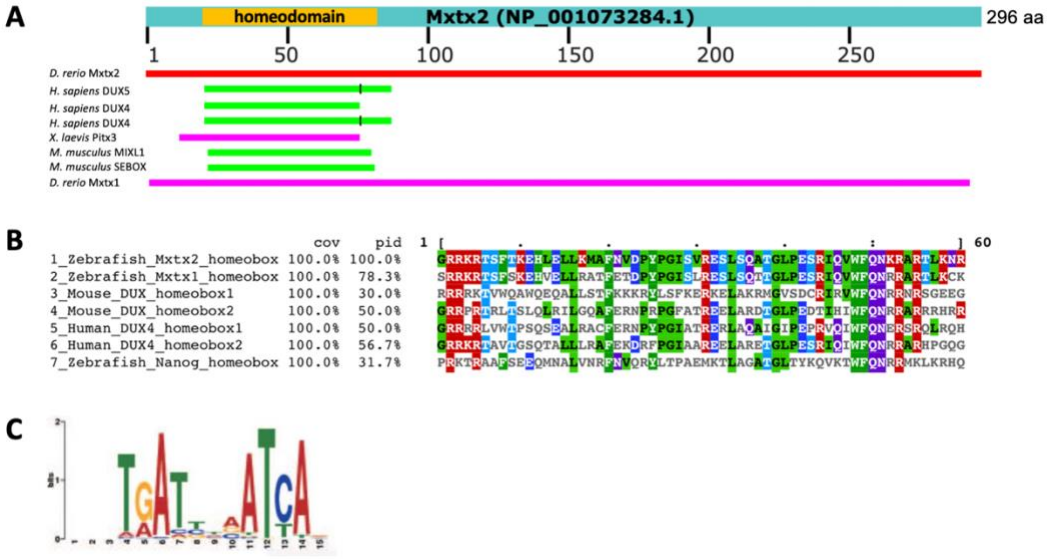


Fig. 1.8. Zebrafish Mxtx2 holds similarity to human DUX4

(A) Results of BLASTP (blast.ncbi.nlm.nih.gov) alignment of zebrafish Mxtx2 amino acid sequence to the non-redundant GenBank protein database. Default alignment settings were used, with the exception of limiting the search to model organisms: human, mouse, *Xenopus laevis* and *tropicalis*, *Drosophila melanogaster* and zebrafish. Results were sorted by the percent identity score. Among the top 50 proteins by percent identity, several proteins of interest were selected (including Mxtx2 as a control) and Graphical Summary of these alignments is presented. Coloured lines depict positions at which respective proteins align and share identity with Mxtx2. Colour of the line indicates the BLASTP alignment score (higher score = higher sequence identity), with green indicating score of 50-80, pink indicating score of 80-200, and red indicating score of 200 or higher. (Altschul *et al.*, 1997) (B) Multiple sequence alignment (T-Coffee) of amino acid sequences of homeobox domains of previously mentioned proteins. In line 7, domain of a more distant member of the homeobox family - zebrafish Nanog - was aligned. Zebrafish Nanog provides information about conservation of the domain within zebrafish, showing identity in several residues, acting as a protein family and species 'control'. Coloured background denotes identity of a residue at that position to the Mxtx2 homeobox domain. "cov" refers to percentage coverage of alignment (it is 100% as all homeobox domains are 60 amino acids), while "pid" refers to the calculation of percentage identity between amino acids sequences, as calculated by the T-Coffee algorithm. (C) The consensus sequence of Mxtx2 DNA binding site, resulting from aligning the 1000 top DNA binding regions of an Mxtx2 ChIPseq analysis done in *mxtx2-Myc* mRNA injected embryos in dome (4.33 hpf) stage. Image from Xu *et al.* 2012, Figure 6B.

1.3.2 Mxtx2 Function and Relation to the Nodal Pathway

Interestingly, zebrafish Mxtx2 has been shown to act downstream and be regulated by maternal Nanog in early development. Maternally deposited Nanog is required for transcription of *mxtx2* in the early embryo and a Nanog binding site has been found at the *mxtx2* locus (Veil *et al.*, 2018; Xu *et al.*, 2012). Several roles of *mxtx2* have been found using a study with morpholino oligonucleotide knockdown of its function. Mxtx2 was found to be a factor necessary and sufficient for induction of one of the first separated cell fates in the embryo – that of yolk syncytial layer (YSL), which will be described later in this thesis. ChIP-seq studies using an overexpression of *mxtx2* found that this transcription factor is able to bind 1751 genes at dome stage (after ZGA), and binds approximately 44% of all YSL specific genes (Xu *et al.*, 2012). The same study found that Nanog functions as an upstream activator of *mxtx2* expression, and that Mxtx2 binds Nodal-related effectors (*ndr1* and *ndr2*) in the zebrafish embryo, likely being responsible for activating their transcription. It was suggested that a direct Nanog-Mxtx2-Nodal pathway of transcriptional activation is responsible for expression of Nodal genes in the YSL (Fig. 8D), which would explain its implication in specification of mesendoderm, as well as the epiboly-specific phenotype.

Nodal is a member of the transforming growth factor beta family of signalling proteins and has been shown to play a pivotal and essential role in formation of a precursor to primary germ layers of mesoderm and endoderm – the mesendoderm (Bennett *et al.*, 2007; Conlon *et al.*, 1994; Schier, 2009). In simple terms, Nodal is a secreted protein that, after being modified by convertase enzymes is a ligand to

membrane receptors, that also act as serine/threonine kinases. In zebrafish, there are 3 secreted Nodal proteins – nodal-related 1 (*ndr1)*, nodal-related 2 (*ndr2)*, and nodal-related 3/southpaw (*ndr3/spaw*). *ndr1* is provided maternally, *ndr2* is an early-expressed purely zygotic gene, and *ndr3/spaw* is only expressed after gastrulation (Xing *et al.*, 2022). Nodal-activated receptors in turn phosphorylate a number of intracellular targets, a main group of which are Smad proteins. Phosphorylation of Smad2/Smad3 transcription factors by activins results in their association with Smad4 protein and translocation to the nucleus, where Smad2/3 can bind their genomic targets and regulate transcription, but also cooperate with multiple binding partners including other transcription factors (Hill, 2018; Ross and Hill, 2008; Schier, 2009). Active Smad complexes are very tightly regulated through these partnerships, and can function through chromatin remodelling and histone modifications to modify transcriptional output of many genes (Ross *et al.*, 2006).

Effects of Nodal signalling overlap functionally with function of Eomesodermin A (*eomesa*), a maternally provided T-box transcription factor implicated in epiboly initiation and progression, as well as mesoderm and endoderm specification, similarly to *mxtx2* (Nelson *et al.*, 2014). Eomesa likely forms a complex with Smad2 in zebrafish, responding to Nodal signalling. Interestingly, Eomesa has been found to bind the cis-regulatory modules of *mxtx2* and suggested to be an activator of its expression (Nelson *et al.*, 2017; Xu *et al.*, 2014). Additionally, overexpression of *eomesa* led to ectopic activation of *mxtx2* expression, with presence of both Eomesa and Mxtx2 required for correct progression into epiboly (Bruce *et al.*, 2005). Taken

together, this suggests a potentially interesting regulatory loop in the regulation of mesendoderm induction and epiboly by those factors.

1.4 Epiboly and the Yolk Syncytial Layer

After the main wave of zygotic genome activation in zebrafish embryos, cells start to acquire changes in morphology and motility and start migrating, forming the first structures and beginning gastrulation. Those first cell movements indicate a start of epiboly, an evolutionarily conserved first morphogenetic event in embryo development. Fundamental work discovering and describing this process was done in a teleost *Fundulus heteroclitus* (Trinkaus, 1984), however most understanding came from studies in zebrafish embryos (Carvalho and Heisenberg, 2010).

Epiboly is the first major event in zebrafish development after ZGA and plays a crucial role in coordinated cell movements leading to gastrulation. Epiboly cell movements start approximately 4.3 hpf, just as dome stage forms, and carry on until bud stage (10 hpf). Onset of epiboly requires zygotic transcription taking place, inhibition of zygotic transcription leads to cells continuing to divide (for a period) at the animal pole past sphere stage, but not beginning epiboly (Kane *et al.*, 1996).

Epiboly in zebrafish has been a subject of interest, and recent developments have been reviewed (Bruce and Heisenberg, 2020). Here, I will give a brief overview of the epiboly events applicable to the subject of this thesis.

Just before epiboly starts, the zebrafish embryo consists of approximately 4000 cells sat on top of the yolk at the animal pole (sphere stage). This group of cells is referred to as blastoderm, with individual cells referred to as blastomeres. Surface blastomeres form a thin sheet of cells with an epithelial cell fate that already began

differentiation called the enveloping layer (EVL), covering the cell mass. The yolk cell is multinucleated, and the nuclei form a syncytium within the yolk (membrane-less multinucleated phase with shared cytoplasm), a layer just below the cell mass – YSL. The individual nuclei forming YSL are known as yolk syncytial nuclei (YSN) (Kimmel *et al.*, 1995). First panel of Fig. 1.9 provides a schematic overview of the embryo anatomy before the onset of epiboly.

As epiboly starts, the yolk cell domes, EVL cells begin to flatten and spread along the surface, deep cells begin migrating towards the embryo surface in the animal "hemisphere" of the embryo (radial intercalation), and YSL begins to narrow along its latitude. The YSL layer begins to form 2 distinct populations. External YSL (E-YSL) is located on the margins of the spherical embryo along the latitude, forming an interface with EVL, while internal YSL (I-YSL) forms an interface with deep cells and begins to involute underneath them. As epiboly progresses, I-YSL expands (Fig. 1.9 - dome and 50% epiboly panels). Radial intercalation of deep cells in the first stages of epiboly has been associated with the establishment of the first morphology changes in these blastomeres, as well as the first axis polarity gradients (Bruce and Heisenberg, 2020; Morita *et al.*, 2017).

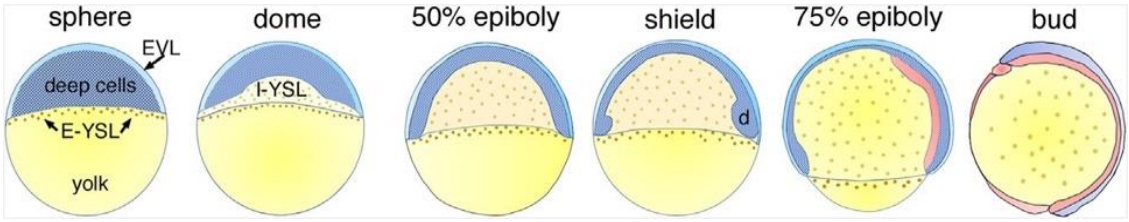


Fig. 1.9. Schematics of zebrafish epiboly.
Lateral views of embryos at selected stages of epiboly. Sphere stage (4 hpf) – blastoderm is present at the animal pole of the embryo and forms a flat interface with the yolk cell (yellow) along the embryo longitude. Enveloping layer (EVL) is marked in light blue and deep cells in dark blue, yolk syncytial nuclei are marked with brown dots, and external yolk syncytial layer (E-YSL) is indicated with arrows.

Fig. 1.9. (cont.) Dome stage (4.3 hpf) - epiboly starts with yolk cell doming and deep cells start their radial intercalation towards the animal pole. Yolk syncytial nuclei start to form the internal yolk syncytial layer (marked I-YSL) underneath the deep cells. During epiboly progression, the yolk syncytial layer and blastoderm move vegetally. 50% epiboly (5.3hpf) - the blastoderm has reached the embryo equator. Shield stage (6hpf) – embryonic shield forms on the dorsal side of the embryo (marked with "d") 75% epiboly (8hpf) - blastoderm and yolk syncytial layer continue to progress towards the vegetal pole, with the EVL and blastoderm spreading across the yolk cell. Mesendoderm (marked in red) is internalized underneath the migrating blastoderm on the dorsal side. Bud stage (10 hpf) - epiboly and gastrulation are complete with the entire yolk cell encapsulated and the YSL and blastomeres closing together at the vegetal pole. Mesendoderm is marked in red, while ectoderm is marked in blue. Figure adapted from (Bruce and Heisenberg, 2020).

As epiboly progresses, external yolk syncytial layer and blastoderm gradually move towards the vegetal pole, with pulling forces generated by yolk microtubules and actomyosin filaments (Bruce and Heisenberg, 2020). An actomyosin ring forms in the E-YSL, contracting around the circumference of the embryo and generating a tension gradient pulling the attached EVL, causing it to spread around the surface (Bonneau *et al.*, 2011; Hernández-Vega *et al.*, 2017). Migration of deep cells and their intercalation and spreading across the surface reduces the thickness of this layer to just 2-3 cells. After 50% epiboly, at which point the E-YSL have reached the embryo equator, embryonic shield forms indicating the first distinctively dorsal structure. Dorso-ventral axis, however, starts to be formed in the embryo by Wnt/ß-catenin signalling much earlier in development, through actions of both maternal and zygotically expressed factors (Langdon and Mullins, 2011).

Yolk nuclei continue to travel towards the vegetal pole while constricting the actomyosin ring, while a subset of migrating deep cells on the dorsal side of the embryo involutes underneath the surface near the blastoderm margin, eventually forming mesendoderm (Fig. 9- 75% epiboly panel) and later mesoderm and endoderm (Warga and Kimmel, 1990). At bud stage, epiboly completes with the yolk

completely enclosed, and gastrulation completes with establishment of 3 germ layers – endoderm, mesoderm, and ectoderm (Fig.9 – bud stage panel).

Several studies have shown that the epiboly movements are dependent primarily on YSL and does not depend on EVL movements or deep cell movements (Trinkaus, 1951) (Bruce and Heisenberg, 2020; Morita *et al.*, 2017). The coordination between all those cell populations is not yet fully understood.

While gastrulation will not be an important theme of this thesis, it is an important process to contextualise the processes of zygotic genome activation and epiboly in the morphology changes during zebrafish development and the temporal positioning of those. Epiboly, however, is a process highly intertwined with maternal-to-zygotic transition and ZGA.

mxtx2 has been implicated as a gene required for expression of multiple factors present in the yolk syncytial layer. From sphere stage, *mxtx2* is most expressed in the YSL and at the blastoderm margin (Du *et al.*, 2012). Maternally provided transcription factor eomesodermin A, a regulator of *mxtx2* expression, plays an important role in the process of doming at the initiation of epiboly and formation of yolk microtubules, but does not affect subsequent epiboly progression (Du *et al.*, 2012). In contrast, *mxtx2*-depleted cells initiate epiboly movements, but have a disordered structure of YSL nuclei and fail to successfully complete epiboly progression, causing a yolk burst phenotype through the premature contraction around the margin (Wilkins *et al.*, 2008). This is however counterintuitive with findings showing that the F-actin ring formed around YSL is disrupted by *mxtx2* depletion (Wilkins *et al.*, 2008). The exact mechanism of how this contraction takes place in *mxtx2* morphants is unknown. The role of *mxtx2* in epiboly correlates well with its expression pattern in late blastula

following the ZGA, however it is still unknown what its role is in the early embryo, before the onset of epiboly.

1.5 Aims of This Thesis

The overall aim of this project is to contribute to the understanding of the mechanisms of zygotic genome activation in a vertebrate embryo model. I set out to focus on the role that early transcribed genes play in the zygotic genome activation and corresponding transformation of fully differentiated, transcriptionally quiescent cells into nearly totipotent stem cells. I specifically focused on the mechanisms of the main wave of genome activation in the nuclei of the yolk syncytial layer, which is a unique structure which represents, alongside the enveloping layer cells the first symmetry breaking and cellular differentiation events in the developing embryo and which is essential for coordination of axis determination, morphogenesis, and nutrient management of the developing embryo cells. My hypothesis has been that the distinct morphological characteristics of the YSL which appear during the start of the main wave of genome activation indicate distinct zygotic genome activation regulation mechanisms shaping these cells from that of the rest of the embryo. I aimed to study the global process of transcription regulation in this extra-embryonic structure and to better understand the regulatory pathways involved in zygotic activation of genes active in the YSL, in order to address the mechanisms of the main wave of genome activation. My second hypothesis has been that the *mxtx2* transcription factor, which is expressed during the minor wave of ZGA and becomes specific to the YSL during the main wave of ZGA represents a candidate

regulator of YSL genome activation transcription. In order to test these hypotheses, I set the following specific aims:

1. Generation and characterisation of loss of function of *mxtx2* gene to study its roles in early development, particularly during ZGA and in establishment of yolk syncytial layer:
    1.1. Generation and validation of loss of function of *mxtx2* in early development using gene editing and mRNA targeting approaches;
    1.2. Generation and characterisation of the null mutant phenotypes by studying embryonic and YSL morphogenesis, transcription, and transcriptomes.
2. Analysis of the dynamics of genome activation in the yolk syncytial layer:
    2.1. Temporal regulation of global genome activation in the yolk syncytial layer;
    2.2. Comparison of genome activation in the yolk syncytial layer to that in the embryonic blastomeres.

Chapter 2. Materials and Methods

2.1 General Molecular Biology Techniques

2.1.1 Nucleic Acids Extraction

2.1.1.1 Phenol-chloroform DNA/RNA purification

Embryos, collected at appropriate stages in appropriate numbers, were placed in 1.5 ml microfuge tubes. As much liquid as possible was removed from the embryos using a Pasteur pipette, and 50 µl of Embryo lysis buffer (10 mM Tris pH 8, 10mM EDTA, 200 mM NaCl, 0.2% Triton-X-100) was added to the microfuge tubes, and embryo structure was disturbed by pipetting up and down with the plastic tip. These were then incubated at 55˚C for 2-3 hours. After that, lysed embryos formed a cell extract, that was further used for nucleic acid extraction. Nucleic acids were phase-separated from embryo cellular extract in phenol-chloroform gradients. An equal volume (1:1) of basic phenol-chloroform (pH 8) for DNA or acidic phenol-chloroform (pH 6.5) for RNA was added to the cellular extract. After mixing, the solution was spun down at 12000 x g for 6 minutes or until the phase separation was achieved. The aqueous (top) phase was transferred to a new 1.5 ml Eppendorf tube and an equal volume of chloroform was added. After mixing and spinning down, 2 volumes of isopropanol and 1/10 of the volume of ammonium acetate 7.5 M (Sigma-Aldrich, A2706, UK) were added to the aqueous phase. DNA/RNA were then chilled over night at -20°C and precipitated in a cold centrifuge for 30 minutes at maximum speed. The pellet was washed once with 70% ethanol and resuspended in the desired amount of water.

2.1.1.2 DNA extraction

DNA was extracted from pooled embryos selected at an appropriate stage. For embryos co-injected with sgRNA targeting *gol* gene, selection of embryos was based on the levels of eye pigmentation observed at 48hpf.

DNA extraction was done using Qiagen DNeasy Blood & Tissue kit. Embryos were dechorionated and transferred to the ATL buffer solution containing proteinase K and lysed at 56°C. Remaining steps were done according to kit manufacturer's manual. Unless stated otherwise in the kit manual, centrifugation steps were done at 11000 x *g.* Concentration of extracted DNA was quantified immediately after the elution step. Extracted genomic DNA was stored long-term at -20°C.

2.1.1.3 RNA extraction

Total mRNA was extracted from pooled embryos collected at an appropriate stage. For early stage microinjected embryos, selection was based on the fluorescence coming from the dye included in the microinjection solution. For collections at sphere and dome stages, 30 embryos were pooled from each sample.

RNA extraction was done using Qiagen RNeasy Micro Kit. Selected embryos are transferred to the RLT buffer containing 2-mercaptoethanol (1% v/v) and disrupted by multiple passages through a fine (27G) hypodermic needle. Remaining steps of the procedure were done in accordance with the kit manufacturer's manual. DNase I digestion step, marked as optional in the manual, was included in every instance of RNA extraction. Unless stated otherwise in the kit manual, centrifugation steps were done at 11000 x *g*. Concentration of extracted total mRNA was quantified immediately after the elution step. Extracted RNA was stored long-term at -80°C.

2.1.2 Quantification of Nucleic Acids

Concentration of nucleic acids from all standard preparations was measured using a NanoDrop 2000 spectrophotometer, according to manufacturer's instructions. The concentration is determined based on 260 nm absorbance, while the absorbance curve and the 260/280 and 260/230 ratios were used to assess the sample purity (presence of contaminants).

2.1.3 Polymerase Chain Reaction (PCR)

PCR reactions were performed for purposes of amplification of nucleic acids. Reactions were performed using DNA Engine Tetrad 2 (Bio-Rad) thermocycler. For most applications, a Taq DNA polymerase (MyTaq Red Mix, Bioline) was used. For high fidelity purposes requiring low error rates, such as annealing sgDNA templates or amplification of fragments for Sanger sequencing, Q5 Hotstart High Fidelity Polymerase (New England BioLabs) was used. This system contains a proprietary polymerase with higher processivity and vastly reduced error rates compared to Taq polymerases. Reactions were prepared according to manufacturers' manuals. Annealing temperatures and extension times were adjusted based on the primers and template DNA properties.

A standard PCR reaction would be set up to consist of a hot start step (98°C), followed by 25-40 cycles of denaturation (in 98°C), annealing (in temperature adjusted to primers, most commonly 60°C) and extension (in 72°C). After cycles, template extension would be allowed to finish through a prolonged step (2-10 min, 72°C) and

the resulting DNA product would be applied to further techniques or stored short-term at 4°C.

2.1.3.1 PCR primers design

Primer pairs for PCR reactions were designed using Primer3 software (Untergasser *et al.*, 2012). Typically, 18-22 nt primer pairs with comparable GC content not exceeding the range of 30-70% (with the optimum of 50%), and with comparable melting temperature (ranging from 55-65˚C, with an optimum of 60˚C, not exceeding 2˚ difference within a pair) would be selected. Otherwise, standard settings for primer search were used in the software.

2.1.3.2 PCR clean-up

Following every PCR reaction, DNA products were purified using a NucleoSpin Gel and PCR Clean-Up Kit (Macherey-Nagel, UK). By column purification, it is ensured that excess dNTPs and enzyme are removed from the DNA product.

2.1.4 Agarose Gel Electrophoresis

Gel electrophoresis was used to determine nucleic acid quality and size. 1.0-2.0 % agarose gels were prepared by melting agarose powder in 1X TAE Buffer (40 mM Tris acetate, 1 mM EDTA, pH 8.3). 0.5 µg/mL of Ethidium Bromide (Sigma-Aldrich, E1510) was added to liquid agarose to allow visualisation of nucleic acids under UV light. Electrophoretic migration of DNA was conducted in a tank containing 1X TAE Buffer by applying 60-150V current until proper band separation was achieved.

2.1.5 Sanger Sequencing

Genomic DNA was collected from 5 selected embryos from an injection round 48 hpf using the Qiagen DNEasy Blood and Tissue Mini kit (Qiagen, 69504). Embryo selection was based on rhodamine fluorescence intensity or *gol* phenotype level. Regions of interest were amplified by a PCR reaction using primers spanning the region of interest.

Resulting DNA amplicons were sequenced using standard Sanger sequencing (service provided by Source Bioscience, UK). Sequencing runs utilised a custom sequencing primer designed within the amplicon sequence, positioned 200-600 bp distant from Cas9 target site in order to remain unaffected by indel mutations, but to allow high quality reads. Resulting sequencing files (.ab1) and read quality were analysed using SnapGene software.

2.2 Fish Husbandry and Embryo Methods

2.2.1 Zebrafish Strains

All embryos used were wild-type embryos used are derived from AB and AB* strains unless specifically stated.

2.2.2 Zebrafish Husbandry

Zebrafish were kept in the University BMSU facilities and maintained according to UK Home Office regulations by members of staff, as set out by Animals (Scientific Procedures) Act 1986. I was licenced for use of zebrafish in biological experiments

under personal licence I23FED1ED, under PIL schedules A and B., while experiments were carried out under an appropriate project licence. Between 20 and 30 adult individuals were kept in tanks in a recirculating housing system (ZEBtec, Tecniplast). Zebrafish were maintained with a regular day/night cycle with 14 hours of light and 10 hours of darkness and fed 3 times a day. Water temperature was maintained at 26°C.

2.2.3 Zebrafish Breeding

Zebrafish breeding was carried out personally in the BMSU facilities. In a breeding round, up to 20 adult fish were placed pair-wise in standard breeding tanks overnight, separated by a barrier. In the morning, barriers were pulled out and pairs were allowed to mate. Adult fish were separated from fertilised eggs by a bottom mesh in the breeding tank, which allows rapid embryo collection and prevents adult fish from eating their eggs.

Eggs were naturally laid and fertilised, and collected from the breeding tanks shortly after laying by filtering the tank water through a net. Collected eggs were transferred to a Petri dish containing E3 buffer (5 mM NaCl, 0.17 mM KCl, 0.33 mM $CaCl_2$, 0.33 mM $MgSO_4$, 0.1% Methylene blue).

2.2.4 Embryo Maintenance

Embryos were kept in Petri dishes containing E3 buffer in an incubator set to maintain constant temperature of 28°C. Zebrafish embryos were grown up to a maximum of 5 days post fertilisation. Medium was changed every day and dead embryos were removed to avoid bacterial infections.

Zebrafish embryo growth stages were assessed according to Kimmel *et al.*, 1995.

2.2.5 Dechorionation

Chorions of embryos before 48hpf were removed by adding 1ml of stock solution (10mg/ml) of protease from *Streptomyces griseus* enzyme (pronase) to 10 ml of E3 medium. This solution was then added to a Petri dish containing embryos and gently stirred to allow the enzymatic proteolysis to occur. In order to prevent embryos from damage, embryos were observed under a microscope throughout the reaction and mixed with an excess amount of E3 media as soon as the chorions appeared to break down and washed three times to remove the enzyme and broken chorions.

Alternatively, embryos were dechorionated manually using a pair of fine, sharp forceps to physically remove the chorion.

2.3 Genome Editing Methods

2.3.1 CRISPR-Cas9

2.3.1.1 Target site selection and reagent design

Appropriate target sites for CRISPR/Cas9 genome editing were determined using CHOPCHOP software, searching for sites consisting of 18-20 nucleotides (excluding PAM motif) in the regions of interest using default settings for CRISPR-Cas9 knock-out, using zebrafish GRCz11 genome assembly as reference. Appropriate candidates were selected from generated outputs based on the scoring algorithm of the application and assessment of potential off-targets (with the optimum of 0), as well as the position of target site with regards to the gene.

For *gol (slc24a5)* mosaic knockout and sgRNA targeting we have used the previously described *gol* crispr cutter B (ccB) target site (Burger *et al.*, 2016).

For *mxtx2*, target sites were selected from the 5'- and 3'-end regions on each side of the gene, containing sequences up- and downstream of the gene respectively.

2.3.1.1.1 In-house single guide RNAs production

Single guide RNAs (sgRNAs) targeting sequences of interest were generated in-house by *in vitro* transcription of designed DNA sequences. DNA oligonucleotides containing a fusion of T7 polymerase promoter (5'-GAAATTAATACGACTCACTATAGGG-3'), target site (excluding PAM) and a *S. pyogene*s Cas9-backbone complementary region (5'-GTTTAAGAGCTATGCTGGAA-3') were ordered from Sigma.

The oligos were then annealed to a reverse complement of a *S. pyogenes* Cas9-specific backbone and amplified in a PCR reaction using proofreading Q5 DNA polymerase to produce a single DNA product – a DNA template of sgRNA.

Following the annealing, the sgRNA template DNA was purified using NucleoSpin PCR purification kit (Macherey-Nagel) and its concentration quantified.

400 ng of purified sgRNA template DNA was further *in vitro* transcribed following the kit protocol of HiScribe T7 High Yield RNA Synthesis Kit (NEB) to produce sgRNA. 10 μl of reaction product was then incubated with 0.5 μl (1U) of Turbo DNase (Invitrogen, AM2238) for 30 minutes in 37°C to ensure degradation of DNA template. sgRNA product was later purified using Monarch RNA Cleanup Kit (NEB) and eluted in 20 μL RNase-free H_2O.

2.3.1.1.2 Synthetic duplex guide RNAs

Alternatively to in-house produced sgRNAs, custom synthetic CRISPR RNAs were ordered from Sigma-Aldrich. These RNA oligos (crRNAs) are synthesised by the producer with a proprietary modified backbone that stabilises RNA molecules and are specific to a Cas9 nuclease strain used, in case of this project this was *S. pyogenes* Cas9 protein.

HPLC-purified and lyophilised ordered crRNAs were dissolved in 10mM Tris-HCl (pH=7.0, prepared with nuclease-free water) to the stock crRNA concentration of 50 µM, and then stored in -20°C.

Custom guide RNAs were then mixed with 50µM *S. pyogenes* trans-activating RNA (tracrRNA) in 1:1 ratio to form functional duplex guide RNA (dgRNA) capable of directing Cas9 to a target site.

For preparation of injections containing more than one crRNA, these were mixed with tracrRNA at a constant final concentration of 25 µM in the mix (see Table 2.1)

Table.2.1 Concentrations of dgRNA components in stock solutions prepared before microinjections.

Component	**1 target site**	**2 target sites**
tracrRNA	25 µM	25 µM
crRNA 1	25 µM	12.5 µM
crRNA2	–	12.5 µM

2.3.1.2 Microinjections

Fertilised zebrafish embryos were microinjected using a pressure-controlled microinjector. Microinjection needles were pulled on a Flaming Brown needle puller.

Zebrafish eggs were collected during 1-cell stage and immediately transferred to Petri dishes. For every injection round, 2-3 μl of injection solution was transferred into a pulled needle, and the needle loaded to a microinjector. The desired volume of microinjection was adjusted by controlling the air pressure and size of the needle, with volume calculations based on the droplet size measured using a microscope graticule. In all experiments presented below, the approximate volume of solution injected per embryo was 1 nl.

Injections were made directly into the embryo cell or into the yolk sac in the animal pole. Approximately 80-100 chorionated embryos were injected with a single injection solution in any experiment. Successful injections were detected using phenol red dye in the injection mix and later on by using rhodamine fluorescent dye or *gol* sgRNA phenotype.

For sgRNA microinjections, injection solutions were prepared directly before zebrafish breeding and microinjection rounds. Solutions contained EnGen *S.pyogenes* Cas9-NLS protein (NEB) mixed with 1 or 2 *in vitro* transcribed sgRNAs, potassium chloride, phenol red and rhodamine dyes.

Recipes of injection solutions were subject to optimisation in early stages of the project, with the optimised solution used in nearly all experiments presented in table 2.2.

Table 2.2 Composition of microinjection mixes used in CRISPR/Cas9 gene targeting by sgRNAs.

Component	Concentration
EnGen *S. pyogenes* Cas9 protein with NLS	1120 ng/µl
sgRNA 1	~97.5 ng/µl*
sgRNA 2	~97.5 ng/µl*
KCl	0.26 mol/l
Phenol Red	0.5% (v/v)
Rhodamine	0.5% (v/v)
Nuclease-free H_2O	X µl**

*dependent on the molecular weight of the sgRNA, calculations made using CrispantCal software.

** variable, dependent on calulations made using CrispantCal software.

For dgRNA microinjections, injection solutions differed slightly from sgRNAs, and were prepared directly before zebrafish breeding and microinjection rounds. Solutions contained EnGen *S. pyogenes* Cas9-NLS protein (NEB) mixed with an existing crRNA:tracrRNA mix solution, NEBuffer 3.1, phenol red dye and nuclease-free water (see table 2.3).

Table 2.3 Composition of microinjection mixes used in CRISPR/Cas9 gene targeting by dgRNAs.

Component	Stock concentration	Volume used*	Final concentration
EnGen *S. pyogenes* Cas9 protein with NLS	25 µM	1 µl	2.5 µM
crRNA(s):tracrRNA mix	tracrRNA: 25µM crRNA: n **	1 µl	tracrRNA: 2.5 µM crRNA: n/10 **
NEBuffer 3.1.	10X	1 µl	1X
Phenol Red	5% (v/v)	1 µl	0.5%
Nuclease-free H_2O	–	6 µl	–

* in a typical microinjection experiment

** where x denotes concentrations of individual crRNAs in the mix, equal to 25µL/c, where c is the number of crRNAs mixed with tracrRNA.

2.4 Quantitative real-time reverse transcription polymerase chain reaction (qRT-PCR)

2.4.1. Reverse Transcription

For each experiment, a standard amount of RNA was established based on the lowest available RNA concentration in a group. Typically, no less than 50 ng and no more than 1 µg of RNA per sample was used. Then, volumes of extracted RNA corresponding to the set RNA amounts were transferred into 0.2 ml PCR tubes, and reverse transcribed using SuperScript IV reverse transcriptase (Thermo Fisher) reaction kit (reaction in Table 2.4 below).

Table 2.4. Reverse transcription reaction composition.

SuperScript IV 5X reaction buffer (Thermo Fisher)	4 µl
SuperScript IV reverse transcriptase enzyme (Thermo Fisher)	1 µl
DTT (0.1 M)	1 µl
RNasin (RNase inhibitor) (Promega)	1 µl
Random Hexamers (Promega)	1 µl
dNTPs (10 mM)	1 µl
RNA template	X µl (depending on sample RNA conc.)
Nuclease-free H_2O	11-X µl
TOTAL	**20 µl**

2.4.2 Quantitative Real-Time Polymerase Chain Reaction

qRT-PCR reactions were set up to the total of 10 µl in 96-well plates with equalised template concentration. Reactions were carried out with PowerUP SYBR Green

Master Mix (Applied Biosystems) and oligonucleotide primers designed to target cDNA of genes of interest. Eukaryotic translation elongation factor 1-alpha (*eef1a1a*) gene was used as a housekeeping control based on (McCurley and Callard, 2008).

qRT-PCR reactions were run in technical triplicate using a recommended cycling protocol using QuantStudio 5 Real-Time PCR system (Thermo Fisher). Reaction data, including melting curve information, was obtained through the Thermo Fisher Cloud applications.

Fold difference in expression of genes of interest was calculated using the $2^{-\Delta\Delta C(q)}$ method of data normalisation (Livak and Schmittgen, 2001), where C(q) indicates a number of PCR cycles required to reach a threshold in fluorescence intensity of the sample, normalised to the intensity of passive reference dye (ROX).

2.5 Whole Mount *in situ* Hybridisation

Whole-mount *in situ* hybridisation (WISH) is a method commonly used for visualisation of mRNA transcripts of particular genes in whole embryos. It can be used as a method to analyse gene expression patterns or as a qualitative detection method for mRNAs of interest (Thisse and Thisse, 2008). The technique utilises an antisense RNA probe, transcribed from cDNA of gene of interest, labelled with digoxygenin-modified nucleotides. The probe is then added to a permeabilised embryo and hybridised to the transcript of interest. The probe is then stained with an alkaline phosphatase-conjugated antibody against digoxigenin and visualised using a chromogenic substrate. The protocol used for WISH was based on Thisse and Thisse, 2008.

2.5.1 Probe Design and Production

Antisense probes were designed to target mRNAs of choice. First, PCR primer pairs specific for the target gene cDNA were selected, spanning between 400-1000bp. Sequences of primers were combined with sequences for a 5' SP6 and 3' T7 promoters and those oligonucleotides were ordered. Primers targeting the mRNA of interest containing the phage promoter sequences were then used to PCR amplify their target sequence from total zebrafish cDNA, producing a DNA template of WISH probe. The templates were then *in vitro* transcribed using a Digoxigenin (DIG) RNA Labelling Kit (SP6/T7) (Roche) using the appropriate promoter according to the manufacturers protocol, with T7 for antisense (mRNA targeting) probes and SP6 for sense (non-targeting) probes.

2.5.2 Fixation of Embryos

Embryos were collected at the appropriate stages and transferred to 4% (v/v) solution of paraformaldehyde (Fisher, 28908) diluted in PBS and pH adjusted to 7.4. After 30-minute incubation in room temperature or, alternatively, overnight incubation at 4°C, the embryos are washed 3 times with PBST and manually dechorionated. Following dechorionation embryos are dehydrated by washing with gradually increasing concentrations of methanol in PBST (25%, 50%, 75%), and finally washed and then stored overnight in pure methanol (100%) at 4°C.

2.5.3 Pre-treatment of Embryos

Embryos were rehydrated by washes in successively diluted methanol gradient (75%, 50%, 25% in PBST) and washed 3 times with PBST to remove all methanol traces. Washes are done at room temperature with gentle agitation on a nutator (<60RPM).

Pre-hybridisation treatment of embryos followed the protocol (Thisse and Thisse, 2008).

2.5.4 Hybridisation of Probe

After pre-hybridisation step, the buffer was removed from the embryos and swiftly replaced by approx. 200 µl of pre-warmed hybridisation buffer containing 30-50ng of antisense DIG-labelled RNA probe. The probe was hybridised to its targets in the embryos during an overnight incubation (at least 16 hours) at 65°C.

2.5.5 Removal of Excess Probe

Following removal of the probe, the embryos were subject to a series of washes with a variety of buffers that gently and gradually replaced the pre-hybridisation buffer as per the protocol (Thisse and Thisse, 2008). Finally, buffers are replaced by PBST, in which embryos are washed 3 times, 10 minutes each.

2.5.6 Antibody Labelling

Embryos were then incubated in >1ml antibody blocking solution (DIG Wash and Block Buffer Set, Roche) – 1 ml 10X Blocking Solution, 1 ml 10X Maleic Acid Buffer, 8 ml ddH_2O – with gentle agitation for 1 hour at room temperature to prevent non-specific antibody binding.

Anti-DIG-AP antibody was diluted 1:10000 in the blocking solution and added to embryos after blocking step. Embryos were incubated with the antibody overnight at 4°C.

After antibody labelling, the solution was removed and embryos are washed 8 times with DIG Washing Solution (Roche), each wash taking 15-30 minutes at room temperature.

2.5.7 Colour Development

Embryos were washed 3 times with Detection Buffer (Roche), for 10 minutes each time at room temperature. Embryos were then moved to a multiwell plate, and incubated in 1 ml of solution containing 3.5 µl/ml BCIP and 4.5 µl/ml NBT in the Detection buffer. The reagents stain the anti-DIG-AP antibody, with a colorimetric reaction occurring. Colour development was monitored closely under a light microscope or with naked eye. Embryos were kept in dark conditions in between observations to allow accurate staining with light-sensitive reagents.

Once the desired colour intensity is reached, samples were washed multiple times in PBST at room temperature, with continuous protection from light.

2.5.8 Fixation of Stained Embryos

After PBST washes, embryos were fixed by incubation in 4% paraformaldehyde solution (in PBS) for 2 hours at room temperature. Fixed samples were washed with PBST and then dehydrated by series of incubations with increasing methanol gradient (as described before). Following that, embryos were washed 3x with PBST at room temperature and kept in PBST at 4C until use.

2.5.9 Mounting and Imaging

For imaging on slides, embryos were delicately placed on glass microscopy slides (SuperFrost®, Thermo Fisher). Adhesive tape was used on edges of a slide to create

elevation, so that the embryo was not damaged or crushed by the coverslip. Embryos were then covered with a coverslip that was fixed in place using an adhesive.

Alternatively, embryos were moved to a small Petri dish, surface of which was covered with 1% agarose (made in PBS), and placed on the agarose surface in a PBS solution, and imaged.

2.6 Visualisation and Analysis of Transcription in Embryos

2.6.1 Immunostaining

Embryos were collected at appropriate stages and fixed for 30 mins at room temperature with a 4% solution of paraformaldehyde in PBS.

Embryos were then rinsed in PBS, dechorionated (if not dechorionated before fixation), and washed 4 times with 0.1% PBS-Triton-X-100 solution (0.1% PBS-Triton).

Embryos were then incubated for 1h at room temperature with 5% solution of Normal Goat Serum (abcam, ab7481) in 0.1% PBS-Triton.

In the meantime, a 1:500 solution of primary antibody: mouse Elongating RNA polymerase II Anti-RNA polymerase II CTD repeat YSPTSPS (phospho S2) antibody [H5] (abcam, ab24758) was prepared in the 5% normal goat serum solution.

After incubation, blocking solution was removed and replaced with the primary antibody solution, in which embryos were incubated overnight at 4C. Following that, primary antibody solution was removed and embryos were washed 4 times, 10 minutes each, with the 0.1% PBS-Triton solution.

In the meantime, 1:1000 solution of the secondary antibody: Goat anti-Mouse IgG (H+L), Superclonal Recombinant Secondary Antibody, Alexa Fluor 488 (Invitrogen, A28175) was prepared in the 5% normal goat serum solution.

Embryos were then incubated in the secondary antibody solution for 2 hours at room temperature, protected from light.

Secondary antibody was then removed and replaced by 0.1% PBS-Triton solution. Embryos were washed 4 times 10 minutes each with 0.1% PBS-Triton. The solution was then removed and replaced with PBS. Embryos were kept protected from light and stored in the dark at 4C.

For staining of nuclei, NucBlue™ Live ReadyProbes™ Reagent (Hoechst 33342) (Thermo Fisher, R37605) was used according to the attached protocol, by addition of 2 drops of the reagent per 1mL of PBS, and incubating the samples at room temperature for 30 mins, before removing the solution and rinsing the samples 2x in PBS. Following that, samples were kept at 4C in the dark for a maximum of 1 week until use.

2.6.2 Light Sheet Microscopy

A Zeiss Lightsheet Z.1 microscope was used for all light sheet microscopy applications. Embryos were submerged in molten 1% agarose solution and taken up into a glass capillary, in which the agarose solution was allowed to solidify. Glass capillaries can be mounted into the chamber of the microscope, with the embryos placed into the view of the objective embedded in a cylinder of agarose and held vertically.

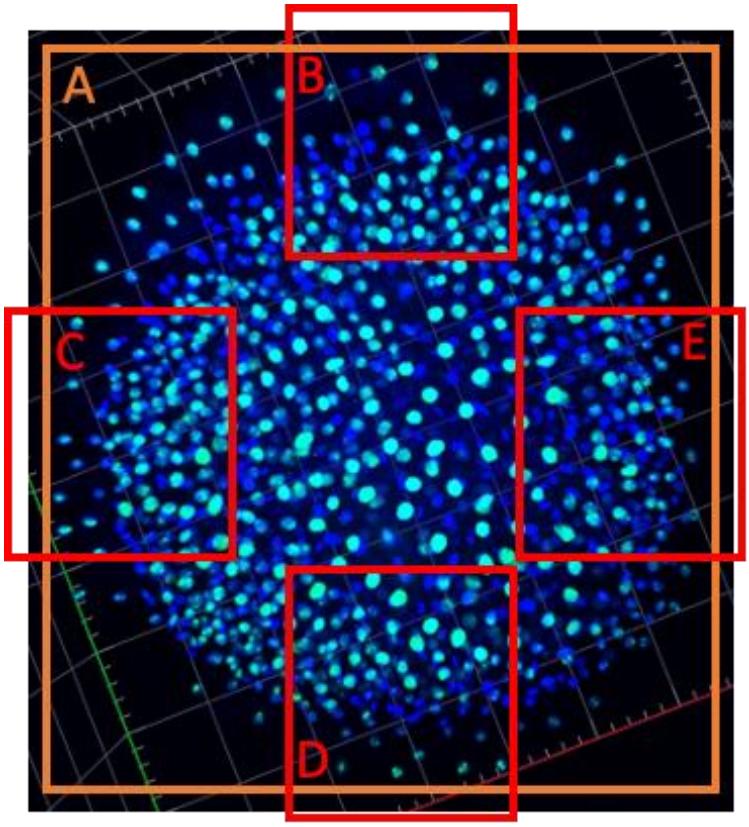


Fig. 2.1. A schematic overview of the imaging strategy for visualisation of transcription in embryos. Embryos were immunostained for elongating RNA pol II (pSer2). After mounting in the Z.1 LSFM, embryos were positioned with animal pole facing forward, towards the camera. A screenshot of 3D view of an oblong stage embryo with labelled elongating RNA Polymerase II (green) and nuclei (blue) is shown with the approximate positions of each of the imaging sets taken in the systematic imaging, with (A) showing the approximate field of view for Z-stack imaging of whole embryo at 10x magnification (20x objective with 0.5x optical zoom), and (B-E) showing approximate fields of view for Z-stack imaging – "north, west, south and east" views – at high magnification of 30x (20x objective, 1.5x optical zoom).

Embryos immunostained with antibody against RNA pol II were imaged with positioning and selected fields of view standardised and set according to the imaging strategy outlined in Fig. 2.1.

Standard settings were used for imaging, with variable Z-stacks depending on the depth of embryo required to be imaged, with Z-slices set for 1.00 µm (overview-Fig.2.1 A) or 0.60µm (high magnification views, Fig.2.1 BCDE). For 405nm laser illumination (for Hoechst) – 3.0% laser power, 30ms exposure settings were used.

For 488nm laser illumination (for AF488 and RNA Pol II) – 2.0% laser power, 30ms exposure settings were used. 20x objective was used in all experiments, with 1.5x or 0.5x optical zoom depending on the required field of view.

2.6.3 Segmentation of 3D Images

All segmentation of 3D images and imaging analysis was conducted by Haseeb Qureshi, with me taking part in selection of embryo images for analysis.

2.6.3.1 Image selection and generation of blastomere and YSL image subsets

For image segmentation and subsequent quantification analysis, Z-stack images focused to individual quadrants of the embryo and YSL ring were selected ("North, South, West, East" views of the embryo imaged at 30x magnification). These Z-stacks are further referred to as "views". Lower magnification images of whole embryo were not further analysed.

Segmentation of images would allow to isolate and normalise the information coming from the light signal detected in each embryo and to eliminate a significant amount of noise from the images. In this case, each segment would be defined as a boundaried region exhibiting fluorescence in a defined channel. In simplified terms, segmentation allows isolation of objects from the raw imaging data.

First, obtained raw images were sorted through individually and manually to discern images of the highest quality. This involved examining each view Z-stack in the Zeiss ZEN 2 Blue Edition (ZEN Digital Imaging for Light Microscopy) software. Each view was scrolled through with an automatic setting of colour look up table histogram

("Auto"), in order to adjust for varying levels of brightness between individual Z slices in the stack. 3 embryos with highest quality were selected per developmental stage, with each view (4 per embryo) containing 2 subsets of nuclei we were interested in comparing in further analysis - blastoderm and YSL. At this point, there was a combined total of 36 imaging files (3 developmental stages x 3 embryos x 4 views).

First, regions not showing RNA Pol II staining were cropped out to reduce the amount of memory each image stack requires for processing, and only the RNA Pol II and DAPI channels were exported to allow higher computational throughput, 5-EU incorporation (in embryos labelled with both RNA Pol II antibody and 5-EU) was not included in analysis at this point.

The complete Z-stacks of each embryo were split into 2 subset stacks approximately at the interface between YSL and blastoderm, as determined by observation of embryo anatomy in 3D view (although in some embryos blastomeres appeared alongside YSL nuclei and were included in the YSL subset), producing 2 output image files, which consisted of 2 fluorophore channels – one for blastoderm, one for YSL. YSL was determined by the spread distribution of fluorophore in the syncytium rather than by taking on the shape of defined cell membranes, as is the case in blastoderm. The syncytium was most clear in the 488nm illumination channel (Alexa Fluor 488 – labelling the RNA Polymerase II antibody) and was characterised with an absence of spherical membrane geometries highlighted by fluorophore.

This process was repeated for each view for each embryo. This totalled to 2 stacks per each view, 4 views per embryo, 3 embryos per stage and 3 stages total, producing 72 stacks to process in 3D. Not every stack produced outputs, as some YSL views did not show any RNA Pol II staining as they were undergoing metaphase. To produce

isolated 3D RNA Pol II segments for analysis, each subset stack was processed using Icy Bioimage software (v2.4.2.0) (de Chaumont *et al.*, 2012).

A workflow was designed for this task where nuclear signal (405nm-Hoechst) would be segmented first to be able to link regions of segmented RNA Pol II to specific nuclei, and then RNA Pol II would be segmented within each 'nuclear' segment.

