UNIL | Université de Lausanne

Unicentre
CH-1015 Lausanne
http://serval.unil.ch

Year : 2023

The timing of zygotic transcription is encoded by transcription factor sensitivity

Ng Hann Shen

Ng Hann Shen, 2023, The timing of zygotic transcription is encoded by transcription factor sensitivity

Originally published at : Thesis, University of Lausanne

Posted at the University of Lausanne Open Archive http://serval.unil.ch
Document URN : urn:nbn:ch:serval-BIB_DA12352896BB3

Droits d’auteur
L'Université de Lausanne attire expressément l'attention des utilisateurs sur le fait que tous les documents publiés dans l'Archive SERVAL sont protégés par le droit d'auteur, conformément à la loi fédérale sur le droit d'auteur et les droits voisins (LDA). A ce titre, il est indispensable d'obtenir le consentement préalable de l'auteur et/ou de l’éditeur avant toute utilisation d'une oeuvre ou d'une partie d'une oeuvre ne relevant pas d'une utilisation à des fins personnelles au sens de la LDA (art. 19, al. 1 lettre a). A défaut, tout contrevenant s'expose aux sanctions prévues par cette loi. Nous déclinons toute responsabilité en la matière.

Copyright
The University of Lausanne expressly draws the attention of users to the fact that all documents published in the SERVAL Archive are protected by copyright in accordance with federal law on copyright and similar rights (LDA). Accordingly it is indispensable to obtain prior consent from the author and/or publisher before any use of a work or part of a work for purposes other than personal use within the meaning of LDA (art. 19, para. 1 letter a). Failure to do so will expose offenders to the sanctions laid down by this law. We accept no liability in this respect.

UNIL | Université de Lausanne

Faculté de biologie
et de médecine

Institut de Centré Intégratif de Génomique

The timing of zygotic transcription is encoded by transcription factor sensitivity

Thèse de doctorat ès sciences de la vie (PhD)

présentée à la

Faculté de biologie et de médecine
de l'Université de Lausanne

par

Hann Shen Ng

Biologiste diplômé ou Master de l'Université de Leeds.

Jury

Prof. Tanja Schwander, Présidente
Prof. Nadine Vastenhouw, Directrice de thèse
Prof. Maria Cristina Gambetta, Experte
Prof. Frank Buchholz, Expert

Lausanne
(2023)

UNIL | Université de Lausanne

Faculté de biologie
et de médecine

Ecole Doctorale

Doctorat ès sciences de la vie

Imprimatur

Vu le rapport présenté par le jury d'examen, composé de

Président·e	Madame	Prof.	Tanja	**Schwander**
Directeur·trice de thèse	Madame	Prof.	Nadine	**Vastenhouw**
Expert·e·s	Madame	Prof.	Maria Cristina	**Gambetta**
	Monsieur	Prof.	Frank	**Buchholz**

le Conseil de Faculté autorise l'impression de la thèse de

Hann Shen Ng

Master - Integrated Masters in Biological Sciences (Molecular Medicine), University of Leeds, Royaume-Uni

intitulée

The timing of zygotic transcription is encoded by transcription factor sensitivity

Lausanne, le 1 décembre 2023

pour le Doyen
de la Faculté de biologie et de médecine

T. Schwander

Prof. Tanja Schwander

Acknowledgments

I dedicate this labour of excitement, struggle, love, and pain to the following wonderful people in my life:

To my parents. For all your hard work, love and sacrifice that provided me with the tools to achieve what I have achieved today. And to my dear mother, who is one of the strongest women I know, thank you for all your nuggets of advice over the years. You may not know it, but your pieces of advice stick with me strongly as I go through the struggles of adult life. Thank you to my siblings for their love and care, and their ability to always make fun of me even though I am only ever home once a year.

To Adam. Thank you for being my rock on the toughest of days and making me laugh even through it all. And even more so for regularly reminding me that there are always things to be positive about. My muddly brain tends to forget it.

To my supervisor, Nadine. What a journey it has been! Thank you for your patience, guidance, and support over the years. This project has not been an easy one. But thank you for continuing to sit through this rollercoaster ride with me.

To my dearest dearest labmates. Thank you for making the lab such a fun place to work. The regular failures of research would be so much more painful without you all to cushion the blows. Special shoutout to Edlyn, Gilles, Ksenia (yes you!), Noémie, Ramya and Shivali, for always being there, with support, snacks, jokes, valuable advice and useful scientific discussions.

To Juan, Marcell and Victor. Thank you for the fun times and sticking with me through all the tough stuff. I am blessed to have you guys in my life.

To my lovely flatmates Rémi, Marine and Iuliana. Thank you for dealing with me the past few weeks as this thesis was being written! Our delicious dinners will definitely be one of the things I will miss the most after I leave Lausanne!

Thank you to all the friends that have been there with me along this journey. In particular, Chiara, Enrico, Elisa and Tommaso (This unintentionally ended up being a list of Italians).

And finally, a MASSIVE thank you to the MPI-CBG staff, the International office, the MPI-CBG fish facility, the CIG staff, the CIG fish facility, and the GTF for always being there to provide scientific advice and support.

Abstract (English)

Transcription activation in a developing embryo is a well-timed process. The embryo begins its life in a transcriptional quiescent state but with time, acquires its own transcriptional competence. This molecular 'coming-of-age' of the embryo depends on a defined series of events that need to take place for zygotic transcription to be possible. Central to this, is the recruitment of transcription factors (TFs) onto the DNA template for initiating transcription. In zebrafish, we now know that enough transcriptional activators need to be accumulated in the embryo for zygotic genome activation (ZGA) to take place. Correspondingly, concentrations of nucleosome-forming histones in the nucleus must deplete, to give way to TFs for access to their cognate TF binding sites. This relationship has been shown to broadly determine the timing of ZGA. However, ZGA is a gradual and temporally ordered process whereby specific genomic regions can turn on several cell cycles earlier than others. In zebrafish, the *mir430* genomic locus is the first gene to be activated during embryo development. Interestingly, this genomic locus has the unique ability to recruit large amounts of nuclear transcriptional activators and transcriptional machinery. How this locus can activate so early during ZGA and, how this relates to its strong ability to recruit transcriptional activators and machinery is still unclear. Here, via in-depth characterisations and targeted long read sequencing, I show that the *mir430* locus is unique in that it is extremely repetitive. Contrary to the 16 kbp long representation of the *mir430* locus on the reference genome, I found that this locus is in fact at least 150 kbp in size (and potentially even larger) consisting mainly of stereotypic 1.7 kbp repeats. These repeats contain functional TF binding sites and I show that collectively, they ensure the early activation of the *mir430* locus. Isolated *mir430* loci with lower numbers of repeats lose this competitive advantage and are activated only later during development, despite having identical sequences. Mechanistically, I show that the competitive advantage conferred by higher repeat numbers is a higher sensitivity for TF binding, likely because of the higher number of TF binding sites. In the context of the competitive relationship between histones and TFs for access to the DNA, a mega-repetitive locus like *mir430* could facilitate localised out-competition of TFs against histones, despite the generally repressive nuclear environment. These findings provide a deeper insight into what genetic features might define when a gene turns on during development and, could be a generalisable way of understanding how TFs interact with their target binding sites.

Abstract (Français)

L'activation de la transcription dans un embryon en développement est un processus bien programmé. L'embryon commence sa vie dans un état de quiescence transcriptionnelle, mais avec le temps, il acquiert sa propre compétence transcriptionnelle. Ce "passage à l'âge adulte" moléculaire de l'embryon dépend d'une série définie d'événements qui doivent avoir lieu pour que la transcription zygotique soit possible. Le recrutement des facteurs de transcription sur la matrice d'ADN pour initier la transcription est au cœur de ce processus. Chez le poisson zèbre, nous savons maintenant que suffisamment d'activateurs de transcription doivent être accumulés dans l'embryon pour que l'activation du génome zygotique (ZGA) ait lieu. En conséquence, les concentrations d'histones formant des nucléosomes dans le noyau doivent s'épuiser pour permettre aux TF d'accéder à leurs sites de liaison. Il a été démontré que cette relation dynamique détermine largement le moment de l'AGZ. Cependant, l'AGZ est un processus graduel et temporellement ordonné par lequel des régions génomiques spécifiques peuvent s'activer plusieurs cycles cellulaires plus tôt que d'autres. Chez le poisson zèbre, le locus génomique mir430 est le premier gène à être activé pendant le développement de l'embryon. Il est intéressant de noter que ce locus génomique a la capacité unique de recruter de grandes quantités d'activateurs transcriptionnels nucléaires et de machinerie transcriptionnelle. On ne sait toujours pas comment ce locus peut s'activer si tôt au cours de la ZGA et comment cela est lié à sa forte capacité à recruter des activateurs et une machinerie transcriptionnels. Ici, grâce à des caractérisations approfondies et au séquençage ciblé de longues lectures, je montre que le locus mir430 est unique en ce sens qu'il est extrêmement répétitif. Contrairement à la représentation longue de 16 kbp du locus mir430 sur le génome de référence, j'ai découvert que ce locus a en fait une taille d'au moins 150 kbp (et potentiellement encore plus grande) consistant principalement en des répétitions stéréotypées de 1,7 kbp. Ces répétitions contiennent des sites de liaison de TF fonctionnels et je montre que, collectivement, ils coopèrent pour assurer l'activation précoce du locus mir430. Les loci mir430 isolés avec un nombre inférieur de répétitions perdent cet avantage compétitif et ne sont activés que plus tard au cours du développement, bien qu'ils aient des séquences identiques. D'un point de vue mécanique, je montre que l'avantage compétitif conféré par un plus grand nombre de répétitions est une plus grande sensibilité à la liaison des TF, probablement en raison du plus grand nombre de sites de liaison des TF. Dans le contexte de la relation compétitive entre les histones et les TF pour l'accès à l'ADN, un locus méga-répétitif comme mir430 pourrait faciliter la compétition localisée des TF contre les histones, malgré l'environnement nucléaire généralement répressif. Ces résultats permettent de mieux comprendre quelles caractéristiques génétiques peuvent définir le moment où un gène s'active au cours du développement et constituent probablement un moyen très généralisable de comprendre comment les TF interagissent avec leurs sites de liaison cibles.

1 Table of Contents

Table of Figures

Introduction

1.1 Temporal control of biological events

Well-timed processes occur across many different scales in biology. Precision in timing these processes ensures the development, physiology, and survival of organisms. On broader timescales, biological clocks maintain the circadian rhythms that underlie organismal behaviour during day and night states. On finer timescales, molecular timers can direct bursts of gene activity that occur over longer or shorter periods, with higher or lower frequencies. With advancements in techniques to visualise and perturb these well-timed processes, we now know that they are governed by elaborate networks of interactions between molecules which culminate in a temporal order. Changes in the numbers, types, chemical structures and interactomes of these molecules would, therefore, directly impact downstream events. In this way, well-timed processes are not measured in absolute time but rather, the temporal order that is established by the underlying molecular interactions. Achieving this temporal order is further complicated by stochasticity in these molecular interactions - deriving both from intrinsic and extrinsic noise. Thus, in addition to being elaborate, these molecular interactions need to be robust.

Gene transcription in the nucleus is regulated via complex interactions between *cis* or *trans* regulatory elements in the genome and, regulatory molecules such as the transcriptional machinery, transcription factors (TFs) and nuclear architectural proteins. These groups of molecules and genomic elements can come together to orchestrate a temporal order of transcription too. This is best illustrated during embryonic development, where the embryo starts off as a single cell that is transcriptionally inactive. Over time, as cell divisions occur and transcription programs are put in place, the embryo develops into a multicellular organism with complex body plans and cell identities. This developmental process relies on the ability of interactions between the genome and regulatory molecules to express the right genes at the right time and place during development. However, even prior to establishing complex transcriptional programs, the zygote must first overcome a major obstacle – turning on its own genome. This developmental checkpoint, often referred to as zygotic genome activation (ZGA), is a well-timed event within species. We now know that the timing of ZGA is broadly regulated by interactions between the genome, histone proteins and TFs. These interactions, however, only go so far to explain how a developing embryo achieves temporal regulation of gene activation.

In this thesis, I will explore how information encoded in the genome can engage with TFs to ultimately result in a temporal order of gene activation during ZGA in zebrafish. To do so, I will first introduce the core concepts that regulate transcription. Next, I will provide a state of the art on ZGA and the proposed mechanisms that control it. This will be followed by a more detailed summary of our current knowledge of zebrafish ZGA. Finally, I will discuss the work done for this thesis to identify core principles that define temporal control of gene activation during zebrafish ZGA.

1.2 Transcriptional regulation – a canonical perspective

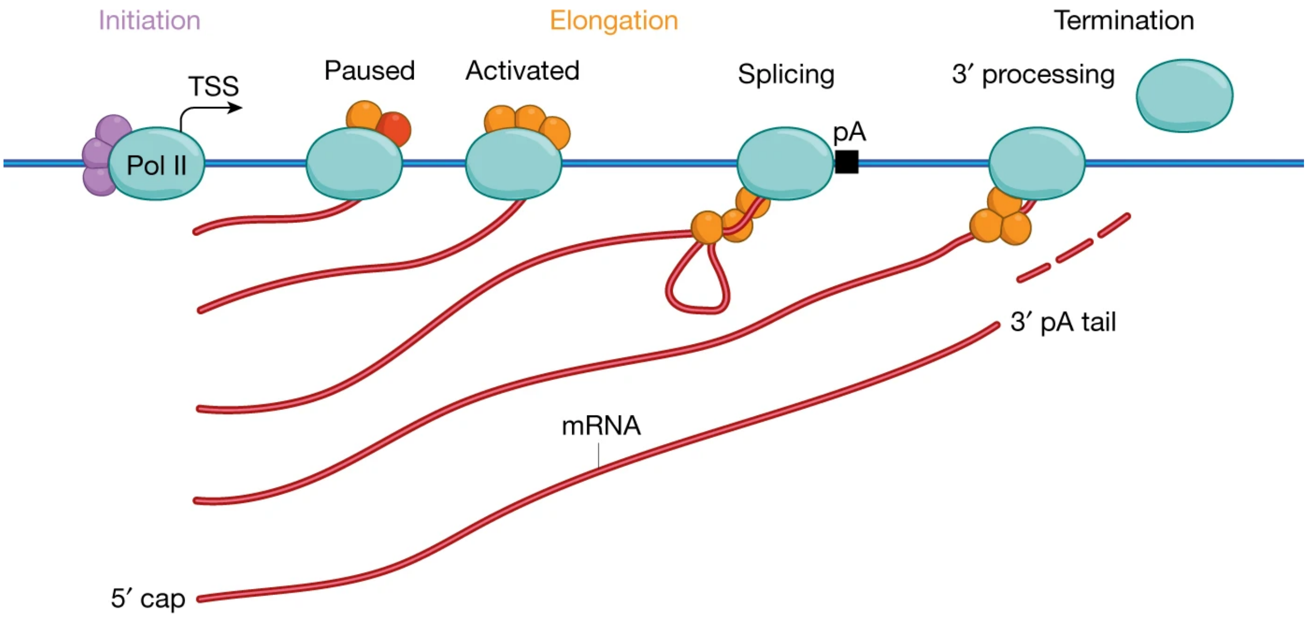

Figure 1.2.1: The transcription cycle (adapted from Cremer (2019)

Transcription is the process by which RNA polymerase complexes read the DNA template while synthesising the complementary mRNA molecule. This process of transcription occurs over multiple steps that are well-characterised and, involve a diverse suite of proteins each with distinct functions (Fig. 1.2.1). In this section, I will focus mainly on the current understanding of eukaryotic RNA pol II transcription since this is the major polymerase that transcribes coding genes.

For transcription to initiate, transcription initiation factors and the RNA polymerase II complex must assemble as a pre-initiation complex (PIC) on gene promoters. Transcription initiation factors consist of TFII proteins A-H (Fig. 1.2.2). Studies into the structures of PICs have shown that the complex formed by the transcription initiation factors function as a bridge between the promoter DNA and the RNA pol II complex (Kostrewa et al., 2009). In addition, transcription initiation factors also associate with TATA-binding proteins (TBPs) which binds to the TATA-box present on core promoter sequences to position the PIC upstream of the TSS (Cramer, 2019). The stabilisation of the PIC at gene promoters further relies on integrated signals from specific TFs which bind at either upstream *cis*-regulatory elements (CREs) or distal enhancers (Fig. 1.2.2). Rather than playing a general role in gene transcription, specific TFs activate limited sets of genes, depending on the presence of their cognate TF motifs. Specific TFs bound at CREs or distal enhancers contact mediator which in turn directly associates with TFIIB and TFIIH to stabilise the PIC (Cramer, 2019). In this way, mediator functions as a bridge between the PIC and specific TFs (Soutourina, 2017). The resultant stabilised PIC engaged on the gene promoter also induces unwinding of promoter DNA via the TFIIH factor (Cramer, 2019).

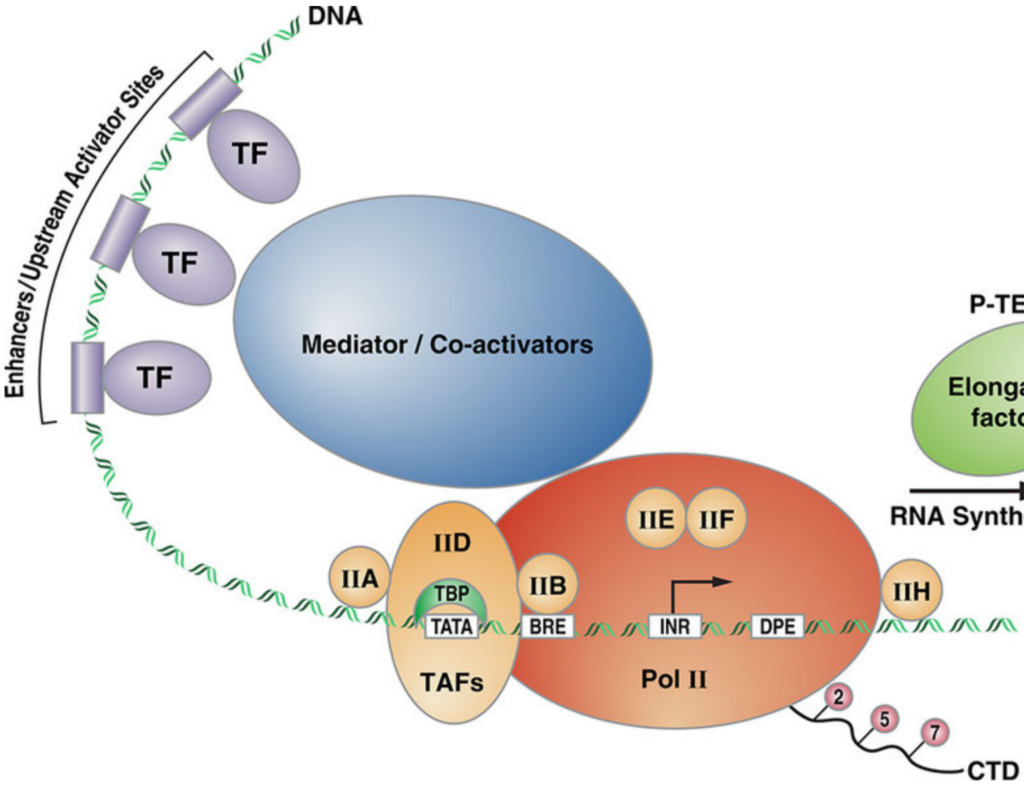

Figure 1.2.2 The pre-initiation complex (Adapted from **Gottesfeld, 2019)**

The pre-initiation complex involves the assembly of multi protein complexes on the core promoter. This involves transcription initiation proteins TFIIA-H. Additionally, the mediator complex acts as a bridge between TFs binding at distal enhancers and the PIC. This interaction stabilises the PIC at the core promoter. When transcription is ready to start, the protein P-TEFb mediates phosphorylation of the RNA pol II CTDs and factors that regulate Pol II pausing.

For the engaged gene to be transcribed, the PIC-RNA pol II complex must enter the elongating phase. This transition from initiating RNA pol II to elongating RNA pol II is also a regulated multistep process. The initiated RNA pol II is first cleared from the promoter – it transcribes along the DNA template, synthesising a short mRNA molecule 20-100 nucleotides long. Promoter clearance requires the CDK7 phosphorylation of the Ser5 residue on the heptapeptide repeats (Tyr1-Ser2-Pro3-Thr4-Ser5-Pro6-Ser7) of the RNA pol II C-terminal domain (CTD), resulting in weakening of the interaction between mediator and the RNA pol II complex (Jeronimo and Robert, 2014; Wong et al., 2014). Following the synthesis of this short transcript, the RNA pol II complex is stalled – a process known as promoter-proximal pausing. Promoter-proximal pausing has come to be appreciated as an obligate step during transcription and, the levels of pausing, as observed by RNA pol II pileup at promoters, are likely representative of the equilibrium between transcription initiation rates and pause-release rates (Core and Adelman, 2019). Pause-release of the RNA pol II requires the phosphorylation of the pause-inducing proteins DSIF and NELF by CDK9 (Fujinaga et al., 2023). At the same time, CDK9 also phosphorylates the Ser2 of the heptapeptide repeats on the RNA pol II CTD (Fujinaga et al., 2023). These events transition the paused RNA pol II into an elongation competent form, allowing productive transcription to begin.

Transcription typically terminates upon the detection of a polyadenylation signal (PAS; AAUAAA) on the mRNA by termination-associated proteins CPSF and CstF (Porrua and Libri, 2015). A long-standing 'torpedo' model proposes that following detection, cleavage occurs at the PAS site of the mRNA and polyadenylation occurs on the mRNA 3' end. The remaining run-off mRNA still associated with the RNA pol II is degraded by the XRN2 exonuclease which continues to kick-off RNA pol II. Other models propose that conformational changes upon PAS detection result in the dissociation of RNA pol II. The exact mechanism by which termination occurs following PAS detection remains unclear.

1.3 The maternal-to-zygotic transition (MZT)

Early embryonic development is characterised by a stepwise progression of events both at the morphological and at the molecular level. While the absolute timings of these events vary from organism to organism, common themes are conserved. In morphology, embryos undergo synchronous cell divisions, gastrulation and eventually the formation of the basic multi-cellular body plan. At the molecular level, early embryogenesis is characterised by an absence of transcriptional activity in the zygotic genome. Development of the embryo during this time is highly dependent on maternally loaded gene products, such as mRNAs and proteins. Over time, these maternally loaded gene products are either degraded or titrated. The clearance of maternal gene products coincides with the timing at which the zygotic genome gradually becomes transcriptionally active – an event termed zygotic genome activation (ZGA). This entire process has been aptly named the Maternal-to-zygotic transition (MZT); whereby developmental control is passed on from mother to zygote. In this introductory chapter, I will provide a historical perspective on MZT discoveries and how it has led to our current understanding of this complex transition. I will also discuss proposed models to explain the timing of ZGA in different organisms and, how ZGA is regulated at the different levels of genome organisation.

1.4 A historical perspective of the MZT

Many of the earliest discoveries relating to the MZT were found in non-mammalian model organisms. In the South African clawed toad (*Xenopus laevis*), researchers observed that the *Xenopus* embryos underwent 12 rounds of rapid and synchronous divisions succeeded by a period of asynchronous divisions where cell cycle lengths showed larger variation and became longer (Graham and Morgan, 1966; Newport and Kirschner, 1982a). Similar observations were also made before in sea urchin and drosophila embryos (Hinegardner et al., 1964; Rabinowitz, 1941). Early biochemical characterisations of the MZT were done by measuring the levels of incorporation of radioactively labelled nucleotide precursors into nascent RNA transcripts (Brown and Littna, 1964; Emerson and Humphreys, 1970; Hinegardner et al., 1964; Newport and Kirschner, 1982b; Zalokar, 1976). These experiments showed that the first instances of RNA synthesis in the embryo coincide with the timings when the cell cycles become asynchronous. In drosophila, direct visualisation of RNA synthesis by RNA polymerases on the DNA template upon ZGA was observed using transmission electron microscopy on drosophila embryo chromatin spreads (McKnight and Miller, 1976). These findings converged on the idea that many cellular changes occur at a defined timepoint during embryonic development, generally termed the midblastula transition (MBT). These changes include cell cycle lengthening, cell cycle asynchrony, gains in cell motility and activation of transcription.