2.6.3.2 Nuclear segmentation

The 16bit grayscale Hoechst channel of the stack was run through K-means thresholding using 13 bins, producing a separate K-means thresholded image stack. A blanket watershed was applied to this processed stack, eliminating all bins below the 6^{th} or 7^{th} bin to eliminate non-nuclear noise, depending on the noisiness of the image, as determined by the isolation of individual nuclei. Too low a threshold would result in conjoined nuclei, as noted by bi-lobular structures which could be determined numerically by filtering for convexity – a well segmented nucleus should have a high convexity.

This produced a binary image stack where all pixels above the thresholded bin were white, and all below were black. The binary image was exported back onto the original 16bit stack as labelled regions of interest, for each nucleus. The precision of the nuclear shape was not of paramount importance, as the goal was just to produce local zones for each nucleus, within which RNA Pol II segmentation could be processed. Using this methodology, nuclear regions of interest (Nuclear ROI) were segmented from all 72 stacks.

2.6.3.3 RNA Pol II segmentation

Segmenting the RNA Pol II channel to isolate regions of visualised elongating RNA polymerase II was a more complex task, as it requires isolation of a significantly larger number of labelled regions due to the smaller, fragmented nature of RNA polymerase II staining, with multiple foci present in nuclei, in contrast to the fewer whole nuclei in the field of view and their defined appearance in the stack. Due to the high requirement for memory in complex segmentation processes, the previously segmented Nuclear ROI were selected, and an inversion operation blacked out all regions outside of the nuclei, excluding any non-nuclear regions in the image stacks from RNA Pol II segmentation.

Analogically to the nuclear segmentation, a 13 bin K-Means thresholder was run and binary thresholding on bin 8-10 (depending on noise) was run on the K-means output. This labelled output was transferred back to the original 16bit image to form 3D regions of interest (ROIs) covering nuclear regions with RNA Pol II signal. RNA Pol II ROIs below 15-25 pixels in volume were filtered as noise, and any nucleus that was cut off at the edges of the field of view was also filtered out.

2.6.4 Imaging Analysis and Statistical Analysis

The remaining nuclear and Pol II 3D ROIs were exported with their metadata containing information regarding geometry, location, and intensities. This produced a spreadsheet which listed every ROI and its corresponding metadata. An inclusion analysis was performed via a downloadable plugin for Icy Bioimage, which linked RNA Pol II ROIs with the nuclear ROIs within which they were contained. This has output a spreadsheet, which listed each nuclear ROI and each RNA Pol II ROI within it.

The ROI metadata and output for inclusion analysis were exported to Microsoft Excel (Microsoft Corporation), where the ROI metadata and Inclusion analysis data were linked using an INDEX:MATCH formula between the 2 datasets.

This linked data for every YSL and blastoderm region for each developmental stage and/or embryo was exported into Graphpad Prism 9 (GraphPad Software), where intensities, volume and surface areas of the RNA Pol II ROIs were output as numerical values and plotted on graphs.

Distributions of intensity, volume and surface area of elongating RNA polymerase II (as extracted from the ROI Metadata output) were compared between identified YSL and blastomere cell populations in all embryos from each stage, and separately within data from individual embryos. These distributions were then statistically compared using a nonparemetric Kolmogorov-Smirnov test within GraphPad Prism 9, with descriptive statistics also produced.

2.7 Live 3-dimensional Imaging of Cell Cycle and Transcription Dynamics

2.7.1 Microinjections

For *in vivo* imaging aiming to visualise cell cycle and transcription dynamics, embryos were injected at 1-cell stage with a solution containing 5µM of fluorescently labelled miR-430-FITC morpholino oligonucleotide (Gene Tools LLC) and 400ng/µl solution of a fluorescently labelled mRuby3:H2B protein. Embryos were incubated in 28.5C until imaging.

2.7.2 Light Sheet Microscopy (Live)

Zeiss Lightsheet Z.1 microscope was used for all light sheet microscopy applications. Chorioned embryos were mounted in agarose cyllinders as described in 2.8.3. Imaging settings were varied, adjusted for maximising possible framerate and visibility of signal. Embryos were imaged in 3-dimensional stacks across time.

2.7.3 Tracking of Nuclei, Cell Cycle and Transcription Lengths

2.7.3.1 Manual tracking

Maximum Intensity Projections (MIPs) were produced from whole embryo live imaging datasets, converting the 4-dimensional images (3 dimensions + time) to 3-dimensional images (2D+time). The maximum intensity projections were loaded into the FIJI ImageJ software (Schindelin *et al.*, 2012) to visualise both imaging channels and create a merge channel.

2.7.3.2 Automated tracking – conducted by Haseeb Qureshi

MIPs produced from whole embryo live imaged datasets were used for automated tracking analysis (Hadzhiev *et al.*, 2019). MIPs were converted from 16bit to 8bit to reduce memory consumption and improve processing speed. Spot detection plugin (Icy Bioimage software) was run on each of the miR-430-FITC morpholino and mRuby:H2B channel respectively, to produce 2D ROIs which could be tracked from frame to frame. Tracks were produced by following the movement of an individual ROI over time, and labelling all of the ROIs for a given nucleus track and its progenies track using a custom script written in javascript. This script would group ROIs over frames into a group of ROIs from different time points to represent the movement

of a given nucleus over time. The same was done for each incidence of miR-430 per nucleus, and these were grouped with their corresponding nuclear tracks too.

These track ROIs were exported into Microsoft Excel, and the timings of each track for miR-430 and nuclear signal could be quantified and displayed in charts produced in Microsoft Excel. Timing of each tracked nucleus between mitosis events was used as an approximation of cell cycle length. The period of transcriptional activity of cells was approximated from the period of miR-430 signal appearing within the embryo.

2.8 Light Microscopy

A Nikon SMZ 745T light microscope was used for embryo observation and staging, as well as to determine *gol* phenotypes and for imaging whole-mount *in situ* hybridisation.

A Zeiss AxioZoom.V16 light microscope was used for observation of fluorescence phenotypes, high magnification observation and imaging of embryos.

Chapter 3: Development of a Genetics Toolset for Disruption of *mxtx2* Expression

3.1 Introduction

Recent years have provided a variety of new tools in genome engineering, with the most major development being the Nobel prize-winning CRISPR-Cas9 genome editing method. Since its discovery and widespread application, a plethora of new tools and methods have been described and new applications such as engineered Cas proteins or use of alternative classes of the Cas system have been developed (Adli, 2018; Liu *et al.*, 2022b). Using gene knockouts can be a very informative approach in discerning gene function, and site-specific approaches such as CRISPR-Cas9 can vastly improve confidence in the achieved results.

In this chapter, I discuss the adaptation of those new methods to generate independent tools to disrupt *mxtx2* expression. This was attempted using 2 main approaches: "classic" CRISPR-Cas9 knockout by targeted gene mutation and using a Cas13d nuclease to knock down expression of *mxtx2* mRNA. The primary aim was for the methods of transient gene disruption to serve as a toolset to characterise the effects of *mxtx2* depletion in the early embryo, and to inform the possibility of creating a stable *mxtx2* mutant line that could be used in further studies of its phenotype, mechanisms of action, and effects on the zygotic genome activation.

mxtx2 has been previously a subject of knockdown studies using morpholino oligonucleotides, showing a severe epiboly phenotype termed yolk burst (Fig. 3.1) and caused embryonic lethality (Wilkins *et al.*, 2008). This suggested that *mxtx2* might play a crucial role in some of the processes in the early zebrafish embryo. However, it has been shown that morphant phenotypes are often caused by off-target effects stemming from the technique (Kok *et al.*, 2015) or can often be more severe than knockout mutants due to the lack of genetic compensation (Rossi *et al.*, 2015). Therefore, it was an opportunity to revisit *mxtx2* and its phenotype, and further explore the mechanisms in which this early zygotic gene is involved.

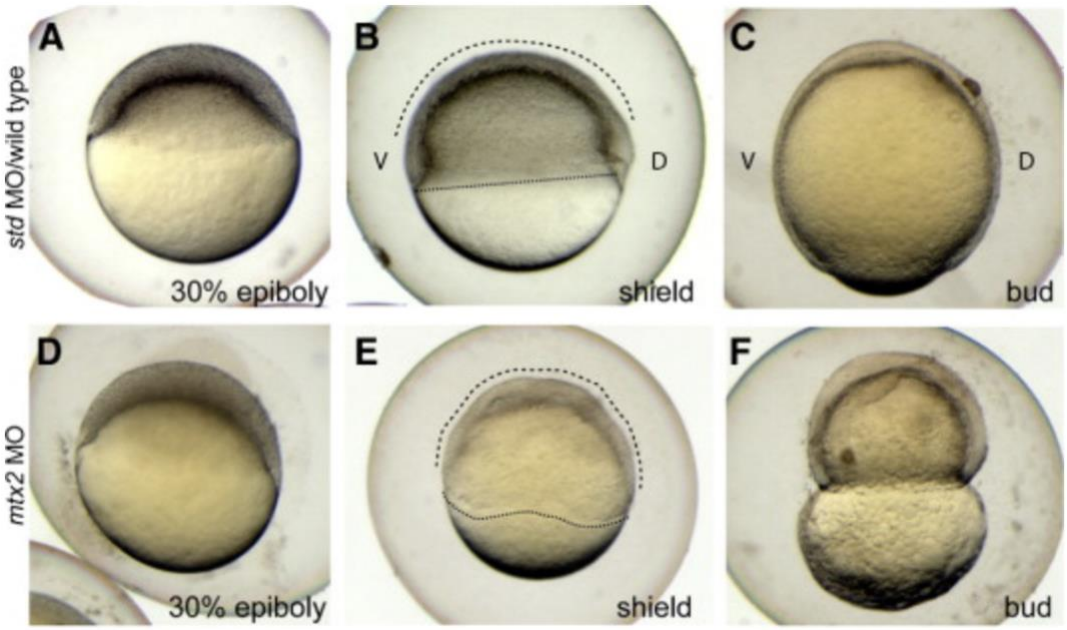


Fig. 3.1 Phenotype of *mxtx2* morpholino knockdown in zebrafish embryo.
Lateral views of normally developing wild-type embryos at 30% epiboly (A), shield (B) and bud stages (C), with dorsal and ventral sides of the embryo highlighted with V and D, respectively. Embryos injected with the *mxtx2* MO proceed into epiboly (D) but show particular abnormalities such as furrowing of the blastoderm cap and margin (E - dashed and dotted lines) and further progression into epiboly causes the marginal blastomeres and YSL to prematurely constrict around the yolk, leading to disruption of the yolk cell membrane and its subsequent burst (F), causing embryo to lyse. Images adapted from (Wilkins *et al.*, 2008).

3.1.1 CRISPR-Cas9 approach

Cas9 nuclease was utilised for targeted genome editing to generate genomic deletions leading to knockout mutations of targeted genes (Hwang *et al.*, 2013; Li *et al.*, 2016). Zebrafish embryos need to be injected with the injection mixture containing Cas9 and guide RNAs (gRNAs) at the 1-cell stage. Injected fish (F0 generation) are grown to adulthood and screened for germline mutations – i.e. if cells in the germline were mutated and therefore if the fish are able to transmit the mutation to the next generation. This is most often done by genotyping the offspring of an F0 fish crossed with a wild type. If the mutation is transmitted, the F0 fish is kept as a line founder, and its genotyped offspring from the cross are grown to adulthood, heterozygous for the mutation of interest. The in-cross of second generation of fish (F1) produces offspring from which 25% would be homozygous for mutation of interest, if Mendelian inheritance rates are maintained (Li *et al.*, 2016). The process of generating a stable knockout mutation line in zebrafish therefore requires growing of 2 generations of fish. Initial steps in generating the mutation in F0 embryos by CRISPR-Cas9 are therefore crucial, because any inaccuracies at that stage will be carried over across 2 generations of fish, and can affect many months of work.

In order to reach that point, first a successful mutation needs to be introduced in F0 embryos.

3.1.1.1 Constraints of CRISPR-Cas9 site-specific mutagenesis

Injection of Cas9 and gRNAs into the 1-cell embryo does not guarantee a complete knockout. Firstly, the double strand break at the target site will not always cause an identical mutation – DNA repair mechanisms like NHEJ lead to a degree of

randomness, and alternative pathways of repair such as homologous recombination or microhomology mediated end joining can also be used to repair the target site (Ata *et al.*, 2018; Li *et al.*, 2016). Therefore, the outcome of the introduced double strand break has a degree of randomness, and could be a random insertion or deletion of a number of base pairs (indel). The break could also be repaired correctly by chance or sequence homology.

Secondly, there is an issue of mosaicism. The double strand break formation and its repair needs to occur during the first cell cycle and be propagated to daughter cells for a chance of a complete gene knockout in the F0 generation. If this does not happen, the induced mutation would be mosaic, with the number of affected cells depending on when the double strand break was generated. Additionally, as the cells contain 2 alleles of its gene, biallelic mutagenesis might be necessary to efficiently knock out gene expression (Burger *et al.*, 2016).

Thirdly, guide RNAs are not perfect. Their binding to targets is subject to multiple variables including binding energy and RNA secondary structure, as well as access to chromatin. All gRNAs also must target sequences preceding an obligatory PAM motif, that ensures Cas9 nuclease binding. This can have a significant effect on efficiency of Cas9-induced mutagenesis. Multiple bioinformatics tools, such as CHOP-CHOP or CRISPRScan have been developed to facilitate the search for optimal target sites for sequences, assessing the potential efficiency of mutagenesis and suggesting guide RNA sequences, as well as scanning the genome for potential off-target sites (Labun *et al.*, 2021; Labun *et al.*, 2019; Moreno-Mateos *et al.*, 2015).

Finally, the delivery of guide RNAs and Cas9 to the embryo is an additional important step, which is most subject to human error. Delivery of optimal concentrations of

Cas9 and guide RNAs in equimolar ratio, optimally buffered can affect mutagenesis efficiency (Burger *et al.*, 2016; Hoshijima *et al.*, 2019). Controls are therefore crucial to ensure that no errors have been made and that RNA-protein complex of Cas9-gRNA is delivered to the cell and active.

3.1.2 Cas13d – a CRISPR RNA-directed RNA nuclease

Among the new developments in the CRISPR approaches, most focus on the Cas9 nuclease – with modifications, catalytic changes and improvements in efficiency. However, Cas9 is a member of a larger family of RNA-directed nucleases, many of which have been explored as potential tools for targeted gene disruption. Most of CRISPR family nucleases target DNA, but class 2 type VI of those nucleases – Cas13 proteins – are able to bind specific RNAs instead (O'Connell, 2019). Targeted RNAs are cut by the Cas13 endonuclease and subsequently targeted for degradation, causing a reduction in target gene translation. Similarly to Cas9, Cas13 also requires a guide RNA to targets a sequence of choice, however the structure of Cas13 gRNAs is distinct to that of Cas9. Similarly to Cas9 nucleases, there are many variants of Cas13 nucleases, often coming from different bacterial species, with the Cas13 protein family containing Cas13a, Cas13b, Cas13c and Cas13d proteins. Cas13 proteins require pre-assembly with a CRISPR-RNA (crRNA) for function – these act very similarly to gRNAs – but there is no PAM motif requirement and nucleases can bind to mRNAs homologous to the crRNA sequence (Huynh *et al.*, 2020).

Cas13 RNA endonuclease therefore provided a potential tool for programmable RNA interference. They have been used to good effect, binding and causing degradation of target RNAs in yeast and mammalian cell lines (Abudayyeh *et al.*, 2017; Cox *et al.*,

2017). RNA knockdowns through one of the types of active Cas13 nucleases, Cas13d, isolated from the bacterium *Ruminococcus flavefaciens*, were shown to be more effective and specific than RNAi in mammalian cells (Konermann *et al.*, 2018). With RNAi not being commonly used in zebrafish, and certain controversies surrounding the use of morpholino oligonucleotides for generation of gene knockdowns, Cas13 appeared to be an exciting alternative. Indeed, it has soon been adapted for use *in vivo* with zebrafish (Kushawah *et al.*, 2020) (Fig.3.2). In a proof of principle study by Kushawah *et al.* the group shows very efficient degradation of several target RNAs by Cas13d. Other variants of type VI Cas13 nucleases, Cas13a (isolated from *Leptotrichia wadei*) and Cas13b (isolated from *Prevotella sp. P5-P125* and from *Porphyromonas gulae*) were also tested but caused unspecific embryo deformations during the first day of development. Cas13d was shown to have a low proportion of unspecific deformations, and was subsequently used to specifically knock down *tbxta* (causing the well-described no-tail phenotype), *dnd1* (disrupting germ cell survival), *nanog* and pigmentation enzyme tyrosinase (*tyr*). This indicated that both maternal and zygotic mRNAs can be efficiently targeted by Cas13d in zebrafish embryos.

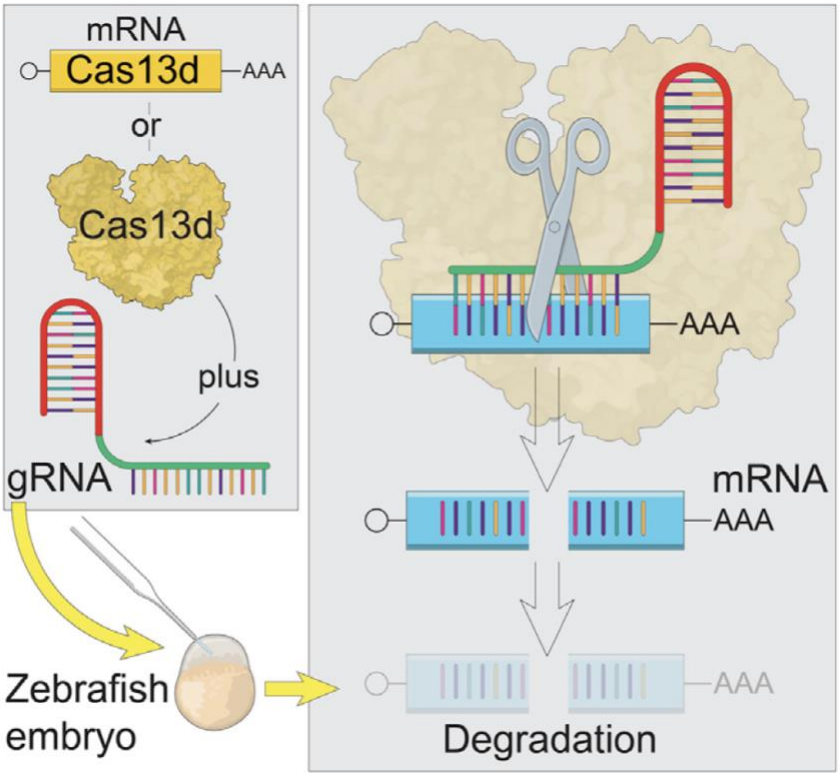


Fig. 3.2. Schematic depiction of the principle of Cas13d-dependent gene targeting zebrafish. In vitro-transcribed mRNA of Cas13d or purified Cas13d protein is microinjected into the zebrafish embryo at one-cell stage, alongside the guide RNA (gRNA) specific to the gene of interest. In the embryo, the ribonucleoprotein complex of Cas13d and gRNA is able to specifically bind to the mRNA of interest and introducing a cut in its structure. There are no PAM motif requirements for Cas13d binding to RNA. The schematic of where the nuclease generates a lesion inthe RNA chain is not to scale. Damaged mRNA is targeted for degradation, causing a gene knockdown and preventing translation. Figure from: (Kushawah *et al.*, 2020).

In context of this study it was therefore an good opportunity to adapt this technique for my research in search for an alternative method of *mxtx2* depletion. Using CRISPR-Cas13d would allow efficient, but transient, knockdown of *mxtx2*, potentially circumventing the need to generate a stable transgenic line.

3.1.3 Aims of this chapter

I set out with the goal of characterising the role of *mxtx2* in ZGA, and for that I needed a reliable set of tools to disrupt its expression. With that, generation of a stable

transgenic line lacking *mxtx2* expression would be possible. However, gene disruption even in F0 generation can be informative of gene function, therefore allowing to ask further questions.

In order to achieve the *mxtx2* gene disruption, I set out to first familiarise myself with the CRISPR-Cas9 method and its controls in zebrafish embryos. Following that, my primary objective was obtaining a reliable and efficient guide RNAs targeting *mxtx2*. I designed sets of guide RNAs, verified and tested them, attempting several methods of guide production and delivery.

The final aim was to adapt the CRISPR-Cas13d technology in order to produce a reliable knockdown of expression of zebrafish *mxtx2* in the embryo.

3.2. Optimisation of CRISPR-Cas9 Approach to Disrupt *mxtx2*

3.2.1 *gol* as a control for Cas9-mediated mutagenesis efficiency

As previously mentioned, one of the biggest challenges in mutagenesis in zebrafish embryos is the stochasticity of Cas9 nuclease activity. High levels of inefficiency could render experiments unreliable, and pose difficulty in distinguishing an embryonic phenotype from experimental noise – damaged embryos, unrelated developmental disorders occurring by chance, or cellular toxicity of Cas9 or other elements of the injection solution. It was therefore pertinent to seek out a positive control of the efficiency of Cas9 activity. Ideally, such a control could also act as a readout of competency in zebrafish embryo microinjections, which can be an important factor in overall efficiency of the technique. Therefore, such control should be a guide targeting a gene with minor or no overall effect on zebrafish development, while simultaneously having a visible, and ideally scorable, phenotype. This phenotype should also be visible relatively early in development to allow quick verification of the methods.

Previous research has found examples of such genes, and one that was selected was *slc24a5*, coding for a cation exchanger protein, responsible for a *golden (gol)* phenotype in zebrafish embryos (Burger *et al.*, 2016; Jao *et al.*, 2013). *gol* will be used further in this thesis to denote *slc24a5* gene*,* as both gene annotations are in common use by the zebrafish community. Mutations in *gol* lead to an easily observable phenotype: depletion in pigmentation of skin melanocytes and retinal

pigment epithelium (Lamason *et al.*, 2005). *gol* is not essential for development, with embryos lacking *gol* surviving to adulthood (Burger *et al.*, 2016). Loss of pigmentation can be observed at 48 hours post fertilisation, as before that point pigmentation is still developing, making phenotype assessment difficult. At 48 hpf, retinal pigmentation is nearly complete in the embryo, allowing for an easy readout of *gol* targeting (Dahlem *et al.*, 2012).

Therefore, *gol* was fitting most of the criteria for a good control for Cas9 mutation efficiency. Burger *et al.* have been able to develop an efficient guide RNA against *gol* – *gol* ccB gRNA (Fig. 3.3 A), targeting the 5' UTR region of the gene. I have decided to utilise it as a control for Cas9 efficiency.

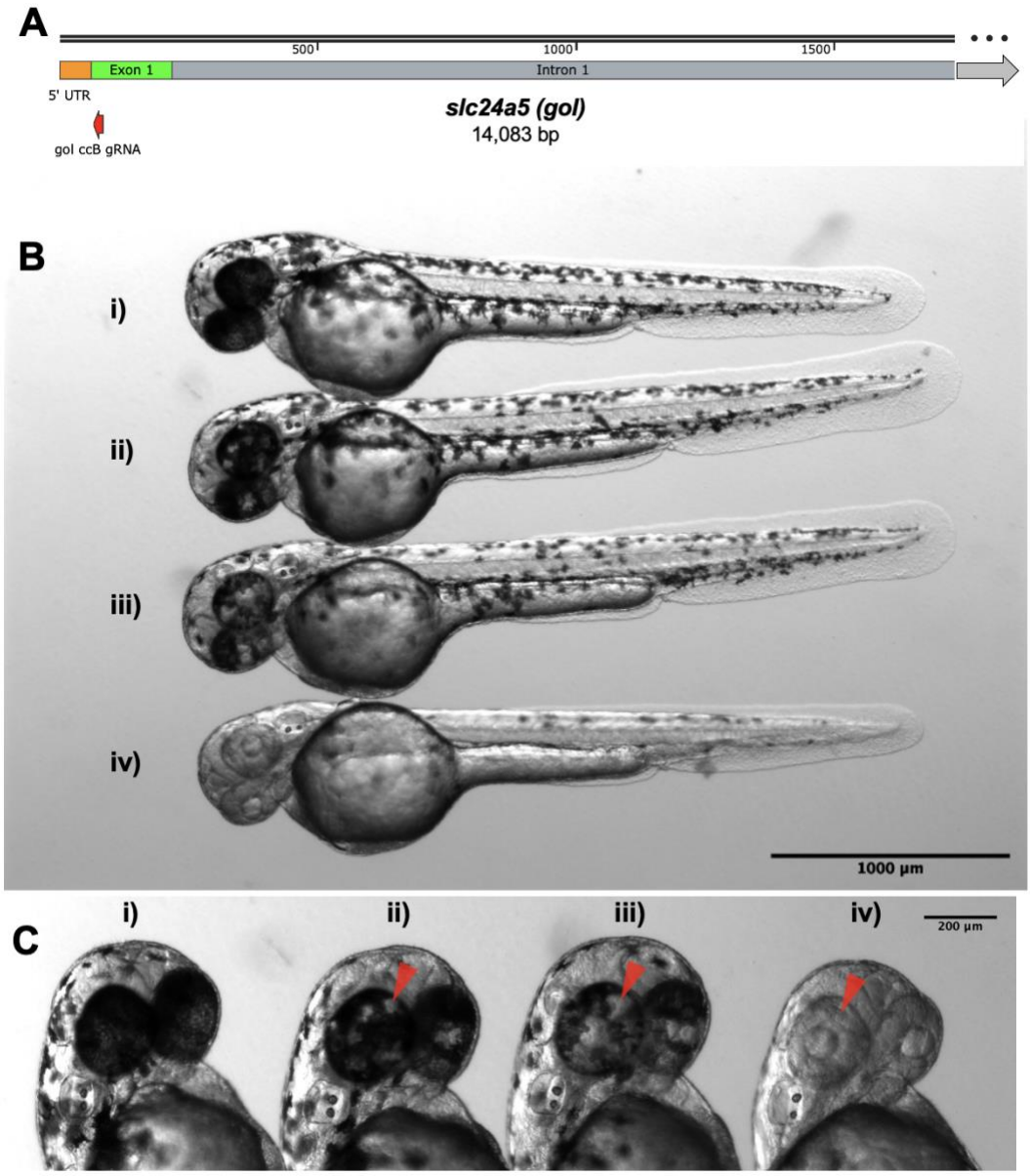


Fig. 3.3 *gol* knockout as a scorable control of Cas9-directed mutagenesis

(A) Genomic region of slc24a5 on chromosome 18, with the *gol ccB* sgRNA target (Burger *et al.* 2016) highlighted in red. (B) A microscope image of 48hpf zebrafish that were microinjected with *gol ccB* sgRNA and Cas9 at 1-cell stage. Each depicted zebrafish larva shows different level of eye pigmentation loss due to Cas9 targeting of *gol* and is an example of pigmentation loss phenotype scoring: (i) is a non-injected wild-type control showing no pigment loss; (ii) shows low levels of pigment loss; (iii) shows over 50% loss of pigment in the eye (moderate loss); and (iv) shows almost no pigmentation in the eye (high loss of *gol*). (C) A zoomed-in microscope image of heads of zebrafish larvae in (B), depicting the mosaic phenotypes generated by injections of Cas9 and *gol*-targeting sgRNA. Scoring of (i)-(iv) larvae as described above. Red arrows point to a small (ii), moderate (iii), and large (iv) areas of mosaic lack of pigment, indicating an earlier editing event in (iii), and a much earlier in (iv).

Embryos were injected at one-cell stage with the in-house *in vitro* transcribed *gol* ccB gRNA, alongside Cas9 protein, salts and a protein buffer. With the obtained mosaic knockouts of *gol,* I have established a scoring system for the mosaic phenotypes (Fig. 3.3 B-C). Pigmentation was assessed only in the retina of the embryo (Fig. 3.3 C), as this is the only area fully covered by pigmented cells at 48 hpf, and assessing random pigmentation of the melanocytes on the trunk was not feasible. By observing pigmentation of retinas of both eyes in every embryo under the light microscope, I have established four categories. First was full pigmentation of the eyes indicating no *gol* loss of function (Fig. 3.3 C i). Second was low levels of *gol* loss of function, with up to 50% of pigment loss in the eyes, indicating high mosaicism (Fig. 3.3 C ii). Between 50% and 90% of pigment loss in the eyes indicated a moderate *gol* loss of function. Finally, over 90% of pigmentation loss (approximately – by a qualitative observation of only single pigmented cells in the retina, if any) indicated high levels of *gol* loss of function, with none or nearly none pigmented retinal cells. It is important to note, however, that embryos indicating moderate or high levels of *gol* loss of function display a smaller size. This is not a phenotype typically associated with *gol* loss, and could indicate a developmental delay of these embryos, as all scoring was done at the same absolute time, but embryos were not segregated by stage. A developmental phenotype, whether caused by Cas9 injections, CRISPR-Cas9 targeting, or any other factors, could influence the degree of pigmentation of the embryo, thus influencing the scoring. This is a major limitation of the phenotype scoring approach and should be taken into account.

Use of *gol* as a control alongside or co-injected with other targeting guide RNAs allowed me to gain confidence and competence in zebrafish embryo microinjections,

allowing me to act to minimise the effect of embryo microinjection variability on further experiments. It also served as an aforementioned positive control, a clear and easy to spot phenotype indicating that injected Cas9 nuclease is active and efficient. It also allowed us to make improvements to the microinjection protocol (test conducted by Lucy Wheatley, data not shown), with changes to the concentration of Cas9 and other components of the injection mix allowing equimolarity between Cas9 protein and guide RNAs. Use of bioinformatics tools like CrispantCal allowed calculations of exact concentrations of injection mix components needed for optimal Cas9 efficiency. Use of the "improved" injection mix led to a significant increase in targeting efficiency (Fig. 3.4), with much fewer embryos exhibiting no *gol* loss of function (14.2% compared to initial 52.1%) and an increase in the proportion of embryos with moderate and high *gol* loss of function.

An additional step in the technical optimisation of use of Cas9 in zebrafish embryos was a switch from *in vitro* transcribed guide RNAs to chemically synthesised guide RNAs. This 2-component system consists of chemically altered RNAs: a targeting crisprRNA (or crRNA) and a trans-activating RNA or tracrRNA. These 2 RNAs together form a duplex (dgRNA) that acts as a complete guide RNA – forming a complex with Cas9 and directing mutagenesis – and has been shown to offer an improved efficiency over *in vitro* transcribed single guide RNAs (Jacobi *et al.*, 2017).

Here, use of synthetic guide RNAs has shown a significantly better targeting efficiency despite exactly same *gol* ccB guide sequence, with below 10% negative (no LOF) embryos, and increased proportion of embryos with moderate and high levels of *gol* loss of function to around 40%. Overall, these developments contributed to an increase in efficiency of Cas9-directed gene disruption, as well as to overall

replicability of experiments. It was decided that synthetic guide RNAs will be further used in the loss of function experiments.

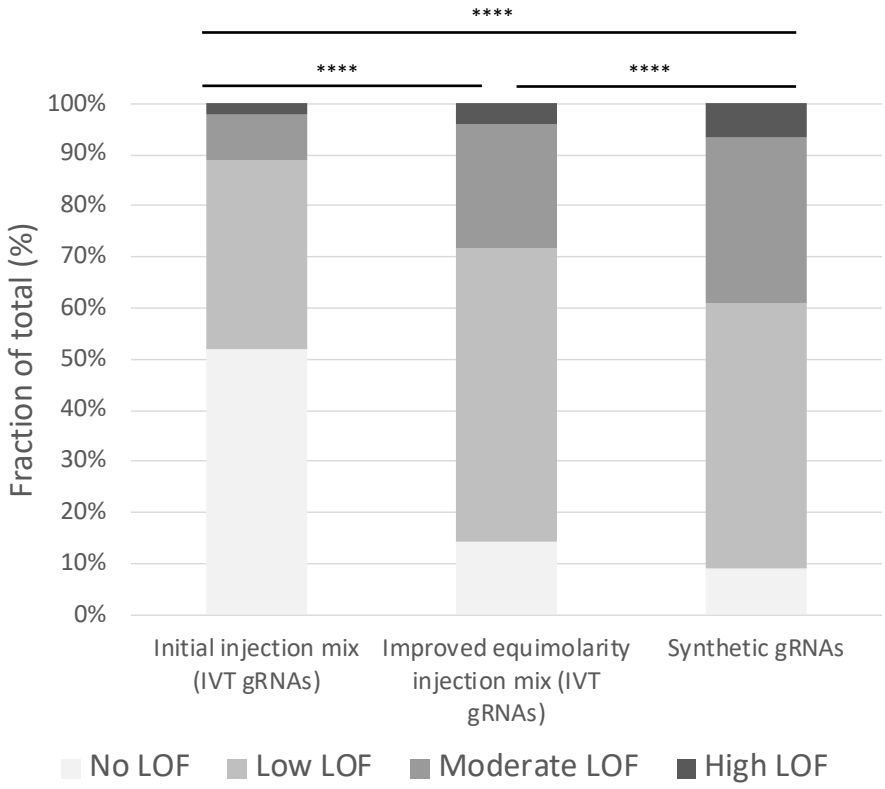


Fig. 3.4. Improvements in Cas9 efficiency illustrated by *gol* phenotype scoring.
Bar charts showing the comparison of percentages of *gol* phenotype scores achieved using the initial injection mix (left), the optimised injection mix based on Cas9 and guide RNA equimolarity (middle), and injection mix based on synthetic guide RNAs (right) across all injection rounds targeting *gol*. Phenotype scoring as in Fig. 3.3. Number of embryos scored for initial injection mix n=1142 (from 9 independent experiments); number of embryos scored for improved equimolarity injection mix n=1950 (from 12 independent experiments), number of embryos scored for synthetic gRNAs n=849 (from 7 independent experiments). Statistical analysis: chi-squared test, (****) denotes p<0.0001.

3.2.2 Design of guide RNAs targeting mxtx2

With increased confidence in the efficiency of Cas9-mediated mosaic knockouts in F0 generation, I proceeded with designing guide RNAs targeting *mxtx2*. I was aiming to both disrupt the expression of the gene by targeting the promoter, in order to generate a loss of function mutant that would not trigger genetic compensation, and to later generate a whole gene locus deletion. Therefore, I designed 4 gRNAs targeting the promoter region on the 5' end of mxtx2 gene (Fig. 3.5A-B) ("L" guides), as well as 4 guides targeting the region on the 3' end of mxtx2 gene (Fig. 3.5A) ("R" guides).

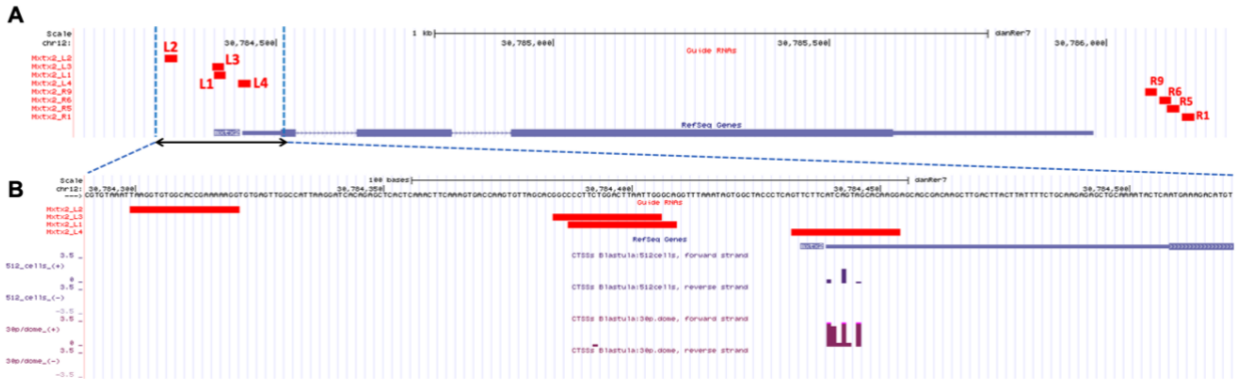


Fig. 3.5 First design of guide RNAs flanking the *mxtx2* locus.
(A) Visualisation of guide RNAs produced for deletion of Mxtx2 locus, flanking it from both 5' and 3' ends, on a genomic map (adapted from UCSC Genome browser).
(B) Expanded view of the 5' UTR and upstream region of mxtx2, showing positions of sgRNA targets for the promoter region (L1-L4) highlighted in red. CAGE-seq transcription start sites are shown on tracks below mxtx2 gene (in blue) – TSSs observed in 512-cell stage embryo (2.75 hpf, 9th cell cycle) are highlighted in purple, while TSSs observed in dome-stage embryo (4.33 hpf) are highlighted in magenta.

Sanger sequencing traces revealed that 3/4 guides designed for the 5' end of the mxtx2 gene were able to generate mutations. None (0/4) of the guides designed for the 3' end of mxtx2 gene generated indel mutations. The CHOP-CHOP bioinformatics tool was unable to find additional candidate guide RNAs with promising efficiency

scores and no off-targets that would target the 3' region of the *mxtx2* locus. With additional difficulties in amplifying the 3' region of *mxtx2* by PCR due to unexplored reasons – either repetitive region present in the vicinity of the 3' UTR of *mxtx2* disrupting the PCR, or errors in primer design for that region – I made a decision to abandon attempts of generating a genomic deletion of the whole gene locus by injecting the embryo with 1 guide RNA targeting the 5' end and 1 targeting the 3' end of the *mxtx2* gene.

In order to verify the targeting by the designed guide RNAs for the 5' region of the *mxtx2* locus, zebrafish embryos were injected at one cell stage with Cas9, *gol* ccB gRNA and individual *mxtx2*-targeting sgRNAs. Injection of none of the individual guides led to a generation of a significant visible developmental phenotype and embryos developed normally when observed until 48 hpf (Fig. 3.6A).

Following the results of Sanger sequencing, guide L3 was considered as non-efficient and was not subsequently used. Lack of discernible developmental phenotype with any of the gRNAs targeting the promoter region of *mxtx2* was discouraging, especially as these were considered efficient by the bioinformatics tools. However, it was unlikely, but not impossible that *mxtx2* expression was disrupted, but produced no phenotype. However, this would make generation of a stable transgenic line nearly impossible, and I was relying on the previous studies that loss of *mxtx2* would disrupt embryonic development and cause developmental defects.

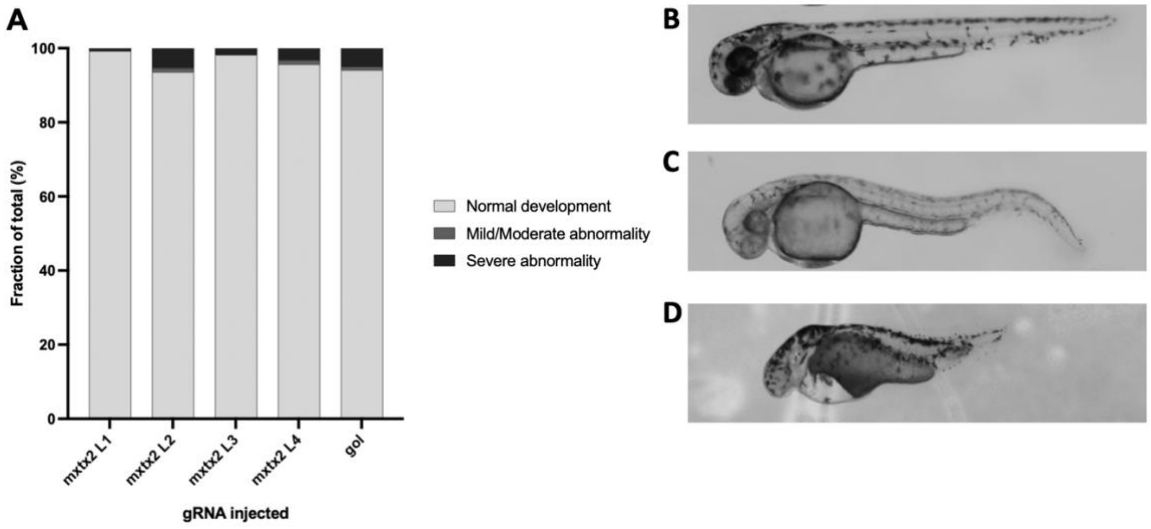


Fig. 3.6 Injections of individual *mxtx2* guide RNAs did not produce a discernible phenotype.
Scoring of developmental abnormalities in embryos injected with individual gRNAs targeting the *mxtx2* 5' region (A). Embryos were observed at 24h post fertilisation. Number of embryos observed: *mxtx2* L1 n=131; *mxtx2* L2 n=93; *mxtx2* L3 n=55; *mxtx2* L4 n=92; *gol* n=303. (B-D) representative examples of scored developmental phenotypes with (B) normal development, (C) mild abnormality and (D) severe abnormality.

Based on the findings of the study by Wu *et al.* (Wu *et al.*, 2018) I decided to inject combinations of pairs of guides in order to increase the probability of mutations being generated in the promoter region as well as create a possibility of a deletion in case of both guide RNAs efficiently directing Cas9 to cut, and thus increase the probability of mosaic knockout of *mxtx2* expression. Since 3 sgRNAs – L1, L2 and L4 – had a potential to generate indels based on Sanger sequencing, there were 3 possible permutations of guide RNA pairs (Fig. 3.7 A) - pair 1: L1+L2, pair 2: L1+L4 and pair 3: L2+L4.

Embryos injected with Cas9 and pairs of *mxtx2* 5'-targeting guides also did not show a 'yolk burst' phenotype, with very few developmental abnormalities observed (Fig. 3.7 B), and the abnormalities being unspecific (Fig. 3.7 C). Therefore, to assess

whether indel mutations introduced by injected Cas9 affect *mxtx2* expression, qRT-PCR was used to measure steady-state levels of *mxtx2* mRNA.

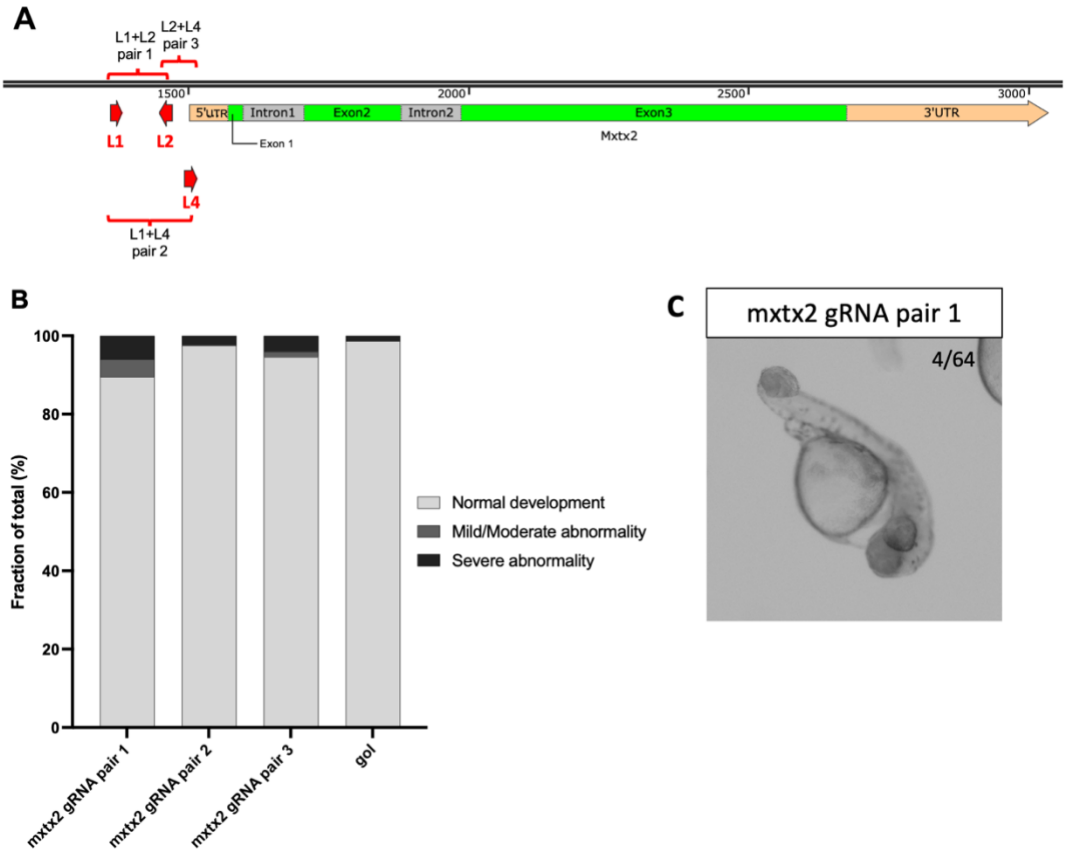


Fig. 3.7 Overview of the experimental design for pair-wise targeting of *mxtx2* promoter by CRISPR-Cas9.

(A) A simplified depiction of the *mxtx2* gene with highlighted sites targeted by mxtx2 gRNAs L1, L2 and L4, and the combinations of pairs that were injected together. Picture obtained from SnapGene® software (Insightful Science). (B) Scoring of developmental abnormalities in embryos injected with pairs of gRNAs targeting the mxtx2 5' region. Number of embryos assessed for each group: pair 1 n=67; pair 2 n=81, pair 3 n=74, gol gRNA n=74. (C) An example of a severe developmental abnormality scored in (B) in embryo injected with Cas9 and mxtx2 gRNA pair 1.

Paired injections of guide RNAs alongside Cas9 have been able to significantly reduce the relative abundance of *mxtx2* RNA in the embryos at high stage (3.33 hpf). gRNA pairs 1 and 2 were able to produce approximately 50% reduction in mxtx2 levels as tested by 2 independent primer sets targeting different parts of the mxtx2 cDNA (Fig.3.8 A-B). This was an encouraging step forward in verification of these guide RNA pairs as efficient. On the other hand, while significant, the reduction in *mxtx2* expression was still limited, perhaps explaining why there was no visible phenotype. Additionally, as the RNA collected for qPCR was pooled from multiple embryos, the obtained results are an average from multiple embryos with stochastic levels of mosaicism.

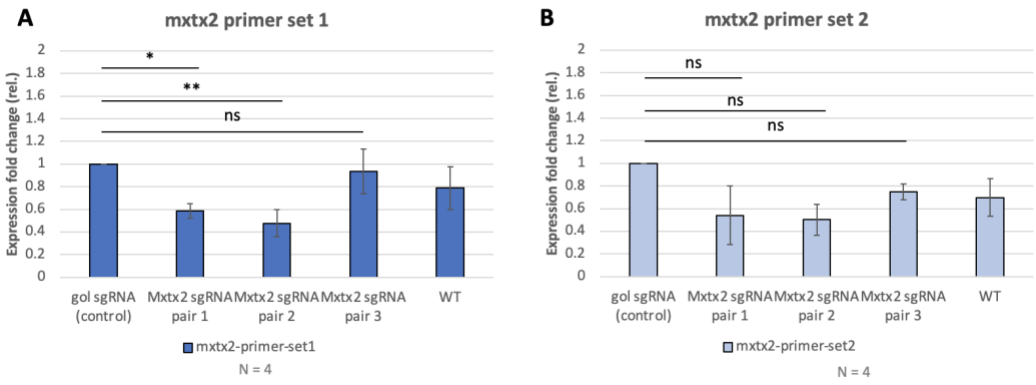


Fig. 3.8 qPCR reveals reduction of *mxtx2* expression in embryos injected with Cas9 and pairs of gRNAs.

Both (A) and (B) are bar charts of fold change of *mxtx2* expression in pools of embryos injected with Cas9 and specified guide RNAs, or in an uninjected wild-type control (WT), amplified using 2 independent sets of primers. Fold change of expression of *mxtx2* was normalised to the *eef1a1a* (*elfa)* housekeeping gene and presented as a fold change relative to the expression in embryos injected with Cas9 targeting *gol* (gol sgRNA – labelled as control). Results presented as mean of N=4 independent biological replicates ± SD. Statistical significance was tested using multiple one-way ANOVA tests with Dunnet's multiple comparisons test, using ΔCq values ($Cq_{[mxtx2]} - Cq_{[elfa]}$) between experimental groups and control (*gol*). ns p>0.05; *p<0.05; **p<0.01.

In order to further verify the effect of injected pairs of guide RNAs on *mxtx2* expression in individual embryos, *mxtx2* mRNA was visualised in embryos using whole-mount in-situ hybridisation (WISH) (Fig. 3.9 A). Mosaicism was assessed by comparing the expression patterns to the uninjected wild-type controls, which showed *mxtx2* expressed in a previously indicated ring-like pattern around the embryo margin (Fig. 3.9 A – WT panel) Only a small proportion of embryos exhibited a mosaic of complete loss of *mxtx2* expression, with 13/59 (22%) mosaic embryos injected with *mxtx2* L1 gRNA only. In embryos injected with the paired L1 and L2 gRNAs (pair 1) the proportion of embryos with mosaic mxtx2 expression was lower than in case of L1 gRNA, suggesting that in that pair it is only L1 gRNA causing Cas9 to generate an indel efficiently. There was a proportion of Cas9 mxtx2 targeted embryos showing no mxtx2 expression (Fig. 3.9 B), however these could likely be a

result of technical errors, as a similar proportion of unstained embryos was observed in the *gol* sample, not targeting *mxtx2*.

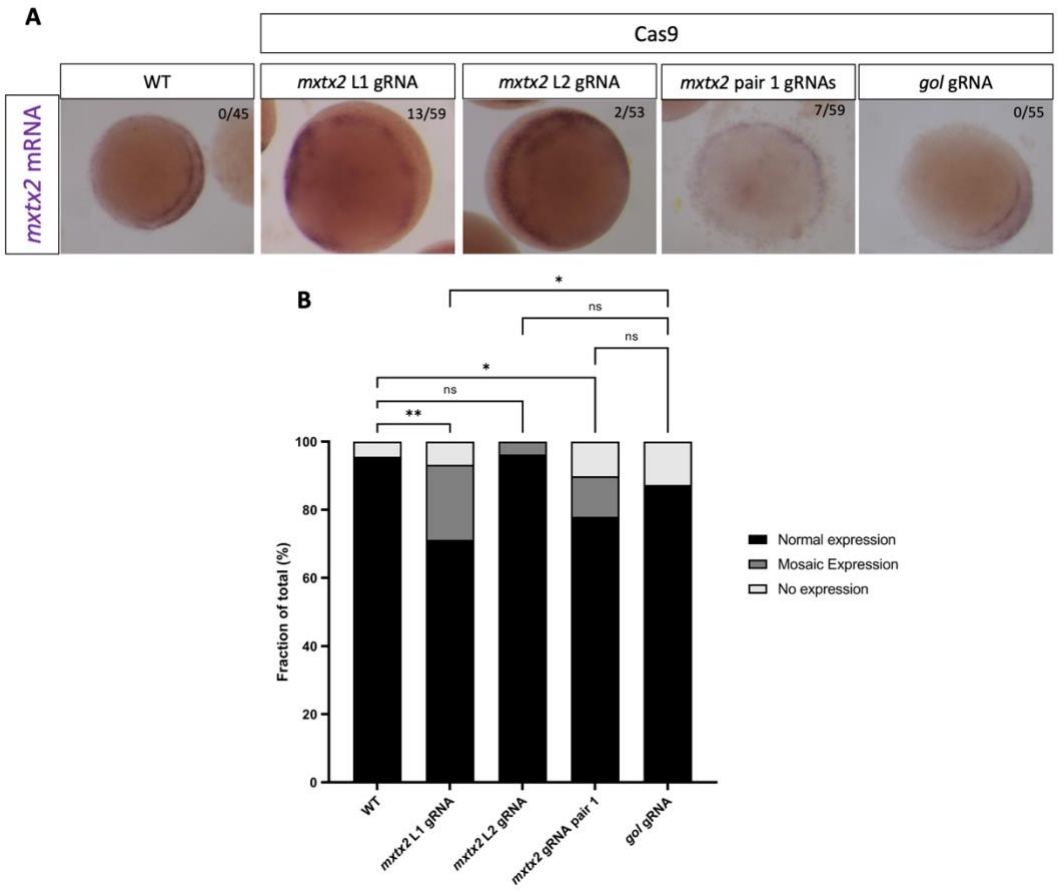


Fig. 3.9 Whole-mount in situ hybridisation of *mxtx2* mRNA in embryos injected with Cas9 targeting the promoter region of *mxtx2*.

(A) A panel of light microscopy images showing colourimetric visualisation of *mxtx2* mRNA in sample embryos from each indicated experimental group and uninjected wild-types (WT). Numbers in the top-right corner of images indicate the number of embryos with mosaic expression out of total from each group. Embryos for the WISH experiment were fixed at 50% epiboly stage. (B) Bar chart showing the fraction of embryos with particular pattern of expression of *mxtx2* (normal, mosaic, no detected expression) as a percentage of all assessed embryos. Statistical significance: Chi-square test on paired datasets, with significance denoted above the compared pairs. P-values in order from left to right: P=0.0014 (**), P=0.8672 (ns), P=0.0114 (*), P=0.0353 (*), P=0.0924 (ns) and P=0.1917 (ns).

In summary, the results obtained through scoring developmental phenotypes, Sanger sequencing, qPCR and WISH show that expression of *mxtx2* can be slightly disrupted by targeting the promoter using individual or paired guides. However, while the reduction of *mxtx2* expression is significant and reaches approximately

50%, the embryos continue to develop normally and do not exhibit the previously reported yolk burst phenotype. WISH visualisation of *mxtx2* expression in the tested samples also showed low numbers of embryos exhibiting mosaicism in *mxtx2* expression or its loss, further indicating that targeting the promoter might be an inefficient strategy of inducing loss of function of *mxtx2*, and that more reliable options should be explored.

3.3. Generation and Verification of Mosaic *mxtx2* knockout

Due to the limited reliability of the promoter-targeting guide RNAs in producing an *mxtx2* knockout, I have decided to restart the process and instead of targeting the promoter of *mxtx2*, expand the search for guide RNAs to the whole gene sequence. In addition, I have employed the use of CRISPRscan software (Moreno-Mateos *et al.*, 2015), as the algorithm has been shown to find guide RNAs that were highly efficient in F0 gene knockouts in zebrafish (Kroll *et al.*, 2021) and was designed for use in zebrafish transgenesis by a group specialising in the maternal-to-zygotic transition (Giraldez group). A group of 4 new guide RNAs with the highest reported efficiency scores was selected (Fig. 3.10). Interestingly, all of those guides targeted exon 2 of the *mxtx2* gene (Fig. 3.10 A), with 3 out of 4 targeting the sequence coding for the homeobox domain of *mxtx2* (Fig. 3.10 B), and one targeting the sequence directly preceding the domain.

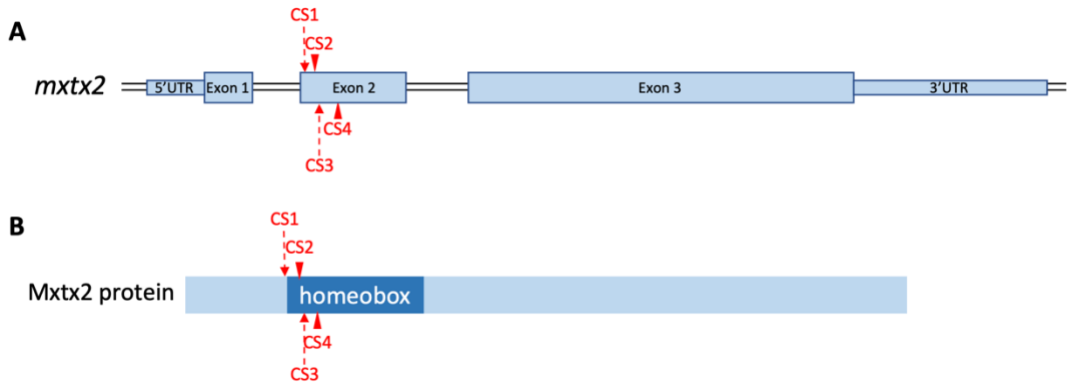


Fig. 3.10 Schematic of design of guide RNAs targeting the exon 2 of *mxtx2*.
(A) A simplified diagram depicting the mxtx2 gene and distinguishing the 5' UTR, exons, introns and 3'UTR. Approximate positions targeted by newly designed guide RNAs (CS1-CS4) are marked with red arrows. (B) A very simplified diagram of the Mxtx2 protein amino acid sequence with a highlighted position of the homeobox DNA binding domain (dark blue). Approximate positions that would be disrupted by guide RNA targeting are marked with red arrows annotated with the associated guide RNA.

Embryos were injected with Cas9 and individual guides at 1-cell stage and collected in batches of 3 for each sample at 24 hours post fertilisation. Following that, genomic DNA of embryos was extracted and sent for sequencing. Efficiency of indel formation by Cas9 targeting by guide RNAs was verified by analysis of Sanger sequencing traces over the amplified region of exon 2 of *mxtx2*, with none of the guide RNAs producing a significant disruption in the genomic sequence (Fig. 3.11 A-D). Guides CS2 and CS4 showed very minor disruptions, but these results indicated low guide targeting efficiency in the injected embryos. This could be due to the timing of collection of embryos for DNA extraction at 24 hpf. Severe developmental phenotypes and embryo death, that could be caused by loss of Mxtx2, would occur earlier, therefore the batches are unnecessarily biased towards healthy, unaffected embryos containing low or no Mxtx2 loss of function. Due to this issue I decided to proceed with observations of embryos with disrupted *mxtx2* expression using these guide

RNAs (individually), to establish whether these are sufficient to produce a developmental phenotype.

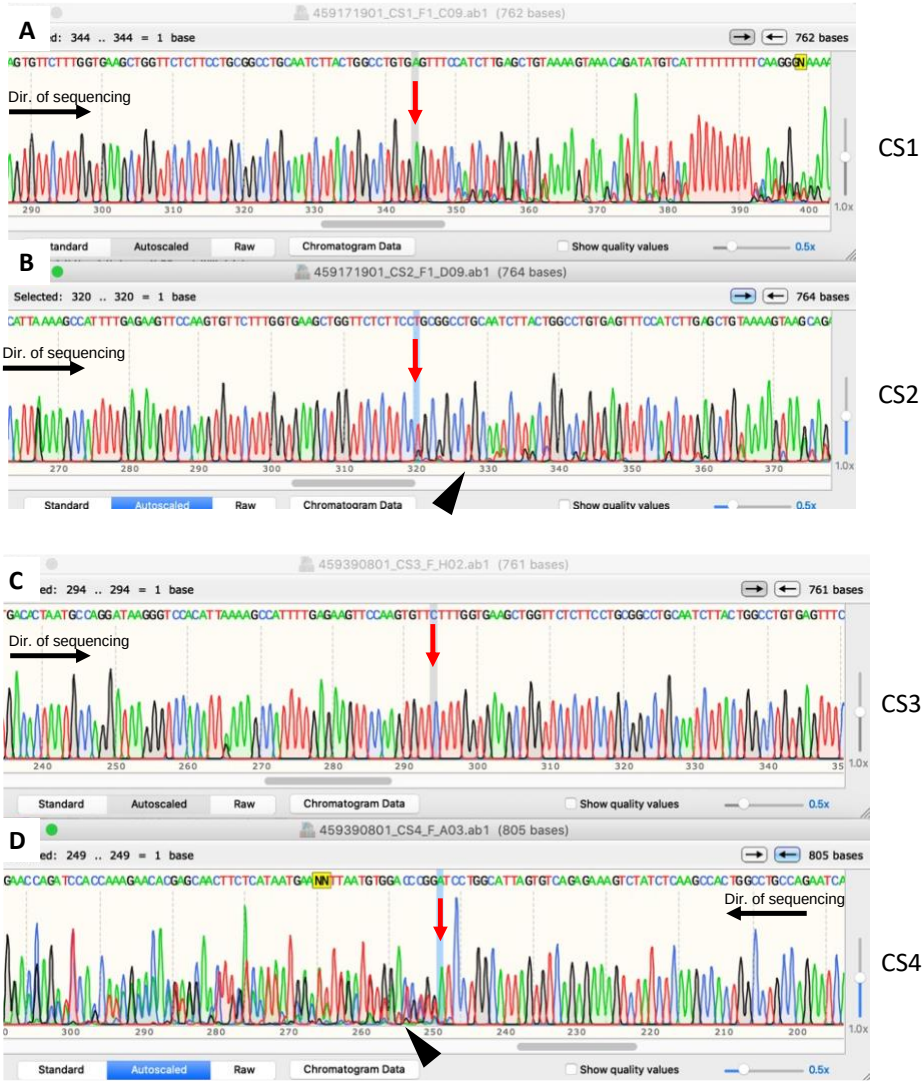


Fig. 3.11 Sanger sequencing of the genomic region of *mxtx2* in embryos injected with Cas9 targeted to the coding sequence of the gene.

Sanger sequencing traces of the region surrounding the mxtx2 exon 2 where all of the CS guides are targeted, used as a readout indicating Cas9 targeting and introduction of indel mutations (or lack of it) in embryos injected with individual guide RNAs designed to target sites indicated in Fig. 3.10. Red arrows denote approximate positions of Cas9 nuclease target sites, while black arrowheads denote the small disruption in genome sequencing indicative of Cas9-induced indels. Screenshots of the sequencing traces were obtained from SnapGene® software (from Insightful Science; available at snapgene.com).

Surprisingly, that was the case, as embryos injected at one-cell stage with Cas9 and individual guides targeting exon 2 of *mxtx2* exhibited a significant number of developmentally disrupted, as well as dead embryos. The pattern of developmental disruption was consistent in the groups with *mxtx2* targeting by Cas9, with a large proportion of embryos arresting approximately around bud stage and failing to develop further when observed at 24 hpf. All groups also had a significant proportion of embryo death that was not present in the *gol*-targeted control embryos not targeting *mxtx2*. This indicated that the presented phenotypes can likely be associated with the effect of *mxtx2* disruption by Cas9. Presence of similar phenotypes in 4 groups, injected independently, was unlikely to be attributed to a technical error or toxicity of the injected components. Additionally, injections of CS4 guide RNA without Cas9 did not result in an increased embryonic lethality, indicating that this effect was Cas9-dependent, while lack of the phenotypes in the *gol*-targeted control confirmed that the effect could not be due to the toxicity of Cas9 nuclease itself. Therefore, it could be concluded that the visible phenotypes and embryo death were induced by the Cas9 targeting *mxtx2*. The effects of guide RNAs CS2 and CS4 stood out, with CS2 causing the highest proportion of developmentally disrupted embryos (27/85, 37.8%), while CS4 guide RNA had by far the highest proportion of induced embryo lethality (47/51, 92.2%). It was still not known however if the guide RNAs truly cause disruption to *mxtx2* expression.

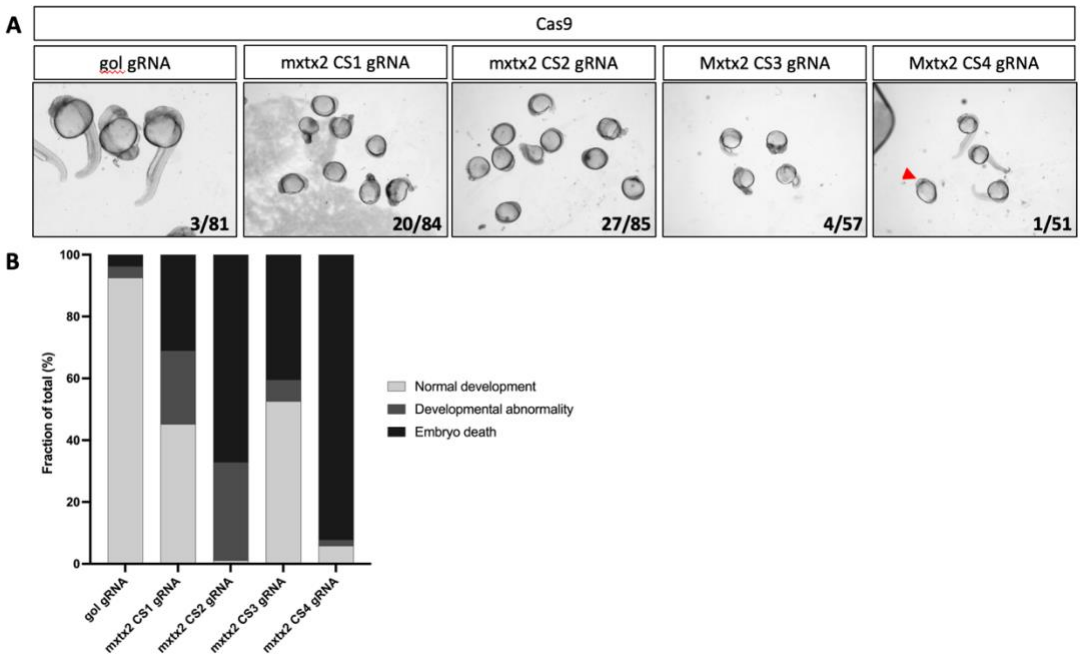


Fig. 3.12 Developmental defects in embryos injected with Cas9 targeting mxtx2 exon 2.
(A) A panel of light microscopy images depicting representative developmentally abnormal embryos observed at 24 hpf in groups injected with Cas9 and indicated guide RNAs. Numbers in the bottom right indicate the number of observed developmentally abnormal embryos out of total. In the mxtx2 CS4 gRNA injected group (right-most panel), the image contains all surviving embryos, with the developmentally abnormal marked with a red arrowhead.
(B) Scoring of developmental abnormalities in embryo groups depicted in (A) - injected with Cas9 and indicated gRNAs targeting the mxtx2 exon 2 and Cas9 and *gol* gRNA as a non-targeting control. Number of embryos assessed for each group: *gol* n=81; mxtx2 CS1 n=84, mxtx2 CS2 n=85, mxtx2 CS3 n=57; mxtx2 CS4 n=51.

Therefore, I have decided to visualise *mxtx2* expression using WISH, as before. Embryos were injected with Cas9 alongside individual guide RNAs and fixed at 30% epiboly stage, following which a digoxygenin-labelled probe was used to bind *mxtx2* mRNA. Anti-digoxygenin antibody was used to bind and visualise the probe bound to the *mxtx2* mRNA. Through that, I could observe the patterns of *mxtx2* expression and identify embryos where this pattern was disrupted or lost. Embryos injected with Cas9 and guides targeting exon 2 of *mxtx2* exhibited very high levels of disruption of expression. All of the guides induced some level of disruption in at least 80% of all assessed embryos (Fig. 3.13). From all of the gRNAs tested, CS4 had the most striking

effect, with 36/42 (85.7%) embryos having no detectable expression of *mxtx2*. These results should be taken with a caveat that 7/39 (17.9%) of control embryos with targeted *gol* instead of *mxtx2* showed no *mxtx2* signal (Fig.3.13). This was most likely due to the human error along the protocol and uneven staining between embryos. However, detected *mxtx2* depletion was much larger in case of all of the experimental samples, suggesting that it was likely not due to experimental error. Embryos also expected a varying degree of mosaic *mxtx2* depletion, suggesting that specific subpopulations of cells were selectively affected, and a Cas9-induced lesion was introduced in those cell populations. These results, taken alongside the phenotype penetrance, suggested that *mxtx2* is efficiently depleted in zebrafish embryos through use of these guide RNAs.

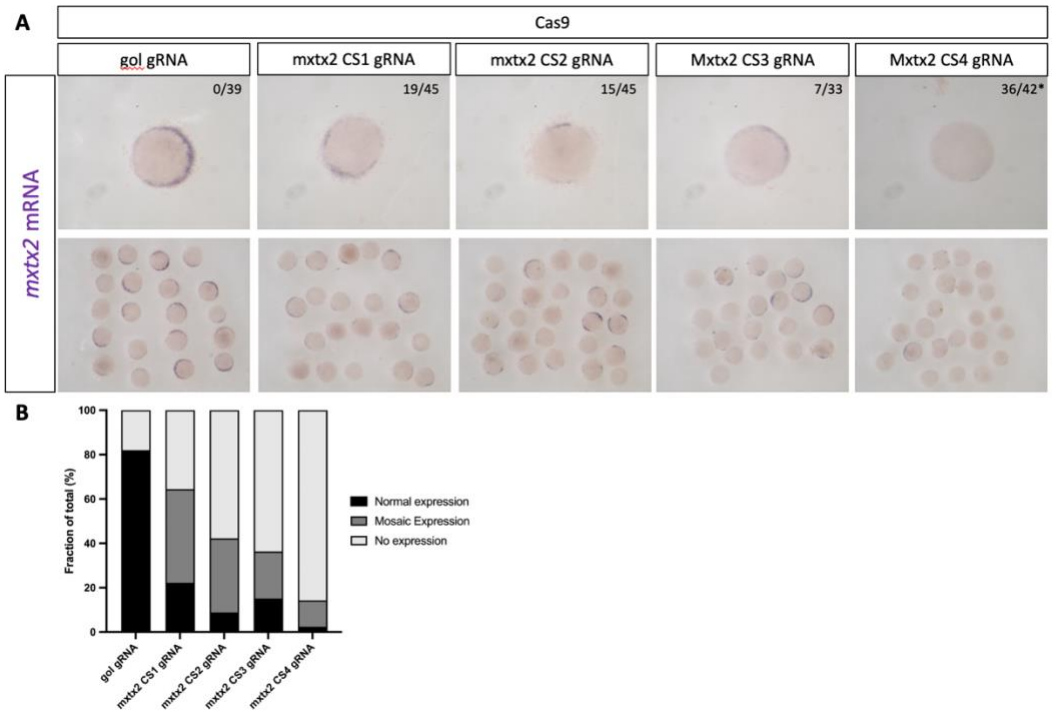


Fig. 3.13 Whole-mount in situ hybridisation of mxtx2 mRNA in embryos injected with Cas9 targeting exon 2 of mxtx2.

(A) A panel of light microscopy images showing colorimetric visualisation of *mxtx2* expression in representative sample individual embryos from each indicated experimental group (top row) and groups of embryos (bottom row). Numbers in the top-right corner of images indicate the number of embryos with mosaic *mxtx2* expression out of total from each group (*in case of CS4 gRNA, the representative picture shows an embryo with no expression of mxtx2, and the number indicates the proportion of embryos with no *mxtx2* expression out of total). Embryos for the WISH experiment were fixed at 30% epiboly stage. (B) Bar chart showing the fraction of embryos with each of the scored patterns of expression of *mxtx2* (normal, mosaic, no detected expression) as a percentage of all assessed embryos. N=1.

To further verify that *mxtx2* is indeed targeted by Cas9 and exon 2 guide RNAs, I have followed with qPCR assay, quantifying *mxtx2* expression. I have selected guide RNAs CS2 and CS4 for this analysis, as guide CS4 has shown a high level of mosaic knockout in WISH, and CS2 guide had a more moderate effect, with more mosaic embryos. qPCR was proceeded with as before. Embryos were injected with Cas9 and guide RNAs at one cell stage, and pools of 25 embryos for each group were collected for RNA extraction at the 30% epiboly stage (4.66 hpf). This stage was selected due to high levels of expression of *mxtx2* reported in RNAseq studies, allowing for a greater

dynamic range in the quantification of *mxtx2* loss. Relative expression was assessed with use of 2 independent primer pairs amplifying separate regions of *mxtx2* (Fig. 3.14 A).

The results of qPCR analysis did not match the significant depletion phenotype shown by WISH (Fig. 3.14 B). Targeting *mxtx2* with guide RNA CS2 showed no significant change to the relative abundance of *mxtx2* mRNA in the embryos. Guide CS4, assessed by WISH to be more efficient, showed a larger degree of depletion of *mxtx2* mRNA (up to 40%), but only with one of the primer pairs, overall indicating a not significant depletion of *mxtx2* mRNA in the injected embryos. This was an unexpected result, as qPCR analysis of guide RNAs with no developmental phenotype and small effect on *mxtx2* expression assessed by WISH showed a higher degree of *mxtx2* depletion. On the other hand, qPCR was conducted on pooled sets of embryos, and it is possible that penetrance of *mxtx2* targeting and degree of mosaicism varied between each of the biological replicates. Human error in the qPCR protocol – such as pipetting errors in loading the 96-well plate - also may have been a source of variability and would in extent explain the variability of obtained results. Nevertheless, assessment of *mxtx2* mRNA expression by qPCR failed to verify the guide RNAs. It must be noted that primer pairs used in this verification were located downstream of potential Cas9-induced lesions, which should make them good indicators of *mxtx2* expression disruption. Ideally, qPCR primer pair located near the Cas9 targeted site should be used to accurately detect expression disruption. Despite the non-significant disruption of expression of *mxtx2* measured by qPCR, and due to limitations of qPCR experiments mentioned above, I decided to follow the results of

WISH and use guide CS4 as a tool for generating a *mxtx2* mosaic knockout in the zebrafish embryo.

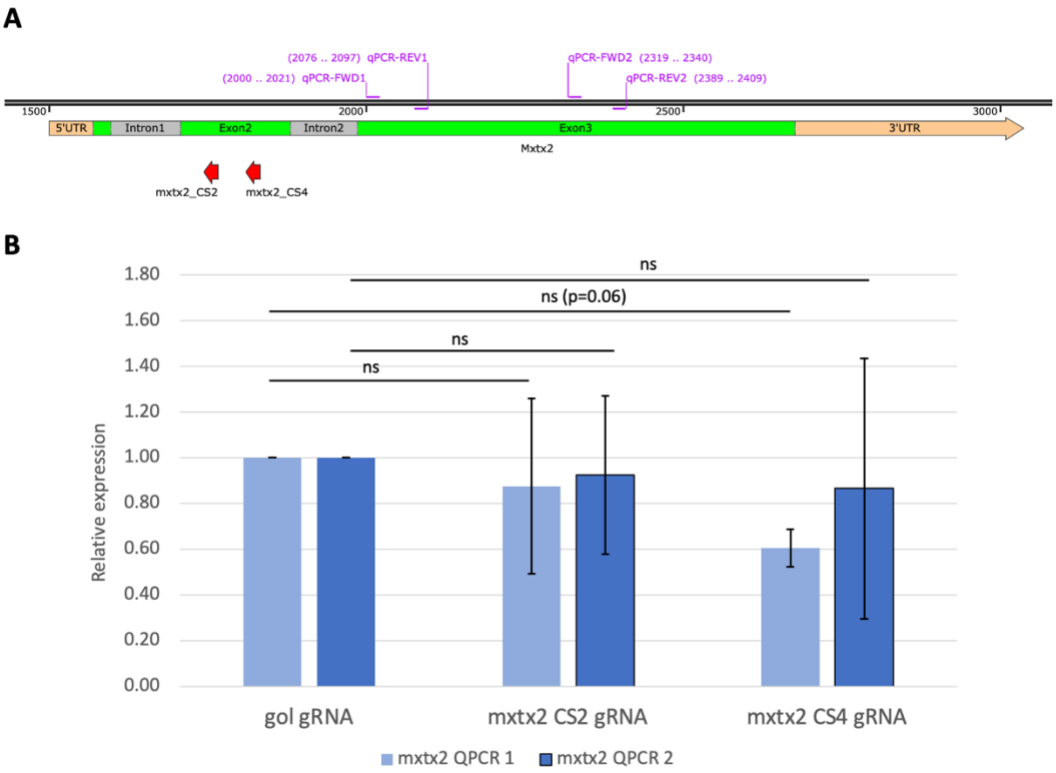


Fig. 3.14 Analysis of mxtx2 expression in embryos injected with Cas9 and guide RNAs targeting the mxtx2 exon 2 by qPCR.

(A) A schematic diagram of *mxtx2* locus, showing locations of CS2 and CS4 guide RNAs (red arrows) and positions of primers used for qPCR analysis (purple lines). (B) A bar chart of fold change of *mxtx2* expression in pools of embryos injected with Cas9 and specified guide RNAs amplified using 2 independent sets of primers. Fold change of expression of *mxtx2* was normalised to the *eef1a1a* (*elfa)* housekeeping gene and presented as a fold change relative to the expression in embryos injected with Cas9 targeting *gol*. Results presented as mean of N=4 independent biological replicates ± SD. Statistical significance was tested using multiple one-way ANOVA tests with Dunnet's multiple comparisons test (ns $p>0.05$).

3.4. Adaptation of Cas13d – RNA Targeting CRISPR Tool

CRISPR-Cas9 offered an opportunity to generate mosaic mutations, and, in perspective, a stable mutant line with disrupted expression of the target gene. However, in my study I wanted to target genes disruption of which was likely to produce a severe embryonic phenotype, if not embryonic lethality. Additionally, generation of stable lines is a laborious and time-consuming process, and as shown above, CRISPR-Cas9 does not always offer exceptionally high efficiency of mutagenesis. Therefore, as outlined in the introduction to this chapter, I was looking for independent methods of disrupting the *mxtx2* gene. One such option would be morpholino oligonucleotide, however the observations of *mxtx2* disruption using that method have already been reported (Wilkins *et al.*, 2008; Xu *et al.*, 2012).

An alternative form of RNA interference presented itself with the advancement in use of CRISPR-Cas13. As the method was adapted with success to use in zebrafish embryos(Kushawah *et al.*, 2020), I have decided to use it to efficiently knock down *mxtx2* expression in embryos. Success with this method could result in subsequent use of this approach to knock down other early zygotic genes, significantly simplifying the process of obtaining loss of function phenotypes. This would be additionally helped by the fact that RNA interference mutants could be studied in F0 generation – immediately after delivery of CRISPR-Cas13.

Thus, I have proceeded to first verify that the method and protocol of Cas13d-dependent RNA targeting in zebrafish embryos established by Kushawah *et al.* is efficient (Hernandez-Huertas *et al.*, 2022; Kushawah *et al.*, 2020). This would serve

as a basis for further application of this method and establish a baseline expected efficiency.

The first objective was to obtain and produce components of the CRISPR-Cas13d complex. In the study, the group uses 2 methods of Cas13d nuclease delivery to the embryo. First as a capped mRNA that is then translated in the embryo, and second as purified protein.

Initially, I obtained a plasmid containing a construct of Cas13d and a fluorescent reporter (EGFP), with a strong elongation factor 1 alpha (EF-1a) promoter (Fig. S2). The 2 protein coding sequences are separated by a linker containing a P2A peptide, which causes cleavage during translation by ribosome skipping (Liu *et al.*, 2017). This allows Cas13d and EGFP to form separate proteins post-translationally. Fusion of Cas13d and EGFP could lead to an unlikely event in which its nuclease activity or ability to bind RNAs is disrupted.

Cas13d and EGFP have been cloned out of this plasmid and inserted into a more convenient expression plasmid used by the Müller group (Hadzhiev *et al.*, 2019): a modified version of pCS2 (Fig. S3). This plasmid contains a T7 promoter that would be used in *in vitro* transcription, generating capped mRNA of Cas13d-2A-EGFP, which will subsequently be injected into embryos.

Plasmid DNA can be injected into zebrafish embryos to cause ectopic expression (Koster and Fraser, 2001). To verify if Cas13d can be transcribed and translated in embryos, I injected one-cell stage embryos with 2ng/µl of the pCS2+T7-Cas13d-2A-EGFP plasmid (Fig. 3.15). Expression of the plasmid was driven by a cytomegalovirus (CMV) promoter. Since EGFP does not contain a separate promoter to Cas13d, green fluorescence should indicate that the whole construct can be transcribed and

translated in the embryo. Indeed, I have observed ectopic and mosaic expression of GFP as early as the sphere stage and 30% epiboly (Fig. 3.15 B-C) when viewing the embryos under a fluorescence microscope. Mosaic expression was expected from the plasmid DNA, however presence of GFP in early development indicated that Cas13d could be translated early enough in the embryo to be of use in disruption of not only zygotic, but potentially also maternal transcripts.

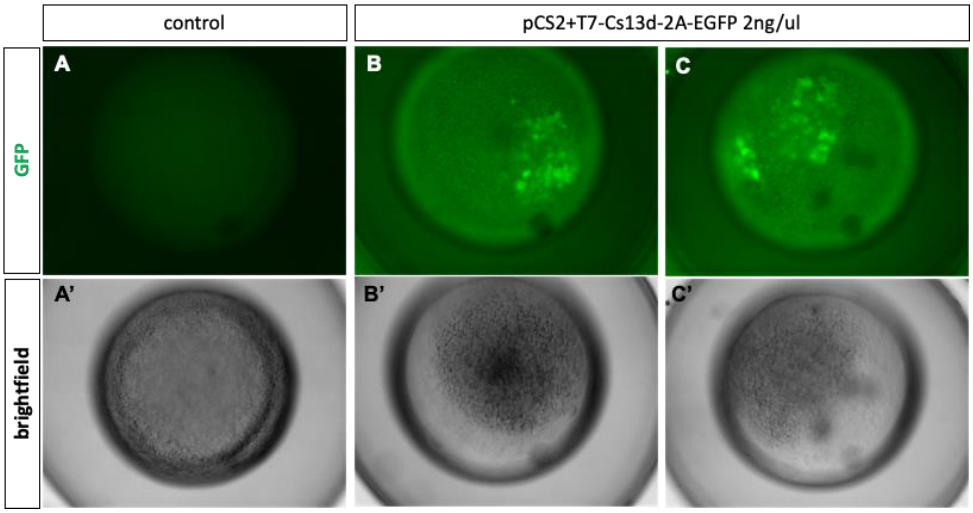


Fig. 3.15 Injections of plasmid DNA carrying the Cas13d-2A-EGFP show expression of the green fluorescent marker at 50% epiboly.
A panel of light microscopy images showing the green fluorescence channel (A, B, C)) and brightfield view (A', B', C') of embryos injected at 1-cell stage with the indicated solutions (control = nuclease-free water). Images were taken at 5.66 hpf. 1 representative control embryo and 2 separate representative embryos injected with the Cas13d-2A-EGFP plasmid are presented.

Following that, I have *in vitro* synthesised capped mRNA of the Cas13d-2A-EGFP construct for further use in RNA targeting in embryos. Simultaneously, I have designed guide RNAs for use with Cas13d. This was based on the protocol established by Kushawah *et al.* and a simplified cartoon depicting guide RNA design and production is presented in Fig. 3.16.

RNAfold software was used to predict the structure of the target mRNA (Lorenz *et al.*, 2011). Unlike Cas9, Cas13d does not require a PAM motif, therefore the potential sequence of target sites is unrestricted. The target site should contain 22 nucleotides, and the candidate sites were selected based on regions with the lowest base-pairing probability based on minimum free energy predictions.

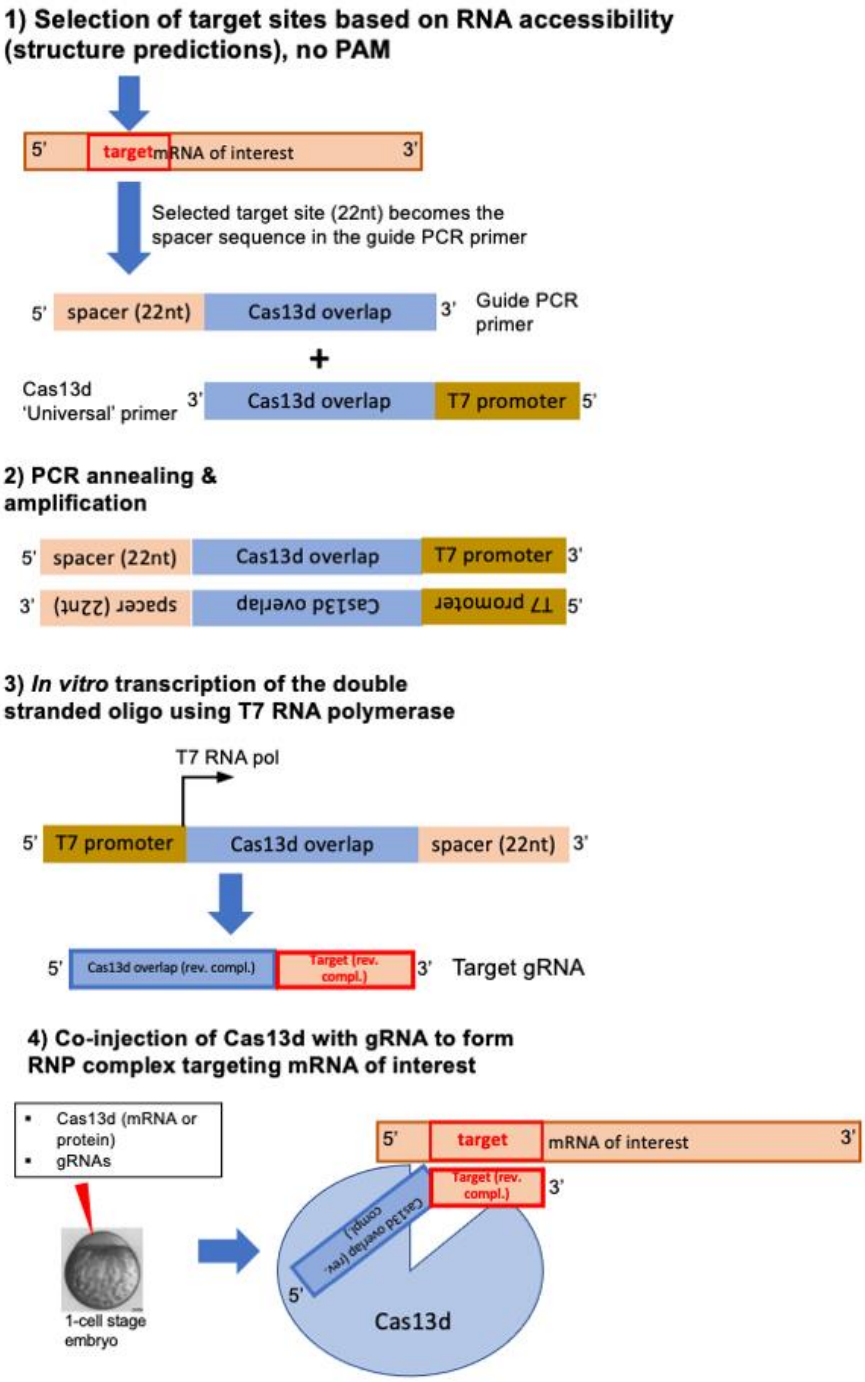


Fig. 3.16 Simplified overview of Cas13d guide RNA design, production and delivery.

Following that, a DNA oligonucleotide containing the targeted sequence (spacer) and a Cas13d overlap (a direct repeat forming a stem loop that allows Cas13d-gRNA complex formation) was ordered. Then, this oligo was PCR annealed and amplified with a "universal" oligo, containing a T7 promoter and a complementary Cas13d overlap sequence. Then, obtained double-stranded DNA oligos serve as a template for the T7 RNA polymerase in an in-vitro transcription reaction. The result of it is a guide RNA, containing a reverse complement sequence of the Cas13d repeat and a 22 nucleotide sequence complementary to the mRNA of interest. Guide RNAs can then be microinjected alongside Cas13d mRNA or protein into the 1-cell stage zebrafish embryo, where the Cas13d-gRNA ribonucleoprotein complex would target the mRNAs of interest.

In the adaptation of this technique, it was found that injection of multiplexed guide RNAs targeted to the same mRNA of interest led to a higher efficiency of mRNA targeting (Konermann *et al.*, 2018; Kushawah *et al.*, 2020). I have therefore also used multiplexed guide RNAs in groups of 3.

3.4.1. Cas13d mRNA and protein injections produce an unspecific phenotype

Initially, I used guide RNAs designed and tested by Kushawah *et al.* as positive controls for the mRNA targeting in the embryo. Guide RNAs targeting *tbxta* (*ntl*) gene were used as an example of targeting one of the key genes involved in early development, one with a distinguishable phenotype, and expressed in the embryo approximately around the same time as the peak of expression of *mxtx2*. Embryos were injected with the Cas13d mRNA at a reported optimal concentration of 200ng/µl (200 pg per embryo), and a combination of 3 previously reported efficient

guide RNAs at a combined concentration of 600 ng/µl (600 pg per embryo). The obtained results were not expected. Cas13d targeting of *tbxta* failed to produce the easily distinguishable no-tail phenotype, instead causing an unspecific phenotype in injected embryos that was dose-dependent and was not a result of *tbxta* targeting, as embryos injected with only the Cas13d mRNA and no guide RNAs displayed the same unspecific phenotype (Fig. 3.17). Embryos with the unspecific phenotype displayed several characteristics and the degree of severity of this phenotype varied between embryos. There was a visible increase of cell death in the epithelium, disruptions to the anterior-posterior and dorso-ventral body axes, multiple severe developmental defects including defective somitogenesis, defects in brain formation, and in most severely affected embryos cyclopia (Fig. S4) or even headlessness. This effect has also been observed in injections with other guide RNAs. It was therefore impossible to utilise Cas13d mRNA as a tool of disruption of expression of developmental genes due to the severe developmental phenotype effectively masking any of the phenotypes induced by true gene depletion.

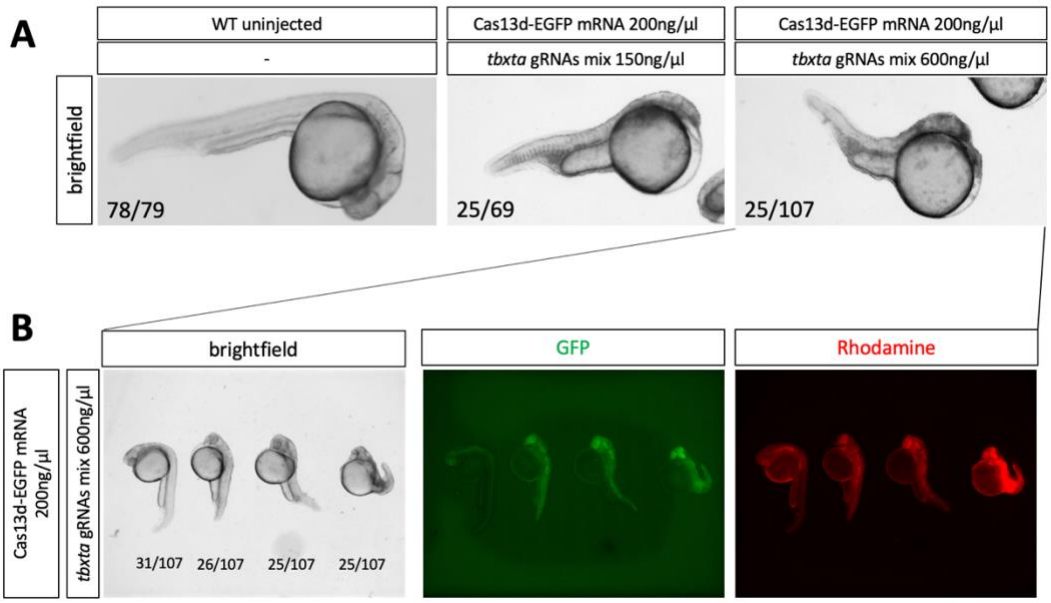


Fig. 3.17 Injections of Cas13d mRNA caused a non-specific phenotype in zebrafish embryos.
(A) Images of representative embryos injected with the denoted concentrations of Cas13d-EGFP mRNA and guide RNAs targeting *tbxta* at 2 concentrations, and wild-type uninjected control. Embryos were imaged at 24hpf. In the left panel, an embryo representative of wild type uninjected controls is presented. Embryos representative of what has been considered a moderate non-specific phenotype have been presented alongside their proportion in the injected group (N=1).
(B) Images of embryos injected at 1-cell stage with the Cas13d-EGFP mRNA (200ng/µl) and a mix of guide RNAs targeting *tbxta* (600ng/µl total gRNA concentration). (Left panel) Embryos were assorted based on severity of the unspecific phenotype and counted, with representative embryos for each group and the number of embryos from total assessed presented in the brightfield panel. Embryos were imaged at 24hpf. (Middle panel) Image showing the green fluorescence of EGFP (translated alongside Cas13d in embryos) in representative embryos from (Left panel). (Right panel) Image showing the fluorescent marker rhodamine that was added to the injection mix to serve as a marker of injected embryos and an indicator of the distribution of the injection mix in the embryo.

Aiming to determine whether this severe unspecific phenotype was caused by the method of delivery of Cas13d as a capped mRNA, I have sought out to use Cas13d protein instead. Cas13d protein was produced from the pET-28b-Cas13d-his plasmid (Fig. S4). Plasmid production and purification was conducted by Dr Neville Gilhooly, who also kindly gifted the purified Cas13d protein and the dialysis buffer used in protein purification for use as a control. Injections of the Cas13d protein into the one cell embryos at the concentration suggested by Kushawah *et al.* – 250ng/µl, and

without targeting guides resulted in a severe embryo toxicity and lethality at 24 hours post fertilisation, and serial dilutions of the protein resulted in a decrease in toxicity, but an increase in the proportion of embryos with the unspecific phenotype (Fig. 3.18). Embryonic lethality of the Cas13d seemed to be dose-dependent.

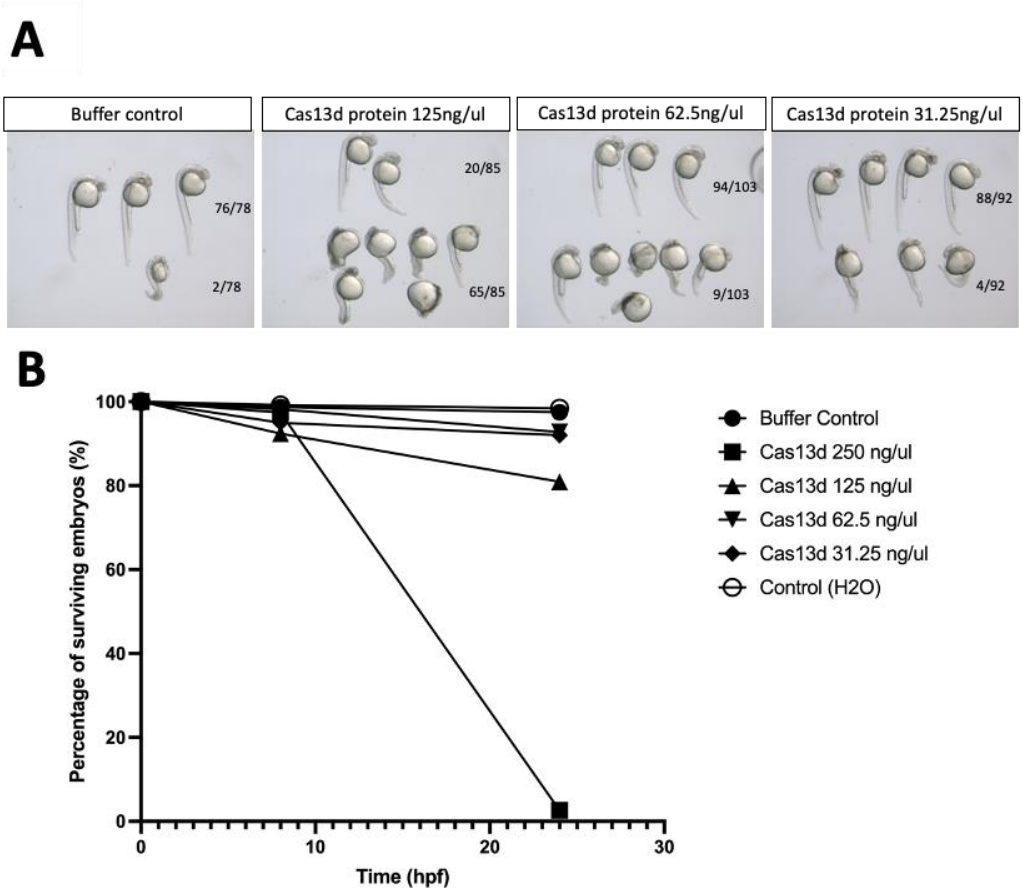


Fig. 3.18. Injection of purified Cas13d protein causes embryonic lethality and unspecific developmental phenotypes in 24hpf embryos.