Here, an important distinction needs to be made between the MZT and the MBT. The MZT refers to a continuous phase throughout early development whereby developmental control is passed from mother to zygote. The MBT refers to a precise timepoint during embryonic development when changes in cell cycle dynamics, motility and transcription occur.

The discovery of the MBT complemented studies showing that RNAs inherited via the oocyte remain in the embryo during the early stages of embryogenesis but eventually are lost at the MBT (Crippa et al., 1967; Humphreys, 1971; Sagata et al., 1980). These maternally provided mRNAs contribute heavily to protein synthesis throughout pre-MBT stages (BRAVO and KNOWLAND, 1979; Humphreys, 1969). The eventual loss of maternal RNAs indicated a timed mechanism that ensures clearance of maternal gene products. On a broader scale, the MBT represented a change in regime during embryo development from maternal control to zygotic independence. Extensive research has since followed to understand the molecular sequence of events that determine when this transition occurs in various metazoic organisms following fertilisation. The cumulative findings have established that the timing of the MBT varies across species. In *Xenopus*, the MBT occurs 7 hours post fertilisation (hpf). The same is observed at 3 hpf in *Drosophila* and zebrafish, and around 20 hpf for sea urchins. Despite these differences, they all follow similar stepwise phases of maternal control leading into zygotic independence.

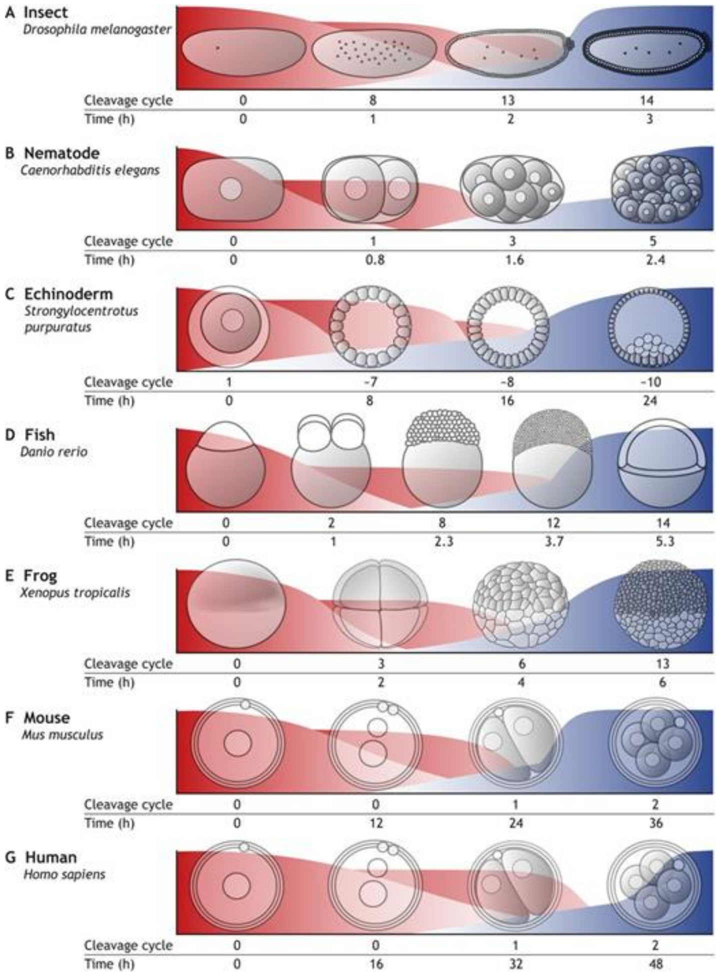

Figure 1.4.1 The maternal-to-zygotic transition (Adapted from Vastenhouw et al, 2019)

The maternal-to-zygotic transition generally begins with a transcriptionally silent embryo. During this time, the embryo is loaded with maternal mRNAs (in red) that direct development. Over time, maternally loaded mRNAs are degraded and zygotic transcription gradually begins. The timing of this transition varies across various organisms.

1.5 Maternal control

The embryo inherits large amounts of maternal gene products, in the form of mRNAs and proteins, from the oocyte. Maternal mRNAs encode important proteins that are required for the faithful development of the embryo, including proteins involved in transcription regulation, cell cycle progression and DNA replication (Aanes et al., 2011; Chan et al., 2019; Eckersley-Maslin et al., 2019; Harrison et al., 2011; Iaco et al., 2019; Lee et al., 2013a; Leichsenring et al., 2013a; Liang et al., 2008a; Veenstra et al., 1999a). In the absence of zygotic transcription, the embryo depends on the maternal mRNA derived proteins to carry out the respective processes. Inhibition of translation early during embryogenesis by cycloheximide treatment in drosophila, xenopus and zebrafish embryos precludes ZGA (Chan et al., 2019; Chen and Good, 2022; Edgar and Schubiger, 1986a; Lund and Dahlberg, 1992). This means that the machinery required for the activation of the zygotic genome is encoded in the maternal mRNAs and need to be expressed sufficiently early for ZGA to happen.

While maternal mRNAs need to be translated at the right time, they also need to be cleared subsequently. Interestingly, maternal mRNAs clearance is partly accomplished by zygotic gene products. In zebrafish, the microRNA miR430 is zygotically transcribed and, functions to degrade at least 40% of maternally loaded transcripts (Giraldez et al., 2006). Similar functions of microRNAs have been detected in *Drosophila* by the zygotically transcribed miRNA miR309 (Bushati et al., 2008). In addition to zygotically transcribed mechanisms of maternal mRNA clearance, studies have also found maternally encoded factors that adopt this role too. In *Drosophila*, clearance of maternal transcripts is also partly driven by the proteins smaug and PAN GU (PNG), which together destabilise maternal mRNAs via poly(A) tail shortening (Eichhorn et al., 2016; Tadros et al., 2007). In both zebrafish and *xenopus*, the terminal uridylyltransferases TUT4 and TUT7 serve to uridylate short-tail maternal mRNAs to target them for degradation (Chang et al., 2018).

These studies together highlight that maternally loaded mRNAs are vital for producing the necessary machinery for ZGA and normal embryo development. However, they function exclusively during the pre-MBT stages, after which, mechanisms encoded both in the maternal mRNAs and the zygotic genome ensure timely clearance of maternally loaded mRNAs.

1.6 Zygotic independence

The hallmark of zygotic independence is the ability of the zygote to begin transcribing its own genome. From a transcriptional regulation perspective, the switch of the zygotic genome from an inactive state to an active state provides a unique opportunity to study the mechanisms that regulate transcription in general. Early studies in metazoan embryos showed that ZGA begins at the MBT, where injected radioactively labelled nucleotides were shown to be incorporated into nascent transcripts (Brown and Littna, 1964; Emerson and Humphreys, 1970; Hinegardner et al., 1964; Newport and Kirschner, 1982b; Zalokar, 1976). These findings, however, merely show presence of transcription but lack information about identity of the zygotic genes and, the transcriptionally active fraction of the genome it represents.

Modern RNA sequencing approaches have gone above-and-beyond in identifying these ZGA genes with higher sensitivity and time resolution. These findings uncovered that in many model organisms, ZGA happens even prior to the MBT. Furthermore, they provide a decisive view that zygotic genes do not all turn on at a single time point. Rather, the zygotic genome activates gradually, with specific genes being turned on at distinct timepoints during development (Bhat et al., 2023; Collart et al., 2014; Heyn et al., 2014; Lott et al., 2011; Mathavan et al., 2005; Owens et al., 2016; Sandler and Stathopoulos, 2016; White et al., 2017). In zebrafish, the earliest known timepoint of zygotic transcription is at the 64-cells stage (2 hpf) where the microRNA gene cluster mir430 is transcribed (Hadzhiev et al., 2019; Heyn et al., 2014; Kuznetsova et al., 2023; Lee et al., 2013a). In *Xenopus*, ZGA begins around the 64/128-cells stage (6 hpf) with the earliest known zygotic gene being those involved in mesendoderm induction (Skirkanich et al., 2011). In *Drosophila*, embryos do not cellularise until 3 hpf. Prior to this time, the nuclei divide within a syncytium and ZGA is known to occur around 1-2 hpf (De Renzis et al., 2007; Edgar and Schubiger, 1986b; Kwasnieski et al., 2019). Finally, ZGA in mammals, such as mice and humans, occurs much later following fertilisation – just before 24 hpf when the embryo is still in its 1-cell stage.

An issue of contention within the field of ZGA is whether ZGA occurs in waves – historically referred to as the minor and major waves. The idea of multiple waves likely came about because of the differences in sensitivity of zygotic transcription detection methods between early and recent studies. Many early studies, such as that of Newport and Kirchner (1982b), refer to the first time point when incorporation of radioactive nucleotides could be observed in nascent transcripts as the timing of ZGA, which coincides with the MBT. The identification of zygotic transcription prior to the MBT using modern sequencing approaches led researchers to define transcription during the pre-MBT cleavage stages as "minor wave" and, transcription during post-MBT stages when cell cycles become asynchronous as the "major wave". At least in the case of organisms such as *Xenopus*, *Drosophila*, and zebrafish, splitting ZGA into 2 waves is arbitrary given that highly sensitive sequencing approaches have shown that gene activation occurs in a continuum throughout development (Bhat et al., 2023; Chen and Good, 2022; Heyn et al., 2014; Kwasnieski et al., 2019; Vastenhouw et al., 2019; White et al., 2017).

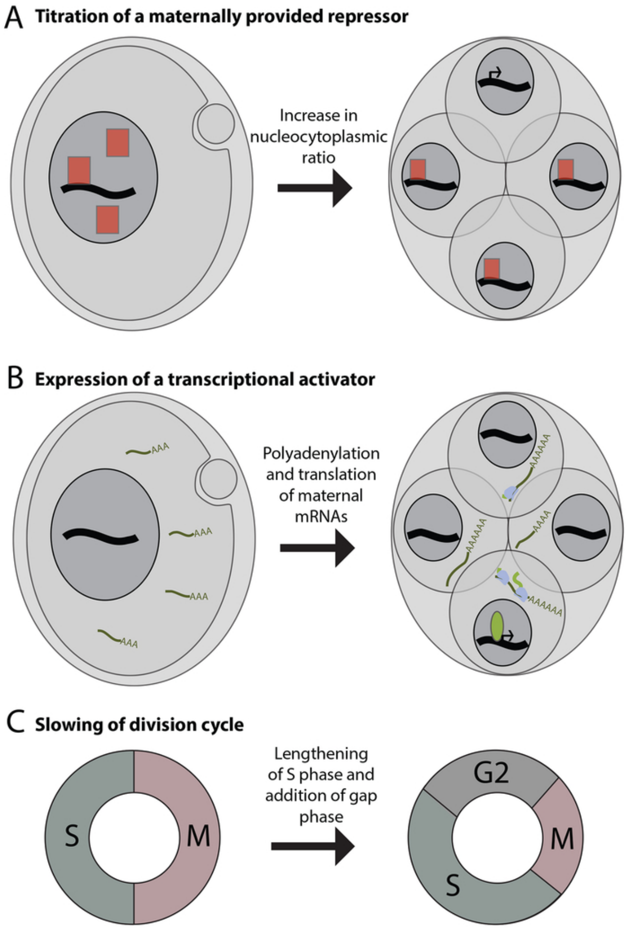
A Titration of a maternally provided repressor
Increase in nucleocytoplasmic ratio
B Expression of a transcriptional activator
AAA
AAA
AAA
AAA
Polyadenylation and translation of maternal mRNAs
AAAAAA
AAAA
AAAAAA
AAAAAA
C Slowing of division cycle
S
M
Lengthening of S phase and addition of gap phase
G2
M
S

Figure 1.6.1: The different models of ZGA (Adapted from Schulz and Harrison, 2019)

A: The nucleocytoplasmic ratio model posits that an excess repressor is initially abundant in the embryo. Over time, increase in DNA amounts with each cell cycle titrate out this excess repressor, allowing transcription to begin.
B: Maternally loaded mRNAs encode transcriptional activators. When sufficient amounts have been accumulated, transcription can begin. Not shown here is the histone-TF competition model, which draws elements from both A and B.
C: During early embryo development, cell cycles are rapid and may not allow sufficient time for elongative transcription. At the MBT, when cell cycles length, transcription may have more time to proceed.