(A) Panel of images showing representative embryos from each experimental group at 24hpf. "Buffer control" embryos were injected with the dialysis buffer used for Cas13d protein storage at an equal volume. Representative embryos developing normally (top) and abnormally (bottom) were selected, with numbers indicating the proportion of each group as part of all alive embryos. (B) Graph displaying survival of embryos from marked groups injected with serially diluted concentrations of Cas13d at 1 cell stage, as measured at 8hpf and 24hpf. "Control (H2O)" group were embryos injected an equal volume of nuclease-free water.

Injecting embryos at 1-cell stage with Cas13d protein and guide RNAs targeting the pigmentation gene tyrosinase *tyr* (replicating the method used in (Kushawah *et al.*, 2020)), a gene which should not have any effect on embryo development, showed the same embryonic lethality at a high dose of Cas13d (250 ng/µl) and an increase in embryonic lethality when injected at the lowest tested protein concentration (31.25 ng/µl) (Fig. 3.19B). This indicates that the embryonic lethality could be resulting from the Cas13d protein activity in the embryo, and not solely its presence. This does not apply to the highest tested concentration, which caused total lethality with both untargeted and *tyr* targeted Cas13d.

Cas13d targeting of *tyr* mRNA did not result with an observable loss of pigmentation in embryos at 48hpf, in contrast to the results obtained by (Kushawah *et al.*, 2020) (Fig. 3.19 A). At 24hpf 25% (14/55) of embryos injected with the low dose of Cas13d protein and *tyr* guide RNAs exhibited the same unspecific phenotype as described before, with varying degrees of severity (Fig. 3.19A, rightmost panel). These results indicated that observed toxicity of the Cas13d protein injection could potentially be correlated with its enzymatic activity, and was independent of the method of nuclease delivery.

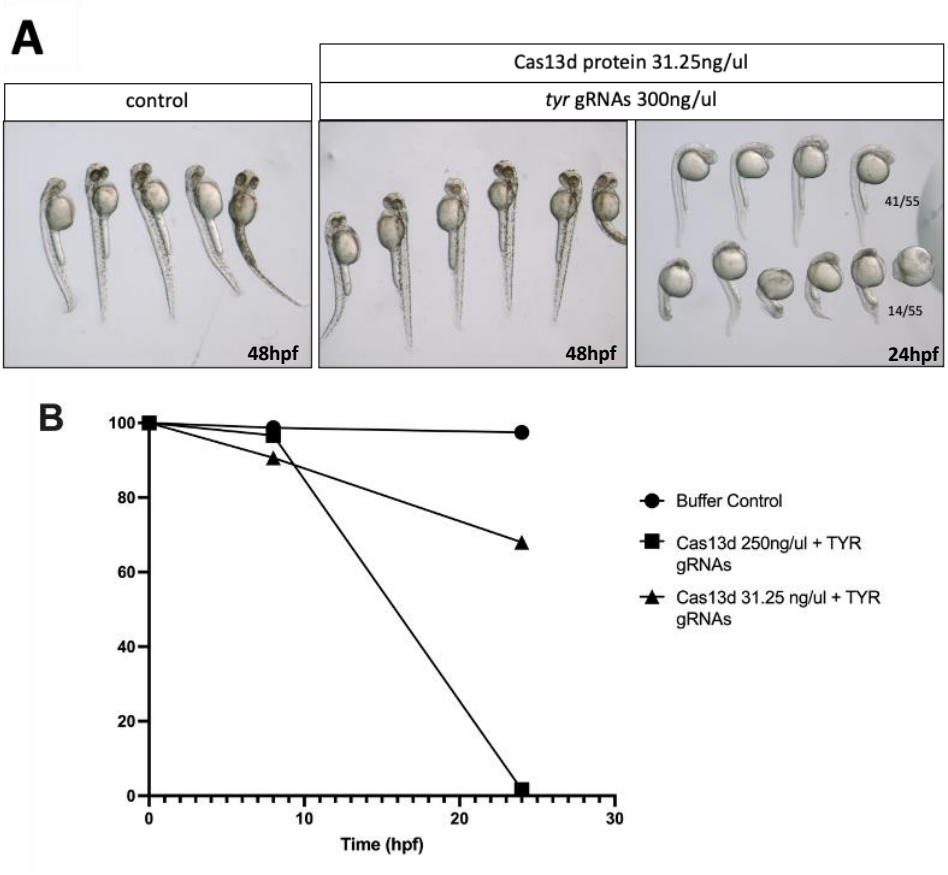


Fig. 3.19. Injection of purified Cas13d protein targeting the *tyr* mRNA does not reduce embryo pigmentation; causes embryonic lethality and unspecific developmental phenotypes in 24hpf embryos.

(A) Panel of images showing representative embryos from each experimental group at the time indicated in the bottom right. Control embryos were injected with the dialysis buffer used for Cas13d protein storage at an equal volume. In the left and middle panel, representative embryos were selected at 48hpf to compare pigmentation levels. In the right panel, representative embryos developing normally (top) and abnormally (bottom) are presented, with numbers indicating the proportion of each group as part of all alive embryos. (B) Graph displaying survival of embryos from marked groups injected with indicated injection solutions at 1 cell stage, as measured at 8hpf and 24hpf.

3.4.2. Targeting mRNA with Cas13d fails to efficiently disrupt gene expression.

The results presented above provided sufficient evidence of Cas13d having an unspecific effect in the embryos, causing a severe developmental phenotype and embryo death. Nevertheless, the question remained whether Cas13d was targeting mRNA in the injected embryos. Results from Fig. 3.19A showed that tyrosinase was not efficiently knocked down by Cas13d. In confirmation of these results, when targeting *mxtx2* in embryos using a mix of 3 guide RNAs designed and produced in-house and Cas13d mRNA (that was better tolerated by the embryos in terms of survival), expression of *mxtx2* is not changed. Visualising the *mxtx2* mRNA by WISH (Fig. 3.20) showed that Cas13d did not disrupt expression of *mxtx2* in 30% epiboly embryos. Bearing in mind that Cas13d could only gradually reduce expression instead of creating mosaic patches of cells with no expression like Cas9, I have used a different scoring system for assessment of the Cas13d *mxtx2* knockdown by WISH, based on the intensity and completeness of the staining around the embryo margin. Embryos stained by WISH were separated into 4 groups – 1) with complete loss of expression; 2) weak, but discernible staining, 3) strong, but partial margin staining, 4) strong and complete margin staining (Fig. 3.20 C). There were no differences in the proportions of each expression scoring groups among all of the experimental samples (including a sample with a doubled concentration of guide RNAs, tested to check if it could potentially influence efficiency). Therefore it can be concluded that Cas13d targeting to the *mxtx2* mRNA did not induce any change to its abundance, expression or stability in the embryo.

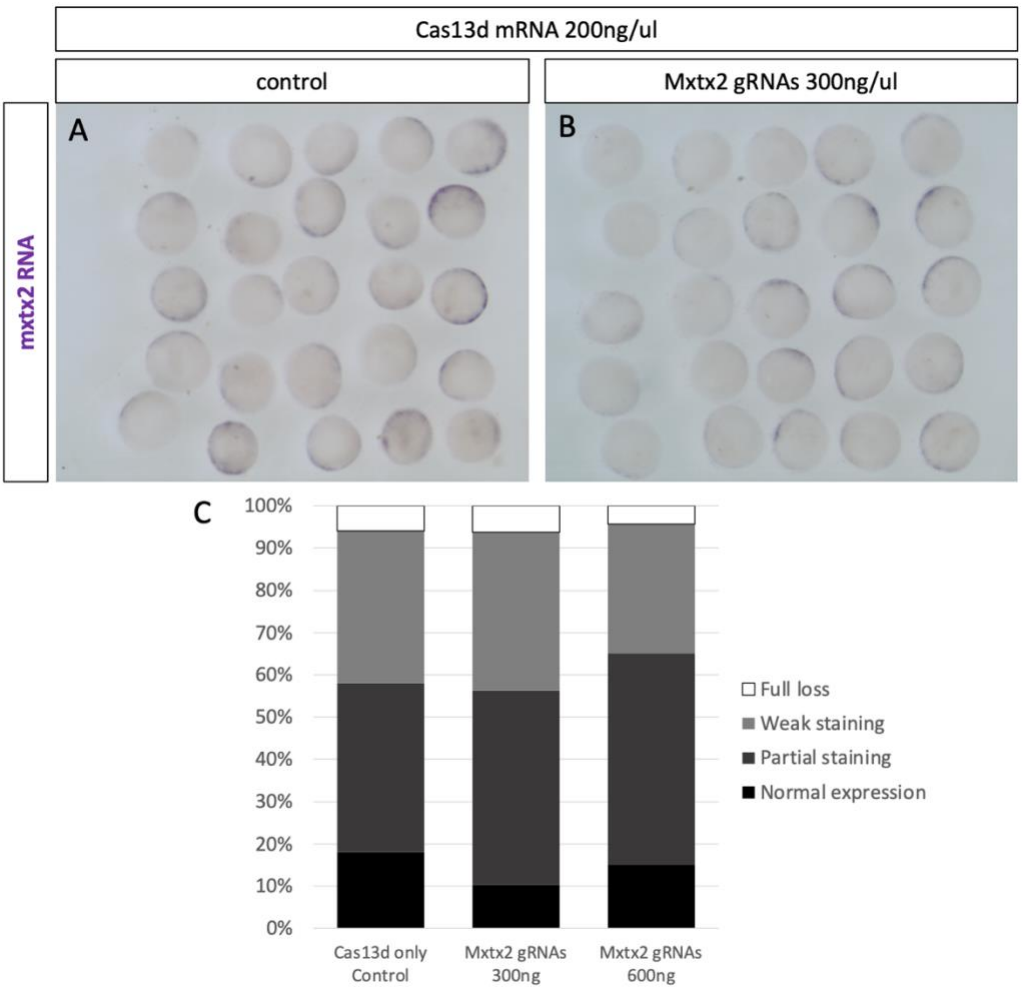


Fig. 3.20 Cas13d targeting of *mxtx2* mRNA did not cause a change in *mxtx2* expression.
(A-B) Light microscopy images showing colorimetric visualisation of *mxtx2* mRNA in groups of embryos injected with the indicated solutions at 1-cell stage. Embryos for the whole mount *in situ* hybridisation experiment were fixed at 30% epiboly stage. (C) Bar chart showing the fraction of embryos with each of the scored patterns of expression of *mxtx2* (normal, weak, and partial staining, and no detected expression) as a percentage of all assessed embryos. N=1; Cas13d mRNA control – n=50; *mxtx2* gRNAs 300ng/µl – n=45; *mxtx2* gRNAs 600 ng/µl – n=42.

3.5 Discussion

The primary aim set out for this chapter was to establish a set of tools for efficiently disrupting *mxtx2* in the embryo. Initial designs and optimisation of the targeting of *mxtx2* required a significant amount of trial and error, each round of experiments informing the next, leading to gradual improvements, but also inconsistencies between approaches. This can result in a lack of systematic datasets. First designed guide RNAs targeting the *mxtx2* 5' promoter region and 3' region were not efficient. This was likely due to a combination of reasons. Firstly, I have restricted the design to targeting the promoter and flanking the gene, therefore significantly restricting the base of potential guide RNAs. Secondly, human errors and injection efficiency may have influenced the obtained results. Thirdly, the protocol for CRISPR-Cas9 disruption of genes was still being established in the laboratory group, with several optimisations. Finally, some *in vitro* transcribed guide RNAs have been found to lack the mutagenic activity at particular sites and unable to efficiently mutagenize their targets, with significantly reduced endonuclease activity of Cas9 (Burger *et al.*, 2016; Gagnon *et al.*, 2014; Thyme *et al.*, 2016).

The improvement of transgenesis efficiency with use of synthetic guides – duplexes of crRNA and tracrRNA – was an important change, aiding the attempts of mosaic knockout of *mxtx2* in embryos. Previous studies in mouse and zebrafish embryos have shown significantly higher mutagenicity of the dgRNAs (duplex guide RNAs) when compared to *in vitro* transcribed single guide RNAs (Hoshijima *et al.*, 2019; Jacobi *et al.*, 2017). This is most likely due to modifications to the RNA backbone improving chemical stability of the guide RNAs, as well as standardised production

leading to reduction of impurities that could hinder formation of guide RNA-Cas9 ribonucleoprotein complex, however this hypothesis has been contested by (Hoshijima *et al.*, 2019). Additionally, they show that single guide RNAs suffered from a reduced efficiency most likely due to supernumerary 5' guanine nucleotides, added by the activity of T7 promoter used for their transcription, and lack of those in dgRNAs contributes to their increased efficiency In this study, sgRNAs were indeed transcribed from a T7 promoter, indicating it as one of possible reasons of low mutagenic efficiency.

Use of WISH and qPCR to assess mxtx2 expression has showed that guides targeting the promoter region of *mxtx2* were able to reduce its expression, but that was not observable by mRNA visualisation. That approach of *mxtx2* knockout also did not lead to a visible phenotype reported in the studies using a morpholino knockdown (Wilkins *et al.*, 2008). Guides targeting the promoter region of mxtx2 were not able to efficiently disrupt its expression, despite use of several combinations.

Using guide RNAs targeted to the functional domain of *mxtx2*, I was able to mosaically knock out this gene in embryos. Efficiency of the gene disruption was supported by WISH and a penetrant phenotype of arrest at the end of epiboly and high levels of embryonic lethality, with both correlating with previously reported roles of *mxtx2* in epiboly. However, quantitative measurement of *mxtx2* expression showed very limited change in embryos where the mosaic knockout was induced. This is a confusing result, however there is a possible explanation. Embryos injected with Cas9 and *mxtx2* guide RNAs exhibit a developmental delay, particularly at the onset of epiboly (own observations, data not shown). In qPCR assays presented in this chapter, I have collected embryos based on the morphology staging. Therefore,

embryos injected with *mxtx2* guides were collected later than the associated controls. While both groups were at the same stage morphologically, it is not impossible that due to this temporal difference, *mxtx2* was expressed to a higher degree in targeted embryos than in controls, creating a significant amount of noise in qPCR data. I have not, however, gathered any evidence that would support this speculation.

An alternative explanation could be that the very limited disruption of *mxtx2* expression by the CS2 and CS4 guide RNAs was true, and even such a small degree of disruption is able to cause a significant phenotype. This, however, does not correlate with WISH assay showing almost complete loss of *mxtx2* expression at 30% epiboly.

Efficiency of Cas9 disruption reported here was not as high as that reported by (Wu *et al.*, 2018) or (Hoshijima *et al.*, 2019), where mutagenesis efficiencies and subsequent reduction in gene transcription ranged from 80 to 99%. This was despite "CS" guides all being used as more efficient dgRNAs with Cas9. Nevertheless, mosaic loss of *mxtx2* as observed by whole mount *in situ* hybridisation was high enough to justify the use of the mxtx2 CS4 guide RNA as a tool for F0 mosaic knockouts of *mxtx2* in further study. Additionally, lesions created by Cas9-gRNA complexes might not have always caused a truncation or nonsense mutation that led to an impact on *mxtx2* mRNA stability. These cases would be detected using WISH, as the lesion could affect probe binding to a sufficient degree, but would not be detected using qPCR, as the primer pairs were not designed to flank the guide RNA target sites.

Adaptation of Cas13d to target mxtx2 was unsuccessful, as I was not able to replicate the results achieved by Kushawah *et al.* despite using the same reagents and following the protocol that this group have established (Hernandez-Huertas *et al.*,

2022; Kushawah *et al.*, 2020). The non-specific phenotype of Cas13d had a striking resemblance to the phenotype shown by use of double stranded RNAs in RNA interference (Oates *et al.*, 2000; Zhao *et al.*, 2001). This suggests that the observed non-specific phenotype could be caused by formation of double-stranded RNA in the embryos causing activation of the antiviral immune responses in the embryo, leading to subsequent cell apoptosis. This effect was likely mediated by Cas13d, as injection of guide RNAs without the protein did not cause the phenotype. Therefore, in this study I have shown a potential unspecific effect of Cas13d nuclease in zebrafish embryos, leading to developmental defects, increased apoptosis, abnormal axis formation, cyclopia and embryo death. While this should not discourage the attempts to establish Cas13d as a method of RNA disruption in zebrafish embryos, those unspecific effects should be taken into account in further research and help inform the potential controls. Potentially, an alternative Cas13 protein, one not inducing a response form the embryo, could be used with the efficiency reported by (Kushawah *et al.*, 2020), while new studies advance the design of highly efficient guide RNAs and develop bioinformatics tools for guide RNA efficiency prediction (Wessels *et al.*, 2020).

In conclusion, while the goal of establishing of a genomic toolset for *mxtx2* disruption was not achieved, in this study I have successfully generated an efficient method of generating zebrafish *mxtx2* mosaic crispants using Cas9 and guide RNA targeted to the homeobox domain of the gene. This could be a potentially useful tool in further studies of *mxtx2* function and serve as a basis for generation of a mutant line. However, due to the remarkable efficiency of depletion of *mxtx2* mRNA as

established by WISH, this guide RNA could also be used for exploration of Mxtx2 function in mosaic, F0 embryos.

With higher efficiency tools, several targeted mutations of *mxtx2* could be developed based on targeting particular sites (such as the homeobox domain), targeted deletions of exons or UTRs or deletion of the whole gene,

Additionally, in this chapter I have shown that Cas13d can cause unspecific developmental abnormalities in zebrafish embryos, an effect present in both mRNA and purified protein injections of this nuclease, and this effect was highly dependent on the co-injection of Cas13d nuclease with guide RNAs. The unspecific phenotype displayed remarkable similarity to that of dsRNA injection into zebrafish embryos. This study failed to replicate results of Kushawah *et al.* and these negative results could be highly informative for other researchers aiming to use Cas13d to disrupt mRNAs in zebrafish embryos. This also poses a question whether Cas13d should be used as a method of gene knockdown in zebrafish, and if so, strict controls ensuring the specificity of phenotypes should be established as standard, similar to morpholino oligonucleotides (Stainier *et al.*, 2017).

Chapter 4: Characterisation of Developmental Defects in Zebrafish Embryos Lacking Mxtx2 Transcription Factor

4.1 Introduction

The preliminary work on transgenesis tools yielded an efficient way of targeting *mxtx2* using CRISPR-Cas9. There was therefore a possibility of expanding on the knowledge already gathered about this transcription factor. Results of RNA sequencing that place *mxtx2* as one of the highest expressed genes at 256 and 512 cell stages in the zebrafish embryo (Heyn *et al.*, 2014) have raised a number of important questions. The main one is why is this gene, with a clearly specified function later in development, expressed so early? What is the need for *mxtx2* to be present in early wave of transcription, instead of during the main wave of genome activation? Is *mxtx2* involved in main-wave ZGA, bringing its function closer to its mammalian relatives?

Most previous studies focused heavily on *mxtx2* function post-ZGA – in epiboly. A large amount of information about its function has already been discovered through use of the morpholino oligonucleotide knockdown. Embryos lacking *mxtx2* fail to progress through epiboly, displaying a yolk burst phenotype – this has been associated with a disorganisation of cells in the yolk syncytial layer, as well as loss of the F-actin contractile ring present in E-YSL (Wilkins *et al.*, 2008). Later, it was found

that expression of *mxtx2* is required for YSL-specific expression of several genes, including a nodal effector *ndr2* and a transcription factor *sox32*. *mxtx2* morphants fail to induce mesoderm and endoderm, mediated by the aforementioned *ndr2* and *sox32* (Hong *et al.*, 2011; Xu *et al.*, 2012).

Perhaps the most informative was the *mxtx2* overexpression ChIP-seq in dome-stage embryos, showing that *mxtx2* has a broad transcriptional activation profile, binding over 1700 gene loci, and approximately 44% of all YSL specific genes (Xu *et al.*, 2012). Expression of mxtx2 has since been placed as a direct target downstream of Nanog pioneer factor (Hong *et al.*, 2011; Xu *et al.*, 2012), and its expression in YSL was dependent also on translation of maternal *eomesa* (Xu *et al.*, 2014). *Mxtx2* was thus shown to be a member of Nodal signalling pathway – downstream of Nanog and Eomesa, but upstream of Nodal effectors. To further confirm this, it was shown that expression of *mxtx2* was lost in *Nanog* mutants, however ectopic expression of *mxtx2* was able to rescue expression of YSL-specific genes (Gagnon *et al.*, 2018; Veil *et al.*, 2018; Xu *et al.*, 2012), and *mxtx2* has been placed by gene regulatory network analysis as one of the key elements in establishing mesendodermal fate in blastomeres (Nelson *et al.*, 2017).

However, whether or not *mxtx2* had an effect on ZGA has not been studied. ChIP-seq analysis has been done in dome stage embryos, after the main wave of ZGA and at the onset of epiboly, and role of *mxtx2* in epiboly has been elucidated. However, as the RNAseq of nascent transcripts (Heyn *et al.*, 2014) and whole mount in situ hybridisation (Fig. 4.1) show, mxtx2 mRNA can be detected in embryos very early. While WISH did not detect *mxtx2* transcripts in 256-cell (in contrast to the undeniably more sensitive RNAseq), they were present in a 1k-cell stage embryo (at the onset of

ZGA) and were distributed ubiquitously across blastomeres. Spatial distribution of *mxtx2* expression then became more restricted, localising to the blastoderm margin and YSL, with limited expression elsewhere. At dome stage and 30% epiboly, expression becomes restricted to the E-YSL ring on the embryo margin, with few marginal blastomeres also expressing the gene. After that, expression of *mxtx2* is not maintained throughout epiboly, with no expression in 75% epiboly embryos (8hpf) (Fig. 4.1.)

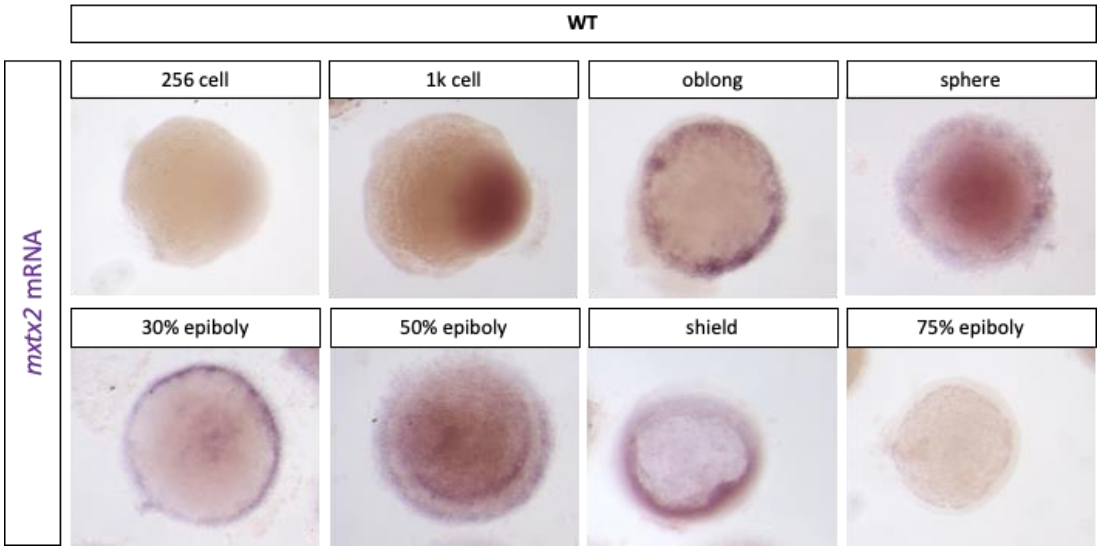


Fig. 4.1. Expression of *mxtx2* in wild-type embryos across early development stages.
A panel of light microscope images showing representative embryos stained using WISH for *mxtx2* mRNA at indicated stages. Embryos presented in lateral view, except 256- and 1k-cell stages, where embryos imaged in side-view, animal pole facing to the left.

To characterise the role of *mxtx2* in ZGA, it would be a necessary step to ensure that its expression is reliably disrupted. While the generation of efficient guide RNAs for use in CRISPR-Cas9 was helpful, the unspecific phenotype caused by Cas13d posed a big issue. As one of the objectives was to produce descriptive information about *mxtx2* without the use of contested morpholinos, Cas13d would have posed a perfect independent method of producing *mxtx2* knockdowns, and a fairly equivalent tool to MO-dependent RNA interference.

Without it, there was a need for a more penetrant method of gene disruption. That was because perhaps the best method of generating exploratory datasets that could inform us of the early function of mxtx2 would be by transcriptomic analysis (e.g. RNAseq). In order for the next-generation sequencing to produce reliable and, most importantly, significant data, differences between treatment datasets should be maximised (Conesa *et al.*, 2016). Induction of mosaic mutations may have shown a remarkable depletion of *mxtx2* mRNA by WISH, but qPCR quantification shown a more moderate depletion. Sequencing RNA from a mosaic mutant would thus have limited power in answering the posed biological questions, mainly if *mxtx2* affects expression of ZGA genes, and if so, what its targets are.

Therefore, I aimed to use the existing gRNA tool that I have obtained to generate *mxtx2* null mutant fish. A standard approach would utilise the F0 mosaic fish grown to adulthood, and screening their offspring embryos for germline transmission of the mutation. If that is the case, F0 founder fish (those showing germline transmission of the mutant) would be outcrossed with wild-type fish to generate F1 heterozygotes. F1 heterozygous adults, when in-crossed, should produce F2 offspring with Mendelian proportion of mutation inheritance, with 25% of embryos having a biallelic mutation in the gene of interest (Li *et al.*, 2016).

However, it was unlikely for this standard approach to be effective. Screening for F0 founders can be done in two major ways – either by genotyping, or by phenotyping the embryos (Kosuta *et al.*, 2018; Li *et al.*, 2016). Phenotyping raised concerns due to the already severe embryonic phenotype of *mxtx2* morphants, which raised the question if *mxtx2* mutations could be embryonic lethal, making it unable to raise mutant embryos to adulthood. Mutations in upstream *Nanog* showed viability and

fertility in heterozygotes, suggesting that similar case could be made for *mxtx2*. F1 embryos could be genotyped by PCR amplification of the mutated site, however this process would require multiple days of isolation of crossed F0 pairs (to ensure that germline-transmission-positive ones can be isolated from negative). This strategy was made impossible by severe restrictions of occupancy and access at the animal facilities, imposed due to the COVID-19 pandemic.

Therefore, a different strategy of knockout generation would have to be employed.

4.1.1 Knock-in-knockout methods

One perfect possibility was by using a reporter knock-in that would disrupt the *mxtx2* gene. That would greatly reduce the time required for embryo screening, as a fluorescent reporter could be injected, allowing easy selection of both F1 and F2 embryos containing the mutation.

Precise insertions of exogenous sequences into cleavage sites produced by site-specific nucleases (knock-ins) have been commonly used in multiple model organisms and utilised for multiple purposes, including gain-of-function mutations and so-called “knock-in knockouts” of genes (Hoshijima *et al.*, 2016; Li *et al.*, 2019; Peng *et al.*, 2014; Prykhozhij and Berman, 2018; Ranawakage *et al.*, 2021). Knock-ins are based on the principle of utilising homology-directed repair (HDR) of DNA at the site of induced double strand break, by providing a repair template that contains homology arms flanking the site targeted by the nuclease. The insertion can be a short DNA fragment or a whole gene, with sequences coding for fluorescent proteins frequently utilised as inserts, as they offer a quick and visible readout of efficient insertion.

One of such methods that have been established in zebrafish is GeneWeld (Wierson *et al.*, 2020), which has been selected here due to its reported efficiency, establishment in zebrafish embryos, readily available reagents, and relative ease of use. Contrary to typical knock-in approaches, GeneWeld is based on the microhomology-mediated end-joining (MMEJ) pathway of DNA repair, which requires much shorter homology arms provided by a repair template. That significantly reduces the complexity of generating the insertion vector. The system contains several components (Fig. 4.2). Cas9 nuclease and guide RNA targeting the gene of interest need to be provided.

Additionally, embryos need to be injected with an insert donor plasmid that has been modified to contain short (24-48bp) homology arms flanking the site targeted by the guide RNA, as well as a cassette containing in order: a 2A peptide, the "Cargo" or the sequence to be inserted, and a poly-A signal. The cassette and homology arms are flanked by the target sites to the Universal guide RNA (UgRNA), a highly efficient guide RNA with no predicted target sites present in the zebrafish, pig or human genomes. UgRNA utilises Cas9 to release the insert construct, which then can be integrated to the gene of interest with aid of the short homology arms.

The group reported up to 70% of F0 embryos that were injected with the system components contained cells positive for the insertion, and high probabilities of germline transmission. Additionally, a software package aiding the design of homology arms to ensure in-frame insertions (gTagHD, (Wierson *et al.*, 2020)), as well as a range of easy-to-use vectors with a choice of reporter genes was provided. Therefore, it was decided that this method will be used as means of generating a *mxtx2* null mutant, by insertion of a EGFP cassette into the site targeted by the

verified guide RNA. Upon insertion, only a short truncated fragment of *mxtx2* could be produced that would be unlikely to be functional due to guideRNA targeting to the functional domain. Poly-A signal located towards the 3'-end of the sequence would prevent transcription past the insertion site. The truncated *mxtx2* fragment which would be followed by a 2A-peptide and the fluorescent EGFP reporter, leading to both being translated separately, and EGFP expression would be solely dependent on the activity of *mxtx2* promoter, as no promoter sequence is present in the insert cassette.

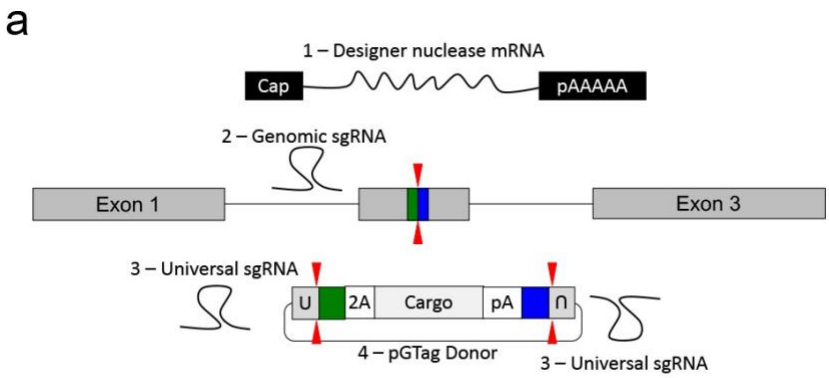


Fig. 4.2 GeneWeld knock-in reagent components.
(1) A designer nuclease, namely a TALEN or Cas9, can be provided as mRNA (showed) and also as a protein. (2) Genomic sgRNA – a guide RNA targeting the nuclease to a sequence in the gene of interest, marked with red arrowheads on the schematic gene sequence. (3) Universal sgRNA (UgRNA) target sites (red arrowheads) on the insert donor vector (4). The insert donor vector contains sequences flanking the insertion cassette, with short homology arms to the gene target (green and blue). Image adapted from (Wierson *et al.*, 2020).

4.1.2 Aims of this chapter

Having successfully generated reliable mosaic knockout of *mxtx2* in F0 embryos using Cas9 targeted to the *mxtx2* functional domain, I was able to start asking questions about the role and function of *mxtx2* in the zebrafish development.

In this chapter, I aimed to characterise the observed phenotype of the *mxtx2* mosaic mutant, as well as use the Cas9 targeting in combination with the GeneWeld knock-in method described above to generate a stable line of *mxtx2* EGFP-knock-in mutant, which would be used in further studies, establishing a phenotype of a full *mxtx2* knockout (and not morphant), as well as aiding further characterisation of the *mxtx2* function in ZGA through use of transcriptomic approaches (RNAseq), use of which would not be feasible without a complete knockout of the gene. Failure to generate an *mxtx2* mutant line would significantly hinder the application and impact of the research so far, and make it impossible to answer the biological question of the role of *mxtx2* in ZGA.

4.2 Results

4.2.1 Mxtx2 mosaic knockout leads to epiboly defects

Initially, I have set out to characterise the mosaic knockout mutant of *mxtx2*. Embryos targeted by the Cas9 and gRNAs targeting the homeodomain have shown a reduction in expression, as well as developmental delays and embryo lethality, but it was still uncertain if this is due to the epiboly failure phenotype previously described in *mxtx2* morphants (Wilkins *et al.*, 2008; Xu *et al.*, 2012), or a different event. For that, I observed embryos at 9 hours post fertilisation. Targeting the *mxtx2* promoter by Cas9 caused developmental defects in embryos, and a significant proportion of embryos showed a yolk burst phenotype (Fig. 4.3 A-B). 49/282 (17.4%) of all scored embryos that were injected with Cas9 and the CS4 gRNA targeting the functional domain of *mxtx2* showed this terminal epiboly defect. In addition, many of the embryos observed at 9hpf were not bursting, but were observed showing a phenotype termed as “moderate epiboly defects” (33.0% - 93/282 scored embryos). These embryos show an irregular structure of blastoderm and furrowing of the embryo surface, they fail to progress in epiboly with the YSL and margin still positioned at approximately 30% of the yolk sphere, and show very abnormal structure of the animal cap. Thanks to the transparency of the embryo, it can be seen that deep cells appear to have migrated into the space beneath the epiblast, indicating either a defect with yolk doming or that the lack of migration of YSL and blastoderm margin leads to an abnormal pattern of deep cell migration in embryos (Fig. 4.3 A).

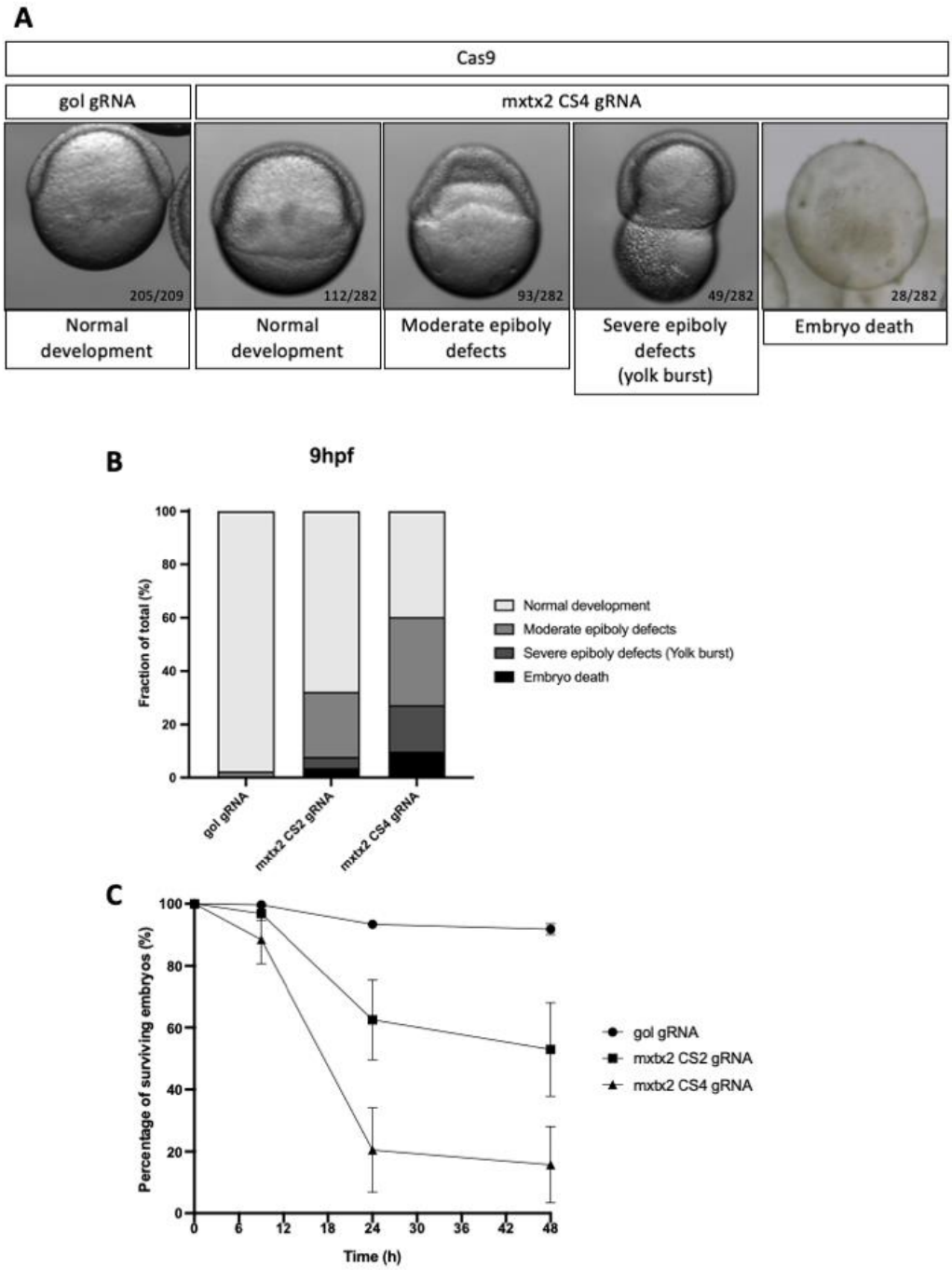


Fig. 4.3 Developmental defects and yolk burst phenotype in *mxtx2* mosaic mutant embryos.
(A) Brightfield microscopy images showing representative embryos displaying phenotypes of (left to right): normal development (in control *gol* mosaic mutants and *mxtx2* mosaic mutants), moderate epiboly development, severe epiboly development and yolk burst. Numbers in the bottom right indicate the proportion of each phenotype as part of all embryos scored, for each treatment group. (B) Bar chart showing the percentages of each phenotype scored at 9hpf as fraction of all scored embryos for each of the treatment groups – embryos injected at one cell stage with Cas9 and the indicated gRNAs. *gol* n=209; *mxtx2* CS2 n=220; *mxtx2* CS4 n=242. (C) Percentages of survival of each of the treatment groups scored for phenotypes in B, with survival measured at 9hpf, 24hpf and 48hpf. Data presented as mean of 3 biological replicates ± standard deviation.

Interestingly, embryos with the yolk burst phenotype the epiblast appears normal and with less furrowing, similar to that of embryos developing normally, but with a visible contraction of the E-YSL/margin ring.

These phenotypes can be attributed the *mxtx2* targeting, as control embryos also injected with Cas9, but targeting *gol* did not exhibit the yolk burst phenotype, while injections of a less efficient guide RNA targeting *mxtx2* – CS2 gRNA – showed a reduced penetrance of the same phenotype, with a small population of embryos exhibiting yolk burst (Fig. 4.3B). Additionally, this gRNA showed reduced mortality compared to the CS4 gRNA (Fig. 4.3C), giving a strong indication of correlation between guide RNA efficiency and severity of *mxtx2* knockout phenotype.

Additionally, I conducted time-lapse imaging on a brightfield microscope, imaging embryos from 3 hours post fertilisation (~1k cell stage), where no indication of a phenotype is present, over the epiboly period up to 14 hpf. Movie S6 (Appendix) shows normal development of embryos injected at 1-cell stage with Cas9 targeting *gol* from 1k cell stage to the bud stage, with no visible defects in epiboly or development. Movie S7 (Appendix) shows the development of embryos injected with Cas9 targeted to *mxtx2* by CS4 gRNA in the analogical period. It shows that *mxtx2* mosaic mutants exhibit a developmental delay, only initiating epiboly and reaching dome stage around 6.5hpf (compared to 5hpf in control embryos). The time-lapse also shows that the "moderate epiboly defects" phenotype, earlier seen in 9hpf embryos, later resulted in embryonic lethality by yolk burst. Interestingly, all of the *mxtx2* mosaic mutant embryos (15/15), even ones appearing normal at 9hpf, exhibited the yolk burst phenotype and died between 9hpf and 14hpf. It can be an indication that embryonic lethality is induced in embryos with varying degrees of

mxtx2 mosaicism, with loss of expression in a small subset of cells leading to the phenotype developing. This can be compared to 0/15 embryos dying in the control *gol* mosaic mutant group.

In order to verify whether in addition to phenocopying the *mxtx2* morphant, mosaic mutants are also able to show an effect on *Nodal* signalling, I have conducted a qPCR experiment assessing the expression of *ndr2*, the zygotically expressed *Nodal* effector in zebrafish, in 30% epiboly embryos (Fig.4.4). *mxtx2* mosaic mutants induced by gRNAs CS2 (moderate penetrance) and CS4 (high penetrance) showed a large decrease in *ndr2* expression, with up to 70% loss of expression compared to *gol*-targeting controls. *ndr1* would be an interesting additional target to investigate changes of expression in response to *mxtx2* mosaic knockout, however due to time constraints I made a decision to omit it, as it was unlikely that Mxtx2 would have a vast impact on expression of a maternally-provided factor, and maternal contribution of *ndr1* could increase noise and require additional controls in the qPCR experiment. Additionally, I have tested whether *mxtx2* mosaic mutants can influence expression of other early wave genes, choosing a zinc-finger transcription factor *klf17*, a zebrafish orthologue of mammalian KLF4 stem cell pluripotency factor. Klf17 has been shown to be expressed as early as 512 cell stage, and plays a role in primitive erythropoiesis and development of an embryo hatching gland (Hadzhiev *et al.*, 2021; Suzuki *et al.*, 2019). However, *mxtx2* mosaic mutations have shown no effect on *klf17* expression as visualised by WISH (Fig. 4.5). *klf17* expression was not likely to be affected by knockout of *mxtx2*, and was used here to confirm whether any effect that *mxtx2* mosaic mutation had on transcription was global, rather than specific to Mxtx2 targets. Ideally, qPCR and WISH assays for both *ndr2* and *klf17* would be completed,

however time constraints of the project made it impossible to do so. I have decided that *ndr2*

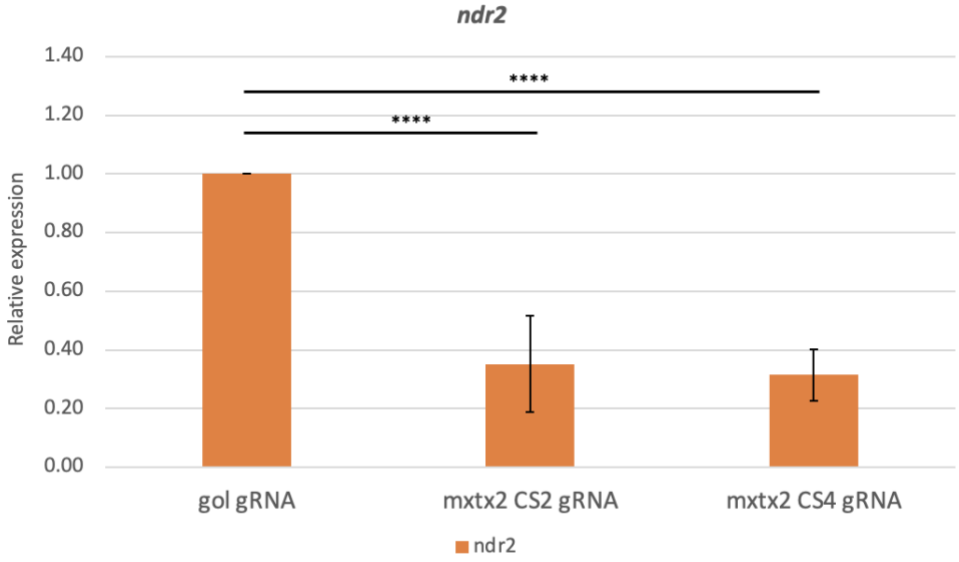


Fig. 4.4 Assessment of ndr2 expression in mxtx2 mosaic mutants by qPCR.
A bar chart showing relative abundance of *ndr2* mRNA as measured using qPCR in the indicated treatment groups at 30% epiboly stage. *ndr2* expression was normalised to that of the *eef1a1a* housekeeping gene, and it is presented relative to the expression in the control embryos (*gol* gRNA), hence the 1.00 relative fold change in expression. Data is presented as mean relative fold change, ±SD, from N=4 independent biological replicates. Statistical significance was tested using multiple one-way ANOVA tests with Dunnet's multiple comparisons test (ns $p>0.05$, *$p<0.05$, **$p<0.01$, ***$p<0.001$, ****$p<0.0001$).

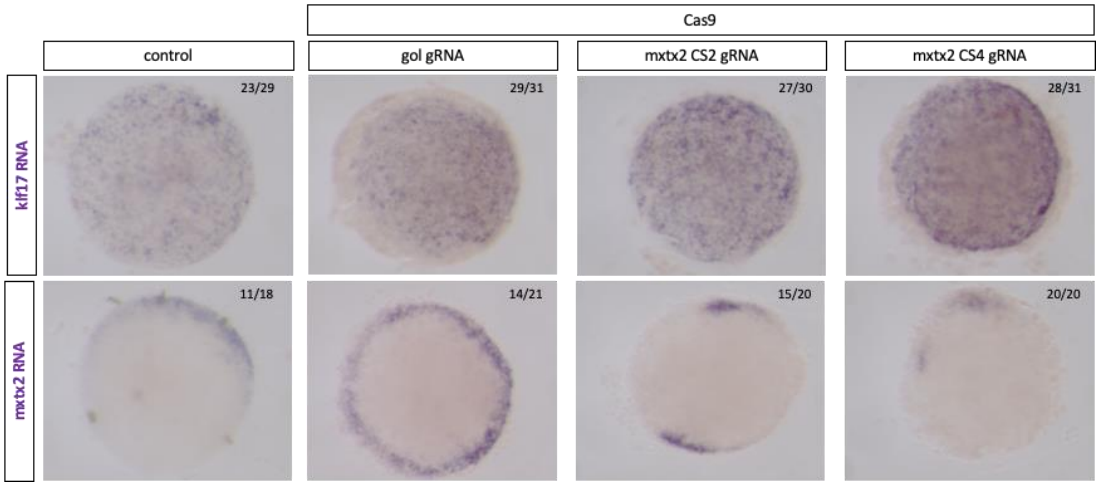


Fig. 4.5 Assessment of *klf17* expression in *mxtx2* mosaic mutant embryos by WISH
A panel of images showing visualised mRNA of *klf17* (top) and *mxtx2* (bottom) as detected by WISH in indicated treatment groups (injected into embryos at 1-cell stage) in 30% epiboly embryos. Representative embryos shown. Control - wild-type, uninjected embryos. Numbers in the top right corner indicate the proportion of embryos with the represented phenotype as part of all assessed embryos.

With the mxtx2 mosaic mutant confirmed to present the same phenotype as the previously studied morphants, I have proceeded to generation of the *mxtx2* null mutant for further use in exploration of its role in ZGA, as little new information beyond replicating and confirming the morphant effects could be discovered by focusing on its yolk syncytial layer-dependent phenotype, interactions with Nanog and Nodal signalling.

4.2.2 GeneWeld knock-in knockout approach

As described above, the GeneWeld method is an interesting tool for generating targeted gene knock-ins, and one that offers a simple and efficient protocol for an in-frame insertion of fluorescent reporters into the gene of interest. Additionally, a protocol for its use in zebrafish has been established by (Wierson *et al.*, 2020), therefore the method was simple to replicate.

Firstly, I selected the site on the *mxtx2* gene to which I would target the insertion. This was chosen to be the CS4 gRNA target site, as it not only offered the highest efficiency in the Cas9 complex, but also targeted the *mxtx2* homeodomain, ensuring that the insertion of the fluorescent reporter into the site would guarantee a functional disruption and loss of function. Following that, I followed the established protocol (Welker *et al.*, 2021; Wierson *et al.*, 2020) and used the associated gTagHD software to design oligos containing the 48bp homology sites to the *mxtx2* gene flanking the CS4 gRNA cutting site (Fig. 4.6 B). The associated software designs the 48bp homology sequences with additional base pairs allowing in-frame insertion at the target site, and adds overhangs specific to the specified insertion vector, to allow for easy ligation at specified restriction enzyme cutting sites where homology arms are inserted. I have selected a vector containing an insertion cassette with EGFP (pGTAG-EGFP-CAAX-SV40 – Fig. S8A), as the aim was to generate a null mutant with an easily distinguishable knock-in reporter. Fig 4.6A shows a simplified diagram of the GeneWeld approach, and the structure of the insertion plasmid and cassette.