1.7 Mechanisms of ZGA

What mechanisms determine the onset of ZGA? Past models proposed involve either the relief of transcription repression or the gain of transcription competence (Fig. 1.6.1). Cumulative findings over many decades have shown that these models are not independent from one another. Rather, they provide complementary views of ZGA regulation from different perspectives. In this section, we will explore how ZGA is regulated from these different perspectives using previously proposed models.

1.7.1 The Nuclear-to-cytoplasmic ratio model

Perhaps the earliest proposed model to explain ZGA onset is the increase in the nuclear-to-cytoplasmic ratio (N/C) (Newport & Kirschner, 1982b, 1982a). Formulation of this model came from 2 experiments in *Xenopus* embryos: 1) Polyspermic embryos (embryos fertilised by multiple sperms and thus carrying more nuclear content) start ZGA earlier, 2) Injection of an excess of exogenous plasmids into the embryo could result in premature ZGA onset (Newport and Kirschner, 1982a, 1982b).These two findings proved that the DNA content in the embryo could influence the timing of ZGA. It was subsequently proposed that during early embryogenesis, the increasing amounts of DNA synthesised with each division could titrate a transcriptional repressor present in the cytoplasm (Newport and Kirschner, 1982a, 1982b). Since these findings, many studies have gone on to show that the nuclear DNA-to-cytoplasmic ratio can regulate the activation of zygotic genes (Chan et al., 2019; Edgar et al., 1986; Jukam et al., 2021; Lu et al., 2009; Syed et al., 2021).

The concept that a threshold amount of DNA was required for ZGA onset made histones an obvious candidate in the identification of the transcriptional repressor (Fig. 1.6.1 A). Indeed, high levels of histones are present in the early embryo (Adamson and Woodland, 1974; Anderson and Lengyel, 1980). Histones have also been shown to be repressive for zygotic transcription in various organisms (Chari et al., 2019; Joseph et al., 2017a; Prioleau et al., 1994; Syed et al., 2021). Thus, in-line with the N/C ratio model, histones (being the excess repressor of the model) initially prevent transcription. Further developments of the N/C ratio model have proposed that rather than just passive titration of histones, active out-competition of histones by transcriptional machinery is required for transcriptional activation to occur (Almouzni & Wolffe, 1995; Joseph et al., 2017; Prioleau et al., 1994). Interestingly, a study in zebrafish has shown that the amount of histones that would be titrated out by DNA at ZGA onset make up only a small fraction of the massive amounts of soluble histones present in the early embryo (Joseph et al., 2017). Thus, titration of histones by DNA as proposed in the original N/C ratio model is unlikely to be the only mechanism that drives transcription during ZGA.

Other potential candidates for transcriptional repressors of ZGA have been proposed in *Drosophila*. The transcription factor *tramtrack (ttk)* was found to repress transcription of zygotic genes such as *fushi tarazu* (*ftz*) (Brown and Wu, 1993; Pritchard and Schubiger, 1996). Altering levels of *ttk* was able to advance or delay transcription of *ftz* (Brown and Wu, 1993; Pritchard and Schubiger, 1996). The specific nature of *ttk* in zygotic gene regulation, however, makes it an unlikely candidate for a general repressor of zygotic transcription (Pagans et al., 2002).

An emerging idea that relates to the nuclear DNA-to-cytoplasmic ratio model is that the nuclear volume to cytoplasmic volume ratio also regulates ZGA onset. In early embryos, cells typically divide rapidly without changes in the absolute size of the embryo, resulting in ultimately smaller cells and an increase in the nuclear-to-cytoplasmic *volume* ratio. Recently, it was shown that artificially increasing the nuclear-to-cytoplasmic volume ratio can lead to premature ZGA (Jevtić and Levy, 2017, 2015). A separate study also found that a cell size threshold is predictive of ZGA at a single-cell level in *Xenopus* embryos (Chen et al., 2019). How exactly cell/nucleus sizes affect DNA amounts or nuclear histones levels to influence ZGA remains unclear.

1.7.2 Developmental timer model

The developmental timer model of ZGA posits that ZGA onset occurs when enough time has passed, following fertilisation, to allow the progression of a sequence of biochemical events. One such possibility is the accumulation of sufficient levels of transcriptional activators. In zebrafish, maternally loaded mRNAs encoding pluripotency associated factors such as Nanog, Pou5f3(Oct4) and Sox19b are highly translated pre-ZGA and, sufficient amounts of these TFs need to accumulate for the activation of the zygotic genome (Lee et al., 2013a; Leichsenring et al., 2013a). Similar findings have been shown for chromatin remodellers, such as the p300 histone acetyltransferase (HAT) and Brd4 (Chan et al., 2019). Overexpression of these factors results in earlier ZGA, suggesting that a threshold amount of transcriptional machineries need to be present for ZGA to begin (Chan et al., 2019; Joseph et al., 2017). Similarly, in *Xenopus* embryos, sufficient levels of the TATA-binding protein (TBP) need to be synthesised for zygotic transcription to begin (Veenstra et al., 1999a). In these cases, the rate of synthesis of the transcriptional activators is, at least in part, the limiting factor for ZGA onset.

Other molecular 'timers' that begin at fertilisation have also been shown. For instance, in *Drosophila* embryos, the protein BRAT binds to the 3' UTRs of maternal mRNAs to repress translation (Larson et al., 2022; Sonoda and Wharton, 2001; Wharton and Struhl, 1991).

Phosphorylation of BRAT by the PNG kinase releases the mRNAs from repression, and this mechanism has been shown to regulate when maternally loaded mRNAs encoding the protein Zelda is translated (Larson et al., 2022). In *Drosophila,* the protein Zelda functions as the master activator of ZGA (Harrison et al., 2011, 2010; Liang et al., 2008a; Nien et al., 2011). This sequence of events thus time when *Drosophila* ZGA begins. Interestingly, it was also found that loss of BRAT-repressor activity and therefore, earlier translation of Zelda, did not result in a corresponding advancement of ZGA (Larson et al., 2022). Thus, while synthesis of transcriptional activators is a key regulator of ZGA, other mechanisms like those presented in the other ZGA models could be limiting factors too.

1.7.3 Cell cycle lengthening model

Embryogenesis in many non-mammalian model systems involve a series of rapid and synchronous cleavage divisions. Transcription onset in the past was identified to coincide with the MBT, when cell cycles lengthen and become asynchronous. This led to the idea that the rapid divisions in the early embryo prevent sustained elongating transcription. In support of this idea, studies in *Xenopus* have shown that lengthened cell cycles result in premature ZGA (Collart et al., 2013; Kimelman et al., 1987). Moreover, many of the earliest transcribed genes tend to be short and intron-less, consistent with a need for succinct transcripts (Heyn et al., 2014). The cell cycle model, however, is not generalisable, as the lengthening of the cell cycle did not result in premature ZGA in zebrafish (M. Zhang et al., 2014). In fact, the inverse has been shown in *Drosophila*, whereby zygotic transcription itself is required for inducing cell cycle lengthening at the MBT (Blythe and Wieschaus, 2015; Farrell and O'Farrell, 2013). Thus, the effect of cell cycle lengthening on ZGA may only apply to *Xenopus.*

1.7.4 New frontiers

Recent studies have proposed a novel mode of ZGA regulation via the import of proteins into the nucleus. In both zebrafish and *Xenopus* embryos, studies showed that the timing of recruitment of proteins such as TFs into the nucleus was consistent with their nuclear activity (Nguyen et al., 2022; Shen et al., 2022). For example, nuclear import of the ZGA regulator Nanog in zebrafish embryos increases steadily from the 64-cells stage onwards, when the zygotic genome is known to be active (Shen et al., 2022). The two studies cited above differed, however, in the proposed mechanism by which temporally regulated nuclear import of proteins is achieved. In Shen et al (2022), the authors propose that comprehensive maturity of the nuclear pore complex (NPC) determines the identity of imported proteins. Whereas in Nguyen et al (2022), the authors propose that differential importin affinities determine which cytoplasmic proteins are imported first.

Other novel models of ZGA regulation have also been shown in *Ciona* embryos whereby the de-repression of FGF signalling gives way to zygotic transcription (Treen et al., 2023). While research in both the above-mentioned fields are in their infancy, they provide exciting new perspectives on how ZGA onset may be regulated.

1.8 Master regulators of ZGA

Sufficient studies have been done to say with reasonable certainty that histones play the role of the 'transcriptional repressors' described in the original N/C ratio model by Newport and Kirchner (1982b). As a counterpart to the repressive function of histones, we now also know that transcription factors play a central role in the activation of the zygotic genome. The exact identity of these TFs varies between species. In this section, we will look at the master activators of ZGA identified in different model organisms and the mechanisms by which they are known to act.

1.8.1 Drosophila

The first identified master TF for ZGA was Zelda in *Drosophila*. Zelda has been shown to bind at TAGteam sites (CAGGTAG) proximal to many early zygotically transcribed genes to induce transcriptional activation (Harrison et al., 2011; Liang et al., 2008; Nien et al., 2011). In addition to activating early zygotic genes, Zelda binding at enhancers can induce chromatin accessibility, facilitating the later binding of other patterning-associated TFs such as Dorsal (Dl) and Bicoid (Bcd) (Foo et al., 2014; Mir et al., 2017; Sun et al., 2015; Yáñez-Cuna et al., 2012). By regulating the recruitment of these downstream factors, Zelda can regulate transcription activity of genes associated with its target enhancers (Yamada et al., 2019). These findings led to Zelda being proposed to be a pioneer factor. Pioneer factors are factors which have an intrinsic ability to engage their TF binding sites in regions of 'closed' chromatin to induce chromatin accessibility (Zaret, 2020). These regions are typically inaccessible to other TFs which are unable to overcome the physical barrier posed by the nucleosome.

Recently, Zelda was found to form subnuclear clusters that are chromatin-bound (Mir et al., 2018). These subnuclear Zelda clusters colocalise with Bcd, consistent with its known role in facilitating binding of other TFs at enhancers (Mir et al., 2018, 2017; Yamada et al., 2019). Given the role of Zelda in regulating ZGA, high local concentrations of Zelda in these subnuclear clusters may be required for it to carry out its transcriptional activator roles and/or maintain accessibility at enhancers to allow binding of other TFs.

Other novel activators of ZGA with pioneering activities have also been identified. The GAGA factor (GAF) was found to function both synergistically with Zelda to activate early zygotic genes, and independently to induce widespread genome activation at the MBT (Gaskill et al., 2021). Interestingly, GAF was also found to form subnuclear protein clusters in early embryonic nuclei. Unlike the Zelda clusters, GAF clusters do not seem to have a transcriptional activating role but rather, function to silence satellite repeats (Gaskill et al., 2023a). Additionally, another factor CLAMP which binds to similar motifs as GAF was also found to function together with Zelda to activate zygotic transcription (Duan et al., 2021). Thus, substantial work has been done in *Drosophila* to identify the major activators of ZGA.

1.8.2 Zebrafish

In zebrafish, the pluripotency associated factors Nanog (N), Pou5f3 (P; more commonly known as Oct4), and Sox19b (S) have been shown to be required for ZGA onset (Lee et al., 2013; Leichsenring et al., 2013a). They operate either all together, in specific combinations, or independently to induce chromatin accessibility and activate zygotic genes (Gao et al., 2022; Kuznetsova et al., 2023; Miao et al., 2022; Pálfy et al., 2020; Riesle et al., 2023). These factors may have analogous functions to the master regulators identified in *Drosophila* as they were found to activate the transcription of the earliest zygotic transcripts (Lee et al., 2013a; Leichsenring et al., 2013a). In fact, many parallels may be drawn between NPS and the *Drosophila* activators. For instance, Nanog and Sox19b both form multi-factor subnuclear TF clusters that are associated with zygotic transcriptional regulation (Kuznetsova et al., 2023). Combinatorial or independent activities of NPS have also been shown to have pioneering activity - they are able to induce chromatin accessibility at regions that, without NPS activity, would be highly occupied by nucleosomes (Miao et al., 2022; Veil et al., 2019). Most notably, the combinatorial activity of NPS regulate the transcription of one of the earliest zygotic genes, mir430 (Lee et al., 2013). Of the 3 factors, only Nanog is indispensable for mir430 activation, suggesting that there is a hierarchy of pioneering activity, at least at the mir430 locus (Kuznetsova et al., 2023; Lee et al., 2013a).

1.8.3 Humans and mice

In mammalian systems, a multitude of TFs activate the zygotic genome. In mice, these include the factors Dppa2, Dppa4, Nfy and Dux (Eckersley-Maslin et al., 2019; Iaco et al., 2019). In the case of Dppa2/4, they are required for early transcription of a cluster of genes encoding Dux proteins, a cluster of genes encoding Zscan4 proteins, LINE-1 elements and, MERVL elements (Eckersley-Maslin et al., 2019; Hendrickson et al., 2017; Iaco et al., 2019). Dppa2/4, however, are not directly responsible for bulk ZGA. Rather, the zygotically produced TFs Dux and Zscan4 activates and maintains transcription of many zygotic genes (De Iaco et al., 2017; Eckersley-Maslin et al., 2019; Hendrickson et al., 2017). Past work has also shown evidence that Nfy establishes chromatin accessibility at zygotic gene promoters (Lu et al., 2016). By the strict definition of 'pioneering activity', it remains unclear if these mouse ZGA factors are, in fact, pioneer factors (Zaret, 2020).

Recent work has identified a bona-fide pioneer factor, named Nr5a2, that contributes to mouse ZGA (Gassler et al., 2022). Via its pioneering activity, Nr5a2 directly binds and regulates ~72% of zygotic genes and overexpression of Nr5a2 induces premature ZGA (Gassler et al., 2022). Other recently identified mouse ZGA regulators include the OBOX proteins, which are partly provided as maternally loaded mRNAs (Ji et al., 2023). The OBOX proteins synthesised from these mRNAs activate hundreds of early ZGA genes, including Nr5a2 (Ji et al., 2023). Mechanistically, OBOX proteins bind to their target genes, induce chromatin accessibility, and recruit RNA Pol II in time for transcription at bulk ZGA (Ji et al., 2023). Despite the multitude of factors identified in mice, it remains unstudied whether these factors function cooperatively (as observed in the *Drosophila* and zebrafish activators) or independently. Furthermore, it is also not known if subnuclear clustering is a functional characteristic of these activators.

In humans, DUX4 (the ortholog of murine Dux), is also responsible for ZGA (De Iaco et al., 2017; Hendrickson et al., 2017). However, much less is known about other master regulators in human embryos for the obvious reason that samples are more difficult to acquire.

1.8.4 <u>Xenopus</u>

For the extensive knowledge derived about the MZT from work in *Xenopus* embryos, comparatively less work has been done to identify master regulators of ZGA. Past work has shown that sufficient TBP must be accumulated in the embryo for ZGA to begin (Veenstra et al., 1999). Other proposed specific TFs that regulate ZGA in *Xenopus* include FoxH1, VegT and Otx1 (Charney et al., 2017; Paraiso et al., 2019). These factors likely do so cooperatively by binding to their target sites on genes and enhancers, pre-marking them for future transcription (Charney et al., 2017; Paraiso et al., 2019).

In summary, there is dramatic divergence in the TFs that activate the zygotic genome across species. However, there is a common theme amongst these factors – they are required for establishing chromatin accessibility at their target sites, in preparation for transcription to occur. Clearly, the chromatin template is a major factor that influences transcriptional competence at ZGA.

1.9 Transcriptional regulation at multiple scales during the MZT

Genome architecture is closely tied with gene regulation. During the MZT, the genome develops structures that are defined by its interactions with TFs and transcriptional activity itself. In this following section, we will explore how the master regulators of ZGA restructure the genome, focusing on the multiple levels of genome organisation where they are known to act.

1.9.1 3D genome organisation

Topologically associated domains (TADs) are regions that tend to have high self-interacting frequencies. The formation of TADs along the linear DNA can keep promoters and enhancers apart to ensure gene silencing but can also bring them in close proximity of one another to allow target gene transcription (Rowley and Corces, 2018). Thus, TADs can play important roles in transcriptional regulation. Work in *Drosophila*, zebrafish and mice have shown that prior to ZGA, the genome is typically unstructured (Du et al., 2017; Hug et al., 2017; Ke et al., 2017; Ogiyama et al., 2018; Wike et al., 2021). Accompanying genome activation, TAD boundaries emerge independently of transcription (Hug et al., 2017; Wike et al., 2021). Rather, it is the activity of TFs such as Zelda and GAF in *Drosophila*, and the p300 HAT in zebrafish, that establishes TAD boundaries. Thus, TFs play a vital role in reshaping chromatin architecture during ZGA. The lack of TADs and general higher-order structures prior to ZGA further indicates that transcriptional quiescence prior to ZGA is unlikely to be due to a repressive chromatin architecture.

1.9.2 Chromatin accessibility

Eukaryotic genomes are packaged in the nucleus into chromatin fibres by association of the DNA with histone octamers – 2 each of H2A, H2B, H3 and H4. The resultant structures, nucleosomes, restrict the access of transcription factors that bind to the DNA template for transcription to occur. As established in the previous section, ZGA regulators such as Zelda and Nanog bind to their cognate TF binding sites and induce chromatin accessibility at early zygotic genes (Harrison et al., 2011; Miao et al., 2022; Pálfy et al., 2020; Sun et al., 2015; Veil et al., 2019). How exactly do these factors establish regions of local chromatin accessibility? Here, I will discuss 2 main modes by which these factors could open up the chromatin.

1.9.2.1 Chemical modifications of the chromatin

Histone subunits that make up the core nucleosome each have histone tails carrying lysine or arginine residues that can be subject to chemical modifications. Depending on the nature and position of these chemical modifications, they have been associated either with transcriptional repression or activity. For instance, H3K4 tri-methylation marks have long been associated with active promoters whereas H3K9 tri-methylation marks are typically associated with heterochromatin.

During the MZT, the landscape of these histone modifications change as the zygotic genome is activated. In zebrafish, *Xenopus* and *Drosophila* embryos, H3K4me3 marks are acquired at promoters of genes which are activated during ZGA (Akkers et al., 2009; Chen et al., 2013a; Hontelez et al., 2015; Lindeman et al., 2011; Vastenhouw et al., 2010; Zhang et al., 2018; Y. Zhang et al., 2014). In addition, histone tail acetylation marks such as H3K27Ac, H3K18Ac and many others are gained at enhancers and promoters (Bogdanović et al., 2012; Chan et al., 2019; Li et al., 2014; Miao et al., 2022; Zhang et al., 2018). H3K27Ac marks are typically associated with active enhancers. In the case of zebrafish and *Drosophila*, the deposition of acetylation marks at enhancers seems to be TF dependent. Loss of Zelda in *Drosophila* resulted in reductions in enhancer acetylation marks (Li et al., 2014). Similarly, in zebrafish, loss of NPS resulted in a reduction in acetylation marks and chromatin accessibility at enhancers (Miao et al., 2022). In addition, promoters also lost acetylation marks and transcriptional activity (Miao et al., 2022). These effects could be rescued by targeted recruitment of HAT activity to a genomic locus, even in the absence of NPS (Miao et al., 2022). These findings provide compelling evidence that histone modifications, deposited downstream of ZGA activators, establish chromatin accessibility to allow transcription to occur.

In addition to activating histone modifications, repressive histone marks also show up during ZGA (Akkers et al., 2009; Laue et al., 2019; Lindeman et al., 2011; Vastenhouw et al., 2010). Combinations of active and repressive histone marks can co-occur at 'bivalent' promoters and, may prepare a gene for future transcription (Vastenhouw et al., 2010).

Overall, histone modifications, deposited downstream of ZGA activators, create a more accessible chromatin environment for transcription to occur. Importantly, the global absence of repressive modifications prior to ZGA also suggests that the absence of transcription prior to ZGA is not due to the chromatin being kept in a repressive state (Vastenhouw et al., 2019).

1.9.2.2 *Nucleosome disruption*

The binding of pioneer factors onto their cognate binding sites alone can disrupt a stably formed nucleosome. As described earlier, pioneer factors have the intrinsic ability to overcome the barrier posed by the nucleosome to bind to their target sites. Independent of

ZGA, studies on murine pioneer factors Pou5f3/Oct4 and Sox2 show that these TFs can physically disrupt nucleosome stability by binding to their cognate binding sites near the entry/exit sites of nucleosomes (Michael et al., 2020). For master ZGA activators, work in zebrafish suggests that at regions regulated by NPS, Nanog and Pou5f3 binding to their motifs located within nucleosomes can destabilise the nucleosome (Veil et al., 2019). This destabilised nucleosome is more prone to subsequent NPS binding that maintains chromatin accessibility to allow for transcription (Veil et al., 2019). Indeed, multiple studies have reported that NPS act in synergy (Miao et al., 2022; Pálfy et al., 2020; Veil et al., 2019). Very similar findings have been reported in *Drosophila* whereby Zelda binding to their motifs within enhancers can evict nucleosomes (Sun et al., 2015). Interestingly, the extent of nucleosome depletion at enhancers was found to depend on the number of Zelda binding sites - more Zelda motifs resulted in stronger nucleosome depletion (Sun et al., 2015). This suggests that the cooperative activity of multiple Zelda binding sites could have a stronger effect on chromatin accessibility, and enhancer activity.

In both the above cases, Zelda or NPS binding to their motifs is strongly tied to the affinity of the underlying DNA sequence to assemble into a nucleosome. Rather counter-intuitively, high nucleosome affinity is predictive of stronger Zelda/NPS binding and increased chromatin accessibility following ZGA (Sun et al., 2015; Veil et al., 2019). In both studies, the authors proposed a mechanism by which the occurrence of several TF binding sites (for one or multiple TFs) could facilitate a form of cooperativity between TFs to evict the nucleosome. Such a mechanism of nucleosome-mediated cooperativity has been previously described computationally (Mirny, 2010), and experimentally (Adams & Workman, 1995; Sönmezer et al., 2021). Of note, previously reported cases of nucleosome-mediated cooperativity often involve 'master regulators' of biological processes (Lupo et al., 2023; Sönmezer et al., 2021; Sun et al., 2015; Veil et al., 2019). Potentially, for nucleosome-mediated cooperativity to work, at least 1 TF with pioneering activity is required.

1.9.3 Promoter encoded information that defines early transcription

The DNA encodes valuable information that determines a genes activity ZGA. The importance of DNA encoded information has been alluded in the earlier sections where we have seen that specific TF binding sites can work together to alter chromatin structures via their interactions with TFs. In the context of the gradual activation of the zygotic genome, DNA encoded information can determine when genes are transcribed along this continuum. Here, we will delve deeper into regulation at the DNA level by looking at TF-DNA interactions in the context of ZGA.

The story of how Zelda was discovered as a master activator of ZGA began with the discovery that one of earliest zygotic genes that are transcribed in the zygotic genome encode the sex-determining genes *sis-a* and *sis-b* (Erickson and Cline, 1998, 1993). Cline and colleagues observed that *sis-a* and *sis-b* began to be transcribed as early as the 8th cleavage division of *Drosophila* embryogenesis, 6 nuclear cycles before the MBT (Erickson and Cline, 1993). To understand how these genes may be co-regulated, Cline and colleagues looked for similarities in the promoter sequences of the two genes and, found what we now know to be the TAGteam motif which Zelda binds to (Erickson and Cline, 1998). The TAGteam motif was further identified in the promoters of many early zygotic genes and, the discovery of the protein Zelda came soon after (Liang et al., 2008a; ten Bosch et al., 2006).

This more than a decade long discovery story highlights that early transcription during ZGA follows a specific promoter code. Subsequent work has shown that Zelda binding sites alone do not determine early transcription. Recently, a systematic dissection of the Zelda-regulated *hunchback (hb)* promoter found that the presence of a Zelda binding site and a TATA-box at the *hb* promoter is required for early activation (Ling et al., 2019). This, however, is not a generalisable rule for early transcription. While many early genes were found to carry a TATA-box, not all had a corresponding proximal Zelda binding site (K. Chen et al., 2013; Ling et al., 2019). It is, therefore, possible that other TFs and combinations of binding sites at promoters also drive early activation.