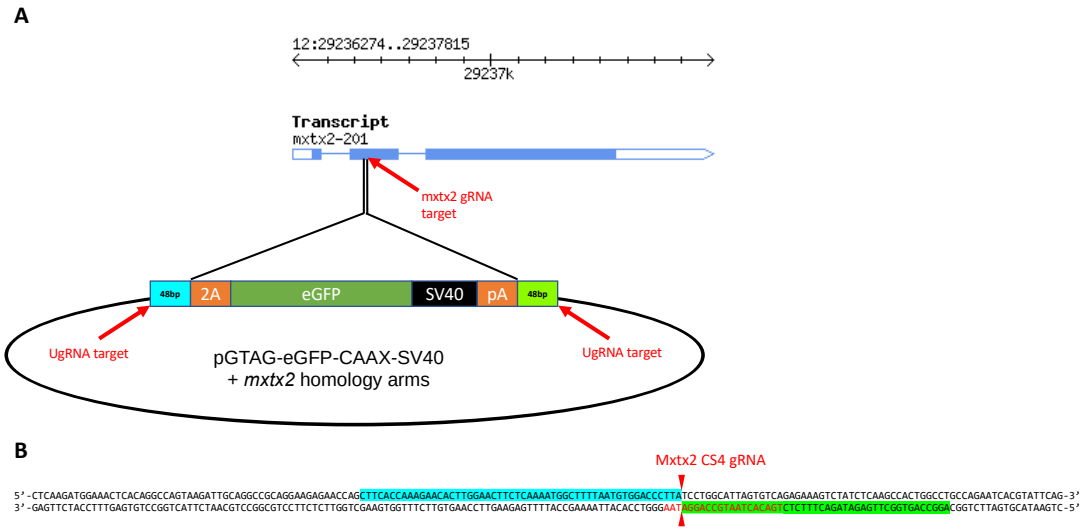


Fig. 4.6 A schematic diagram depicting the strategy of GeneWeld knock-in into the mxtx2 gene at the cutting site of the CS4 gRNA (red arrow). A simplified schematic of the plasmid and its insertion cassette (containing 2A peptide, coding sequence of EGFP gene, SV40 poly-A signal) and the 5' and 3' 48 bp homology sequences, marked with blue and green respectively. Red arrows point to the UgRNA target sequences, positioned to immediately flank the inserted homology arms. (B) fragment of the *mxtx2* DNA sequence with marked CS4 gRNA cut site (in red) and highlighted 5' and 3' homology sequences.

Previously designed 48 bp homology arm oligos for the 5' and 3' flanking sequence of *mxtx2* CS4 gRNA cutting site were PCR annealed to produce dsDNA fragments that could be cloned into the pGTAG plasmid. 2 rounds of restriction digest, ligation using T4 ligase, transformation into competent bacteria, colony plating, growth of plasmid-containing bacterial culture and MidiPrep plasmid isolation were done, one for each homology arm. PCR amplification and Sanger sequencing was used to confirm the presence of homology arms in the vector. All of those steps were conducted in accordance with the protocol established by (Welker *et al.*, 2021). The resulting plasmid (Fig.S8 B) (further referred to as GeneWeld insert plasmid) contained the insert EGFP cassette, flanked with 48bp *mxtx2* homology arms, flanked with UgRNA target sites that will be used by Cas9 nuclease (alongside the Universal guide RNA) to release the insertion cassette. Injection of all of those components – Cas9, *mxtx2*

CS4 gRNA, UgRNA and the GeneWeld insert plasmid – was required for Cas9-mediated knock-in.

4.2.3 EGFP knock-in into the *mxtx2* locus in F0 embryos using GeneWeld

Embryos were injected at the 1 cell stage with injection mixes containing the GeneWeld components. With the primary objective of establishing a mutant line, F0 embryos needed to be verified for mosaic incorporation of the insert. Two Cas9 nucleases were used – first was *S. pyogenes* EnGen Cas9-NLS protein which was used in all previous Cas9 experiments. Second was mRNA of *S. pyogenes* Cas9 fused with a 3' UTR of the *nanos* gene (Cas9-nos mRNA), expression plasmid for which (pCS2-nCas9n-nanos 3'UTR) was kindly donated by Dr Andreas Zaucker. A 3'UTR nanos sequence contains a primordial germ cell localisation signal that would direct the mRNA to the cells with germ cell fate to increase local translation, therefore increasing the probability of knock-in mutation occurring in those cells, therefore increasing the probability of germ-line transmission of mutations from F0 fish to their F1 offspring. This method of localising the Cas9-mediated mutations to the germ line has been previously reported in zebrafish by (Moreno-Mateos *et al.*, 2015). I decided to attempt to use this strategy alongside the "classical" Cas9 protein, as the increased probability of the knock-in occurring in germline could potentially save time and labour in screening of F1 embryos in search for F0 founder fish.

Embryos were injected at 1-cell stage with the components for knock-in strategy (Cas9 protein or Cas9-nos mRNA, GeneWeld insert plasmid with *mxtx2* homology arms, UgRNA and mxtx2 CS4 gRNA) and observed in early development for GFP fluorescence, which would indicate successful incorporation of the insert cassette

into the genome and its expression. In the unlikely event of an off-target incorporation, it would be unlikely that EGFP would be expressed, as the insertion cassette did not contain any promoter that would allow its transcription.

Fig. 4.7 shows the result of those observations in F0 embryos. Approximately 14% of embryos injected with Cas9 protein and the GeneWeld “cocktail” showed some kind of mosaic GFP expression (Fig.4.7 A,B,D), but only in approximately 1.2% of Cas9-nanos mRNA injected embryos (Fig.4.7 D). Controls, including injection of Cas9 with only gRNAs or with only the GeneWeld insert plasmid, showed no EGFP expression, indicating that GFP signal came from the incorporated knock-in, and not autofluorescence or the random incorporation of the plasmid. The presence of EGFP mRNA in embryos was confirmed by WISH at 30% epiboly (Fig.4.7 C). However, in Fig.4.7A it can be seen that the embryo showing the EGFP signal appears to be developmentally delayed and shows similar furrowing of the animal cap as *mxtx2* mosaic mutants. This was of course expected, as *mxtx2* was targeted, however could raise question about survival of *mxtx2* mosaic knock-in mutant embryos.

Nevertheless, 100 of the F0 embryos injected with Cas9 protein and the GeneWeld knock-in “cocktail” were grown to adulthood in order to screen them for germline transmission of the EGFP knock-in into *mxtx2*, with 54 fish surviving.

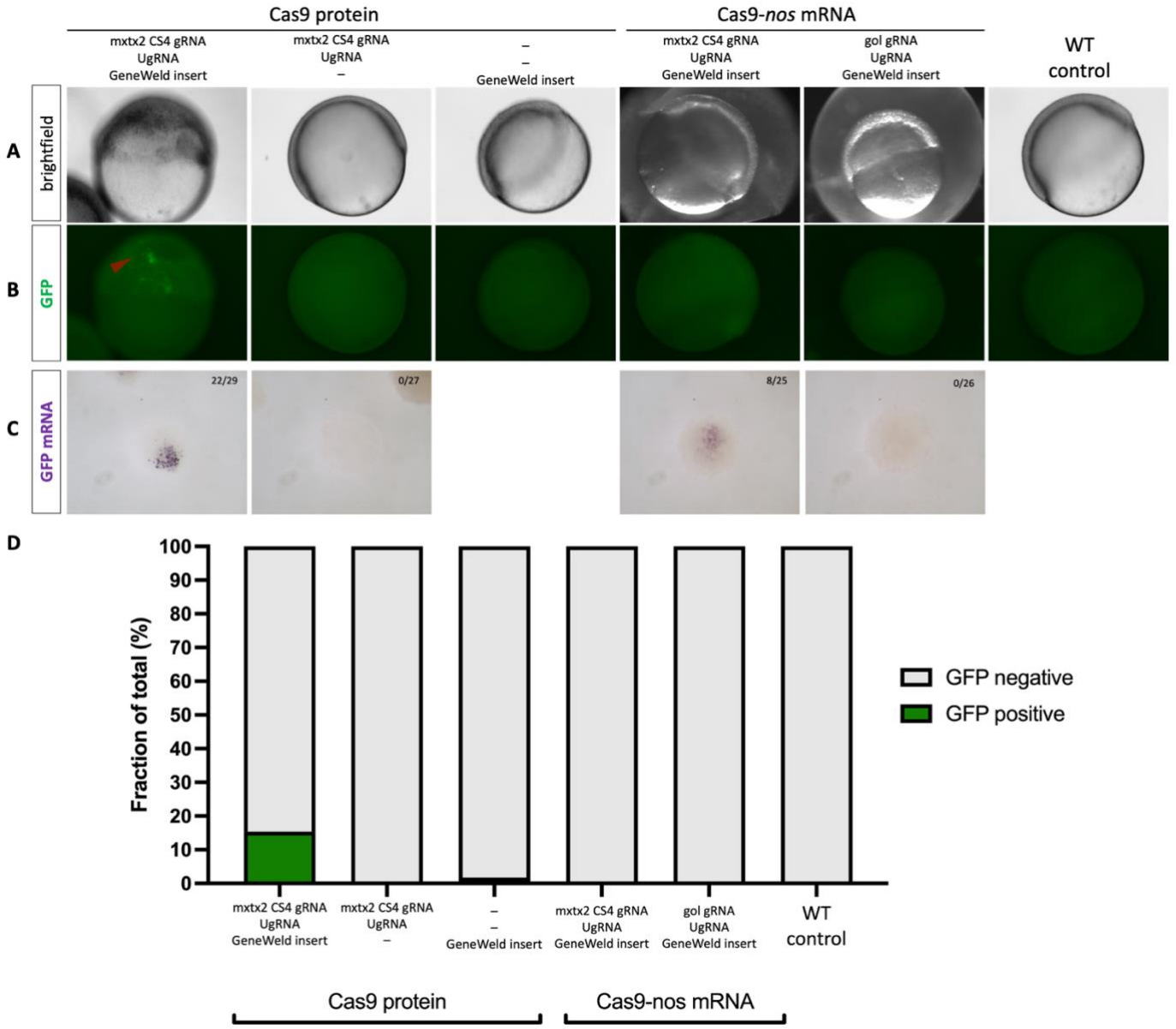

Cas9 protein
Cas9-nos mRNA
mxtx2 CS4 gRNA
UgRNA
GeneWeld insert
mxtx2 CS4 gRNA
UgRNA
–
–
–
GeneWeld insert
mxtx2 CS4 gRNA
UgRNA
GeneWeld insert
gol gRNA
UgRNA
GeneWeld insert
WT
control
A
brightfield
B
GFP
C
GFP mRNA
22/29
0/27
8/25
0/26
D
Fraction of total (%)
100
90
80
70
60
50
40
30
20
10
0
GFP negative
GFP positive
mxtx2 CS4 gRNA
UgRNA
GeneWeld insert
mxtx2 CS4 gRNA
UgRNA
–
–
–
GeneWeld insert
mxtx2 CS4 gRNA
UgRNA
GeneWeld insert
gol gRNA
UgRNA
GeneWeld insert
WT
control
Cas9 protein
Cas9-nos mRNA

Fig. 4.7 GeneWeld knock-in of EGFP into *mxtx2* gene using Cas9 and Cas9-nanos

(figure on previous page)

(A-B) A panel of brightfield (top) and green fluorescence (bottom) light microscopy images showing representative embryos of groups injected with the GeneWeld knock-in components targeted to *mxtx2* or indicated controls. Cas9-nos mRNA-mediated targeting was controlled by using a gRNA targeting another genomic site (*gol)*, replacing the *mxtx2* gRNA.

(C) A panel of images showing the visualisation of EGFP mRNA by WISH at 30% epiboly in indicated groups, with representative embryos shown alongside the proportion of embryos with that mRNA staining pattern from all assessed embryos.

(D) A bar chart showing the percentage of embryos with GFP expression (GFP positive) and without (GFP negative) from all assessed F0 embryos in groups presented in (A).

4.2.4 No germline transmission of mxtx2 EGFP knock-in mutation in F1 embryos

These 54 adult F0 fish were crossed and their offspring, F1 embryos were screened for presence of GFP fluorescence in early development (as EGFP expression would be driven by *mxtx2* promoter, and therefore would coincide with its expression). Adult F0 fish were each crossed at least two times to ensure that pairs producing little of no offspring in one round of screening, could lay fertilised embryos with another partner. 5677 embryos from 53 crossings were screened, with 0 GFP positive embryos. No F0 founder fish were identified (Fig 4.8, Table S9). The knock-in mutation was either not present in the germline of these fish and not transmitted to the offspring, or not present at all. With no F0 founder adults showing germline transmission of EGFP integration into the *mxtx2* gene, F1 generation heterozygotes could not be grown, and a transgenic line could not be established. In-cross of F1 heterozygotes would have yielded 25% of embryos homozygous for the EGFP integration disrupting the *mxtx2* locus, effectively being *mxtx2*$^{-/-}$ zebrafish. As it was unknown whether null mutation of *mxtx2* would be embryonic lethal in early development, those embryos could have been studied in transient, as most of the further scientific enquiries were for the function of *mxtx2* in the earliest transcription events in the embryo, and would not require the embryos surviving past epiboly. However, the mutant line was not established.

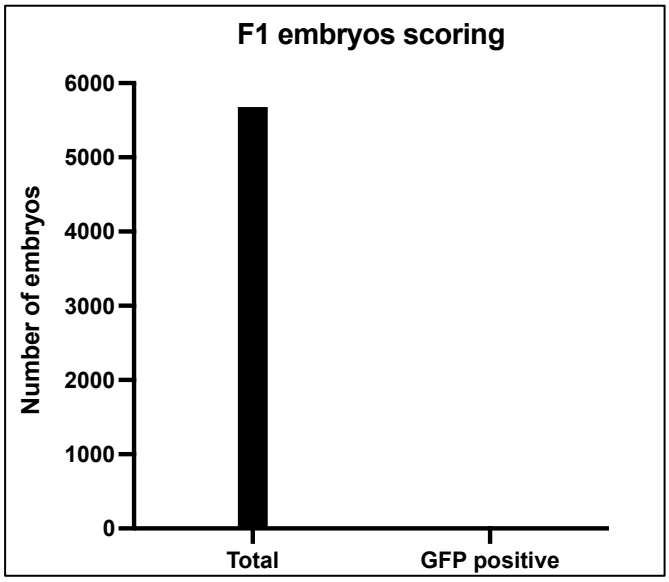


Fig. 4.8. Screening of F1 embryos, offspring of F0 *mxtx2*-EGFP knock-in mutant zebrafish. Embryos were assessed for presence of GFP positive cells at multiple stages of development between sphere and 50% epiboly stages. Bar chart indicates the total number of embryos assessed (N=5677), and the number of GFP positive embryos found in the screening (0/5677).

4.3 Discussion

In this chapter, I have shown that a mosaic knockout of *mxtx2* phenocopies the developmental phenotypes observed in morpholino oligonucleotide knockdown of the gene (Bruce *et al.*, 2005; Hong *et al.*, 2011; Wilkins *et al.*, 2008; Xu *et al.*, 2012). *mxtx2* mosaic knockout is sufficient to cause a penetrant yolk burst phenotype in embryos, with arrested epiboly and defects in deep cell migration towards the epiblast preceding the contraction of the YSL/blastoderm margin around the yolk, bursting it and subsequently causing embryo lysis. Additionally, high degree of embryonic lethality of the *mxtx2* mosaic mutant during the first 24 hours post fertilisation can be attributed to the penetrance of the yolk burst phenotype with embryos dying in developmental stages corresponding to late epiboly or early somitogenesis (9-13 hpf), as shown by time-lapse imaging. Time-lapse imaging also illustrated the process causing the yolk burst phenotype, with heavily pronounced failure of migration of E-YSL and the marginal blastoderm across the surface of the yolk, while migration of deep cells towards the epiblast appears to occur normally, causing a severe disruption to the structure of the embryo. The stalling appears to begin around 30%-50% epiboly stage, with embryos not showing visible dorsal structures beyond slight asymmetry of deep cell migration. This observation differs from the morphant phenotype that showed the *mxtx2*-deficient embryos having stalled epiboly at shield stage (Bruce *et al.*, 2005). Towards the presumed end of epiboly (in temporal terms), a contractile force appears to exert a force on the yolk cell around the equator position, localised to the approximate position of E-YSL and marginal blastoderm. Yolk cell is “squeezed” by the contractile ring before violently

exploding, killing the embryo and causing necrosis of remaining blastoderm cells. This is analogical to the previously reported phenotype (Bruce *et al.*, 2005; Wilkins *et al.*, 2008). Mosaic mutants of *mxtx2* were also shown to have a greatly reduced expression of *ndr2* gene, further confirming the previously observed morphant phenotype and the current understanding of *mxtx2* driving expression of Nodal effectors in zebrafish (Hong *et al.*, 2011).

While these results did not culminate in new findings, they still can serve the zebrafish community, with an independent method of gene disruption confirming the phenotypes found in embryos where *mxtx2* expression was targeted and disrupted by antisense morpholino oligonucleotide. With the increasing scrutiny on phenotypes presented by MO knockdown methods, and indications that targeted gene mutations do not correlate with previously described MO-induced phenotypes (Gerety and Wilkinson, 2011; Kok *et al.*, 2015; Moreno *et al.*, 2018; Stainier *et al.*, 2017), it is an important task to systematically establish if the gene function derived from MO-induced phenotype correlates with the gene mutation. Replicability of scientific results using independent methods often sees little enthusiasm due to limited novelty, yet it remains an important part of the scientific method.

The major aim of providing a further characterisation of *mxtx2* function in the context of zygotic genome activation was not achieved. While the mosaic mutation of this gene has shown high penetrance and a moderate reduction in expression, it would be a challenging tool when used in global transcriptomic analysis. Thus, an objective was set out to develop a knock-in mutation, in which an inserted cassette would disrupt the *mxtx2* locus, leading to expression of a truncated, functionally inactive *mxtx2* alongside a fluorescent reporter which would provide a visual output

of knock-in integration into the genome and allow easy selection of embryos positive for the knock-in mutation (Wierson *et al.*, 2020). I have been able to generate a targeted integration of a reporter insert in the *mxtx2* locus using the short homology arms-mediated method, with a moderate efficiency shown by imaging of the fluorescent reporter in F0 generation embryos. However, screening of the offspring of F0 zebrafish showed that this integration failed to be transmitted in the germline, causing the attempt of generating a stable transgenic line *Tg(mxtx2-2A-EGFP)* to be unsuccessful.

This was probably facilitated by the mortality induced by the *mxtx2* disruption. F0 embryos were screened for presence of GFP signal during epiboly, as the expression of the fluorescent reporter is only driven from expression of *mxtx2*, and *mxtx2* is not expressed post epiboly. Embryos positive for the GFP signal at that point showed developmental disruption associated with the *mxtx2* phenotype, and it is not impossible that F0 embryos with a stronger mosaic incorporation of the insert did not survive to adulthood, while surviving fish had only minor mosaic integration, disrupting mxtx2 in a very limited number of cells and reducing the probability of germline transmission. This was initially addressed by an attempt to target the integration to the germ line using a Cas9-nanos3'UTR fusion, however efficiency of insert integration using that nuclease was too low.

With the observations of the phenotype of the mosaic mutation in *mxtx2*, I can speculate that further attempts at generation of an *mxtx2* mutant line could conclude with a similar outcome. Embryonic lethality of this mutation is very penetrant, and alternative methods of establishment of a transgenic line should be considered. Use of inducible Cas9 systems could be considered (Sun *et al.*, 2019),

however it would have to be combined with strict localisation of Cas9 to the germline, as induction of mutation in a gastrulation embryo (this way it would avoid the epiboly phenotype) would result in a very small subset of cells showing the insert integration. Alternatively, the Cas9-nanos targeting of editing events to germ cells could be repeated, with the caveat of germ cells not expressing *mxtx2,* making pre-screening of F0 embryos for reporter integration near impossible.

The mosaic mutation of *mxtx2* established in this thesis can be used for further studies on the function of the gene with improvements in Cas9-mediated targeting efficiency, with alternatives being knockdown methods such as antisense MO or, if proven to not cause non-specific phenotypes, an alternative Cas13 nuclease.

These could be of use in further exploration of the role of *mxtx2* in activating transcription, and could help elucidate its role (if any) in ZGA. This could be achieved by a transcriptomic analysis: RNA sequencing of *mxtx2*-depleted embryos across several developmental stages, ranging from 512 cell stage (for early wave ZGA), through 1k-cell and oblong stages (for major wave ZGA) to dome stage (for epiboly-specific transcription). This could help establish a global profile of transcription mediated by this transcription factor, as well as identify its temporal profiles during early embryonic development. With the data collected in multiple stages and with an appropriate non-targeting control, differential gene expression analysis could be performed to identify candidate gene targets of Mxtx2 at multiple stages, including during ZGA. This data would provide a global, unbiased view of the Mxtx2 activity, potentially shedding further light on its role in early transcriptional activation, as well as the relationship with members of the Nodal signalling pathway and its role in YSL specification. The latter would be a particularly interesting development, as YSL is

the first lineage-restricted structure in the embryo, with cells forming the yolk syncytial nuclei acquiring the first specific cell fate in the zebrafish embryo(Carvalho and Heisenberg, 2010; Chu *et al.*, 2012; Kimmel *et al.*, 1995).

Alternatively, a panel of candidate Mxtx2 targets, including early wave and main wave ZGA genes could be assessed by qPCR, however this approach would offer limited new information about *mxtx2* function at those stages, and would be heavily biased by the selection of candidate genes.

Depending on the outcomes of the above, ChIP-seq analysis could be repeated(Xu *et al.*, 2012) in *mxtx2*-depleted embryos at appropriate early stages of development to complement the transcriptomic analysis, revealing the pattern of binding of Mxtx2 on the genome and further confirming if the genes identified by RNAseq are direct targets of the Mxtx2 transcription factor.

All in all, further analysis of the *mxtx2* function would be possible with additional time and some improvements to the efficiency of gene disruption techniques used. Further data, generated through next-generation sequencing approaches such as RNAseq, as well as ChIP-seq analysis could allow us to characterise the role of this key transcription regulator in early embryonic development, discern its role in transcription activation beyond it being a direct downstream target of Nanog (Gagnon *et al.*, 2018; Xu *et al.*, 2012), and further elucidate the mechanisms by which the global reprogramming event of zygotic genome activation is achieved.

Chapter 5: Zebrafish Yolk Syncytial Layer Undergoes a Delayed Zygotic Genome Activation

Preamble

Results presented in this chapter are part of a study conducted in collaboration with Haseeb Qureshi. Results presented in this chapter are going to be a part of a manuscript co-first-authored by myself and Haseeb Qureshi, currently in preparation. In this chapter, my main role was in experimental design, zebrafish embryo handling, all laboratory work and data generation. Data analysis was conducted primarily by Haseeb Qureshi, with several exceptions.

5.1 Introduction

5.1.1 YSL formation in zebrafish embryo

Exploration of the role of *mxtx2* has brought our attention to the remarkable structure of yolk syncytial layer. *Mxtx2* was shown to be heavily involved in YSL formation and directing epiboly, activating transcription of over 40% of all identified genes with a YSL-specific expression profile (Xu *et al.*, 2012). However, understanding of mechanisms leading to formation of this syncytial layer in the embryo is still elusive, with only limited descriptive studies exploring this process (Chu *et al.*, 2012). An overview of yolk syncytial layer function has been introduced in Chapter 1, and here I will present a brief summary of the events leading to its formation.

During the cleavage stages of the early blastula, cells of the zebrafish embryo divide without growth while sitting on top of the vegetally positioned yolk cell, eventually forming multiple layers of blastomeres. Interestingly, in the layer positioned directly above the yolk cell (vegetal-most) some cells positioned at the margins of the embryo (marginal blastomeres) - have been found to maintain an interface with the yolk cell through the presence of so-called cytoplasmic bridges, allowing passage of materials and nutrients from yolk to the cells (Kimmel and Law, 1985a; b). After 9 rounds of division (at 512 cell stage) of the zebrafish embryo, these marginal blastomeres collapse into the yolk, depositing its nuclei and cytoplasm into the yolk, giving rise to a non-yolky layer of a multinuclear syncytium - the YSL (Fig. 5.1). Individual nuclei of the YSL are referred to as yolk syncytial nuclei (or YSNs) (Carvalho and Heisenberg, 2010; Kimmel and Law, 1985a; b). First, the nuclei form a single row along the blastoderm/yolk margin, before undergoing 3 rounds of metasynchronous divisions,

after which YSNs become post-mitotic (Carvalho and Heisenberg, 2010; Kane and Kimmel, 1993; Trinkaus, 1993).

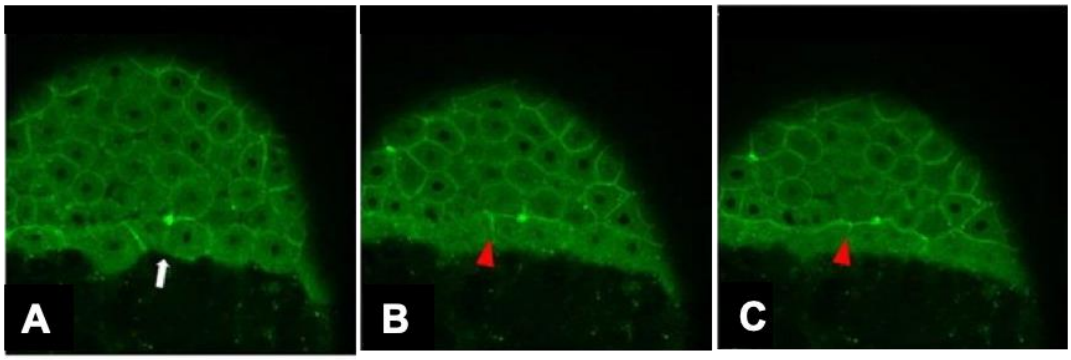


Fig. 5.1 Yolk syncytial layer formation by margin cell collapse into the yolk.
Embryos of the mGFP (plasma membrane-targeted GFP) transgenic line were visualised in time-lapse imaging, images adapted from (Chu *et al.*, 2012) (A) Before 512 cell stage, cells of the blastoderm margin show the same blastoderm morphology as the other blastomeres, with cell membranes separating them from each other (arrow). (B) At 512 cell stage, all blastomeres divide, with marginal blastoderm cells performing incomplete cytokinesis with the cell membranes separating marginal blastomeres regressing (red arrowhead), resulting in (C) their cytoplasm fusing with the animal part of the yolk without any cell membranes between them - forming a syncytium (YSL), separated from the blastoderm with a cell membrane.

This characteristic has been very intriguing and led us to raise several questions. Why are the YSNs divisions synchronous at a time when all of their cousin cells - blastomeres -increase their cell cycle length and lose synchrony? What is the mechanism causing this difference? Are YSL and blastoderm regulated by different transcriptional programmes?

The latter question was supported by the fact that the yolk syncytial layer becomes the first restricted lineage in the embryo with a fate of cells seemingly established before zygotic genome activation even takes place. This raised a hypothesis that yolk syncytial nuclei and their cousin blastomeres could have a differential transcription regulation.

Initial observations of the published *in vivo* imaging datasets of zebrafish embryo development (Keller, 2013) have indicated that in addition to synchrony in YSL embryos, the timing of cell divisions appears shorter than that of blastomeres. However, that was an observation of a single imaged embryo, and could be an artifact.

Nevertheless, this has supported our interest in the subject, as our previous study indicated that in pre-ZGA embryos, appearance of the previously described *miR-430* transcription foci (indicative of zygotic transcription) is coordinated with the length of the cell cycle, with S-phase being permissive to transcription (Hadzhiev *et al.*, 2019). This culminates at MBT (1k cell stage), when blastomere cell cycles become significantly longer and lose their metasynchrony, and zygotic genome activation takes place. However, is this mechanism different in YSL due to their maintaining of synchrony?

Dynamics of transcription in the pre-ZGA zebrafish embryos has been studied, but mostly focused on individual cells of blastoderm, without distinguishing between cell populations (Hadzhiev *et al.*, 2019; Hilbert *et al.*, 2021; Kuznetsova *et al.*, 2022). During late cleavage stages transcription in embryos is concentrated to 2 transcription bodies. These have been characterised by active transcription and concentration of *miR-430*, as well as specific localisation of both initiating and elongating RNA polymerase II to those transcription bodies.

No studies to date have attempted to characterise transcription and genome activation specifically in the newly formed YSL, and very limited information is available about cell cycle dynamics in YSL, apart from the observation of metasynchrony.

5.1.2 Aims of this chapter

With a readily available method of measuring cell cycle lengths and visualising transcription bodies present in pre-ZGA blastomeres (Hadzhiev *et al.*, 2019), we could investigate whether both cell cycle and transcriptional patterns differ between YSL and blastomeres during zygotic genome activation.

In this study, we aim to elucidate why such a distinct pattern of synchronous cell cycles is exhibited by yolk syncytial nuclei after the 1000-cell stage, and whether that can be linked to differences in transcriptional patterns as compared to blastoderm cells. We use newly developed software to track nuclei and transcription foci in both time and space, with use of 4-dimensional lightsheet microscopy. With that, we aimed to answer these specific questions:

How are cell cycles in YSL nuclei different to those of blastomeres?

How do the detectable global features of transcription compare between YSL nuclei and blastomeres? Can predicted differences in transcription dynamics between yolk syncytial nuclei and blastomeres be quantitated?

5.2 YSL Nuclei Have a Separate Pattern of Divisions to Rest of the Embryo

5.2.1 YSN divisions appear coordinated, unlike blastoderm

Initially we set out to confirm whether we can observe the unique cell cycle dynamics using our methods of visualisation. To do that, embryos were injected at 1-cell stage with a fluorescent histone protein mRuby:H2B, to stain chromatin in individual nuclei in the embryo. Embryos were imaged live in 3 dimensions using Light Sheet Fluorescent Microscopy (LSFM) between 64-cell stage and dome stage. As YSL formation occurs at the blastoderm margin, it is easy to distinguish YSNs from the rest of the blastoderm due to its anatomical position. Obtained imaging was later processed and individual nuclei were manually tracked and mitotic events (indicated by the change in nuclear staining appearance during anaphase) were annotated. Our results confirmed that divisions of YSNs are temporally metasynchronous after the 1k-cell stage (Fig. 5.2), with short periods during which mitoses were detected in yolk syncytial nuclei, resembling the pattern observed in blastomeres during pre-ZGA cleavage divisions. In comparison, mitotic events in the blastoderm after 1k-cell stage were stochastic and did not show any temporal coordination.

We also observed that in addition to the temporal coordination of mitoses in the YSL (metasynchrony), YSNs divide in a spatially coordinated “wave-like” pattern (Fig. 5.3 & Fig. 5.4., see also Movie S10).

This could suggest a presence of a particular symmetry axis or gradient defining which nuclei divide first, however we were not able to address this issue further.

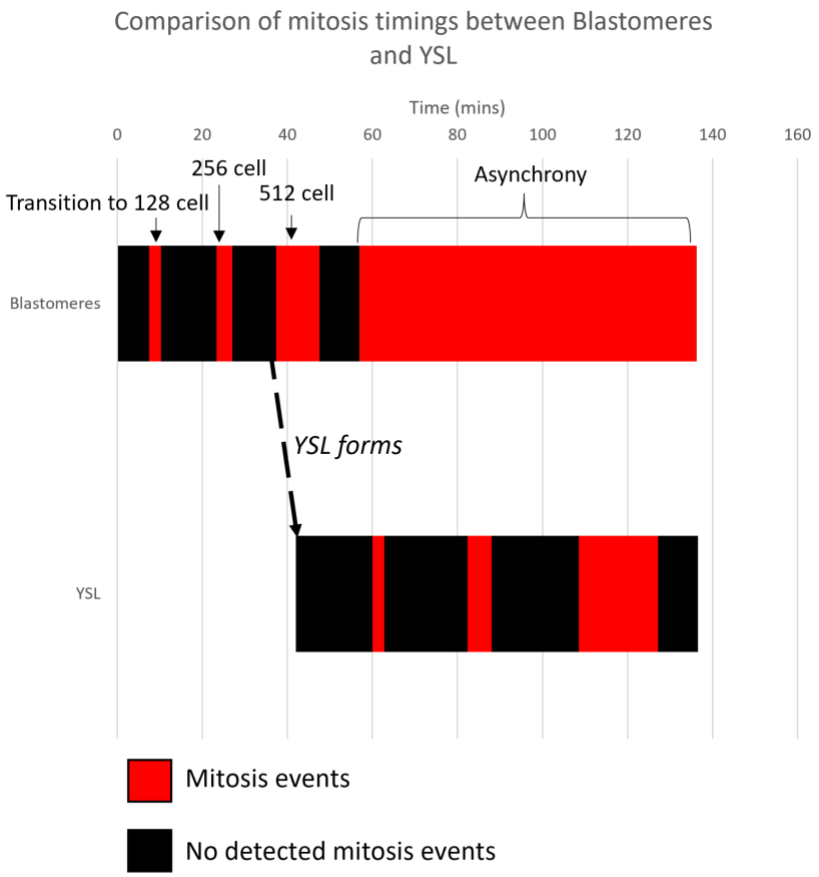


Fig. 5.2 Stacked bar chart showing timings of global mitoses from a single embryo.
Red shows times within which mitoses are detected, and black shows periods when no mitoses are detected. Distinct red periods signify transitions from one developmental stage to another until synchrony in blastoderm is lost. Measurements are based on tracking mitoses in a LSFM imaging dataset labelling H2B. Time as indicated is measured from the beginning of imaging with "0" during the 64-cell stage.

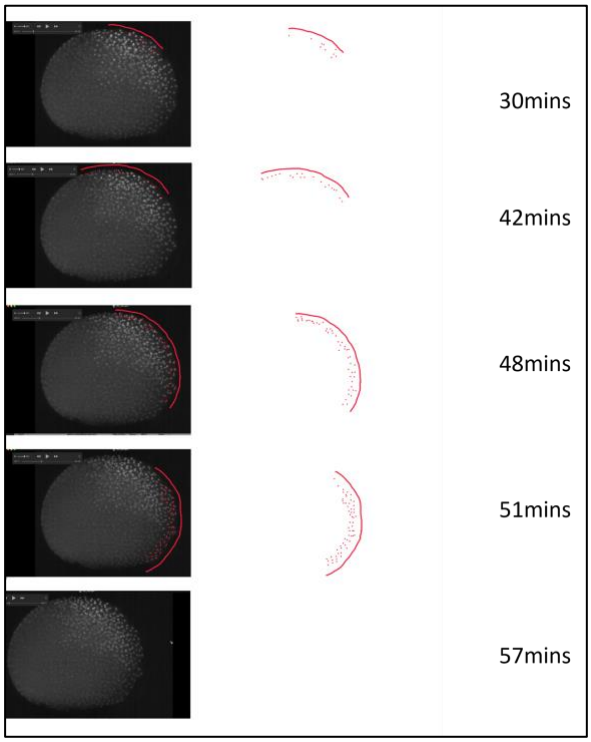


Fig. 5.3 Wave-like cell cycle dynamics of the yolk syncytial layer nuclei at the oblong stage.
LSFM 4D imaging presented in Movie S10 of fluorescently labelled nuclei (microinjected mRuby:H2B protein) was rendered into an animation from animal pole view. YSL nuclei were detectable at the perimeter of the blastoderm as large discs. Condensation of chromatin was used as indication of mitotic phase of the cell cycle and indicated with red dots near the observed condensed chromatin. A red line depicts the region of the YSL where condensed chromatin is detected. Overlaid red dots and lines representing condensed chromatin region are separately shown in the middle panels. Lack of red dots and line on the bottom panel indicate no chromatin condensation in YSL and suggest the finish of mitotic wave of nuclear divisions. On the right, time (minutes) from the start of imaging (64-cell-stage) is shown. Imaging was carried out at 55 seconds/ frame rate. Note, that only nuclei of the right hand side of embryo disc are illuminated sufficiently to detect their chromatin state.

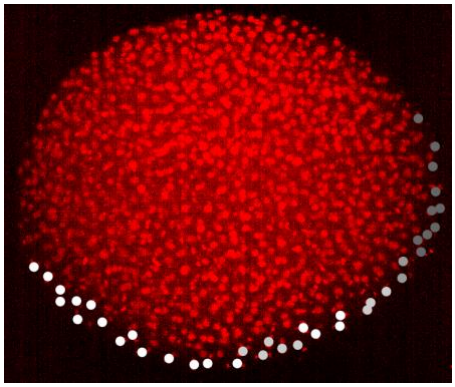


Fig. 5.4. Mitotic events in the YSL show a temporal coordination in a "wave-like" pattern across the embryo.
Embryos were injected at 1-cell stage with mRuby:H2B histone and imaged in a live time-lapse using LSFM with 55 seconds per frame. Individual frames of the time-lapse were manually annotated. Presence of mitotic events was marked in YSL embryos between high and oblong stages, with each spot depicting a mitotic event and its position. A brightness gradient was used to indicate temporal changes, with brightest spots being the earliest events, and darkest spots the latest.

5.2.2 YSN division cycles are shorter than the blastoderm cousin cells

Having confirmed the previously reported metasynchrony of YSNs (Kane and Kimmel, 1993; Trinkaus, 1993), we have set out to determine whether the length of cell cycles differs between blastomeres and YSNs. mRuby:H2B chromatin staining has been used to label nuclei, and miR-430-FITC MO was used to visualise active transcription (transcription bodies in the embryos). Manual tracking of individual cell cycles in blastomeres and YSNs was conducted as before, additionally presence of transcriptional foci in the tracked nuclei was also noted and used as an indication of the period of transcriptional activity of each cell or YSN.

Fig. 5.5 shows a frame-by-frame time-lapse of representative individual blastomere (5.5 A) and YSL nucleus (5.5 B), with both tracked from the 1k-cell stage. Cell cycle in the blastomere was significantly longer than in its cousin YSN, with 31 minute period compared to 23 minutes.

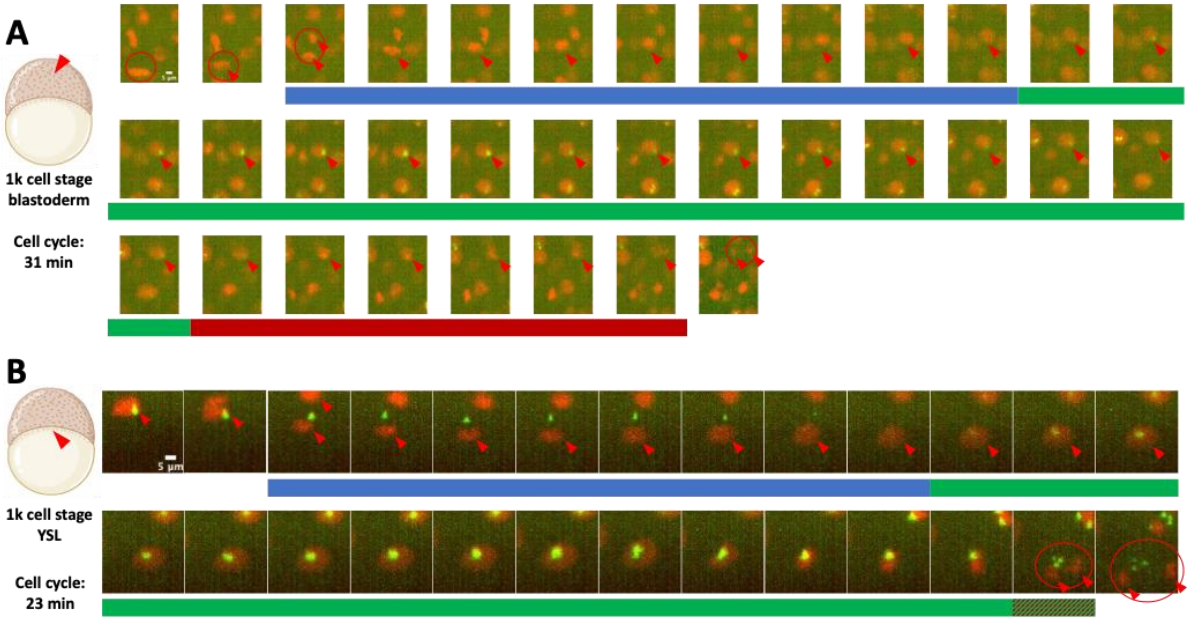


Fig. 5.5 Cell cycles in the YSL nuclei are shorter than in blastoderm cousin cells.
Time-lapse of light sheet imaging of embryos injected at 1-cell stage with mRuby:H2B histone protein (red) and miR-430-FITC morpholino (green), labelling chromatin and highlighting the transcription foci respectively. Embryos were imaged by LSFM with the framerate of 1min/frame. Representative blastomere and YSL nuclei were manually tracked and cell cycle length was measured from anaphase to anaphase. Arrowheads point to the nucleus being tracked and (during mitosis) to its sister nucleus. Markings underneath the images correspond to periods of cell cycle assigned to nuclei, with blue showing the post-mitotic transcription quiescence period, green representing the transcriptional activity period, and red depicting the mitosis stages, with no transcription present. In case of YSL nuclei, red and green was used for the anaphase due to the unexplained persistence of miR-430 transcription foci.

We have quantified the manual tracking of cell cycle lengths in multiple nuclei in the same embryo in an analogical pattern (comparing cell cycles initiated at a similar time, of cousin cells), and this analysis (Fig. 5.5) has confirmed that observation. Cell cycles in YSL were shorter than those of blastomeres at 1k cell stage, during the initiation of main wave of zygotic genome activation. However, when comparing the duration of cell cycles between YSL and blastoderm, an unusual pattern emerged. In simple terms, these data seem to indicate that YSL nuclei are “delayed” by one cell cycle when compared to their blastoderm cousins. This “lag” in YSL persists across

the first 2, out of the total of 3 metasynchronous divisions of YSNs, with both their first and second division cycles resembling the temporally “previous” stage in blastomeres (Fig. 5.5). This observation raises a hypothesis that YSL is developmentally delayed compared to its cousin blastomeres but would need to be confirmed in a larger number of embryos and across a longer developmental period. However, the hypothesis that YSL exhibits developmental delay is aided by the observed phenotypes in YSL being developmentally premature compared to the rest of the embryo - synchronous and shorter cell divisions.

Interestingly, Fig. 5.5 shows a persisting, extranuclear signal coming from the miR-430-FITC morpholino, present after the nuclei have divided and there should be no active transcription. This phenomenon was observed in multiple imaging datasets in both blastomeres and YSL, indicating that it is not specific to one of these groups. It is not currently known why it occurs. One possible explanation is that the fluorescent morpholino forms a separate aggregate that is removed from the nucleus and persists for several minutes before dispersing. Alternatively, this could be a result of a process of removal of miR-430 from nuclei during cell division. It could also be speculated that this is an artifact resulting from flattening of 3D images into a 2D plane, however this does not answer the question of miR-430 signal persisting beyond cell divisions.

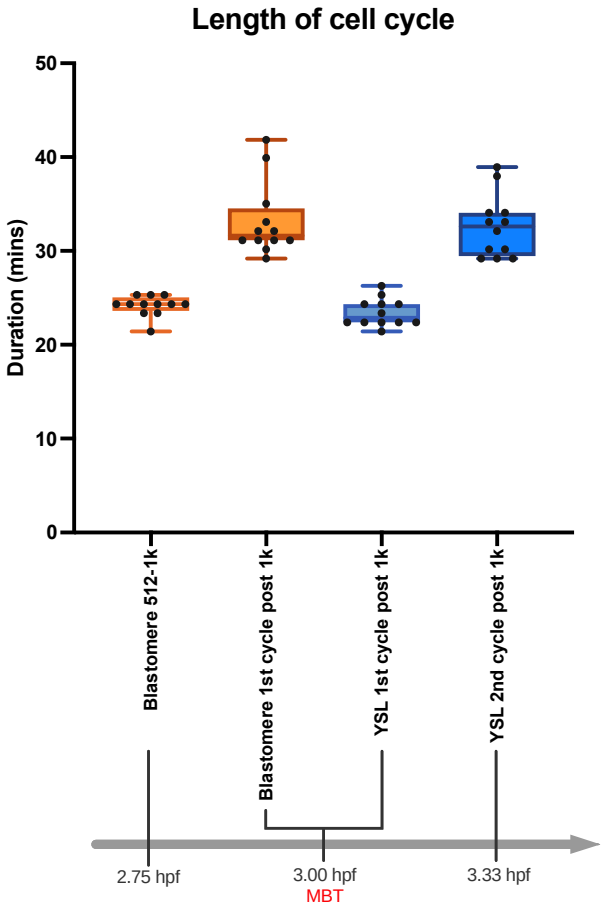


Fig. 5.6 Comparison of cell cycle durations between blastomeres and YSNs.
Quantification of manual tracking of cell cycle lengths in indicated nuclei at indicated stages. Cell cycles were tracked analogically to Fig. 5.4. (anaphase to anaphase) in n=12 nuclei of each group within a single embryo. Measurements were done on a frame by frame basis in an imaging dataset of 1min/frame. Data presented as a box-and-whiskers plot, with whiskers ranging from minimum to maximum, and box denoting 75th percentile, median and 25th percentile. Individual datapoints are presented on the graph.

5.3 Transcriptional Dynamic of YSN Is Delayed In Comparison to the Blastoderm

Having established that YSL nuclei appear to be “delayed”, with their cell cycle length indicating developmentally premature dynamics, we asked whether that delay is also present in their transcription dynamics. Due to the close relationship between cell cycle length and transcription in blastula embryos, transcriptional dynamics in YSL should, in theory, also be developmentally delayed. If that is the case, and transcription profile of YSL is delayed, this could indicate that YSL are undergoing ZGA at a different stage to the blastoderm, a previously unseen phenomenon. Therefore, we have proceeded with characterising the transcription pattern and dynamics in YSL and compared it to that in blastoderm. For that, 2 separate methods of visualising transcription in embryos have been used. In the first, we injected embryos with the fluorescent miR-430-FITC MO to mark the *miR-430* transcription foci appearing in late blastula and conducted *in vivo* time-lapse imaging. Using 2-photon microscopy, we demonstrated that transcription bodies are present in both YSL nuclei and blastoderm at 512 cell stage, with a clear distinction in anatomy between the blastoderm cells and YSL nuclei (Fig. 5.7). During the 512-cell stage, prospective yolk syncytial layer has already fused with the yolk, forming a clearly distinct layer with the specific anatomy (fusion of cells into a single mass) and a different distribution of fluorescence from both blastoderm and yolk cytoplasm.