Different combinations of TF binding sites at promoters may also distinguish early and late activation times. For instance, promoters regulated by CLAMP and Zelda in combination are activated mid – late ZGA (Colonnetta et al., 2021; Duan et al., 2021). Whereas independent GAF promoters regulate late ZGA genes (Gaskill et al., 2021).

In zebrafish, we now know that early genes are regulated by NPS, most notably the mir430 gene cluster (Lee et al., 2013). Here, NPS activate mir430 expression as early as the 64-cells stage of zebrafish embryogenesis. However, NPS motifs alone are unlikely the sole regulators of early genes since many later genes are also regulated by NPS (Lee et al., 2013; Miao et al., 2022; Pálfy et al., 2020; Veil et al., 2019). Recently, it has been proposed that early activated promoters during zebrafish ZGA carry distinct features such as a sharp TSS and a canonical TATA-box (Haberle et al., 2014; Hadzhiev et al., 2023). In contrast, late activated promoters have broadly distributed TSSs and lacked a TATA-box (Haberle et al., 2014; Hadzhiev et al., 2023). Indeed, this study showed that zygotic genes activated before the 512-cells stage (including the earliest transcribed gene, mir430), all had "early" promoter features. This suggests that early and late promoters may utilise distinct sets of core transcriptional machinery for activation. What is unclear, however, is why amongst genes with "early" promoters, some are activated earlier than others. The mir430 gene cluster activates at the 64-cells stage whilst other "early" promoter genes have only been reported to turn on 2-3 cell cycles later (Bhat et al., 2023; Hadzhiev et al., 2023; Heyn et al., 2014). Thus, more work is required to define the genetic features that underlie these differences in activation times.

As in the *Drosophila hb* promoter, a TATA-box alone is unlikely to predict early transcription given the generality of this motif at core promoters. It remains unclear how NPS regulation relates to TATA-box dependent early transcription. Perhaps a specific organisation of motifs (NPS and TATA) upstream of the TSS confers a strong positioning sequence for PIC formation at the TSS. Such a requirement would be supported by findings from dissection of the *Drosophila hb* promoter which showed that a Zelda binding site and a TATA-box positioned -41 to -30 upstream of the TSS were important for *hb* transcription (Ling et al., 2019).

Promoter codes of TF motifs may also determine the usage of enhancers that are active during early ZGA. In the case of zebrafish, NPS are required for enhancer activity (Miao et al., 2022). Nanog has been proposed to cluster at enhancers and, contact RNA pol II at promoters to activate transcription (Pownall et al., 2023). To properly understand how NPS binding relates to genome-wide enhancer-promoter contacts, and transcription, in a gene-specific manner would require simultaneous detection of proteins, DNA-DNA contacts, and RNA in embryonic tissue. Advances in modern imaging approaches have made this possible on a smaller scale, when the underlying sequences of interest are known (Immunofluorescence, DNA and RNA fluorescence *in situ* hybridisation (FISH)). However, no technique exists yet to scale this up for the whole genome.

1.9.4 Nuclear compartments

Transcriptional activity in nuclear space is not homogenously distributed. High local concentrations of transcriptional machinery form subnuclear bodies of active transcription (Cho et al., 2016; Chong et al., 2018; Cisse et al., 2013; Hilbert et al., 2021; Jackson et al., 1993; Kuznetsova et al., 2023; Sabari et al., 2018; Ugolini et al., 2023). Historically, these have been visualised via the radiolabelling of nascent RNA transcripts in in-tact nuclei (Jackson et al., 1993), and more recently, by imaging of fluorescently labelled transcriptional machinery or nascent transcripts (Cho et al., 2016; Cisse et al., 2013; Hilbert et al., 2021; Kuznetsova et al., 2023; Sabari et al., 2018). These transcription bodies also form in nuclei during zebrafish ZGA. Transcription of *mir430* forms two distinct and long-lived bodies enriched with transcriptional machinery and nascent *miR430* transcripts (Hadzhiev et al., 2019; Kuznetsova et al., 2023; Ugolini et al., 2023). Targeted deletion of the *mir430* locus results in a loss of these transcription bodies and global mis-regulation of zygotic transcription, suggesting that the *mir430* transcription bodies may organise transcription in nuclear space (Ugolini et al., 2023). They do so, partly, by sequestering transcriptional machineries within these bodies, preventing ectopic activation of other zygotic genes (Ugolini et al., 2023). Past reports have also suggested that other early zygotic genes may be spatially co-transcribed within the *mir430* transcription bodies (Hadzhiev et al., 2019). Thus, spatial organisation of transcription in the nucleus may represent a previously unappreciated aspect of transcriptional regulation during ZGA.

The formation of transcription bodies is likely seeded by specific transcription factors binding to DNA. In the case of zebrafish ZGA, Nanog forms a cluster that is seeded by the *mir430* locus (Kuznetsova et al., 2023). This initial seeding event results in the sequential clustering of downstream transcriptional machinery, including Sox19b and RNA pol II (first in its initiating, then elongating form). This indicates that Nanog binding to the *mir430* locus is a key step, but probably not the only, in the formation of the *mir430* transcription bodies (Kuznetsova et al., 2023). Interestingly, the *mir430* locus is highly repetitive and contains multiple Nanog binding sites. No published work so far has studied the importance of these multiple Nanog binding sites at the *mir430* locus in seeding the formation of the *mir430* transcription bodies, or even *mir430* transcription in general.

Conceptually similar nuclear bodies have been identified to form during *Drosophila* ZGA too. The histone locus body (HLB) is a nuclear body where the highly repetitive histone locus which encodes genes for all 4 core histones are transcribed. During ZGA, the HLBs occur as two nuclear bodies with high concentrations of TFs and transcriptional machinery (Cho and O'Farrell, 2023; Rieder et al., 2017a; Salzler et al., 2013). Formation of HLBs is seeded by the binding of the TF, CLAMP, on a specific site within the repeat unit of the histone locus (Rieder et al., 2017a; Salzler et al., 2013). The resultant clustering of CLAMP forms a 'proto-HLB' which matures over time, recruiting other HLB factors such as Mxc and FLASH, and eventually, large, and stable clusters of RNA pol II (Cho and O'Farrell, 2023; Koreski et al., 2020; Rieder et al., 2017b; Salzler et al., 2013). Recent studies also suggest that HLBs can sequester transcriptional machinery from a smaller histone locus elsewhere in the zygotic genome (Koreski et al., 2020). To what extent the deletion of the histone locus would globally affect early zygotic transcription remains unknown and would be an interesting direction for future work.

These two examples show that during ZGA, transcription and transcriptional machineries can be compartmentalised in nuclear space. Transcription control, as a direct or indirect result of subnuclear compartments such as transcription bodies, may provide a novel perspective on gene regulation during ZGA.

1.9.5 Repetitive elements and ZGA

Across ZGA in different model organisms, a running theme can be observed – the earlier transcribed genes tend to be highly repetitive. This seems to be the case for the *mir430* locus, which is highly repetitive (at least 8 x 1.7 kb *mir430* repeating units), very early transcribed and seeds the formation of TF clusters. It is also the case for the histone locus (~100 x 5 kb histone gene repeating units) which seeds the formation of HLBs, and is transcriptionally active prior to the *Drosophila* MBT. Moreover, in mice, the *Dux* gene locus which is one of the earliest zygotic genes to be activated is also highly repetitive (at least 28 x 3.3 kb D4Z4 repeating units), spanning ~350 kbp (Grow et al., 2021). What features about these repetitive loci ensure that they are activated early during ZGA before other zygotic genes? One possibility is that *cis*-regulatory information, such as TF binding sites, within the repeat units of these gene clusters could collectively enhance TF binding. Cooperativity between TF binding sites for TF binding is a well-described idea. Studies in enhancers have long proposed that motif numbers, positions, types, and combinations can influence cooperative TF recruitment at enhancers (Spitz and Furlong, 2012). Cooperativity between TF binding sites may underlie the ability of the *mir430* locus and histone locus to seed the formation of TF clusters even under a repressive nuclear environment whereby histones generally preclude TF binding. This localised out-competition of TFs against histones for DNA binding may explain the early activation of these repetitive loci. Determining if this is the case will require uncoupling the repetitive nature of these loci from the *cis*-regulatory activities of individual repeat units – a feat that is not easy.

A major obstacle to studying these repetitive loci is the very fact that these loci are so repetitive. Often, reference genomes assembled using whole genome shotgun (WGS) sequencing do not have sufficient read lengths to resolve highly repetitive loci. As such, mis-assemblies at these loci are commonplace, impacting how we design experiments and interpret results on these loci. In the future, properly understanding how these loci are activated would first require resolving their true structures using long-read sequencing.

Zebrafish ZGA begins with *mir430* transcription

So far, I have provided a broad view of the regulators of ZGA across different model organisms and how they can organise transcription through their unique abilities to reshape the genome. Through different examples, we have also seen that certain genes are known to be activated earlier during ZGA, and they typically display specific features, such as having combinations of TF binding sites or being very repetitive. In this section and for this thesis, we will delve into the case of *mir430* activation during zebrafish ZGA. Here, we will discuss the open questions about *mir430* activation posed at the start of this project and, the aims I formulated to address these questions.

In zebrafish, ZGA first begins at the 64-cells stage of embryonic development (Heyn et al., 2014; White et al., 2017). This initial spark of transcriptional activity is seeded by the *mir430* gene cluster which encodes high copy numbers of genes from the *mir430* gene family (*mir430a, b and c*) (Chan et al., 2019; Hadzhiev et al., 2019; Heyn et al., 2014; Kuznetsova et al., 2023). Transcription of this gene cluster can be visualised as 2 distinct and long-lived transcription bodies, rich in transcriptional machinery and nascent miR430 transcripts (Chan et al., 2019; Hadzhiev et al., 2019; Kuznetsova et al., 2023; Sato et al., 2019). Activation of *mir430* at the 64-cells stage is closely followed by the gradual activation of other zygotic genes in the subsequent cell cycles(Bhat et al., 2023; Hadzhiev et al., 2019; Heyn et al., 2014; White et al., 2017). By the 1k-cells stage, when the MBT occurs, transcription is widespread (Bhat et al., 2023; Chan et al., 2019; Haberle et al., 2014; Hadzhiev et al., 2023; Heyn et al., 2014; Hilbert et al., 2021; White et al., 2017). Thus, ZGA in zebrafish occurs over a continuum, and begins with *mir430* transcription.

How does *mir430* transcription precede transcription of all other zygotic genes? Answering this question was particularly important given the known developmental role of *mir430* in clearing maternally loaded transcripts (Giraldez et al., 2006, 2005). From a transcriptional regulation perspective, the ability of *mir430* to be activated despite a generally repressive nuclear environment during early embryonic stages also suggested that it is regulated differently from the rest of the genome.

Addressing this question, however, was hampered by the fact that the *mir430* locus is repetitive. Reference genome assemblies are typically made from short reads, and may struggle to properly resolve such a locus, casting doubts onto its true structure. Furthermore, it is also not known how such a repetitive locus would be transcribed with respect to the *mir430* repeat units. To this end, the first aim of my thesis was to characterise the *mir430* locus. **Under this aim, I set out to address 2 questions: 1) What is the true structure of the**

mir430 **locus? And 2) What are the transcription products generated from the** ***mir430*** **locus?**

The above aim fed back into an even more pressing question: What drives the early activation of *mir430*? Competition between soluble histones and TFs for DNA binding has been previously shown to regulate the timing of ZGA in zebrafish (Joseph et al., 2017). While this model can explain ZGA timing on broader timescales, it falls short in explaining why *mir430* is activated earlier than other genes if the dynamics of histone-TF competition act upon the whole genome. The logical explanation for this would be that genetic features intrinsic to *mir430* distinguish it from the rest of the genome. **For the second aim of my thesis, I set out to identify these genetic features specific to** ***mir430*** **that determine its early activation during ZGA.**

Overall, these aims contribute towards an understanding of how *mir430* is transcribed at the start of zebrafish ZGA, and the regulatory logics that govern temporally ordered gene activation. Further, they may also provide principles for understanding transcriptional regulation in general.

2 Results

2.1 Pinning down the structure of the mir430 locus

2.1.1 Characterisation of the GRCz11 mir430 locus

To begin to characterise the *mir430* locus, I first studied the structure of the *mir430* locus as represented in the GRCz11 reference genome. Based on the GRCz11 reference genome, the *mir430* locus is a ~16 kbp locus that resides on the chr4 long arm (chr4q) of the zebrafish genome and contains clusters of *mir430* genes. These *mir430* genes are in fact, a family of gene isoforms – *mir430a, mir430b, mir430c and mir430i*, with a,b and c being the most commonly occurring isoforms. These *mir430* genes are further organised within 8 ~1.7 kbp tandemly repeating units (Fig. 2.1.1 A). Each of these repeating units is made up of a 650 bp promoter sequence upstream of 6-8 mir430 genes. Here, I define a repeat unit as the sequence bookended by the start of a *mir430* promoter and the start of the next *mir430* promoter. This sub-structure within *mir430* repeating units has also been reported recently in separate studies (Hadzhiev et al., 2023; Pownall et al., 2023). By this definition of a repeat unit, an alignment of the sequences of the repeat units show that the repeats have striking sequence similarity Fig. 2.1.1 A). The sequence of the *mir430* promoter is highly conserved between repeats. Furthermore, repeats often contain 6 *mir430* genes organised as doublets of *mir430a, c and b*. The promoter + 6 *mir430* genes (2 X a-c-b) configuration make up the repeat consensus sequence of ~1.7 kbp (Fig. 2.1.1 B). Variations do occur with more *mir430* genes being present, resulting in larger repeat lengths of up to 2.3 kb.

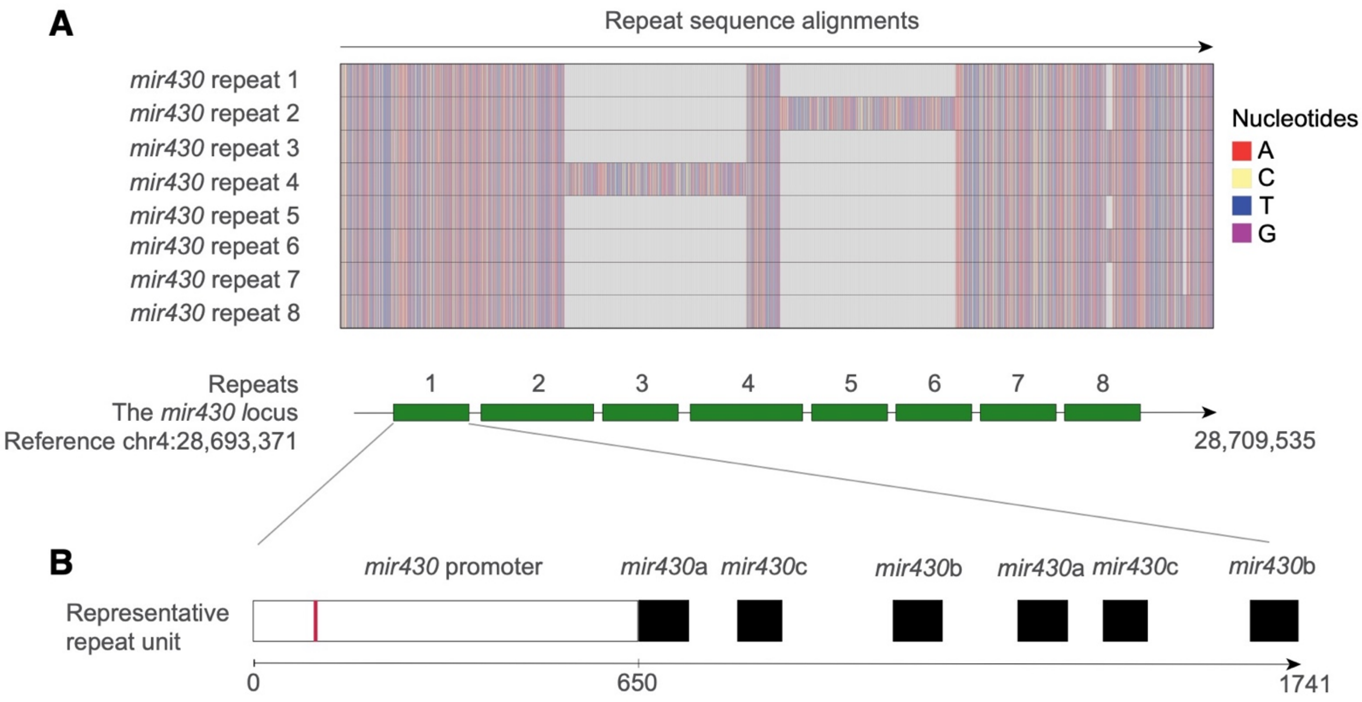

Figure 2.1.1: The GRCz11 mir430 locus:

A) An overview of the structure of the mir430 locus as presented in the GRCz11 reference assembly. The locus is made up of 8 tandemly repeating units. A multi-sequence alignment of the mir430 repeats show that the sequence is highly conserved amongst repeats. B) Consensus sequence of a mir430 repeat showing a single mir430 promoter, and 2 triplets of mir430a, c and b genes. A motif search also identified an AAUAAA poly(A) signal on the left side of the promoter sequence (red bar; chapter 2).

However, this description of the *mir430* locus is inaccurate. The *mir430* locus in the GRCz11 reference genome was assembled using next-gen sequencing (NGS) short reads that are typically much shorter than the *mir430* locus tandem repeat array. NGS short reads do not provide sufficient information to resolve the exact number of repeat copies present. As such, assemblies of repetitive loci using short reads can result in the collapse of the locus into fewer copies and, the loss of valuable information about sequence variation between repeats. Indeed, the region covering the *mir430* locus is marked as "unsure" by the original assemblers of this locus, suggesting that this locus may not be properly resolved.

2.1.2 Xdrop enrichment of the mir430 locus

Given these uncertainties, I set out to re-sequence the *mir430* locus using long read sequencing. The aims were to 1) resolve the true structure of the *mir430* locus, and 2) confirm the sequence of the *mir430* repeating units. To this end, I used a technique called Xdrop developed by the company, Samplix, for targeted enrichment of genomic loci for long read sequencing. Xdrop uses water-oil-water double emulsion droplets to specifically isolate the genomic locus of interest for sequencing (Madsen et al., 2020). By coupling this with long read sequencing, we aimed to get long reads spanning multiple repeat units, without expending sequencing reads on the rest of the genome. The Xdrop targeted sequencing workflow is the following (Fig. 2.1.2):

1) **Droplet PCR (dPCR) to identify DNA molecules harbouring the locus of interest**
   High molecular weight gDNA is packaged into water-oil-water double emulsion droplets alongside reagents for a fluorescence PCR specific to the target locus. Rather than having to PCR-amplify the whole locus of interest, this PCR serves only to generate sufficient amplicons from a short region (100-150 bp) proximal to/within the locus of interest. The resultant newly synthesised dsDNA can be detected by a fluorescent dsDNA dye and, the droplets containing the target locus are sorted by flow activated cell sorting (FACS).
2) **FACS of fluorescent droplets containing target locus of interest**
   Droplets from the dPCR are segregated based on the levels of fluorescence, where fluorescent droplets contain the target locus.
3) **Amplification of enriched target DNA**
   Positive (fluorescent) droplets are broken up. The isolated target DNA is pooled and repackaged again into water-oil-water double emulsion droplets for droplet multiple displacement amplification (dMDA). dMDA involves amplification of enriched DNA with phi29 DNA polymerases. Due to its exceptional strand displacement and proofreading capabilities, phi29 polymerase products have high molecular weight and fidelity. Additionally, isolating MDA reactions into distinct droplets minimises the likelihood of intermolecular chimeras forming during MDA. The resultant DNA is enriched for the target locus.
4) **PacBio long read sequencing of amplified target DNA**.
   The enriched target DNA is used for library preparations for PacBio long read sequencing (HiFi).

I adapted the Xdrop assay to the *mir430* locus by designing primer pairs that uniquely amplify *mir430a* and *c* isoform sequences (Fig. 2.1.3 A). The rationale was to pull out *mir430* containing sequences, wherever they may lie in the genome. In addition, I designed Xdrop assays targeting the regions upstream and downstream of the *mir430* locus based on the GRCz11 reference genome (Fig. 2.1.3 A). These *mir430*-specific Xdrop assays were used to enrich for the *mir430* locus from high molecular weight gDNA isolated from a wildtype female AB strain zebrafish. We were able to specifically isolate droplets containing the *mir430* locus using FACS and following the amplification step, we had more than 100x enrichment of *mir430*-containing DNA for each of the individual assays (Fig. 2.1.3 A). Calculated enrichment over input was higher in the Upstream and Downstream assays compared to the *mir430a* and *mir430c* assays. This is likely due to the already high baseline target detection in the input for *mir430a* and *mir430c*. Overall, these tests confirmed that we were able to specifically enrich the *mir430* locus in our DNA sample.

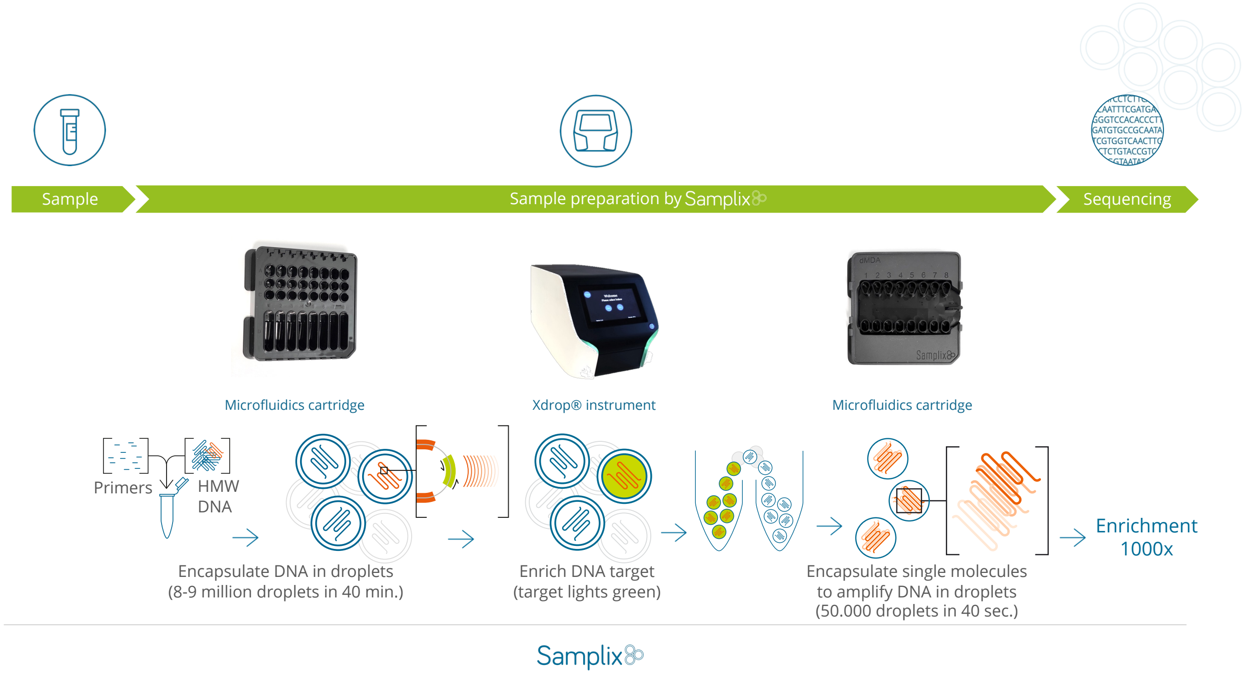
Sample
Sample preparation by Samplix
Sequencing
Microfluidics cartridge
Xdrop® instrument
Microfluidics cartridge
Primers
HMW DNA
Encapsulate DNA in droplets
(8-9 million droplets in 40 min.)
Enrich DNA target
(target lights green)
Encapsulate single molecules
to amplify DNA in droplets
(50.000 droplets in 40 sec.)
Enrichment
1000x
Samplix