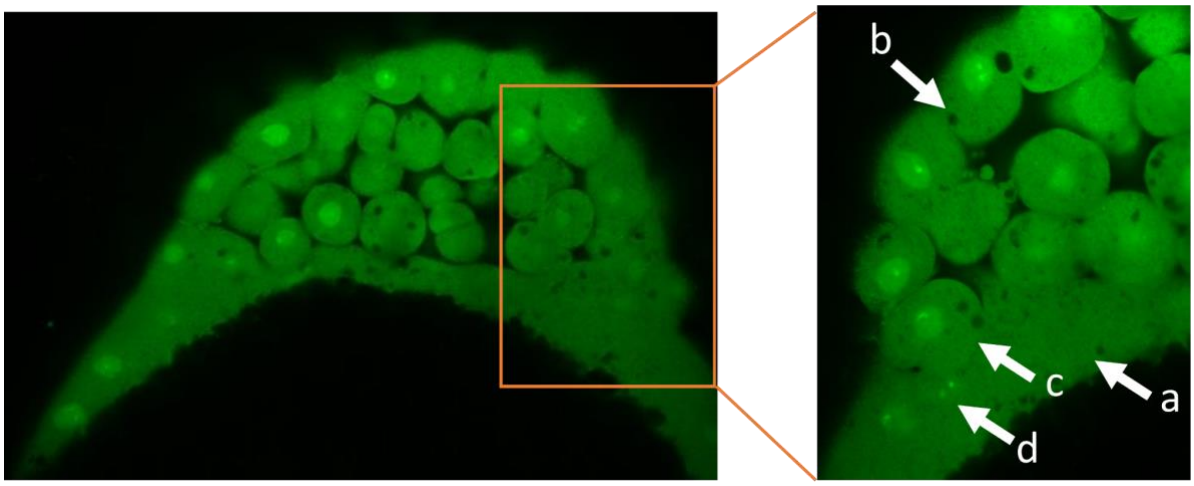


Fig 5.7 miR-430 transcription bodies are present in the yolk syncytial nuclei and blastoderm. *In vivo* 2-photon microscopy image (single z-slice from 3D stack) of zebrafish embryo YSL (**a**) and blastoderm (**b**) taken at 512 cell stage, showing fusion of blastomere with yolk cell (**c**), and nuclei already fused with the yolk syncytium (**d**) where transcriptional foci are present. *Image taken by Haseeb Qureshi.*

The second method of visualising transcription was by immunostaining. Transcription dynamics were visualised through this method by using an antibody against the actively transcribing (elongating) form of RNA Polymerase II (RNA Pol II Ser2P), which is specifically phosphorylated at the C-terminal domain. This way, we can ensure that the visualisation of RNA Pol II imaging would provide a readout for both the extent and localisation of transcription activity in the cell. However, for immunostaining embryos need to be fixed, therefore any imaging would only provide a snapshot of a single timepoint of the highly dynamic process in embryos. Additionally, with the stochasticity of cell cycles at stages after MBT, individual cells within the embryo might not be at comparable stages of the cell cycle, for example one cell being in S-phase and showing high levels of transcription, while its neighbour is undergoing mitosis with no transcriptional output. This has led to development of imaging strategy outlined in section 2.7.3, whereby multiple embryos stained for RNA Pol II

Ser2P are imaged, and 4 standardised and unbiased regions of interest containing both YSL and neighbouring blastoderm are imaged per embryo. Imaging was done using LSFM, with 3-dimensional Z-stacks taken for each defined view in multiple embryos, with embryos stained and fixed at 3 cell stages - 512 cell (2.75hpf), 1k-cell stage (3hpf) and oblong stage (3.5 hpf). Obtained 3D images (examples of which are shown in Fig. 5.8) were then processed using an in-house developed (H. Qureshi) image segmentation pipeline to extract information about the a) volume and b) surface area of the RNA Pol II foci in individual cells, which we used as a readout of transcriptional activity to be compared between 2 distinct populations - YSL and blastomeres – at each of the selected stages.

Using this imaging and image processing analysis strategy, we aimed to answer the question of whether transcription dynamics in YSL differ from those in blastomeres, and if transcription output of YSL nuclei is reduced in YSL, indicating delay.

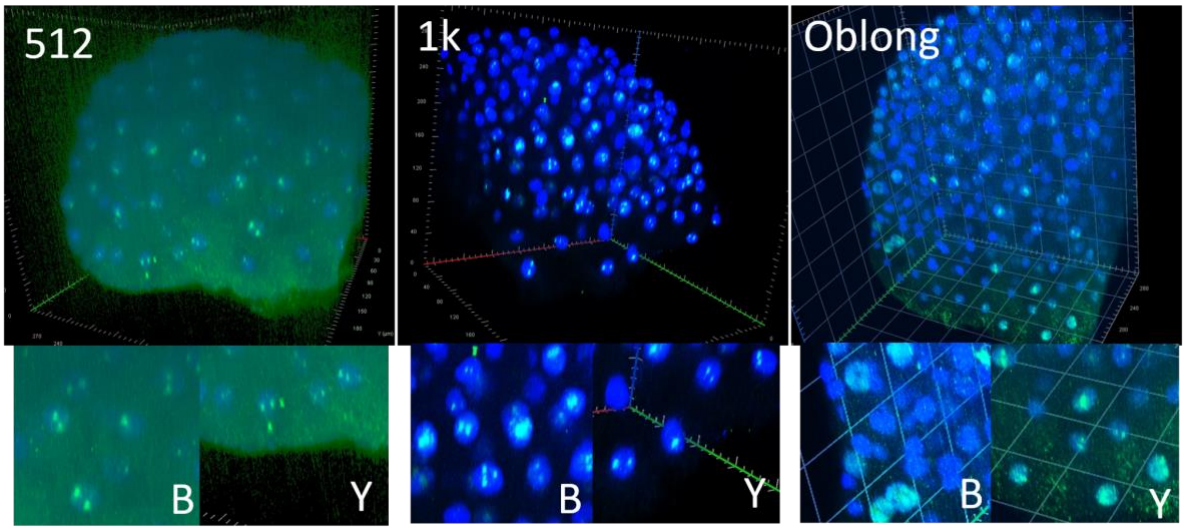


Fig. 5.8 3D views of the LSFM visualisation of active transcription in fixed embryos, imaging data of which was segmented to extract volumes and surface areas of RNA Pol II transcription foci.
From left to right, images show representative views of embryos at 512 cell stage, 1k cell stage and oblong stage (top), with bottom panels presenting zoomed-in view of blastomere RNA Pol II regions (B) and YSL RNA Pol II regions (Y). Elongating RNA Pol II (Ser2P) is labelled in green and nuclei (Hoechst) are labelled in blue. Screenshots of 3D views taken from Zeiss Zen Blue software.

Image analysis based purely on comparison of the RNA pol II foci intensity or distribution between cell populations could be misleading due to the aforementioned stochasticity of cell cycles, as in a fixed image the dynamic process of transcription is reduced to a snapshot. To additionally complicate the comparison, illumination of each nucleus is not equal in the light-sheet microscopy, with the deeper nuclei beneath the surface of the blastoderm showing increased light scattering and therefore reduced intensity of signal - this is due to illumination light having to pass through multiple layers of cells. YSL is positioned on the margins of the embryo, with few layers of cells covering it.

Before moving to quantitative image analysis with segmented RNA Pol II regions, our imaging has revealed certain interesting characteristics of the yolk syncytial. Previous observations by (Kimmel and Law, 1985a; b; Trinkaus, 1993)have shown that the yolk

syncytial layer forms somewhat of a separate phase to the yolk cytoplasm layer, which contains yolk granules. Here, thanks to the undesirable, but specific to the YSL, noisy distribution of the unincorporated secondary antibody (AF488) signal, we show that yolk syncytial nuclei, despite being part of a combined syncytial mass and not possessing cell membranes, form certain niches we dubbed “exclusion zones” that appear to be phase separated from the rest of the yolk cytoplasm (Fig. 5.9). This discovery provides some additional information about the unique anatomy of this structure, suggesting an alternative way to cell membranes in which syncytial units retain spatial independence.

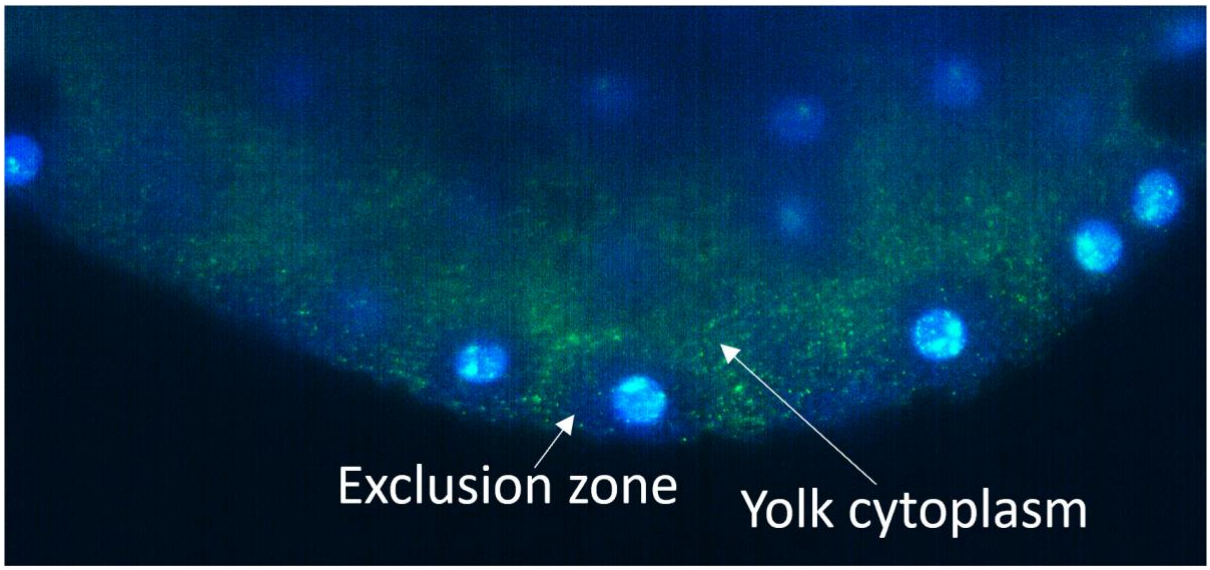


Fig 5.9. 2D Z-slice from 1000 cell stage embryo showing the YSL syncytium and its component nuclei. YSL shows widespread RNA Pol II aggregates with no sharp membrane distinctions as is seen in blastoderm. There is ‘space’ between exclusion zones. Embryos are labelled by RNA Pol II Ser2P antibody (green) and nuclear Hoechst stain (blue). Screenshot of 3D views taken from Zeiss Zen Blue software.

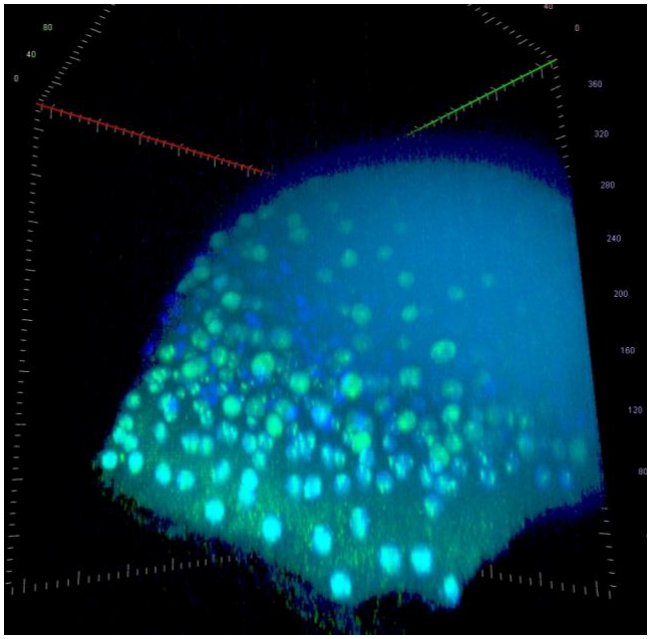


Fig 5.10. 3D view of the LSFM imaging of a fixed embryo at sphere stage (4hpf)
Elongating RNA Pol II (Ser2P) is labelled in green, and nuclei (Hoechst) are labelled in blue. Screenshots of 3D views taken from Zeiss Zen Blue software.

In order to control our imaging strategy, we looked at RNA pol II embryos imaged at later stages, as we expected that widespread genome activation, if delayed in YSL nuclei, must occur before epiboly onset, as we knew that at dome stage (4.33 hpf) hundreds of YSL specific genes are induced by Mxtx2 transcription factor. Indeed, with immunostaining for active RNA pol II we see a widespread profile of genome activation in both YSNs and blastomeres (Fig. 5.10 - green foci colocalising with the entire nucleus and not only *miR-430* transcription bodies) as early as at sphere stage (4.00 hpf). This observation indicated that by sphere stage, YSL and blastomeres show equivalent levels of active transcription, and delay in transcription dynamics becomes minimised at that point. This coincides with YSL nuclei becoming non-mitotic.

With that in mind, we moved on to the quantitative comparison of transcriptional activity between yolk syncytial nuclei and blastoderm nuclei. Imaging data from embryos fixed at 3 stages and stained against the elongating RNA Pol II was collected - 512 cell, 1000 cell and oblong stages. These data were segmented to isolate only regions containing true RNA pol II signal, and these regions were then quantified to produce measurements of RNA polymerase surface area and volume in each of the nuclei. 2 subsets of data were analysed, with one containing measurements in YSL nuclei, and the second measurements in blastoderm nuclei. This was done for each individual embryo, and 3 embryos were analysed for each stage. Total volume and total surface area of RNA Pol II regions has been used as a readout of total transcriptional activity in the analysed nuclei populations.

In Fig. 5.11 we compare the distributions of these 2 readouts within YSL and blastomere nuclei populations in single embryos. We show that across all 3 stages mean volumes of active Pol II remain comparable between blastoderm and YSL, blastoderm nuclei display a much larger range and maxima of the total volume than YSL, suggesting that a subset of blastoderm nuclei displays much higher transcriptional activity. Analogical results have been seen when comparing surface area of active RNA pol II, with much higher maxima displayed by the blastoderm nuclei. Additionally, mean RNA pol II surface areas at 1000-cell and oblong stages are higher in blastoderm, giving a further indication of a smaller transcriptional activity exhibited by YSL nuclei when compared to their contemporaries.

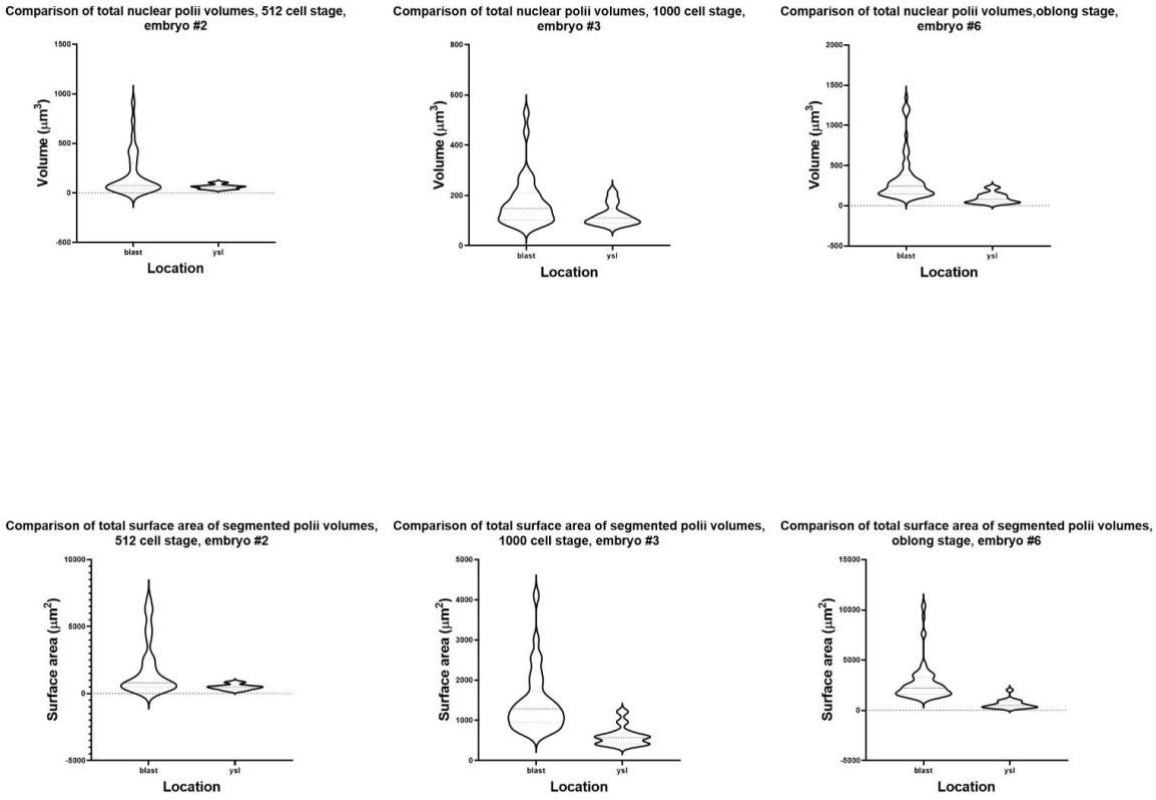


Fig. 5.11. Comparisons of surface area and volume of segmented RNA polymerase II regions between YSL and blastoderm in individual embryos.

Comparisons of surface area (top row) and volume (bottom row) of segmented RNA Pol II regions between YSL and blastoderm. Data presented was collected from individual embryos, displaying information collected from cell populations present within a single embryo, per each developmental stage. Statistical significance tested by unpaired Kolmogorov-Smirnov test. Top left: 512 cell stage volume, not significant difference, bottom left: 512 cell stage surface area, not significant difference. Top middle: 1000 cell stage volume, not significant difference. Bottom middle: 1000 cell stage surface area, $P<0.0001$. Top right: Oblong stage volume, $P<0.0001$. Bottom right: Oblong stage surface area, $P<0.0001$.

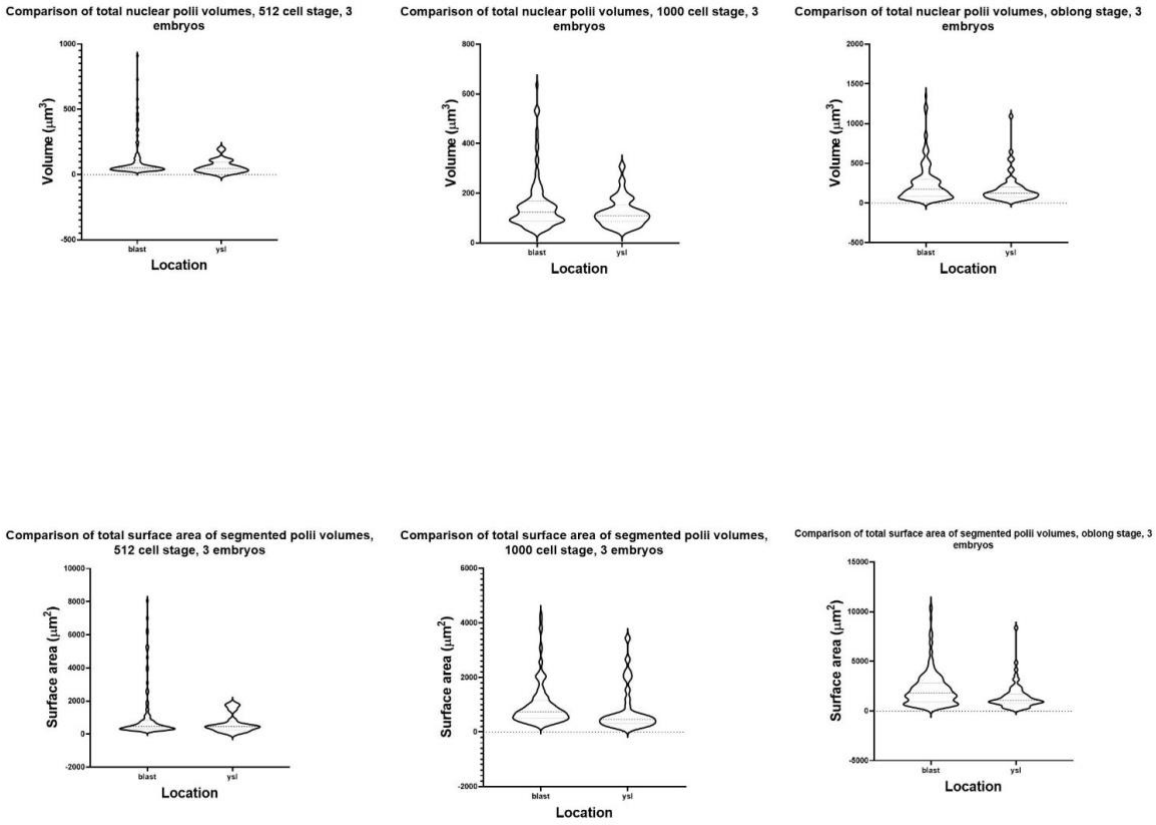


Fig. 5.12. Comparisons of surface area and volume of segmented RNA polymerase II regions between YSL and blastoderm aggregated from imaging multiple embryos.

Comparisons of volume (top row) and surface area (bottom row) of segmented RNA Pol II regions between YSL and blastoderm. Aggregated data from multiple embryos (3 embryos per stage). Statistical significance tested by unpaired Kolmogorov-Smirnov test. Top left: 512 cell stage volume, not significant difference, bottom left: 512 cell stage surface area, not significant difference. Top middle: 1000 cell stage volume, not significant difference. Bottom middle: 1000 cell stage surface area, P=0.0030. Top right: Oblong stage volume, P=0.0253. Bottom right: Oblong stage surface area, P=0.0004.

In Fig. 5.12 we present the analogical analysis, but this time as data aggregated from 3 independently analysed embryos. The obtained results comparing RNA Pol II volume and surface area at 512-cell stage were similar to the findings in single embryos, with differences in distribution and significantly larger maxima of those readouts of transcriptional activity. At 1000-cell and oblong cell stages, the aggregated distributions of RNA Pol II area and volume nuclei populations from 3 embryos showed more similar distributions. However, in those comparisons, means of RNA Pol II area and volume were significantly higher in blastoderm population, indicating that an average blastomere has a higher transcriptional activity than an average yolk syncytial nucleus. An important caveat of this analysis is that varying numbers of nuclei were analysed per embryo, with larger number of blastomeres than YSNs. This could be corrected by selecting a random subset of blastomeres corresponding to the number of available YSNs for analysis.

All in all, analysis of RNA Pol II surface area and volume provided strong evidence that YSL nuclei display lower transcriptional activity than blastomeres. This evidence is in line with our previous observations of delay of YSL nuclei with respect to blastomeres and supports our hypothesis that zygotic genome activation in yolk syncytial layer occurs later than in the blastoderm.

5.4 YSL Isolation by Embryo Sectioning

Further studies comparing YSL and blastomere transcription would likely require specific isolation of those two populations, for example for use in transcriptomic analysis (RNAseq). In anticipation of that, I have explored potential methods of

isolating blastomeres from yolk, and generated preliminary data. I have attempted to manually separate the 2 populations using a hypodermic needle as a blade to remove blastomeres from the surface of yolk. This procedure has been previously utilised in zebrafish embryos, transplanting animal caps in mesoderm induction studies (showing that yolk syncytial layer promotes mesoderm cell fates) (Chu *et al.*, 2012; Ober and Schulte-Merker, 1999).

Through use of this method I have isolated animal caps (blastomeres) and yolk (containing YSNs) at the oblong stage. As the isolation was not perfect due to yolk cell fragility and difficulty in clearing blastomeres from its surface, I have used qPCR to determine whether isolated yolks are enriched for YSL-specific transcripts (Fig. 5.13). Expression of *mxtx1* as zygotic and YSL-specific gene was used, with additional controls being beta-actin (maternal, ubiquitously expressed gene) and *klf17* (zygotic, ubiquitously expressed gene). qPCR has showed indications that *mxtx1* was enriched in the isolated yolks, however this enrichment was not statistically significant (Fig. 5.13). This approach was only preliminary and utilised as a query into the viability of embryo dissection as a method of YSL isolation.

Due to the limited scope and exploratory character of this question, just four genes were included, therefore it is not possible to draw conclusions from these experiments. However, significant improvements to this query would be the use of multiple appropriate housekeeping controls and an inclusion of a larger panel of assayed genes specific to the expression profiles I am attempting to distinguish (maternally provided, zygotic – embryo proper, zygotic – YSL). *mxtx1* was used as a convenient readout of YSL expression and enrichment, but as a gene that is not exclusively expressed in the YSL, it is not a perfect gene for this assay. Additionally,

more practice and repetition of the technique would reduce the variability displayed by the data and potentially lead to better separation of YSL from embryo proper.

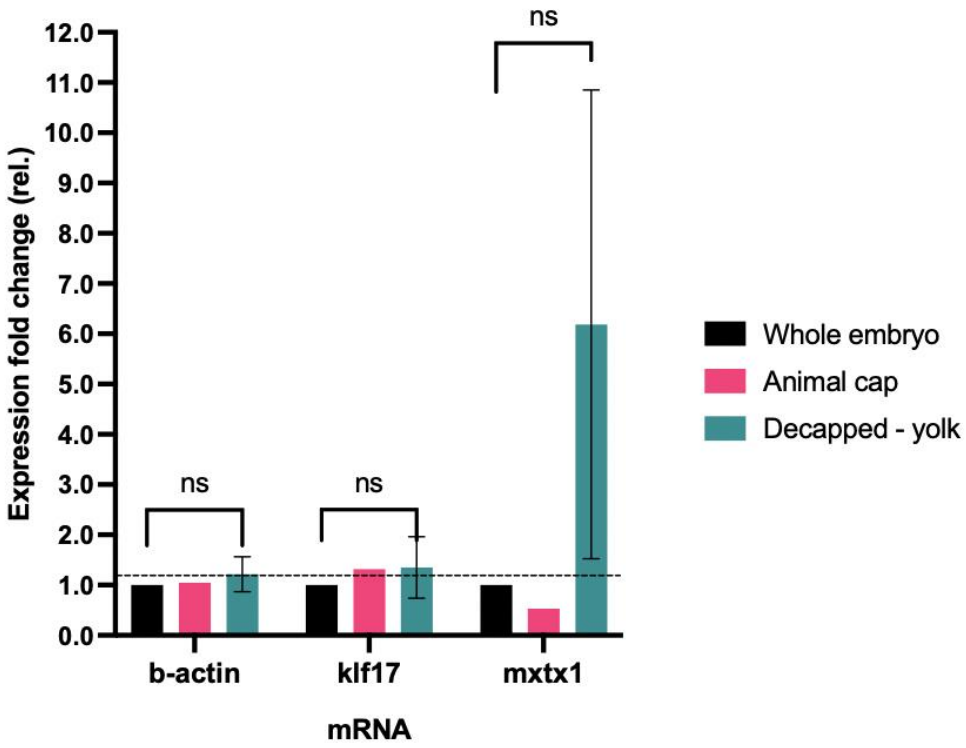


Fig. 5.13 qPCR in mechanically 'decapped' embryo yolks shows limited enrichment of the YSL-specific gene expression.

qPCR analysis of enrichment of *b-actin* (ubiquitously expressed maternal gene), *klf17* (ubiquitously expressed zygotic gene) and *mxtx1* (YSL-only expressed zygotic gene) transcripts in specified mechanically separated cell populations. Cell populations were collected for RNA extraction at oblong stage. Relative transcript enrichment was normalised to the *eef1a1a* housekeeping gene and is presented as relative abundance to the whole embryo. Data presented as mean±SD, N=3 for whole embryo and decapped-yolk samples, N=1 for animal caps. RNA of 15 embryos (or equivalent) isolated for each group in each replicate.

5.5 Discussion

In this chapter, we have provided a detailed characterisation of cell cycle and transcription dynamics in the yolk syncytial layer. We have confirmed that YSL undergoes 3 temporally metasynchronous divisions after its formation at 512 cell stage. Having identified cell cycle synchrony as a potential indicator of developmental immaturity of yolk syncytial nuclei, we have shown that cell cycles are significantly shorter in YSNs when compared to their contemporary blastomere cousin cells. Cell cycle lengths exhibited by YSNs were similar to those of blastomeres in their previous cell cycle, indicating that a delay or "lag" is present in the YSNs, providing further evidence of developmental immaturity. We also show that metasynchronous cell cycles in YSL are both temporally and spatially coordinated.

We have queried whether transcription dynamics of YSL are also subject to the delay. Through a light-sheet fluorescent microscopy imaging strategy of elongating RNA polymerase II combined with image segmentation we showed that yolk syncytial nuclei exhibit lower transcriptional activity than their blastomere contemporaries at 512, 1k, and oblong cell stages. Taking into account all presented evidence, we propose that yolk syncytial layer nuclei are developmentally and transcriptionally immature compared to blastomeres in zebrafish development, and that YSL undergoes zygotic genome activation with a delay, separately to the blastoderm.

Additionally, we have identified an intriguing new characteristic of yolk syncytial nuclei, which appear to be self-contained in what we have dubbed "exclusion zones", which we speculate offers limited mixing with the rest of yolk syncytial cytoplasm.

These have been observed due to non-specific binding of primary or secondary antibody, and the noisy signal being excluded from the area surrounding YSNs. Potentially, this exclusion could be a result of endoplasmic reticulum (ER) surrounding the YSNs, which could be tested using a specific marker for ER.

Our initial observations of cell cycle dynamics, especially cell cycle length comparisons, showing the delay in YSL, have been limited by manual tracking of the embryo and possess limited statistical power, despite quite clear indications of the delayed behaviour of YSNs. To address that, we have generated additional live cell imaging datasets using mRuby:H2B, and we hope to use an automated pipeline to automatically track cell cycles and nuclei lineages in those datasets (Qureshi *et al.*, 2020). This may help provide additional evidence of the delay.

Quantifying active transcription in fixed embryos shows a wide range of transcription values. Fixing embryos to capture transcription at a single time point offers the stability to make reliable measurements but offers the caveat of giving an incomplete view of transcriptional states in the embryo. Primarily, the loss of blastodermal synchrony in contrast with the metasynchronous wave-like divisions of the YSNs (Trinkaus, 1993) means that the data captures a wide range of transcriptional states, from early transcription, to peak activity to the declining phase where transcription peters off (Hadzhiev *et al.*, 2019). This loss of synchrony affects the outcomes from blastodermal transcription disproportionately more in comparison to the more metasynchronous YSL nuclei. Additionally, fixation and immunostaining seems to not have a uniform distribution, with certain spatial biases in staining (Fig. 8 and Fig. S12-14). This spatial bias is not consistent between individual embryos, suggesting that the method of immunostaining could be further optimised to improve on the quality

and distribution. If improvements in quality of immunostaining do not improve the spatial biases, there could be an indication of certain regional subsets within the embryo being more transcriptionally active at certain times.

Because a lack of synchrony provides RNA Pol II volumes of varying stages, the upper quartile of segmented volumes was selected for comparative analysis between blastoderm analysis, as a representation of transcriptional states reaching their maximal activity, in line with the assumption that RNA Pol II volume correlates with levels of its activity (Tadros and Lipshitz, 2009).

5.5.1 Surface area as a proxy for transcriptional activation morphology

Measuring the surface area of segmented volumes was selected as an approximate measure of the shape of transcription volumes. As described in (Hadzhiev *et al.*, 2019), early embryonic transcriptional states in the zebrafish embryo take on a double dot appearance, reflective of the pre-MZT activation states of transcription where fewer genes are actively transcribed. Post MBT, a greater number of genes from across many genomic loci are activated (Newport and Kirschner, 1982a; b), which would result in numerous local speckles across the nucleus. The surface area of segmented RNA Pol II volumes could be used to help gauge whether RNA Pol II activity is scattered and speckled, similar to a post MBT state, by virtue of a larger number of volumes offering a large surface area: volume ratio. This is generally visible from across the measured embryo samples, although low sample sizes for viable YSL nuclei do make robust statistical analysis difficult. The analysed data provides a compelling early glance into the state of RNA Pol II distribution across contemporary YSL and blastoderm nuclei, with a greater overall surface area in blastomere RNA Pol II regions suggesting a more granular topology, which would be

consistent with the notion of distributed RNA Pol II speckling that occurs with wider genome activation.

Conversely, maximally transcribing nuclei may be segmented to show the entire nucleus filled with a singular transcription volume, decreasing the surface area:volume ratio, although the extent of this occurring in our segmented datasets needs to be investigated.

5.5.2 Maxima of segmented RNA Pol II volumes are greater in blastoderm with oblong stage showing more parity between YSL and blastoderm RNA Pol II volumes.

The extent of transcriptional activity, as measured via the volume of segmented active RNA Pol II regions, shows a greater volume in the blastoderm in comparison to the YSL. The greater volume suggests more widespread genomic activation within blastomere nuclei. The disparity, though not consistently deemed significant by the statistical test, is pronounced across 512 and 1000 cell stage but this difference reduces, or even reverses during oblong stage. This may suggest that the delay in genomic activation is 'catching up' by oblong stage. Based on observations from as yet unprocessed sphere stage data (Fig. 5.10), it appears that the levels of YSL transcription do indeed match the levels seen in the wider blastoderm, so it is inferred that this 'catching up' occurs before then, and perhaps the oblong stage is when this begins to occur.

This is consistent with the data from looking at cell cycle mitosis synchrony, where the waves of metasynchronous mitosis of YSL nuclei follow the same pattern as the waves of preMBT blastomere nuclei, lagging behind by approximately 1 stage –

where the first YSL division (512-1k) matches the same level of synchrony of the transition from 256-512cell transition.

5.5.3 Maxima of transcriptional activation are higher in blastoderm compared to YSL

The data often show very similar median values for volume and surface area, particularly at 512 cell stage and oblong stage. The closeness in geometries of RNA Pol II regions between YSL and blastoderm nuclei at 512 cell stage is expected as this is the stage the YSL is formed; the contents of the cell drop into the cytoplasm of the yolk cell and it would seem so according to the conducted imaging. This may result in the newly deposited YSL nuclei carrying largely similar properties to their blastomere contemporaries. To support this idea, we see 'exclusion zones' surrounding YSL nuclei which resemble the cytoplasm of blastomeres (Fig. 5.8). This model doesn't explain how the transcriptional activity of the YSL progressively differentiates from the blastoderm into 1000-cell stage, but this observation may offer a potential lead into how YSL nuclei may be partitioned from each other.

Furthermore, as observed from the aggregated volume and surface area data, even when medians are similar, the peak volumes and surface areas of blastoderm RNA Pol II regions tend to reach higher maxima than YSL across embryos, suggesting an overall greater potential for activation in contemporary blastomeres in comparison to YSL.

All in all, we have shown evidence for cell cycle and transcriptional delay in YSL, suggestive of this layer developmentally lagging behind the blastoderm

contemporaries. With that evidence, we hypothesise that zygotic genome activation is delayed in YSL.

5.5.4 Future studies

In order to verify our proposed hypothesis that the observed differences in temporal dynamics of YSL transcription are a delay of zygotic genome activation, further characterisation of the transcription dynamics in this cell layer would have to be conducted. An absolutely key experiment that would provide evidence to support or reject this hypothesis would be a temporal series of global transcription analysis in YSL by means of transcriptomics. For that experiment, efficient isolation of YSL nuclei with high purity for RNA-seq and for nascent RNA-seq (e.g. GRO-seq) (Lopes *et al.*, 2017) would be required. Alternatively, a method of spatially separating cell populations, for example via cryosectioning of embryos could be employed, followed by RNA sequencing (TOMO-seq) (Holler *et al.*, 2021; Liu *et al.*, 2022a). These experiments would allow direct comparison of the transcriptome profiles to that of embryonic blastomeres and temporal sequence of gene activation events compared. Delayed genome activation in YSL would likely be coupled to delayed activation of major wave ZGA genes in the transcriptome data. Alternative to gene expression analysis could be analysis of chromatin opening at promoters using ATAC-seq. This would require nuclear extracts to be generated from YSL nuclei and compared to blastomere nuclei data. Ideally single cell ATAC could be used which would further increase the cellular resolution of the data, but at a cost of reduced sensitivity at genes which are being opened.

Finally, additional *in vivo* transcription imaging could be conducted. Through imaging of *miR-430* morpholino using the MOVIE technique developed in-house by the Müller lab (Hadzhiev *et al.*, 2019), we are able to observe the earliest transcription events in the embryo at 64 cell stage, and *miR-430* has been shown to be indicative of minor wave transcription. However, while this has indeed allowed us to detect transcription in YSL nuclei in preliminary imaging data, the temporal dynamics of miR-430 transcription bodies are unlikely to reveal any additional information relevant to YSL, as the nuclei forming the YSL at 512 cell stage are minor wave nuclei, with the supposed delay occurring at the major wave of ZGA.

In future studies of YSL cell cycle and transcription, we could address whether synchrony presented by YSL during its 3 metasynchronous cell cycles is similar to the synchrony present in blastoderm during the minor wave of ZGA. Cleavage cycles of blastoderm lack gap phases and DNA damage checkpoints, and if YSL replicates minor wave of ZGA behaviour, it would also need to lack those, providing potential avenues of experimentally testing this hypothesis for example by analysis of DNA damage and cell cycle checkpoint dynamics in YSL and blastoderm, or by attempting to detect markers found in blastomeres during synchronous cell divisions, such as replication factors, which were shown to be enriched in blastoderm nuclei during the minor wave (Kermi *et al.*, 2017).

Another potential question raised by our data is whether the delay exhibited by YSL is dependent on the nuclear:cytoplasmic (N:C) ratio. However, as YSL forms a large syncytium without physical barriers between nuclei, manipulation of either of the components of the ratio may prove difficult. Nevertheless, addressing this question experimentally would require manipulations of DNA content in YSL specifically (to

increase the nuclear component) or global manipulations of the N:C ratio such as tetraploidisation of embryos through heat-shock (Hadzhiev *et al.*, 2019).

In summary, our results have indicated that first cell-type specific differences within the zebrafish embryo occur even before zygotic genome activation, and have opened an avenue for potential future research on cell cycle and transcription dynamics in the yolk syncytial layer that could contribute to understanding not only of this extraembryonic layer, but also to understanding events underlying zygotic genome activation, cell fate acquisition and genome reprogramming.

Chapter 6: Summary and Conclusions

The methodologies for disruption of the function of developmental genes in early embryos are fast developing. Zebrafish has played a key role in pioneering adaptation of those methods. Besides the classical approach of production of targeted lesions in stable mutant lines, new methods utilise direct disruption of gene function in directly injected embryos, without the need of generation of a transgenic line. This is due to the relatively high efficiency of these tools, generating mutations in a large proportion of the cells of the developing embryo. While these directly injected embryos provide immensely useful information for target optimisation and for preliminary data generation, the directly injected reagents do not always give the desired effect and require optimisation of multiple parameters. In my studies I have tested several approaches for disruption of mxtx2 gene function including generation of lesions in the coding sequence by CRISPR/Cas9 targeting, disruption of gene function by targeted insertion (knock-in) and disruption of mRNA production by a mRNA targeting version of the CRISPR-associated nuclease (Cas13). For future exploitation of these tools more optimisation is required, for most of the above technologies. Cas9 lesions are routinely used in loss of function studies, but target optimisation would benefit from further improvement of the technology by increasing the number of cells in which targeting takes place. This could be achieved by improving the synthetic Cas9 protein production, optimisation of microinjection conditions to boost the targeting reaction as early as possible during cleavage to reduce mosaicism, and by further optimising the reagents (variant Cas9,

concentration and structural design of guides). The insertional targeting method recently published by (Wierson *et al.*, 2020) did not yield the desired results, and I was unable to generate a stable transgenic line disrupting the *mxtx2* gene using this approach. Discussion with colleagues and members of the zebrafish research community indicate that the insertional approach is not used routinely by most laboratories, suggesting the need for further optimisation of the protocol. Both the Cas9-mediated lesion and targeted insertion approaches may benefit from the application of easily detected phenotypes such as the use of eye pigmentation defects resulting from *gol* targeting, as demonstrated in this thesis.

In this thesis I have aimed to characterise the role of *mxtx2* in zygotic genome activation. This however was not possible due to the issues with loss of function mutation efficiency outlined above. I have provided additional evidence showing that mosaic Cas9 depletion of *mxtx2* phenocopies its morphant phenotype, providing an independent verification of both tools for targeted disruption of this gene. However, to achieve the aim of characterising the role of *mxtx2* in ZGA, a stable transgenic line would need to be generated, which could be later used in transcriptomic analyses, aiming to address if *mxtx2* plays a part in regulating zygotic transcription at early embryonic stages. As demonstrated in this thesis, colleagues and I have detected transcription dynamics differences between the yolk syncytial layer (YSL) and embryo proper (blastoderm). Our hypothesis is that *mxtx2* plays a role in activation of YSL genes and therefore potentially regulates the correct timing and scale of transcription in the YSL. To test this hypothesis, we would need to characterise the formation and transcription dynamics in the YSL in *mxtx2* mutants using the methods described in Chapter 5 of the thesis.

A key further question is to understand the transcriptional mechanism and the molecular reasons for the observed epiboly defects in *mxtx2* crispants. It is predicted that target genes of *mxtx2* include those that are required for the correct morphogenesis of the YSL and its function in regulating epiboly movements (e.g. Nodal pathway). Analysis of the potential role of Mxtx2 target genes in epiboly and YSL morphogenesis may be considered as a future study direction, including loss of function analysis of candidate target genes to verify their role in YSL formation and function. Transcriptomic characterisation of the *mxtx2* mutant with specific focus in the YSL would be a key aim. This is dependent on the development of a robust method for isolating the YSL and separately measuring transcript dynamics from the embryonic blastomeres.

Finally, in this study myself and colleagues have demonstrated a cell cycle and transcriptional delay in YSL, suggestive of this layer developmentally lagging behind the blastoderm contemporaries. With that evidence, we hypothesise that zygotic genome activation is delayed in YSL.

To verify our proposed hypothesis that the observed differences in temporal dynamics of YSL transcription are a delay of zygotic genome activation, further characterisation of the transcription dynamics in this cell layer would have to be conducted, such as a temporal series of global transcription analysis in YSL by means of transcriptomics or analysis of chromatin openness through ATAC-seq. Additional experiments are needed to improve our understanding of events occurring in YSL, including additional live imaging with transcription visualisation, potential experiments manipulating the nuclear:cytoplasmic ratio in YSL, as well as

development of efficient methods of isolating the YSL from blastoderm. All in all, the findings presented in this thesis provide interesting new information on yolk syncytial layer and its transcriptional dynamics, shedding new light on regulation of zygotic genome activation and providing an intriguing avenue for further studies.

Appendix

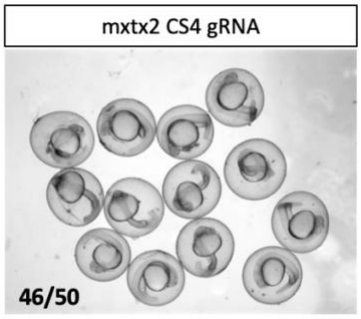


Fig. S1 Additional information to Fig. 3.12.
Embryos injected with only the *mxtx2* targeting CS4 guide RNA, but without Cas9 nuclease did not exhibit the embryonic lethality nor the developmental defects. The micrograph shows a group of those embryos at 24 hpf. Number in the bottom left indicates the number of normally developing embryos from the total scored. Authors note: the remaining 4/50 embryos were found dead at 24hpf with 0/50 abnormally developing.

Fig. S2. Plasmid map of pXR001:Cas13d-EGFP

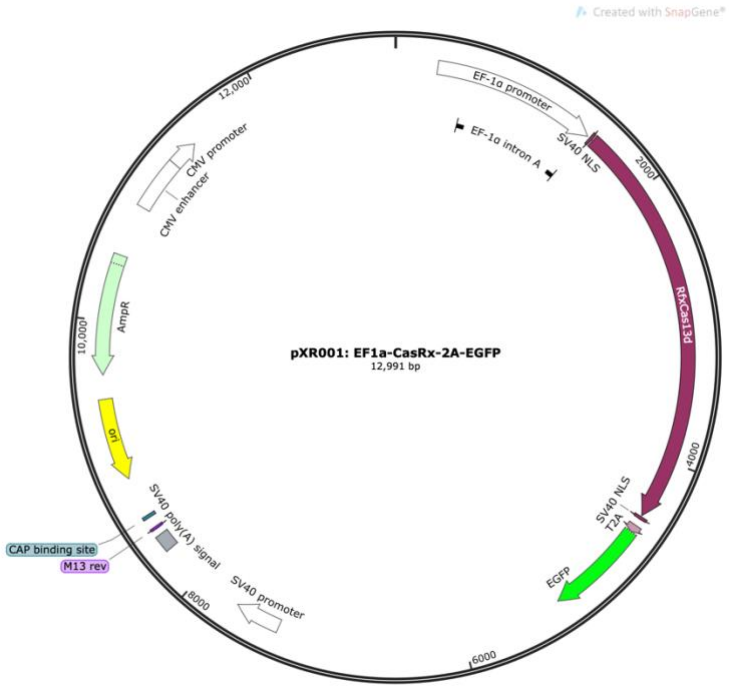


Fig. S3. Plasmid map of pCS2+T7-Cas13d-EGFP

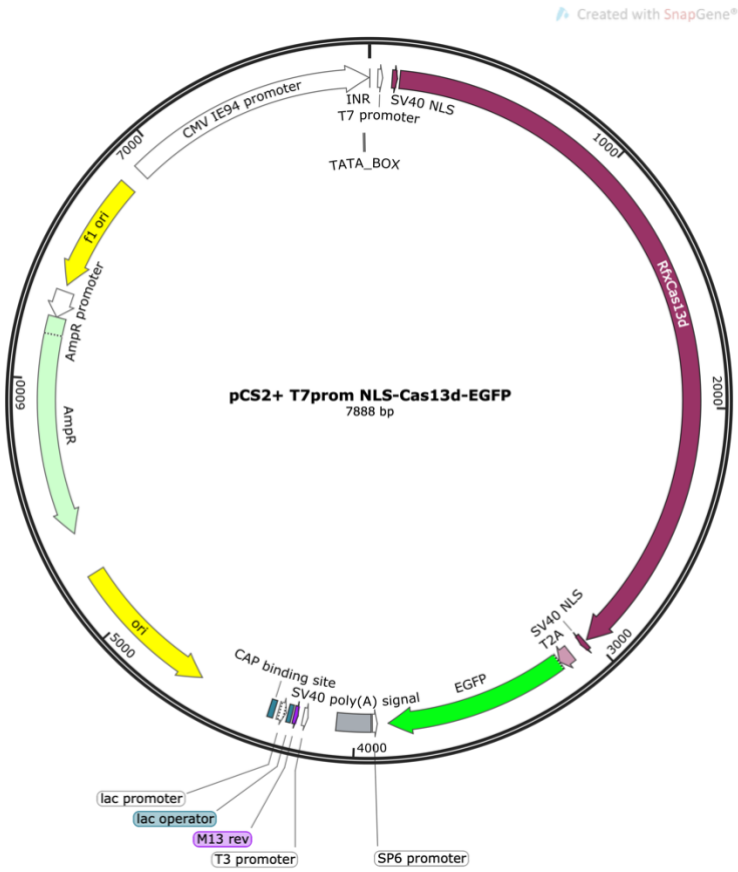


Fig. S4 Cyclopic embryos generated by Cas13d

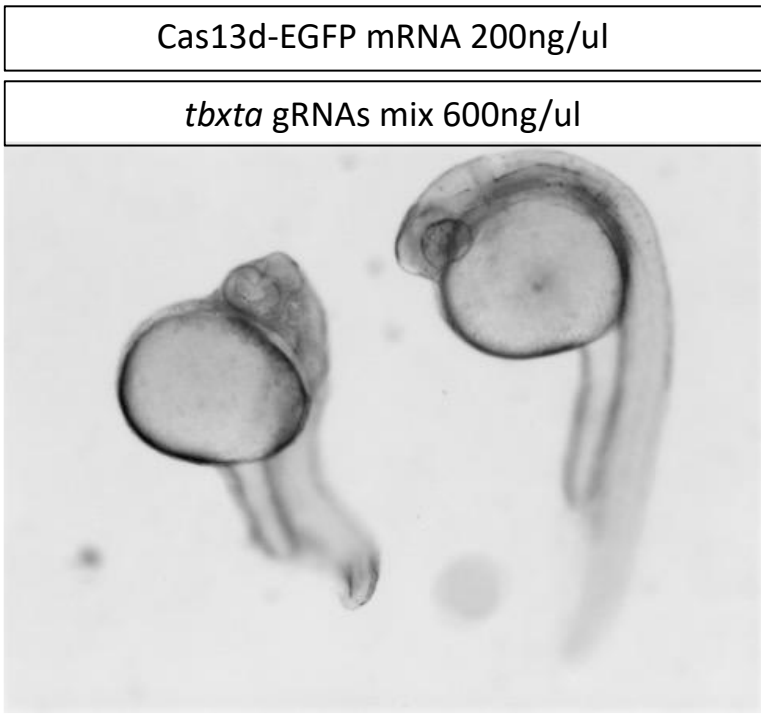
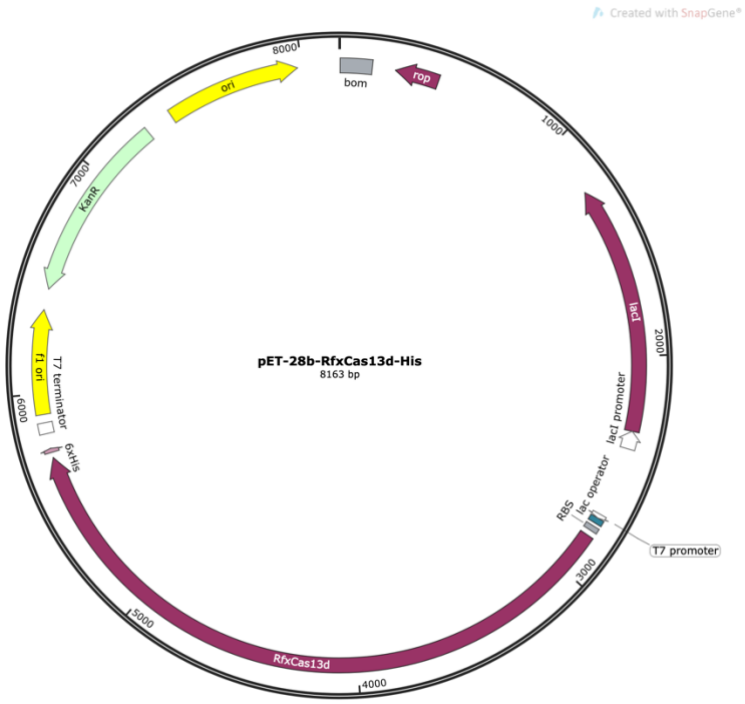


Fig. S5. Plasmid map of pET28b-Cas13d-his



Movie S6 Time-lapse imaging of gol mosaic knockout zebrafish embryo development across epiboly.
For the purposes of easy access, movie was uploaded to the YouTube video sharing platform and is available at: https://youtu.be/ub682kuAWCo
Embryos were injected at 1-cell stage with Cas9 and guide RNA targeted to gol gene. Imaging was started at approximately 3 hours post fertilisation. 135 frames, each taken every 5 minutes are included, with the video sped up to 12 frames per second. Approximate time (in hpf) is shown in the top left corner.

Movie S7 Time-lapse imaging of *mxtx2* mosaic knockout zebrafish embryo development across epiboly

For the purposes of easy access, movie was uploaded to the YouTube video sharing platform and is available at: https://youtu.be/HaIjnGTJiK8

Embryos were injected at 1-cell stage with Cas9 and guide RNA targeted to *mxtx2* gene. Imaging was started at approximately 3 hours post fertilisation. 135 frames, each taken every 5 minutes are included, with the video sped up to 12 frames per second. Approximate time (in hpf) is shown in the top left corner.

Fig. S8 Plasmid maps of pGTAG -EGFP-CAAX-SV40 before (A) and after 48bp homology arms for mxtx2 CS4 gRNA were cloned in (B).

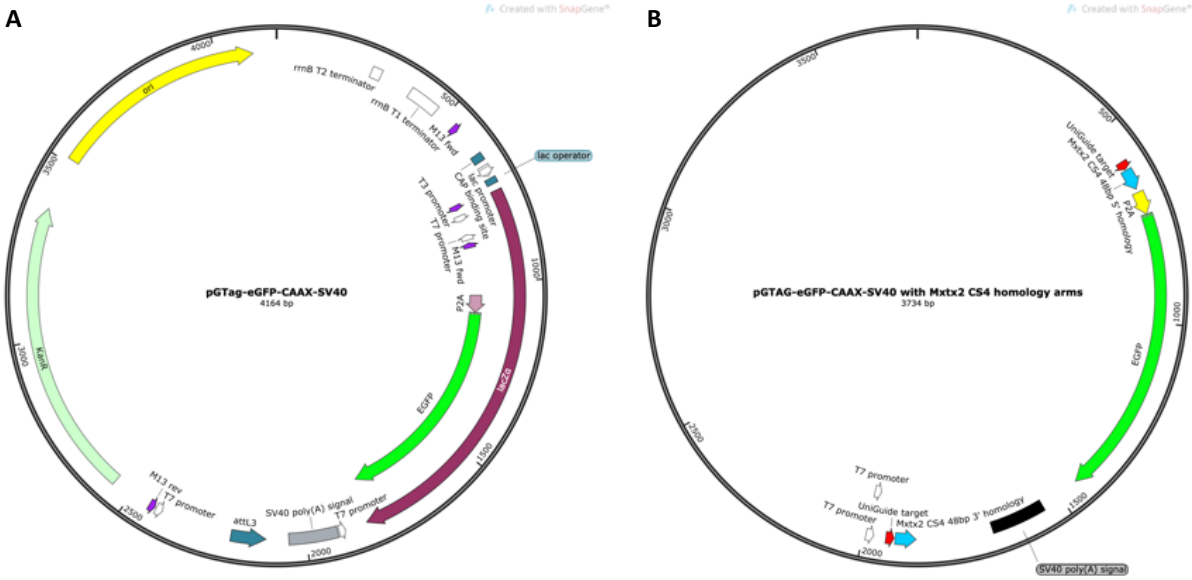


Table S9. Additional data for Fig. 4.9. Summary of screening of F1 embryos, offspring of adult zebrafish injected with *mxtx2* and the associated GeneWeld construct and grown to adulthood.
Across screening rounds, different sets of pairs were crossed to ensure that all of the adult fish were separated and crossed. The same males and females have been crossed at least 2 times to ensure that all of them were screened for the GFP insertion into the *mxtx2* gene. Embryos were screened between dome stage and 75% epiboly for the presence of the GFP positive cells under the fluorescence light microscope.

Screening round 1			
Pair	**Embryos assessed**	**GFP negative**	**GFP positve**
1	106	106	0
2	156	156	0
3	51	51	0
4	121	121	0
5	75	75	0
6	106	106	0
7	34	34	0
8	89	89	0
9	65	65	0
10	178	178	0
11	170	170	0
12	44	44	0
13	62	62	0
14	82	82	0
15	76	76	0
TOTAL	**1415**	**1415**	**0**
Screening round 2			
1	39	39	0
2	170	170	0
3	160	160	0
4	73	73	0
5	231	231	0
6	124	124	0
7	111	111	0
8	75	75	0
9	160	160	0
10	78	78	0
11	163	163	0
12	308	308	0
TOTAL	**1692**	**1692**	**0**

Screening round 3			
1	195	195	0
2	73	73	0
3	38	38	0
4	261	261	0
5	182	182	0
6	81	81	0
7	93	93	0
8	123	123	0
9	110	110	0
10	68	68	0
11	8	8	0
12	31	31	0
13	97	97	0
14	101	101	0
15	4	4	0
16	162	162	0
17	142	142	0
18	59	59	0
19	40	40	0
20	42	42	0
21	4	4	0
22	266	266	0
23	223	223	0
24	73	73	0
25	53	53	0
26	41	41	0
TOTAL	**2570**	**2570**	**0**

Movie S10 – Mitotic events in zebrafish YSL appear coordinated in a wave-like metasynchrony
For the purposes of easy access, movie was uploaded to the YouTube video sharing platform and is available at https://youtu.be/tQavKBo3VIU
Time lapse light sheet imaging of zebrafish embryo nuclei visualised using marking histones with mRuby:H2B protein. Embryos were imaged from 64 cell stage for the duration of 2h and 30 mins, until dome stage. Time-lapse has been generated from a maximum intensity projection of 3D imaging, with framerate of 54 seconds per frame. Nuclear divisions in the YSL are highlighted using white dots.

Primer	Sequence (5'-3')
Mxtx2_FWD_qPCR_1	CAAGAGAGCCAGAACCCTAAAG
Mxtx2_REV_qPCR_1	CCATGTGAGGAGGTAAGAAAGG
Mxtx2_FWD_qPCR_2	TCTCCTGAGTCATGGGATGTAG
Mxtx2_REV_qPCR_2	GGTAAGGAGGTGGTGGATAGA
mxtx1_qPCR_set1_F	TCGCACGTTGAAATGCAAGG
mxtx1_qPCR_set1_R	AGGCGGTGGTGAACATAAAG
ndr2_qPCR_F	TTCATAGCAGGCCCTGGTTG
ndr2_qPCR_R	TGAAAACAGCAGCAGCATGG
actb_qPCR_F	TGACCGAGCGTGGCTGCTACA
actb_qPCR_R	CTTGATGTCACGGACAATTTCTCT
eef1a1a_qPCR_F	GGGCAAGGGCTCCTTCAA
eef1a1a_qPCR_R	CGCTCGGCCTTCAGTTTG

Table S11. Primers used for qPCR experiments.

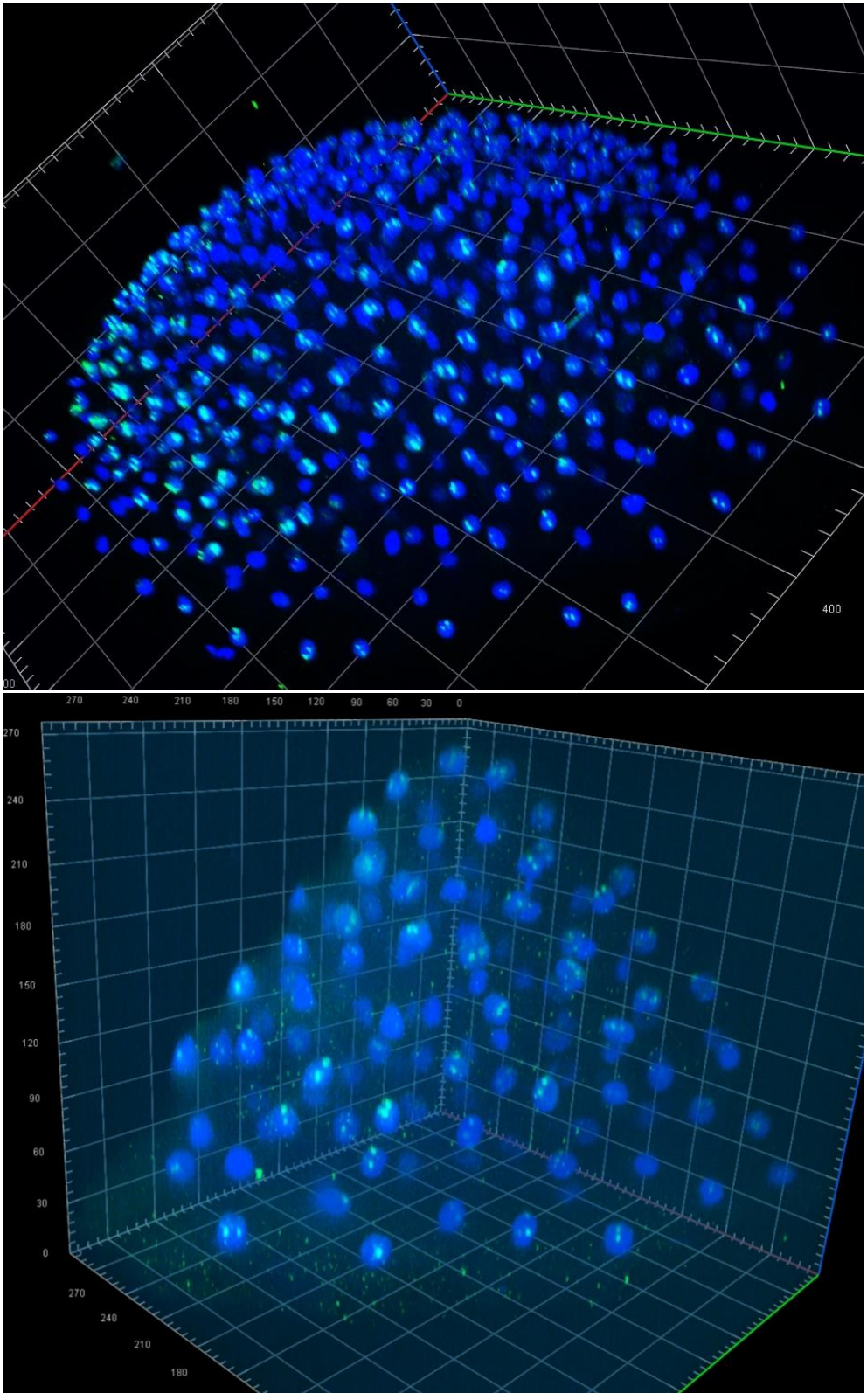


Fig. S12 – additional figures for Fig. 5.8, additional images of RNA Pol II distribution in 512-cell stage embryos.

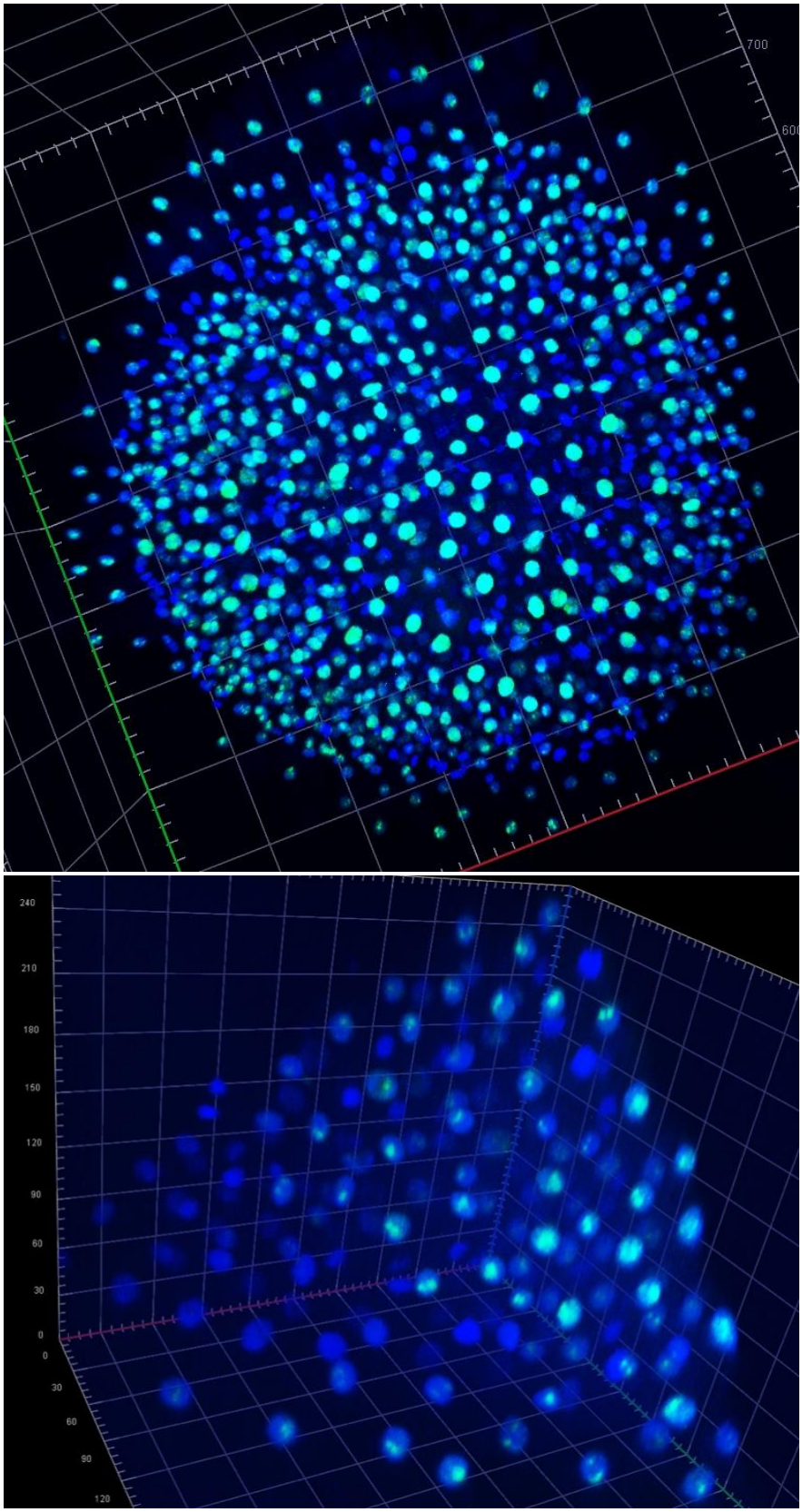


Fig. S13 – additional figures for Fig. 5.8, additional images of RNA Pol II distribution in 1k-cell stage embryos.

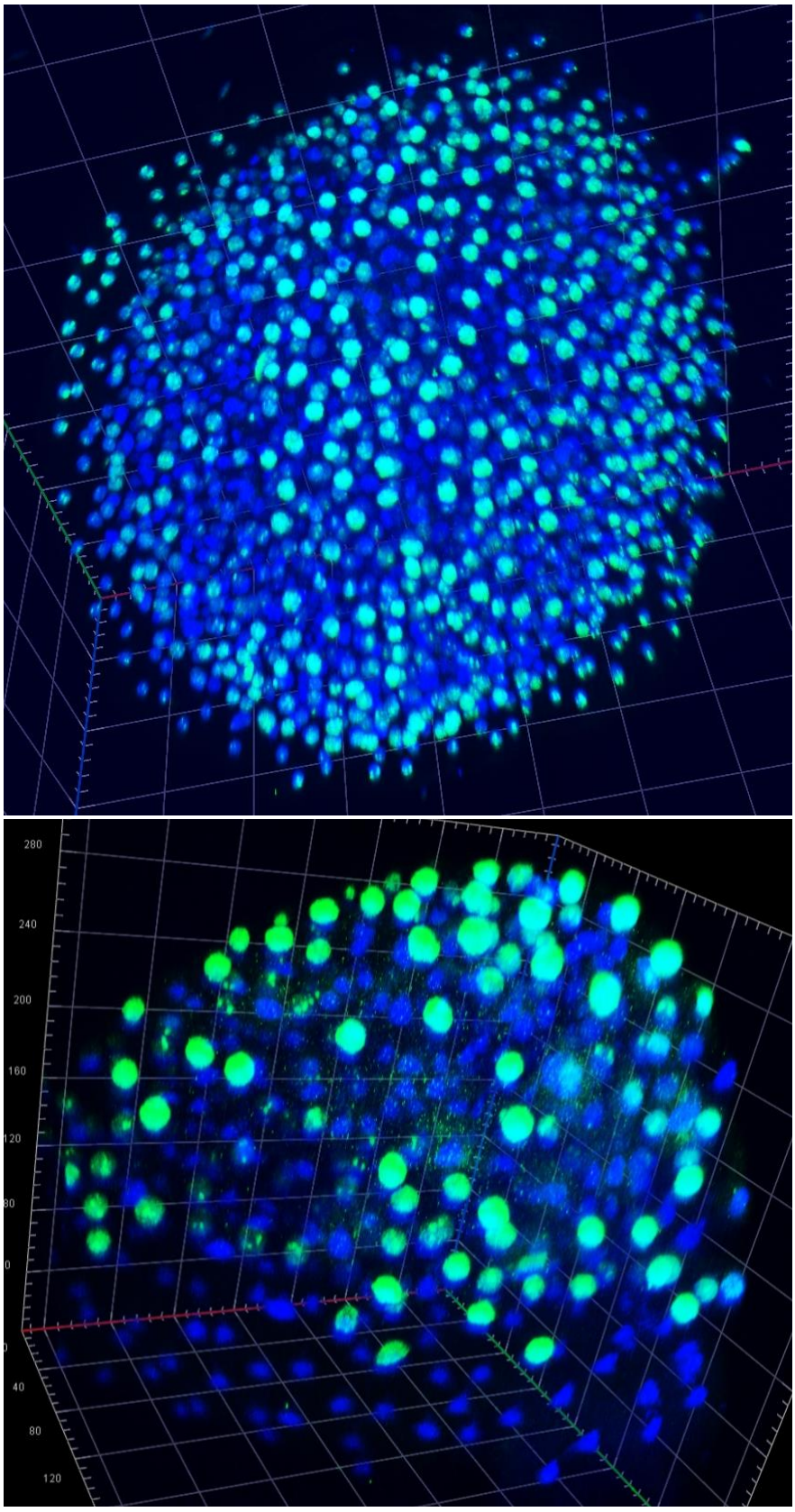


Fig. S14 – additional figures for Fig. 5.8, additional images of RNA Pol II distribution in oblong stage embryos.

Bibliography

Abe, K., Inoue, A., Suzuki, M.G., and Aoki, F. (2010). Global gene silencing is caused by the dissociation of RNA polymerase II from DNA in mouse oocytes. J Reprod Dev *56*, 502-507. 10.1262/jrd.10-068a.

Abe, K.I., Funaya, S., Tsukioka, D., Kawamura, M., Suzuki, Y., Suzuki, M.G., Schultz, R.M., and Aoki, F. (2018). Minor zygotic gene activation is essential for mouse preimplantation development. Proc Natl Acad Sci U S A *115*, E6780-E6788. 10.1073/pnas.1804309115.

Abu-Siniyeh, A., and Al-Zyoud, W. (2020). Highlights on selected microscopy techniques to study zebrafish developmental biology. Lab Anim Res *36*, 12. 10.1186/s42826-020-00044-2.

Abudayyeh, O.O., Gootenberg, J.S., Essletzbichler, P., Han, S., Joung, J., Belanto, J.J., Verdine, V., Cox, D.B.T., Kellner, M.J., Regev, A., *et al.* (2017). RNA targeting with CRISPR–Cas13. Nature *550*, 280-284. 10.1038/nature24049.

Adli, M. (2018). The CRISPR tool kit for genome editing and beyond. Nat Commun *9*, 1911. 10.1038/s41467-018-04252-2.

Akdogan-Ozdilek, B., Duval, K.L., and Goll, M.G. (2020). Chromatin dynamics at the maternal to zygotic transition: recent advances from the zebrafish model. F1000Res *9*. 10.12688/f1000research.21809.1.

Ali-Murthy, Z., Lott, S.E., Eisen, M.B., and Kornberg, T.B. (2013). An essential role for zygotic expression in the pre-cellular Drosophila embryo. PLoS Genet *9*, e1003428. 10.1371/journal.pgen.1003428.

Allen, F., Crepaldi, L., Alsinet, C., Strong, A.J., Kleshchevnikov, V., De Angeli, P., Palenikova, P., Khodak, A., Kiselev, V., Kosicki, M., *et al.* (2018). Predicting the mutations generated by repair of Cas9-induced double-strand breaks. Nat Biotechnol. 10.1038/nbt.4317.

Altschul, S.F., Madden, T.L., Schäffer, A.A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D.J. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res *25*, 3389-3402. 10.1093/nar/25.17.3389.

Amodeo, A.A., Jukam, D., Straight, A.F., and Skotheim, J.M. (2015). Histone titration against the genome sets the DNA-to-cytoplasm threshold for the Xenopus midblastula transition. Proceedings of the National Academy of Sciences *112*, E1086-E1095. doi:10.1073/pnas.1413990112.

Amsterdam, A., Burgess, S., Golling, G., Chen, W., Sun, Z., Townsend, K., Farrington, S., Haldi, M., and Hopkins, N. (1999). A large-scale insertional mutagenesis screen in zebrafish. Genes Dev *13*, 2713-2724. 10.1101/gad.13.20.2713.

Aoki, F. (2022). Zygotic gene activation in mice: profile and regulation. J Reprod Dev *68*, 79-84. 10.1262/jrd.2021-129.

Aoki, F., Hara, K.T., and Schultz, R.M. (2003). Acquisition of transcriptional competence in the 1-cell mouse embryo: requirement for recruitment of maternal mRNAs. Mol Reprod Dev *64*, 270-274. 10.1002/mrd.10227.

Aoki, F., Worrad, D.M., and Schultz, R.M. (1997). Regulation of transcriptional activity during the first and second cell cycles in the preimplantation mouse embryo. Dev Biol *181*, 296-307. 10.1006/dbio.1996.8466.

Ata, H., Ekstrom, T.L., Martínez-Gálvez, G., Mann, C.M., Dvornikov, A.V., Schaefbauer, K.J., Ma, A.C., Dobbs, D., Clark, K.J., and Ekker, S.C. (2018). Robust activation of microy-mediated end joining for precision gene editing applications. PLOS Genetics *14*, e1007652. 10.1371/journal.pgen.1007652.

Bachvarova, R., Davidson, E.H., Allfrey, V.G., and Mirsky, A.E. (1966). Activation of RNA synthesis associated with gastrulation. Proc Natl Acad Sci U S A *55*, 358-365. 10.1073/pnas.55.2.358.

Bazzini, A.A., Lee, M.T., and Giraldez, A.J. (2012). Ribosome profiling shows that miR-430 reduces translation before causing mRNA decay in zebrafish. Science *336*, 233-237. 10.1126/science.1215704.

Bedell, V.M., Westcot, S.E., and Ekker, S.C. (2011). Lessons from morpholino-based screening in zebrafish. Brief Funct Genomics *10*, 181-188. 10.1093/bfgp/elr021.

Bennett, J.T., Joubin, K., Cheng, S., Aanstad, P., Herwig, R., Clark, M., Lehrach, H., and Schier, A.F. (2007). Nodal signaling activates differentiation genes during zebrafish gastrulation. Dev Biol *304*, 525-540. 10.1016/j.ydbio.2007.01.012.

Benoit, B., He, C.H., Zhang, F., Votruba, S.M., Tadros, W., Westwood, J.T., Smibert, C.A., Lipshitz, H.D., and Theurkauf, W.E. (2009). An essential role for the RNA-binding protein Smaug during the Drosophila maternal-to-zygotic transition. Development *136*, 923-932. 10.1242/dev.031815.

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*, 69-77. 10.1089/zeb.2008.0555.

Blitz, I.L., and Cho, K.W.Y. (2021). Control of zygotic genome activation in Xenopus. Curr Top Dev Biol *145*, 167-204. 10.1016/bs.ctdb.2021.03.003.

Blythe, S.A., and Wieschaus, E.F. (2015). Zygotic genome activation triggers the DNA replication checkpoint at the midblastula transition. Cell *160*, 1169-1181. 10.1016/j.cell.2015.01.050.

Bogdanovic, O., Fernandez-Minan, A., Tena, J.J., de la Calle-Mustienes, E., Hidalgo, C., van Kruysbergen, I., van Heeringen, S.J., Veenstra, G.J., and Gomez-Skarmeta, J.L. (2012a). Dynamics of enhancer chromatin signatures mark the transition from pluripotency to cell specification during embryogenesis. Genome Res *22*, 2043-2053. 10.1101/gr.134833.111.

Bogdanovic, O., van Heeringen, S.J., and Veenstra, G.J. (2012b). The epigenome in early vertebrate development. Genesis *50*, 192-206. 10.1002/dvg.20831.

Bonneau, B., Popgeorgiev, N., Prudent, J., and Gillet, G. (2011). Cytoskeleton dynamics in early zebrafish development: A matter of phosphorylation? Bioarchitecture *1*, 216-220. 10.4161/bioa.18116.

Bosnakovski, D., Gearhart, M.D., Ho Choi, S., and Kyba, M. (2021). Dux facilitates post-implantation development, but is not essential for zygotic genome activation. Biol Reprod *104*, 83-93. 10.1093/biolre/ioaa179.

Bouniol, C., Nguyen, E., and Debey, P. (1995). Endogenous transcription occurs at the 1-cell stage in the mouse embryo. Exp Cell Res *218*, 57-62. 10.1006/excr.1995.1130.

Bouniol-Baly, C., Hamraoui, L., Guibert, J., Beaujean, N., Szollosi, M.S., and Debey, P. (1999). Differential transcriptional activity associated with chromatin configuration in fully grown mouse germinal vesicle oocytes. Biol Reprod *60*, 580-587. 10.1095/biolreprod60.3.580.

Boyer, L.A., Lee, T.I., Cole, M.F., Johnstone, S.E., Levine, S.S., Zucker, J.P., Guenther, M.G., Kumar, R.M., Murray, H.L., Jenner, R.G., *et al.* (2005). Core transcriptional regulatory circuitry in human embryonic stem cells. Cell *122*, 947-956. 10.1016/j.cell.2005.08.020.

Bradford, Y.M., Van Slyke, C.E., Ruzicka, L., Singer, A., Eagle, A., Fashena, D., Howe, D.G., Frazer, K., Martin, R., Paddock, H., *et al.* (2022). Zebrafish information network, the knowledgebase for Danio rerio research. Genetics *220*. 10.1093/genetics/iyac016.

Braude, P., Bolton, V., and Moore, S. (1988). Human gene expression first occurs between the four- and eight-cell stages of preimplantation development. Nature *332*, 459-461. 10.1038/332459a0.

Bruce, A.E.E., and Heisenberg, C.P. (2020). Mechanisms of zebrafish epiboly: A current view. Curr Top Dev Biol *136*, 319-341. 10.1016/bs.ctdb.2019.07.001.

Bruce, A.E.E., Howley, C., Dixon Fox, M., and Ho, R.K. (2005). T-box gene eomesodermin and the homeobox-containing Mix/Bix gene mtx2 regulate epiboly movements in the zebrafish. Developmental Dynamics *233*, 105-114. https://doi.org/10.1002/dvdy.20305.

Burger, A., Lindsay, H., Felker, A., Hess, C., Anders, C., Chiavacci, E., Zaugg, J., Weber, L.M., Catena, R., Jinek, M., *et al.* (2016). Maximizing mutagenesis with solubilized CRISPR-Cas9 ribonucleoprotein complexes. Development *143*, 2025-2037. 10.1242/dev.134809.

Bushati, N., Stark, A., Brennecke, J., and Cohen, S.M. (2008). Temporal reciprocity of miRNAs and their targets during the maternal-to-zygotic transition in Drosophila. Curr Biol *18*, 501-506. 10.1016/j.cub.2008.02.081.

Campbell, P.D., Chao, J.A., Singer, R.H., and Marlow, F.L. (2015). Dynamic visualization of transcription and RNA subcellular localization in zebrafish. Development *142*, 1368-1374. 10.1242/dev.118968.

Carvalho, L., and Heisenberg, C.P. (2010). The yolk syncytial layer in early zebrafish development. Trends Cell Biol *20*, 586-592. 10.1016/j.tcb.2010.06.009.

Chan, S.H., Tang, Y., Miao, L., Darwich-Codore, H., Vejnar, C.E., Beaudoin, J.D., Musaev, D., Fernandez, J.P., Benitez, M.D.J., Bazzini, A.A., *et al.* (2019). Brd4 and P300 Confer Transcriptional Competency during Zygotic Genome Activation. Dev Cell *49*, 867-881 e868. 10.1016/j.devcel.2019.05.037.

Chen, H., Einstein, L.C., Little, S.C., and Good, M.C. (2019). Spatiotemporal Patterning of Zygotic Genome Activation in a Model Vertebrate Embryo. Dev Cell *49*, 852-866 e857. 10.1016/j.devcel.2019.05.036.

Chen, H., and Good, M.C. (2020). Imaging nascent transcription in wholemount vertebrate embryos to characterize zygotic genome activation. Methods Enzymol *638*, 139-165. 10.1016/bs.mie.2020.03.002.

Chen, K., Johnston, J., Shao, W., Meier, S., Staber, C., and Zeitlinger, J. (2013). A global change in RNA polymerase II pausing during the Drosophila midblastula transition. Elife *2*, e00861. 10.7554/eLife.00861.

Chen, Z., Xie, Z., and Zhang, Y. (2021). DPPA2 and DPPA4 are dispensable for mouse zygotic genome activation and pre-implantation development. Development *148*. 10.1242/dev.200178.

Chen, Z., and Zhang, Y. (2019). Loss of DUX causes minor defects in zygotic genome activation and is compatible with mouse development. Nat Genet *51*, 947-951. 10.1038/s41588-019-0418-7.

Chu, L.-T., Fong, S.H., Kondrychyn, I., Loh, S.L., Ye, Z., and Korzh, V. (2012). Yolk syncytial layer formation is a failure of cytokinesis mediated by Rock1 function in the early zebrafish embryo. Biology Open *1*, 747-753. 10.1242/bio.20121636.

Collart, C., Allen, G.E., Bradshaw, C.R., Smith, J.C., and Zegerman, P. (2013). Titration of four replication factors is essential for the Xenopus laevis midblastula transition. Science *341*, 893-896. 10.1126/science.1241530.

Collart, C., Owens, N.D., Bhaw-Rosun, L., Cooper, B., De Domenico, E., Patrushev, I., Sesay, A.K., Smith, J.N., Smith, J.C., and Gilchrist, M.J. (2014). High-resolution analysis of gene activity during the Xenopus mid-blastula transition. Development *141*, 1927-1939. 10.1242/dev.102012.

Conesa, A., Madrigal, P., Tarazona, S., Gomez-Cabrero, D., Cervera, A., McPherson, A., Szcześniak, M.W., Gaffney, D.J., Elo, L.L., Zhang, X., and Mortazavi, A. (2016). A survey of best practices for RNA-seq data analysis. Genome Biology *17*, 13. 10.1186/s13059-016-0881-8.

Conlon, F.L., Lyons, K.M., Takaesu, N., Barth, K.S., Kispert, A., Herrmann, B., and Robertson, E.J. (1994). A primary requirement for nodal in the formation and maintenance of the primitive streak in the mouse. Development *120*, 1919-1928. 10.1242/dev.120.7.1919.

Corey, D.R., and Abrams, J.M. (2001). Morpholino antisense oligonucleotides: tools for investigating vertebrate development. Genome Biology *2*, reviews1015.1011. 10.1186/gb-2001-2-5-reviews1015.

Cox, D.B.T., Gootenberg, J.S., Abudayyeh, O.O., Franklin, B., Kellner, M.J., Joung, J., and Zhang, F. (2017). RNA editing with CRISPR-Cas13. Science *358*, 1019-1027. doi:10.1126/science.aaq0180.

Creyghton, M.P., Cheng, A.W., Welstead, G.G., Kooistra, T., Carey, B.W., Steine, E.J., Hanna, J., Lodato, M.A., Frampton, G.M., Sharp, P.A., *et al.* (2010). Histone H3K27ac separates active from poised enhancers and predicts developmental state. Proc Natl Acad Sci U S A *107*, 21931-21936. 10.1073/pnas.1016071107.

Dahl, J.A., Jung, I., Aanes, H., Greggains, G.D., Manaf, A., Lerdrup, M., Li, G., Kuan, S., Li, B., Lee, A.Y., *et al.* (2016). Broad histone H3K4me3 domains in mouse oocytes modulate maternal-to-zygotic transition. Nature *537*, 548-552. 10.1038/nature19360.

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*, e1002861. 10.1371/journal.pgen.1002861.

Darbo, E., Herrmann, C., Lecuit, T., Thieffry, D., and van Helden, J. (2013). Transcriptional and epigenetic signatures of zygotic genome activation during early Drosophila embryogenesis. BMC Genomics *14*, 226. 10.1186/1471-2164-14-226.

de Chaumont, F., Dallongeville, S., Chenouard, N., Herve, N., Pop, S., Provoost, T., Meas-Yedid, V., Pankajakshan, P., Lecomte, T., Le Montagner, Y., *et al.* (2012). Icy: an open bioimage informatics platform for extended reproducible research. Nat Methods *9*, 690-696. 10.1038/nmeth.2075.

De Iaco, A., Planet, E., Coluccio, A., Verp, S., Duc, J., and Trono, D. (2017). DUX-family transcription factors regulate zygotic genome activation in placental mammals. Nat Genet *49*, 941-945. 10.1038/ng.3858.

De Iaco, A., Verp, S., Offner, S., Grun, D., and Trono, D. (2020). DUX is a non-essential synchronizer of zygotic genome activation. Development *147*. 10.1242/dev.177725.

De La Fuente, R., Viveiros, M.M., Burns, K.H., Adashi, E.Y., Matzuk, M.M., and Eppig, J.J. (2004). Major chromatin remodeling in the germinal vesicle (GV) of mammalian oocytes is dispensable for global transcriptional silencing but required for centromeric heterochromatin function. Dev Biol *275*, 447-458. 10.1016/j.ydbio.2004.08.028.

De Renzis, S., Elemento, O., Tavazoie, S., and Wieschaus, E.F. (2007). Unmasking activation of the zygotic genome using chromosomal deletions in the Drosophila embryo. PLoS Biol *5*, e117. 10.1371/journal.pbio.0050117.

Desmarais, J.A., Hoffmann, M.J., Bingham, G., Gagou, M.E., Meuth, M., and Andrews, P.W. (2012). Human embryonic stem cells fail to activate CHK1 and commit to apoptosis in response to DNA replication stress. Stem Cells *30*, 1385-1393. 10.1002/stem.1117.

Du, S., Draper, B.W., Mione, M., Moens, C.B., and Bruce, A. (2012). Differential regulation of epiboly initiation and progression by zebrafish Eomesodermin A. Developmental Biology *362*, 11-23. https://doi.org/10.1016/j.ydbio.2011.10.036.

Dufourt, J., Trullo, A., Hunter, J., Fernandez, C., Lazaro, J., Dejean, M., Morales, L., Nait-Amer, S., Schulz, K.N., Harrison, M.M., *et al.* (2018). Temporal control of gene expression by the pioneer factor Zelda through transient interactions in hubs. Nat Commun *9*, 5194. 10.1038/s41467-018-07613-z.

Eckersley-Maslin, M., Alda-Catalinas, C., Blotenburg, M., Kreibich, E., Krueger, C., and Reik, W. (2019). Dppa2 and Dppa4 directly regulate the Dux-driven zygotic transcriptional program. Genes Dev *33*, 194-208. 10.1101/gad.321174.118.

Ekker, S.C. (2008). Zinc finger-based knockout punches for zebrafish genes. Zebrafish *5*, 121-123. 10.1089/zeb.2008.9988.

El-Brolosy, M.A., Kontarakis, Z., Rossi, A., Kuenne, C., Günther, S., Fukuda, N., Kikhi, K., Boezio, G.L.M., Takacs, C.M., Lai, S.L., *et al.* (2019). Genetic compensation triggered by mutant mRNA degradation. Nature *568*, 193-197. 10.1038/s41586-019-1064-z.

El-Brolosy, M.A., and Stainier, D.Y.R. (2017). Genetic compensation: A phenomenon in search of mechanisms. PLoS Genet *13*, e1006780. 10.1371/journal.pgen.1006780.

Ferg, M., Sanges, R., Gehrig, J., Kiss, J., Bauer, M., Lovas, A., Szabo, M., Yang, L., Straehle, U., Pankratz, M.J., *et al.* (2007). The TATA-binding protein regulates maternal mRNA degradation and differential zygotic transcription in zebrafish. EMBO J *26*, 3945-3956. 10.1038/sj.emboj.7601821.

Flach, G., Johnson, M.H., Braude, P.R., Taylor, R.A., and Bolton, V.N. (1982). The transition from maternal to embryonic control in the 2-cell mouse embryo. EMBO J *1*, 681-686. 10.1002/j.1460-2075.1982.tb01230.x.

Fogarty, P., Kalpin, R.F., and Sullivan, W. (1994). The Drosophila maternal-effect mutation grapes causes a metaphase arrest at nuclear cycle 13. Development *120*, 2131-2142. 10.1242/dev.120.8.2131.

Fu, S., Nien, C.Y., Liang, H.L., and Rushlow, C. (2014). Co-activation of microRNAs by Zelda is essential for early Drosophila development. Development *141*, 2108-2118. 10.1242/dev.108118.

Funaya, S., and Aoki, F. (2017). Regulation of zygotic gene activation by chromatin structure and epigenetic factors. J Reprod Dev *63*, 359-363. 10.1262/jrd.2017-058.

Gagnon, J.A., Obbad, K., and Schier, A.F. (2018). The primary role of zebrafish nanog is in extra-embryonic tissue. Development *145*. 10.1242/dev.147793.

Gagnon, J.A., Valen, E., Thyme, S.B., Huang, P., Ahkmetova, L., Pauli, A., Montague, T.G., Zimmerman, S., Richter, C., and Schier, A.F. (2014). Efficient Mutagenesis by Cas9 Protein-Mediated Oligonucleotide Insertion and Large-Scale Assessment of Single-Guide RNAs. PLOS ONE *9*, e98186. 10.1371/journal.pone.0098186.

Gao, L., Wu, K., Liu, Z., Yao, X., Yuan, S., Tao, W., Yi, L., Yu, G., Hou, Z., Fan, D., *et al.* (2018). Chromatin Accessibility Landscape in Human Early Embryos and Its Association with Evolution. Cell *173*, 248-259 e215. 10.1016/j.cell.2018.02.028.

Gao, M., Veil, M., Rosenblatt, M., Riesle, A.J., Gebhard, A., Hass, H., Buryanova, L., Yampolsky, L.Y., Grüning, B., Ulianov, S.V., *et al.* (2022). Pluripotency factors

determine gene expression repertoire at zygotic genome activation. Nature Communications *13*, 788. 10.1038/s41467-022-28434-1.

Gao, T., Zheng, J., Xing, F., Fang, H., Sun, F., Yan, A., Gong, X., Ding, H., Tang, F., and Sheng, H.Z. (2007). Nuclear reprogramming: the strategy used in normal development is also used in somatic cell nuclear transfer and parthenogenesis. Cell Res *17*, 135-150. 10.1038/cr.2007.2.

Gassler, J., Kobayashi, W., Gáspár, I., Ruangroengkulrith, S., Kümmecke, M., Kravchenko, P., Zaczek, M., Vallot, A., Hernandez, L.G., Rico, L.C., *et al.* (2022). Zygotic genome activation by the totipotency pioneer factor Nr5a2. bioRxiv, 2022.2005.2017.492379. 10.1101/2022.05.17.492379.

Gentsch, G.E., Owens, N.D.L., and Smith, J.C. (2019a). The Spatiotemporal Control of Zygotic Genome Activation. iScience *16*, 485-498. 10.1016/j.isci.2019.06.013.

Gentsch, G.E., Spruce, T., Owens, N.D.L., and Smith, J.C. (2019b). Maternal pluripotency factors initiate extensive chromatin remodelling to predefine first response to inductive signals. Nat Commun *10*, 4269. 10.1038/s41467-019-12263-w.

Gerety, S.S., and Wilkinson, D.G. (2011). Morpholino artifacts provide pitfalls and reveal a novel role for pro-apoptotic genes in hindbrain boundary development. Dev Biol *350*, 279-289. 10.1016/j.ydbio.2010.11.030.

Giacomotto, J., Rinkwitz, S., and Becker, T.S. (2015). Effective heritable gene knockdown in zebrafish using synthetic microRNAs. Nature Communications *6*, 7378. 10.1038/ncomms8378.

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*, 833-838. 10.1126/science.1109020.

Giraldez, A.J., Mishima, Y., Rihel, J., Grocock, R.J., Van Dongen, S., Inoue, K., Enright, A.J., and Schier, A.F. (2006). Zebrafish MiR-430 promotes deadenylation and clearance of maternal mRNAs. Science *312*, 75-79. 10.1126/science.1122689.

Grunwald, D.J., Kimmel, C.B., Westerfield, M., Walker, C., and Streisinger, G. (1988). A neural degeneration mutation that spares primary neurons in the zebrafish. Dev Biol *126*, 115-128. 10.1016/0012-1606(88)90245-x.

Guo, H., Zhu, P., Yan, L., Li, R., Hu, B., Lian, Y., Yan, J., Ren, X., Lin, S., Li, J., *et al.* (2014). The DNA methylation landscape of human early embryos. Nature *511*, 606-610. 10.1038/nature13544.

Gurdon, J.B., Elsdale, T.R., and Fischberg, M. (1958). Sexually mature individuals of Xenopus laevis from the transplantation of single somatic nuclei. Nature *182*, 64-65. 10.1038/182064a0.

Haberle, V., Arnold, C.D., Pagani, M., Rath, M., Schernhuber, K., and Stark, A. (2019). Transcriptional cofactors display specificity for distinct types of core promoters. Nature *570*, 122-126. 10.1038/s41586-019-1210-7.

Haberle, V., Li, N., Hadzhiev, Y., Plessy, C., Previti, C., Nepal, C., Gehrig, J., Dong, X., Akalin, A., Suzuki, A.M., *et al.* (2014a). Two independent transcription initiation codes overlap on vertebrate core promoters. Nature *507*, 381-385. 10.1038/nature12974.

Haberle, V., Li, N., Hadzhiev, Y., Plessy, C., Previti, C., Nepal, C., Gehrig, J., Dong, X., Akalin, A., Suzuki, A.M., *et al.* (2014b). Two independent transcription initiation codes overlap on vertebrate core promoters. Nature *507*, 381-385. 10.1038/nature12974.

Hadzhiev, Y., Qureshi, H.K., Wheatley, L., Cooper, L., Jasiulewicz, A., Van Nguyen, H., Wragg, J.W., Poovathumkadavil, D., Conic, S., Bajan, S., *et al.* (2019). A cell cycle-coordinated Polymerase II transcription compartment encompasses gene expression before global genome activation. Nature Communications *10*, 691. 10.1038/s41467-019-08487-5.

Hadzhiev, Y., Wheatley, L., Cooper, L., Ansaloni, F., Whalley, C., Chen, Z., Gustincich, S., Sanges, R., Burgess, S., Beggs, A., and Müller, F. (2021). The miR-430 locus with extreme promoter density is a transcription body organizer, which facilitates long range regulation in zygotic genome activation. bioRxiv, 2021.2008.2009.455629. 10.1101/2021.08.09.455629.

Harrison, M.M., Botchan, M.R., and Cline, T.W. (2010). Grainyhead and Zelda compete for binding to the promoters of the earliest-expressed Drosophila genes. Dev Biol *345*, 248-255. 10.1016/j.ydbio.2010.06.026.

Harrison, M.M., Li, X.Y., Kaplan, T., Botchan, M.R., and Eisen, M.B. (2011). Zelda binding in the early Drosophila melanogaster embryo marks regions subsequently activated at the maternal-to-zygotic transition. PLoS Genet *7*, e1002266. 10.1371/journal.pgen.1002266.

Hendrickson, P.G., Dorais, J.A., Grow, E.J., Whiddon, J.L., Lim, J.W., Wike, C.L., Weaver, B.D., Pflueger, C., Emery, B.R., Wilcox, A.L., *et al.* (2017). Conserved roles of mouse DUX and human DUX4 in activating cleavage-stage genes and MERVL/HERVL retrotransposons. Nat Genet *49*, 925-934. 10.1038/ng.3844.

Hernandez-Huertas, L., Kushawah, G., Diaz-Moscoso, A., Tomas-Gallardo, L., Moreno-Sanchez, I., da Silva Pescador, G., Bazzini, A.A., and Moreno-Mateos, M.A.

(2022). Optimized CRISPR-RfxCas13d system for RNA targeting in zebrafish embryos. STAR Protoc *3*, 101058. 10.1016/j.xpro.2021.101058.

Hernández-Vega, A., Marsal, M., Pouille, P.-A., Tosi, S., Colombelli, J., Luque, T., Navajas, D., Pagonabarraga, I., and Martín-Blanco, E. (2017). Polarized cortical tension drives zebrafish epiboly movements. The EMBO Journal *36*, 25-41. https://doi.org/10.15252/embj.201694264.

Heyn, P., Kircher, M., Dahl, A., Kelso, J., Tomancak, P., Kalinka, A.T., and Neugebauer, K.M. (2014). The earliest transcribed zygotic genes are short, newly evolved, and different across species. Cell Rep *6*, 285-292. 10.1016/j.celrep.2013.12.030.

Hilbert, L., Sato, Y., Kuznetsova, K., Bianucci, T., Kimura, H., Julicher, F., Honigmann, A., Zaburdaev, V., and Vastenhouw, N.L. (2021). Transcription organizes euchromatin via microphase separation. Nat Commun *12*, 1360. 10.1038/s41467-021-21589-3.

Hill, C.S. (2018). Spatial and temporal control of NODAL signaling. Current Opinion in Cell Biology *51*, 50-57. https://doi.org/10.1016/j.ceb.2017.10.005.

Holler, K., Neuschulz, A., Drewe-Boss, P., Mintcheva, J., Spanjaard, B., Arsie, R., Ohler, U., Landthaler, M., and Junker, J.P. (2021). Spatio-temporal mRNA tracking in the early zebrafish embryo. Nat Commun *12*, 3358. 10.1038/s41467-021-23834-1.

Hong, S.K., Jang, M.K., Brown, J.L., McBride, A.A., and Feldman, B. (2011). Embryonic mesoderm and endoderm induction requires the actions of non-embryonic Nodal-related ligands and Mxtx2. Development *138*, 787-795. 10.1242/dev.058974.

Hoshijima, K., Jurynec, Michael J., and Grunwald, David J. (2016). Precise Editing of the Zebrafish Genome Made Simple and Efficient. Developmental Cell *36*, 654-667. 10.1016/j.devcel.2016.02.015.

Hoshijima, K., Jurynec, M.J., Klatt Shaw, D., Jacobi, A.M., Behlke, M.A., and Grunwald, D.J. (2019). Highly Efficient CRISPR-Cas9-Based Methods for Generating Deletion Mutations and F0 Embryos that Lack Gene Function in Zebrafish. Dev Cell *51*, 645-657 e644. 10.1016/j.devcel.2019.10.004.

Howe, K., Clark, M.D., Torroja, C.F., Torrance, J., Berthelot, C., Muffato, M., Collins, J.E., Humphray, S., McLaren, K., Matthews, L., *et al.* (2013). The zebrafish reference genome sequence and its relationship to the human genome. Nature *496*, 498-503. 10.1038/nature12111.

Hruscha, A., and Schmid, B. (2015). Generation of zebrafish models by CRISPR /Cas9 genome editing. Methods Mol Biol *1254*, 341-350. 10.1007/978-1-4939-2152-2_24.

Huynh, N., Depner, N., Larson, R., and King-Jones, K. (2020). A versatile toolkit for CRISPR-Cas13-based RNA manipulation in Drosophila. Genome Biology *21*, 279. 10.1186/s13059-020-02193-y.

Hwang, W.Y., Fu, Y., Reyon, D., Maeder, M.L., Tsai, S.Q., Sander, J.D., Peterson, R.T., Yeh, J.R.J., and Joung, J.K. (2013). Efficient genome editing in zebrafish using a CRISPR-Cas system. Nature Biotechnology *31*, 227-229. 10.1038/nbt.2501.

Hwang, W.Y., Peterson, R.T., and Yeh, J.R. (2014). Methods for targeted mutagenesis in zebrafish using TALENs. Methods *69*, 76-84. 10.1016/j.ymeth.2014.04.009.

Jacobi, A.M., Rettig, G.R., Turk, R., Collingwood, M.A., Zeiner, S.A., Quadros, R.M., Harms, D.W., Bonthuis, P.J., Gregg, C., Ohtsuka, M., *et al.* (2017). Simplified CRISPR tools for efficient genome editing and streamlined protocols for their delivery into mammalian cells and mouse zygotes. Methods *121-122*, 16-28. https://doi.org/10.1016/j.ymeth.2017.03.021.

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 *110*, 13904-13909. doi:10.1073/pnas.1308335110.

Jiang, M., Xiao, Y., E, W., Ma, L., Wang, J., Chen, H., Gao, C., Liao, Y., Guo, Q., Peng, J., *et al.* (2021). Characterization of the Zebrafish Cell Landscape at Single-Cell Resolution. Frontiers in Cell and Developmental Biology *9*. 10.3389/fcell.2021.743421.

Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J.A., and Charpentier, E. (2012). A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. Science *337*, 816-821. 10.1126/science.1225829.

Joseph, S.R., Palfy, M., Hilbert, L., Kumar, M., Karschau, J., Zaburdaev, V., Shevchenko, A., and Vastenhouw, N.L. (2017). Competition between histone and transcription factor binding regulates the onset of transcription in zebrafish embryos. Elife *6*. 10.7554/eLife.23326.

Jukam, D., Shariati, S.A.M., and Skotheim, J.M. (2017). Zygotic Genome Activation in Vertebrates. Dev Cell *42*, 316-332. 10.1016/j.devcel.2017.07.026.

Kaaij, L.J.T., van der Weide, R.H., Ketting, R.F., and de Wit, E. (2018). Systemic Loss and Gain of Chromatin Architecture throughout Zebrafish Development. Cell Rep *24*, 1-10 e14. 10.1016/j.celrep.2018.06.003.

Kane, D.A., and Kimmel, C.B. (1993). The zebrafish midblastula transition. Development *119*, 447-456. 10.1242/dev.119.2.447.

Kane, D.A., Maischein, H.M., Brand, M., van Eeden, F.J., Furutani-Seiki, M., Granato, M., Haffter, P., Hammerschmidt, M., Heisenberg, C.P., Jiang, Y.J., *et al.* (1996). The zebrafish early arrest mutants. Development *123*, 57-66. 10.1242/dev.123.1.57.

Keller, P.J. (2013). In vivo imaging of zebrafish embryogenesis. Methods *62*, 268-278. https://doi.org/10.1016/j.ymeth.2013.03.015.

Kermi, C., Lo Furno, E., and Maiorano, D. (2017). Regulation of DNA Replication in Early Embryonic Cleavages. Genes (Basel) *8*. 10.3390/genes8010042.

Kim, V.N., Han, J., and Siomi, M.C. (2009). Biogenesis of small RNAs in animals. Nat Rev Mol Cell Biol *10*, 126-139. 10.1038/nrm2632.

Kimelman, D., Kirschner, M., and Scherson, T. (1987). The events of the midblastula transition in Xenopus are regulated by changes in the cell cycle. Cell *48*, 399-407. 10.1016/0092-8674(87)90191-7.

Kimmel, C.B., Ballard, W.W., Kimmel, S.R., Ullmann, B., and Schilling, T.F. (1995). Stages of embryonic development of the zebrafish. Dev Dyn *203*, 253-310. 10.1002/aja.1002030302.

Kimmel, C.B., Kane, D.A., Walker, C., Warga, R.M., and Rothman, M.B. (1989). A mutation that changes cell movement and cell fate in the zebrafish embryo. Nature *337*, 358-362. 10.1038/337358a0.

Kimmel, C.B., and Law, R.D. (1985a). Cell lineage of zebrafish blastomeres: I. Cleavage pattern and cytoplasmic bridges between cells. Developmental Biology *108*, 78-85. https://doi.org/10.1016/0012-1606(85)90010-7.

Kimmel, C.B., and Law, R.D. (1985b). Cell lineage of zebrafish blastomeres: II. Formation of the yolk syncytial layer. Developmental Biology *108*, 86-93. https://doi.org/10.1016/0012-1606(85)90011-9.

Ko, S.-K., Chen, X., Yoon, J., and Shin, I. (2011). Zebrafish as a good vertebrate model for molecular imaging using fluorescent probes. Chemical Society Reviews *40*, 2120-2130. 10.1039/C0CS00118J.

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., *et al.* (2015). Reverse Genetic Screening Reveals Poor Correlation between Morpholino-Induced and Mutant Phenotypes in Zebrafish. Developmental Cell *32*, 97-108. https://doi.org/10.1016/j.devcel.2014.11.018.

Konermann, S., Lotfy, P., Brideau, N.J., Oki, J., Shokhirev, M.N., and Hsu, P.D. (2018). Transcriptome Engineering with RNA-Targeting Type VI-D CRISPR Effectors. Cell *173*, 665-676.e614. https://doi.org/10.1016/j.cell.2018.02.033.

Koster, R.W., and Fraser, S.E. (2001). Tracing transgene expression in living zebrafish embryos. Dev Biol *233*, 329-346. 10.1006/dbio.2001.0242.

Kosuta, C., Daniel, K., Johnstone, D.L., Mongeon, K., Ban, K., LeBlanc, S., MacLeod, S., Et-Tahiry, K., Ekker, M., MacKenzie, A., and Pena, I. (2018). High-throughput DNA Extraction and Genotyping of 3dpf Zebrafish Larvae by Fin Clipping. J Vis Exp. 10.3791/58024.

Kroll, F., Powell, G.T., Ghosh, M., Gestri, G., Antinucci, P., Hearn, T.J., Tunbak, H., Lim, S., Dennis, H.W., Fernandez, J.M., *et al.* (2021). A simple and effective F0 knockout method for rapid screening of behaviour and other complex phenotypes. eLife *10*, e59683. 10.7554/eLife.59683.

Kushawah, G., Hernandez-Huertas, L., Abugattas-Nuñez del Prado, J., Martinez-Morales, J.R., DeVore, M.L., Hassan, H., Moreno-Sanchez, I., Tomas-Gallardo, L., Diaz-Moscoso, A., Monges, D.E., *et al.* (2020). CRISPR-Cas13d Induces Efficient mRNA Knockdown in Animal Embryos. Developmental Cell *54*, 805-817.e807. https://doi.org/10.1016/j.devcel.2020.07.013.

Kuznetsova, K., Ugolini, M., Wu, E., Lalit, M., Oda, H., Sato, Y., Kimura, H., Jug, F., and Vastenhouw, N. (2022). Nanog organizes transcription bodies. bioRxiv, 2022.2006.2013.495463. 10.1101/2022.06.13.495463.

Labelle-Dumais, C., Jacob-Wagner, M., Pare, J.F., Belanger, L., and Dufort, D. (2006). Nuclear receptor NR5A2 is required for proper primitive streak morphogenesis. Dev Dyn *235*, 3359-3369. 10.1002/dvdy.20996.

Labun, K., Krause, M., Torres Cleuren, Y., and Valen, E. (2021). CRISPR Genome Editing Made Easy Through the CHOPCHOP Website. Current Protocols *1*, e46. https://doi.org/10.1002/cpz1.46.

Labun, K., Montague, T.G., Krause, M., Torres Cleuren, Y.N., Tjeldnes, H., and Valen, E. (2019). CHOPCHOP v3: expanding the CRISPR web toolbox beyond genome editing. Nucleic Acids Research *47*, W171-W174. 10.1093/nar/gkz365.

Lai, J.K.H., Gagalova, K.K., Kuenne, C., El-Brolosy, M.A., and Stainier, D.Y.R. (2019). Induction of interferon-stimulated genes and cellular stress pathways by morpholinos in zebrafish. Developmental Biology *454*, 21-28. https://doi.org/10.1016/j.ydbio.2019.06.008.

Lamason, R.L., Mohideen, M.A., Mest, J.R., Wong, A.C., Norton, H.L., Aros, M.C., Jurynec, M.J., Mao, X., Humphreville, V.R., Humbert, J.E., *et al.* (2005). SLC24A5, a putative cation exchanger, affects pigmentation in zebrafish and humans. Science *310*, 1782-1786. 10.1126/science.1116238.

Langdon, Y.G., and Mullins, M.C. (2011). Maternal and zygotic control of zebrafish dorsoventral axial patterning. Annu Rev Genet *45*, 357-377. 10.1146/annurev-genet-110410-132517.

Langenbacher, A.D., Huang, J., Chen, Y., and Chen, J.N. (2012). Sodium pump activity in the yolk syncytial layer regulates zebrafish heart tube morphogenesis. Dev Biol *362*, 263-270. 10.1016/j.ydbio.2011.12.004.

Lecuyer, E., Yoshida, H., Parthasarathy, N., Alm, C., Babak, T., Cerovina, T., Hughes, T.R., Tomancak, P., and Krause, H.M. (2007). Global analysis of mRNA localization reveals a prominent role in organizing cellular architecture and function. Cell *131*, 174-187. 10.1016/j.cell.2007.08.003.

Lee, D.R., Lee, J.E., Yoon, H.S., Roh, S.I., and Kim, M.K. (2001). Compaction in preimplantation mouse embryos is regulated by a cytoplasmic regulatory factor that alters between 1- and 2-cell stages in a concentration-dependent manner. J Exp Zool *290*, 61-71. 10.1002/jez.1036.

Lee, M.T., Bonneau, A.R., and Giraldez, A.J. (2014). Zygotic genome activation during the maternal-to-zygotic transition. Annu Rev Cell Dev Biol *30*, 581-613. 10.1146/annurev-cellbio-100913-013027.

Lee, M.T., Bonneau, A.R., Takacs, C.M., Bazzini, A.A., DiVito, K.R., Fleming, E.S., and Giraldez, A.J. (2013). Nanog, Pou5f1 and SoxB1 activate zygotic gene expression during the maternal-to-zygotic transition. Nature *503*, 360-364. 10.1038/nature12632.

Leichsenring, M., Maes, J., Mossner, R., Driever, W., and Onichtchouk, D. (2013). Pou5f1 transcription factor controls zygotic gene activation in vertebrates. Science *341*, 1005-1009. 10.1126/science.1242527.

Li, M., Zhao, L., Page-McCaw, P.S., and Chen, W. (2016). Zebrafish Genome Engineering Using the CRISPR-Cas9 System. Trends Genet *32*, 815-827. 10.1016/j.tig.2016.10.005.

Li, W., Zhang, Y., Han, B., Li, L., Li, M., Lu, X., Chen, C., Lu, M., Zhang, Y., Jia, X., *et al.* (2019). One-step efficient generation of dual-function conditional knockout and geno-tagging alleles in zebrafish. eLife *8*, e48081. 10.7554/eLife.48081.

Li, X.Y., Harrison, M.M., Villalta, J.E., Kaplan, T., and Eisen, M.B. (2014). Establishment of regions of genomic activity during the Drosophila maternal to zygotic transition. Elife *3*. 10.7554/eLife.03737.

Liang, H.L., Nien, C.Y., Liu, H.Y., Metzstein, M.M., Kirov, N., and Rushlow, C. (2008). The zinc-finger protein Zelda is a key activator of the early zygotic genome in Drosophila. Nature *456*, 400-403. 10.1038/nature07388.

Lindeman, L.C., Andersen, I.S., Reiner, A.H., Li, N., Aanes, H., Ostrup, O., Winata, C., Mathavan, S., Muller, F., Alestrom, P., and Collas, P. (2011). Prepatterning of developmental gene expression by modified histones before zygotic genome activation. Dev Cell *21*, 993-1004. 10.1016/j.devcel.2011.10.008.

Liu, C., Li, R., Li, Y., Lin, X., Zhao, K., Liu, Q., Wang, S., Yang, X., Shi, X., Ma, Y., *et al.* (2022a). Spatiotemporal mapping of gene expression landscapes and developmental trajectories during zebrafish embryogenesis. Developmental Cell *57*, 1284-1298.e1285. https://doi.org/10.1016/j.devcel.2022.04.009.

Liu, G., Lin, Q., Jin, S., and Gao, C. (2022b). The CRISPR-Cas toolbox and gene editing technologies. Mol Cell *82*, 333-347. 10.1016/j.molcel.2021.12.002.

Liu, K., Petree, C., Requena, T., Varshney, P., and Varshney, G.K. (2019). Expanding the CRISPR Toolbox in Zebrafish for Studying Development and Disease. Front Cell Dev Biol *7*, 13. 10.3389/fcell.2019.00013.

Liu, M.M., Davey, J.W., Jackson, D.J., Blaxter, M.L., and Davison, A. (2014). A conserved set of maternal genes? Insights from a molluscan transcriptome. Int J Dev Biol *58*, 501-511. 10.1387/ijdb.140121ad.

Liu, W.-Y., Wang, Y., Sun, Y.-H., Wang, Y., Wang, Y.-P., Chen, S.-P., and Zhu, Z.-Y. (2005). Efficient RNA interference in zebrafish embryos using siRNA synthesized with SP6 RNA polymerase. Development, Growth & Differentiation *47*, 323-331. https://doi.org/10.1111/j.1440-169X.2005.00807.x.

Liu, Y., Zhu, Z., Ho, I.H.T., Shi, Y., Li, J., Wang, X., Chan, M.T.V., and Cheng, C.H.K. (2020). Genetic Deletion of miR-430 Disrupts Maternal-Zygotic Transition and Embryonic Body Plan. Frontiers in Genetics *11*. 10.3389/fgene.2020.00853.

Liu, Z., Chen, O., Wall, J.B.J., Zheng, M., Zhou, Y., Wang, L., Ruth Vaseghi, H., Qian, L., and Liu, J. (2017). Systematic comparison of 2A peptides for cloning multi-genes in a polycistronic vector. Scientific Reports *7*, 2193. 10.1038/s41598-017-02460-2.

Loh, Y.H., Wu, Q., Chew, J.L., Vega, V.B., Zhang, W., Chen, X., Bourque, G., George, J., Leong, B., Liu, J., *et al.* (2006). The Oct4 and Nanog transcription network regulates

pluripotency in mouse embryonic stem cells. Nat Genet *38*, 431-440. 10.1038/ng1760.

Lopes, R., Agami, R., and Korkmaz, G. (2017). GRO-seq, A Tool for Identification of Transcripts Regulating Gene Expression. Methods Mol Biol *1543*, 45-55. 10.1007/978-1-4939-6716-2_3.

Lorenz, R., Bernhart, S.H., Honer Zu Siederdissen, C., Tafer, H., Flamm, C., Stadler, P.F., and Hofacker, I.L. (2011). ViennaRNA Package 2.0. Algorithms Mol Biol *6*, 26. 10.1186/1748-7188-6-26.

Lu, F., Liu, Y., Inoue, A., Suzuki, T., Zhao, K., and Zhang, Y. (2016). Establishing Chromatin Regulatory Landscape during Mouse Preimplantation Development. Cell *165*, 1375-1388. 10.1016/j.cell.2016.05.050.

Meehan, R.R., Dunican, D.S., Ruzov, A., and Pennings, S. (2005). Epigenetic silencing in embryogenesis. Exp Cell Res *309*, 241-249. 10.1016/j.yexcr.2005.06.023.

Meyers, J.R. (2018). Zebrafish: Development of a Vertebrate Model Organism. Current Protocols Essential Laboratory Techniques *16*, e19. https://doi.org/10.1002/cpet.19.

Miao, L., Tang, Y., Bonneau, A.R., Chan, S.H., Kojima, M.L., Pownall, M.E., Vejnar, C.E., Gao, F., Krishnaswamy, S., Hendry, C.E., and Giraldez, A.J. (2022). The landscape of pioneer factor activity reveals the mechanisms of chromatin reprogramming and genome activation. Mol Cell *82*, 986-1002 e1009. 10.1016/j.molcel.2022.01.024.

Moreno, R.L., Williams, K., Jones, K.L., and Ribera, A.B. (2018). Investigation of Islet2a function in zebrafish embryos: Mutants and morphants differ in morphologic phenotypes and gene expression. PLoS One *13*, e0199233. 10.1371/journal.pone.0199233.

Moreno-Mateos, M.A., Vejnar, C.E., Beaudoin, J.-D., Fernandez, J.P., Mis, E.K., Khokha, M.K., and Giraldez, A.J. (2015). CRISPRscan: designing highly efficient sgRNAs for CRISPR-Cas9 targeting in vivo. Nature Methods *12*, 982-988. 10.1038/nmeth.3543.

Morita, H., Grigolon, S., Bock, M., Krens, S.F.G., Salbreux, G., and Heisenberg, C.-P. (2017). The Physical Basis of Coordinated Tissue Spreading in Zebrafish Gastrulation. Developmental Cell *40*, 354-366.e354. 10.1016/j.devcel.2017.01.010.

Muller, F., Lakatos, L., Dantonel, J., Strahle, U., and Tora, L. (2001). TBP is not universally required for zygotic RNA polymerase II transcription in zebrafish. Curr Biol *11*, 282-287. 10.1016/s0960-9822(01)00076-8.

Muller, F., Zaucker, A., and Tora, L. (2010). Developmental regulation of transcription initiation: more than just changing the actors. Curr Opin Genet Dev *20*, 533-540. 10.1016/j.gde.2010.06.004.

Naert, T., Tulkens, D., Edwards, N.A., Carron, M., Shaidani, N.I., Wlizla, M., Boel, A., Demuynck, S., Horb, M.E., Coucke, P., *et al.* (2020). Maximizing CRISPR/Cas9 phenotype penetrance applying predictive modeling of editing outcomes in Xenopus and zebrafish embryos. Sci Rep *10*, 14662. 10.1038/s41598-020-71412-0.

Nasevicius, A., and Ekker, S.C. (2000). Effective targeted gene 'knockdown' in zebrafish. Nat Genet *26*, 216-220. 10.1038/79951.

Nelles, D.A., Fang, M.Y., O'Connell, M.R., Xu, J.L., Markmiller, S.J., Doudna, J.A., and Yeo, G.W. (2016). Programmable RNA Tracking in Live Cells with CRISPR/Cas9. Cell *165*, 488-496. 10.1016/j.cell.2016.02.054.

Nelson, A.C., Cutty, S.J., Gasiunas, S.N., Deplae, I., Stemple, D.L., and Wardle, F.C. (2017). In Vivo Regulation of the Zebrafish Endoderm Progenitor Niche by T-Box Transcription Factors. Cell Rep *19*, 2782-2795. 10.1016/j.celrep.2017.06.011.

Nelson, A.C., Cutty, S.J., Niini, M., Stemple, D.L., Flicek, P., Houart, C., Bruce, A.E., and Wardle, F.C. (2014). Global identification of Smad2 and Eomesodermin targets in zebrafish identifies a conserved transcriptional network in mesendoderm and a novel role for Eomesodermin in repression of ectodermal gene expression. BMC Biol *12*, 81. 10.1186/s12915-014-0081-5.

Nepal, C., Hadzhiev, Y., Previti, C., Haberle, V., Li, N., Takahashi, H., Suzuki, A.M., Sheng, Y., Abdelhamid, R.F., Anand, S., *et al.* (2013). Dynamic regulation of the transcription initiation landscape at single nucleotide resolution during vertebrate embryogenesis. Genome Res *23*, 1938-1950. 10.1101/gr.153692.112.

Newport, J., and Kirschner, M. (1982a). A major developmental transition in early Xenopus embryos: I. characterization and timing of cellular changes at the midblastula stage. Cell *30*, 675-686. 10.1016/0092-8674(82)90272-0.

Newport, J., and Kirschner, M. (1982b). A major developmental transition in early Xenopus embryos: II. Control of the onset of transcription. Cell *30*, 687-696. 10.1016/0092-8674(82)90273-2.

O'Connell, M.R. (2019). Molecular Mechanisms of RNA Targeting by Cas13-containing Type VI CRISPR–Cas Systems. Journal of Molecular Biology *431*, 66-87. https://doi.org/10.1016/j.jmb.2018.06.029.

Oates, A.C., Bruce, A.E., and Ho, R.K. (2000). Too much interference: injection of double-stranded RNA has nonspecific effects in the zebrafish embryo. Dev Biol *224*, 20-28. 10.1006/dbio.2000.9761.

Ober, E.A., and Schulte-Merker, S. (1999). Signals from the Yolk Cell Induce Mesoderm, Neuroectoderm, the Trunk Organizer, and the Notochord in Zebrafish. Developmental Biology *215*, 167-181. https://doi.org/10.1006/dbio.1999.9455.

Oh, B., Hwang, S., McLaughlin, J., Solter, D., and Knowles, B.B. (2000). Timely translation during the mouse oocyte-to-embryo transition. Development *127*, 3795-3803. 10.1242/dev.127.17.3795.

Oka, Y., and Sato, T.N. (2015). Whole-mount single molecule FISH method for zebrafish embryo. Scientific Reports *5*, 8571. 10.1038/srep08571.

Omura, C.S., and Lott, S.E. (2020). The conserved regulatory basis of mRNA contributions to the early Drosophila embryo differs between the maternal and zygotic genomes. PLoS Genet *16*, e1008645. 10.1371/journal.pgen.1008645.

Onuma, Y., Takahashi, S., Yokota, C., and Asashima, M. (2002). Multiple nodal-related genes act coordinately in Xenopus embryogenesis. Dev Biol *241*, 94-105. 10.1006/dbio.2001.0493.

Ooga, M., Fulka, H., Hashimoto, S., Suzuki, M.G., and Aoki, F. (2016). Analysis of chromatin structure in mouse preimplantation embryos by fluorescent recovery after photobleaching. Epigenetics *11*, 85-94. 10.1080/15592294.2015.1136774.

Owens, N.D.L., Blitz, I.L., Lane, M.A., Patrushev, I., Overton, J.D., Gilchrist, M.J., Cho, K.W.Y., and Khokha, M.K. (2016). Measuring Absolute RNA Copy Numbers at High Temporal Resolution Reveals Transcriptome Kinetics in Development. Cell Rep *14*, 632-647. 10.1016/j.celrep.2015.12.050.

Palfy, M., Schulze, G., Valen, E., and Vastenhouw, N.L. (2020). Chromatin accessibility established by Pou5f3, Sox19b and Nanog primes genes for activity during zebrafish genome activation. PLoS Genet *16*, e1008546. 10.1371/journal.pgen.1008546.

Parichy, D.M. (2015). Advancing biology through a deeper understanding of zebrafish ecology and evolution. Elife *4*. 10.7554/eLife.05635.

Parichy, D.M., Elizondo, M.R., Mills, M.G., Gordon, T.N., and Engeszer, R.E. (2009). Normal table of postembryonic zebrafish development: staging by externally visible anatomy of the living fish. Dev Dyn *238*, 2975-3015. 10.1002/dvdy.22113.

Peng, Y., Clark, K.J., Campbell, J.M., Panetta, M.R., Guo, Y., and Ekker, S.C. (2014). Making designer mutants in model organisms. Development *141*, 4042-4054. 10.1242/dev.102186.

Pereira, L.A., Wong, M.S., Mei Lim, S., Stanley, E.G., and Elefanty, A.G. (2012). The Mix family of homeobox genes--key regulators of mesendoderm formation during vertebrate development. Dev Biol *367*, 163-177. 10.1016/j.ydbio.2012.04.033.

Potok, M.E., Nix, D.A., Parnell, T.J., and Cairns, B.R. (2013). Reprogramming the maternal zebrafish genome after fertilization to match the paternal methylation pattern. Cell *153*, 759-772. 10.1016/j.cell.2013.04.030.

Prioleau, M.N., Huet, J., Sentenac, A., and Mechali, M. (1994). Competition between chromatin and transcription complex assembly regulates gene expression during early development. Cell *77*, 439-449. 10.1016/0092-8674(94)90158-9.

Pritchard, D.K., and Schubiger, G. (1996). Activation of transcription in Drosophila embryos is a gradual process mediated by the nucleocytoplasmic ratio. Genes Dev *10*, 1131-1142. 10.1101/gad.10.9.1131.

Prykhozhij, S.V., and Berman, J.N. (2018). Zebrafish knock-ins swim into the mainstream. Dis Model Mech *11*. 10.1242/dmm.037515.

Qureshi, H.K., Magony, A., Hadzhiev, Y., Wozniak, K., Jasiulewicz, A., Müller, F., and Sik, A. (2020). A lightweight segmentation and lineage tracking tool for noisy, low frame rate 4D fluorescence microscopy data. bioRxiv, 2020.2011.2001.364083. 10.1101/2020.11.01.364083.

Raff, J.W., and Glover, D.M. (1988). Nuclear and cytoplasmic mitotic cycles continue in Drosophila embryos in which DNA synthesis is inhibited with aphidicolin. J Cell Biol *107*, 2009-2019. 10.1083/jcb.107.6.2009.

Ranawakage, D.C., Okada, K., Sugio, K., Kawaguchi, Y., Kuninobu-Bonkohara, Y., Takada, T., and Kamachi, Y. (2021). Efficient CRISPR-Cas9-Mediated Knock-In of Composite Tags in Zebrafish Using Long ssDNA as a Donor. Frontiers in Cell and Developmental Biology *8*. 10.3389/fcell.2020.598634.

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*, e78. 10.1371/journal.pgen.0030078.

Rosa, A., and Brivanlou, A.H. (2009). MicroRNAs in early vertebrate development. Cell Cycle *8*, 3513-3520. 10.4161/cc.8.21.9847.

Ross, S., Cheung, E., Petrakis, T.G., Howell, M., Kraus, W.L., and Hill, C.S. (2006). Smads orchestrate specific histone modifications and chromatin remodeling to activate transcription. EMBO J *25*, 4490-4502. 10.1038/sj.emboj.7601332.

Ross, S., and Hill, C.S. (2008). How the Smads regulate transcription. Int J Biochem Cell Biol *40*, 383-408. 10.1016/j.biocel.2007.09.006.

Rossi, A., Kontarakis, Z., Gerri, C., Nolte, H., Hölper, S., Krüger, M., and Stainier, D.Y. (2015). Genetic compensation induced by deleterious mutations but not gene knockdowns. Nature *524*, 230-233. 10.1038/nature14580.

Rothe, M., Pehl, M., Taubert, H., and Jackle, H. (1992). Loss of gene function through rapid mitotic cycles in the Drosophila embryo. Nature *359*, 156-159. 10.1038/359156a0.

Sakaguchi, T., Kikuchi, Y., Kuroiwa, A., Takeda, H., and Stainier, D.Y. (2006). The yolk syncytial layer regulates myocardial migration by influencing extracellular matrix assembly in zebrafish. Development *133*, 4063-4072. 10.1242/dev.02581.

Samata, M., Alexiadis, A., Richard, G., Georgiev, P., Nuebler, J., Kulkarni, T., Renschler, G., Basilicata, M.F., Zenk, F.L., Shvedunova, M., *et al.* (2020). Intergenerationally Maintained Histone H4 Lysine 16 Acetylation Is Instructive for Future Gene Activation. Cell *182*, 127-144 e123. 10.1016/j.cell.2020.05.026.

Sandelin, A., Carninci, P., Lenhard, B., Ponjavic, J., Hayashizaki, Y., and Hume, D.A. (2007). Mammalian RNA polymerase II core promoters: insights from genome-wide studies. Nat Rev Genet *8*, 424-436. 10.1038/nrg2026.

Sander, J.D., Cade, L., Khayter, C., Reyon, D., Peterson, R.T., Joung, J.K., and Yeh, J.-R.J. (2011). Targeted gene disruption in somatic zebrafish cells using engineered TALENs. Nature Biotechnology *29*, 697-698. 10.1038/nbt.1934.

Santos, F., Peters, A.H., Otte, A.P., Reik, W., and Dean, W. (2005). Dynamic chromatin modifications characterise the first cell cycle in mouse embryos. Dev Biol *280*, 225-236. 10.1016/j.ydbio.2005.01.025.

Sato, Y., Hilbert, L., Oda, H., Wan, Y., Heddleston, J.M., Chew, T.L., Zaburdaev, V., Keller, P., Lionnet, T., Vastenhouw, N., and Kimura, H. (2019). Histone H3K27 acetylation precedes active transcription during zebrafish zygotic genome activation as revealed by live-cell analysis. Development *146*. 10.1242/dev.179127.

Schier, A.F. (2001). Axis formation and patterning in zebrafish. Curr Opin Genet Dev *11*, 393-404. 10.1016/s0959-437x(00)00209-4.

Schier, A.F. (2009). Nodal morphogens. Cold Spring Harb Perspect Biol *1*, a003459. 10.1101/cshperspect.a003459.

Schindelin, J., Arganda-Carreras, I., Frise, E., Kaynig, V., Longair, M., Pietzsch, T., Preibisch, S., Rueden, C., Saalfeld, S., Schmid, B., *et al.* (2012). Fiji: an open-source platform for biological-image analysis. Nature Methods *9*, 676-682. 10.1038/nmeth.2019.

Schultz, R.M., Stein, P., and Svoboda, P. (2018). The oocyte-to-embryo transition in mouse: past, present, and future. Biol Reprod *99*, 160-174. 10.1093/biolre/ioy013.

Schulz, K.N., Bondra, E.R., Moshe, A., Villalta, J.E., Lieb, J.D., Kaplan, T., McKay, D.J., and Harrison, M.M. (2015). Zelda is differentially required for chromatin accessibility, transcription factor binding, and gene expression in the early Drosophila embryo. Genome Res *25*, 1715-1726. 10.1101/gr.192682.115.

Schulz, K.N., and Harrison, M.M. (2019). Mechanisms regulating zygotic genome activation. Nat Rev Genet *20*, 221-234. 10.1038/s41576-018-0087-x.

Sertori, R., Trengove, M., Basheer, F., Ward, A.C., and Liongue, C. (2016). Genome editing in zebrafish: a practical overview. Briefings in Functional Genomics *15*, 322-330. 10.1093/bfgp/elv051.

Sibon, O.C., Laurençon, A., Hawley, R., and Theurkauf, W.E. (1999). The Drosophila ATM ue Mei-41 has an essential checkpoint function at the midblastula transition. Curr Biol *9*, 302-312. 10.1016/s0960-9822(99)80138-9.

Sibon, O.C., Stevenson, V.A., and Theurkauf, W.E. (1997). DNA-replication checkpoint control at the Drosophila midblastula transition. Nature *388*, 93-97. 10.1038/40439.

Skirkanich, J., Luxardi, G., Yang, J., Kodjabachian, L., and Klein, P.S. (2011). An essential role for transcription before the MBT in Xenopus laevis. Dev Biol *357*, 478-491. 10.1016/j.ydbio.2011.06.010.

Stainier, D.Y.R., Raz, E., Lawson, N.D., Ekker, S.C., Burdine, R.D., Eisen, J.S., Ingham, P.W., Schulte-Merker, S., Yelon, D., Weinstein, B.M., *et al.* (2017). Guidelines for morpholino use in zebrafish. PLoS Genet *13*, e1007000. 10.1371/journal.pgen.1007000.

Sun, F., Fang, H., Li, R., Gao, T., Zheng, J., Chen, X., Ying, W., and Sheng, H.Z. (2007). Nuclear reprogramming: the zygotic transcription program is established through an "erase-and-rebuild" strategy. Cell Res *17*, 117-134. 10.1038/cr.2007.1.

Sun, N., Petiwala, S., Wang, R., Lu, C., Hu, M., Ghosh, S., Hao, Y., Miller, C.P., and Chung, N. (2019). Development of drug-inducible CRISPR-Cas9 systems for large-scale functional screening. BMC Genomics *20*, 225. 10.1186/s12864-019-5601-9.

Sun, Y., Nien, C.Y., Chen, K., Liu, H.Y., Johnston, J., Zeitlinger, J., and Rushlow, C. (2015). Zelda overcomes the high intrinsic nucleosome barrier at enhancers during Drosophila zygotic genome activation. Genome Res *25*, 1703-1714. 10.1101/gr.192542.115.

Suzuki, H., Ishizaka, T., Yanagi, K., Sone, R., Sunaga, Y., Ohga, R., and Kawahara, A. (2019). Characterization of biklf/klf17-deficient zebrafish in posterior lateral line neuromast and hatching gland development. Scientific Reports *9*, 13680. 10.1038/s41598-019-50149-5.

Swinburne, I.A., and Silver, P.A. (2008). Intron delays and transcriptional timing during development. Dev Cell *14*, 324-330. 10.1016/j.devcel.2008.02.002.

Tadros, W., and Lipshitz, H.D. (2009). The maternal-to-zygotic transition: a play in two acts. Development *136*, 3033-3042. 10.1242/dev.033183.

Takacs, C.M., and Giraldez, A.J. (2016). miR-430 regulates oriented cell division during neural tube development in zebrafish. Dev Biol *409*, 442-450. 10.1016/j.ydbio.2015.11.016.

Takada, S., Kwak, S., Koppetsch, B.S., and Theurkauf, W.E. (2007). grp (chk1) replication-checkpoint mutations and DNA damage trigger a Chk2-dependent block at the Drosophila midblastula transition. Development *134*, 1737-1744. 10.1242/dev.02831.

Takahashi, K., and Yamanaka, S. (2006). Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. Cell *126*, 663-676. 10.1016/j.cell.2006.07.024.

Tatarakis, D., Cang, Z., Wu, X., Sharma, P.P., Karikomi, M., MacLean, A.L., Nie, Q., and Schilling, T.F. (2021). Single-cell transcriptomic analysis of zebrafish cranial neural crest reveals spatiotemporal regulation of lineage decisions during development. Cell Rep *37*, 110140. 10.1016/j.celrep.2021.110140.

Taubenschmid-Stowers, J., Rostovskaya, M., Santos, F., Ljung, S., Argelaguet, R., Krueger, F., Nichols, J., and Reik, W. (2022). 8C-like cells capture the human zygotic genome activation program in vitro. Cell Stem Cell *29*, 449-459 e446. 10.1016/j.stem.2022.01.014.

Temeles, G.L., and Schultz, R.M. (1997). Transient polyadenylation of a maternal mRNA following fertilization of mouse eggs. J Reprod Fertil *109*, 223-228. 10.1530/jrf.0.1090223.

Thisse, C., and Thisse, B. (2008). High-resolution in situ hybridization to whole-mount zebrafish embryos. Nature Protocols *3*, 59-69. 10.1038/nprot.2007.514.

Thyme, S.B., Akhmetova, L., Montague, T.G., Valen, E., and Schier, A.F. (2016). Internal guide RNA interactions interfere with Cas9-mediated cleavage. Nature Communications *7*, 11750. 10.1038/ncomms11750.

Trinkaus, J.P. (1951). A study of the mechanism of epiboly in the egg of Fundulus heteroclitus. Journal of Experimental Zoology *118*, 269-319. https://doi.org/10.1002/jez.1401180204.

Trinkaus, J.P. (1984). Mechanism of Fundulus Epiboly—A Current View. American Zoologist *24*, 673-688. 10.1093/icb/24.3.673.

Trinkaus, J.P. (1993). The yolk syncytial layer of Fundulus: its origin and history and its significance for early embryogenesis. J Exp Zool *265*, 258-284. 10.1002/jez.1402650308.

Untergasser, A., Cutcutache, I., Koressaar, T., Ye, J., Faircloth, B.C., Remm, M., and Rozen, S.G. (2012). Primer3--new capabilities and interfaces. Nucleic Acids Res *40*, e115. 10.1093/nar/gks596.

Varga, Z.M., Lawrence, C., Ekker, S.C., and Eisen, J.S. (2016). Universal Healthcare for Zebrafish. Zebrafish *13 Suppl 1*, S1-4. 10.1089/zeb.2016.1311.

Vassena, R., Boue, S., Gonzalez-Roca, E., Aran, B., Auer, H., Veiga, A., and Izpisua Belmonte, J.C. (2011). Waves of early transcriptional activation and pluripotency program initiation during human preimplantation development. Development *138*, 3699-3709. 10.1242/dev.064741.

Vastenhouw, N.L., Cao, W.X., and Lipshitz, H.D. (2019). The maternal-to-zygotic transition revisited. Development *146*. 10.1242/dev.161471.

Vastenhouw, N.L., Zhang, Y., Woods, I.G., Imam, F., Regev, A., Liu, X.S., Rinn, J., and Schier, A.F. (2010). Chromatin signature of embryonic pluripotency is established during genome activation. Nature *464*, 922-926. 10.1038/nature08866.

Veenstra, G.J., Destree, O.H., and Wolffe, A.P. (1999). Translation of maternal TATA-binding protein mRNA potentiates basal but not activated transcription in Xenopus embryos at the midblastula transition. Mol Cell Biol *19*, 7972-7982. 10.1128/MCB.19.12.7972.

Veil, M., Schaechtle, M.A., Gao, M., Kirner, V., Buryanova, L., Grethen, R., and Onichtchouk, D. (2018). Maternal Nanog is required for zebrafish embryo architecture and for cell viability during gastrulation. Development *145*. 10.1242/dev.155366.

Veil, M., Yampolsky, L.Y., Gruning, B., and Onichtchouk, D. (2019). Pou5f3, SoxB1, and Nanog remodel chromatin on high nucleosome affinity regions at zygotic genome activation. Genome Res *29*, 383-395. 10.1101/gr.240572.118.

Vejnar, C.E., Abdel Messih, M., Takacs, C.M., Yartseva, V., Oikonomou, P., Christiano, R., Stoeckius, M., Lau, S., Lee, M.T., Beaudoin, J.D., *et al.* (2019). Genome wide analysis of 3' UTR sequence elements and proteins regulating mRNA stability during maternal-to-zygotic transition in zebrafish. Genome Res *29*, 1100-1114. 10.1101/gr.245159.118.

Wagner, D.E., Weinreb, C., Collins, Z.M., Briggs, J.A., Megason, S.G., and Klein, A.M. (2018). Single-cell mapping of gene expression landscapes and lineage in the zebrafish embryo. Science *360*, 981-987. 10.1126/science.aar4362.

Wang, Q.T., Piotrowska, K., Ciemerych, M.A., Milenkovic, L., Scott, M.P., Davis, R.W., and Zernicka-Goetz, M. (2004). A genome-wide study of gene activity reveals developmental signaling pathways in the preimplantation mouse embryo. Dev Cell *6*, 133-144. 10.1016/s1534-5807(03)00404-0.

Warga, R.M., and Kimmel, C.B. (1990). Cell movements during epiboly and gastrulation in zebrafish. Development *108*, 569-580. 10.1242/dev.108.4.569.

Welker, J.M., Wierson, W.A., Almeida, M.P., Mann, C.M., Torrie, M.E., Ming, Z., Ekker, S.C., Clark, K.J., Dobbs, D.L., Essner, J.J., and McGrail, M. (2021). GeneWeld: Efficient Targeted Integration Directed by Short Homology in Zebrafish. Bio Protoc *11*, e4100. 10.21769/BioProtoc.4100.

Welten, M.C., de Haan, S.B., van den Boogert, N., Noordermeer, J.N., Lamers, G.E., Spaink, H.P., Meijer, A.H., and Verbeek, F.J. (2006). ZebraFISH: fluorescent in situ hybridization protocol and three-dimensional imaging of gene expression patterns. Zebrafish *3*, 465-476. 10.1089/zeb.2006.3.465.

Wessels, H.-H., Méndez-Mancilla, A., Guo, X., Legut, M., Daniloski, Z., and Sanjana, N.E. (2020). Massively parallel Cas13 screens reveal principles for guide RNA design. Nature Biotechnology *38*, 722-727. 10.1038/s41587-020-0456-9.

Westerfield, M. (2000). The zebrafish book. A guide for the laboratory use of zebrafish (Danio rerio). *4th ed.* (University of Oregon Press).

White, R.J., Collins, J.E., Sealy, I.M., Wali, N., Dooley, C.M., Digby, Z., Stemple, D.L., Murphy, D.N., Billis, K., Hourlier, T., *et al.* (2017). A high-resolution mRNA expression time course of embryonic development in zebrafish. eLife *6*, e30860. 10.7554/eLife.30860.

Wierson, W.A., Welker, J.M., Almeida, M.P., Mann, C.M., Webster, D.A., Torrie, M.E., Weiss, T.J., Kambakam, S., Vollbrecht, M.K., Lan, M., *et al.* (2020). Efficient targeted integration directed by short homology in zebrafish and mammalian cells. eLife *9*, e53968. 10.7554/eLife.53968.

Wilkins, S.J., Yoong, S., Verkade, H., Mizoguchi, T., Plowman, S.J., Hancock, J.F., Kikuchi, Y., Heath, J.K., and Perkins, A.C. (2008). Mtx2 directs zebrafish morphogenetic movements during epiboly by regulating microfilament formation. Dev Biol *314*, 12-22. 10.1016/j.ydbio.2007.10.050.

Wragg, J., and Müller, F. (2016). Chapter Six - Transcriptional Regulation During Zygotic Genome Activation in Zebrafish and Other Anamniote Embryos. In Advances in Genetics, N.S. Foulkes, ed. (Academic Press), pp. 161-194. https://doi.org/10.1016/bs.adgen.2016.05.001.

Wu, R.S., Lam, II, Clay, H., Duong, D.N., Deo, R.C., and Coughlin, S.R. (2018). A Rapid Method for Directed Gene Knockout for Screening in G0 Zebrafish. Dev Cell *46*, 112-125 e114. 10.1016/j.devcel.2018.06.003.

Xing, C., Shen, W., Gong, B., Li, Y., Yan, L., and Meng, A. (2022). Maternal Factors and Nodal Autoregulation Orchestrate Nodal Gene Expression for Embryonic Mesendoderm Induction in the Zebrafish. Frontiers in Cell and Developmental Biology *10*. 10.3389/fcell.2022.887987.

Xu, C., Fan, Z.P., Muller, P., Fogley, R., DiBiase, A., Trompouki, E., Unternaehrer, J., Xiong, F., Torregroza, I., Evans, T., *et al.* (2012). Nanog-like regulates endoderm formation through the Mxtx2-Nodal pathway. Dev Cell *22*, 625-638. 10.1016/j.devcel.2012.01.003.

Xu, P., Zhu, G., Wang, Y., Sun, J., Liu, X., Chen, Y.G., and Meng, A. (2014). Maternal Eomesodermin regulates zygotic nodal gene expression for mesendoderm induction in zebrafish embryos. J Mol Cell Biol *6*, 272-285. 10.1093/jmcb/mju028.

Xue, Z., Huang, K., Cai, C., Cai, L., Jiang, C.Y., Feng, Y., Liu, Z., Zeng, Q., Cheng, L., Sun, Y.E., *et al.* (2013). Genetic programs in human and mouse early embryos revealed by single-cell RNA sequencing. Nature *500*, 593-597. 10.1038/nature12364.

Yu, C., Ji, S.Y., Dang, Y.J., Sha, Q.Q., Yuan, Y.F., Zhou, J.J., Yan, L.Y., Qiao, J., Tang, F., and Fan, H.Y. (2016). Oocyte-expressed yes-associated protein is a key activator of the early zygotic genome in mouse. Cell Res *26*, 275-287. 10.1038/cr.2016.20.

Yuan, K., Seller, C.A., Shermoen, A.W., and O'Farrell, P.H. (2016). Timing the Drosophila Mid-Blastula Transition: A Cell Cycle-Centered View. Trends Genet *32*, 496-507. 10.1016/j.tig.2016.05.006.

Zhang, M., Kothari, P., Mullins, M., and Lampson, M.A. (2014). Regulation of zygotic genome activation and DNA damage checkpoint acquisition at the mid-blastula transition. Cell Cycle *13*, 3828-3838. 10.4161/15384101.2014.967066.

Zhang, P., Zucchelli, M., Bruce, S., Hambiliki, F., Stavreus-Evers, A., Levkov, L., Skottman, H., Kerkela, E., Kere, J., and Hovatta, O. (2009). Transcriptome profiling of human pre-implantation development. PLoS One *4*, e7844. 10.1371/journal.pone.0007844.

Zhao, X.-F., Fjose, A., Larsen, N., Helvik, J.V., and Drivenes, Ø. (2008). Treatment with small interfering RNA affects the microRNA pathway and causes unspecific defects in zebrafish embryos. The FEBS Journal *275*, 2177-2184. https://doi.org/10.1111/j.1742-4658.2008.06371.x.

Zhao, Z., Cao, Y., Li, M., and Meng, A. (2001). Double-stranded RNA injection produces nonspecific defects in zebrafish. Dev Biol *229*, 215-223. 10.1006/dbio.2000.9982.

Zhu, W., Xu, X., Wang, X., and Liu, J. (2019). Reprogramming histone modification patterns to coordinate gene expression in early zebrafish embryos. BMC Genomics *20*, 248. 10.1186/s12864-019-5611-7.

Zou, Z., Zhang, C., Wang, Q., Hou, Z., Xiong, Z., Kong, F., Wang, Q., Song, J., Liu, B., Liu, B., *et al.* (2022). Translatome and transcriptome co-profiling reveals a role of TPRXs in human zygotic genome activation. Science *0*, eabo7923. doi:10.1126/science.abo7923.