Figure 2.1.2: Xdrop enrichment workflow (As provided by Samplix)

Starting from left to right, high molecular weight gDNA is first packaged into water-oil-water double emulsion droplets together with primers specific to the target site, polymerases, buffers and a dsDNA dye. Each droplet is an isolated PCR. Following dPCR, the droplet containing the locus of interest would light up in green due to synthesis of the PCR amplicon. This specific droplet contains the locus of interest and can be sorted from negative droplets using FACS. The positive droplets are broken up, and the target DNA is then pooled and repackaged again into droplets for dMDA to amplify the amount of enriched target DNA. The final output is the enriched target DNA

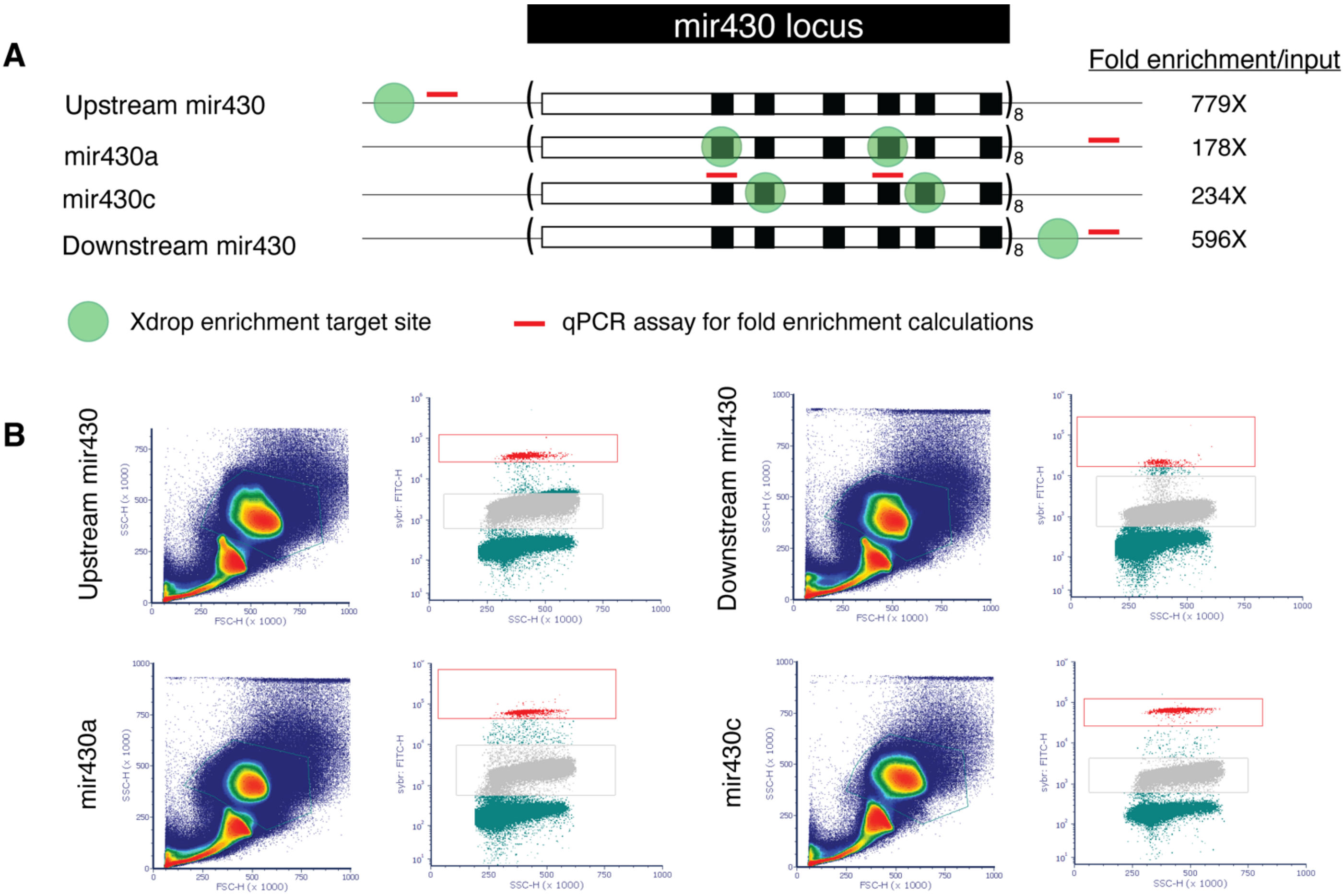

Figure 2.1.3: The mir430 locus target enrichment experimental design

A: Design of Xdrop target sites upstream of the mir430 locus, downstream of the mir430 locus, and at mir430a and mir430c genes. B: FACS density plots showing successful sorting of positive mir430 containing droplets for each target site. Boxed in red are positive droplets while boxed in grey are negative droplets.

2.1.3 PacBio HiFi sequencing of *mir430* enriched DNA

We then performed Pacbio HiFi sequencing on our *mir430*-enriched DNA sample. The PacBio HiFi sequencing generated 746440 circular-consensus (CCS) reads – 210895 from Upstream mir430, 139343 from Downstream mir430, 252171 from *mir430a*, and 144031 from *mir430c*. *mir430*-containing reads made up 2296 (1.1%) of the Upstream *mir430* reads, 2783 (2.0%) of the Downstream *mir430* reads, 17643 (7.0%) of *mir430a* reads, and 4229 (2.9%) of *mir430c* reads. The fraction of *mir430*-containing reads were higher for the *mir430a* and *mir430c* assays compared to the Upstream and Downstream assays. We reasoned that this was due to the multi-copy nature of the *mir430a* and m*ir430c* target sites. In total, *mir430*-containing reads made up 26951 (3.6%) of all reads sequenced. This is higher than previously reported yields for the Xdrop protocol (1.6%) – likely a result of the highly repetitive nature of the Xdrop target sites (Madsen et al., 2020).

Next, we mapped all CCS reads to the GRCz11 reference genome to determine coverage over the *mir430* locus. We saw high coverage of reads mapping over a 100kb region centred on the *mir430* locus (Fig. 2.1.4). The coverage spanned across all 4 Xdrop target sites. Interestingly, the mapped reads identified a 20 kb region upstream of the GRCz11 *mir430* locus where a sudden drop-off in coverage occurs, suggesting a previous mis-assembly. While a portion of the high coverage over the region is resultant from the Xdrop enrichment, we also expected the coverage to be partly due to the pileup of long reads on a collapsed repetitive locus.

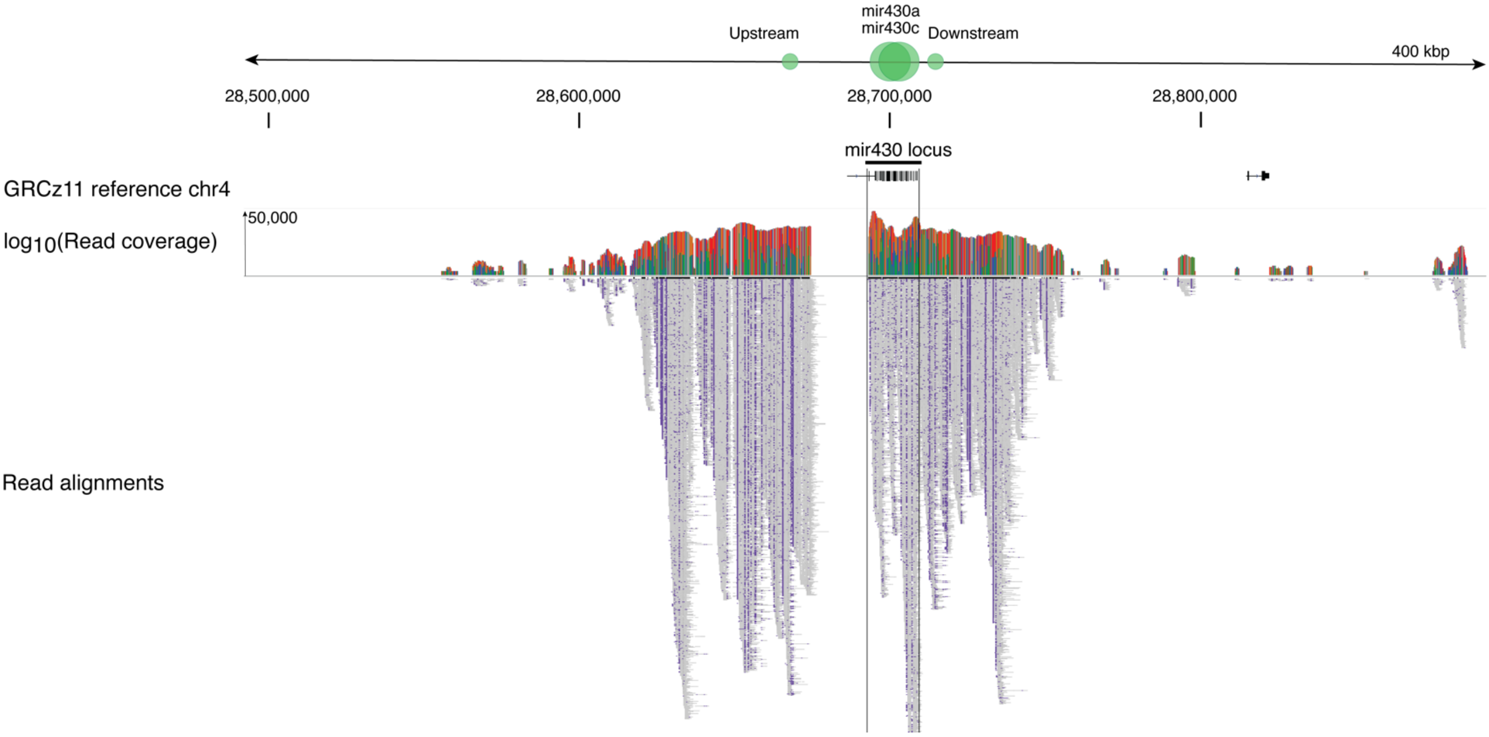

Figure 2.1.4 Coverage of mir430-enriched CCS reads along mir430 locus in the GRCz11 reference assembly

Mir430-enriched CCS reads mapped across a 100 kbp region centered on the GRCz11 mir430 locus. Green circles highlight the Xdrop target sites for Upstream mir430, Downstream mir430, mir430a and mir430c genes. Very high read coverage spanned all 4 designed Xdrop target sites. Mapped reads also identified a 20 kb directly upstream of the mir430 locus with no coverage, suggesting a previous misassembly.

To determine the true structure of the *mir430* locus, we first looked at *mir430*-containing reads. *Mir430*-containing reads had a mean length of 8449.1 bp, with the longest read being 38734 bp (Fig. 2.1.5 A). A closer inspection of the top 1% longest *mir430*-containing reads showed that almost all the reads fully consisted of *mir430* genes. A common problem arising from MDA reactions is the formation of chimeras. While the isolation of MDA reactions into droplets during the dMDA minimises intermolecular chimeras, intramolecular chimeras can still occur within droplets. This can result in aberrations in the sequencing template such as inversions and duplications. To remove the potential chimeras, we split reads at potential chimeric sites using SACRA (Split Amplified Chimeric Read Algorithm). SACRA uses an all-vs-all approach to detect chimeric sites within individuals reads that are not supported by other reads from the sequencing pool (Kiguchi et al., 2021). Thus, only reads that are supported by other reads based on the all-vs-all alignments are left unsplit. Our analysis showed that chimeras were prevalent amongst the sequenced reads. Of the 746440 HiFi reads, only 28080 (3.76%) of reads were left unsplit. The remaining 718360 reads were split on average 13.7 times per read. While the fraction of split reads was much higher than anticipated, we decided to continue with the stringent chimeric site detection parameters to ensure accurate re-assembly of the locus. Following SACRA, *mir430*-containing reads had a mean length of 1038.7 bp with the longest read being 11857 bp (Fig. 2.1.5 B). All reads above 1kb were subsequently kept for de novo contig assembly using Hifiasm.

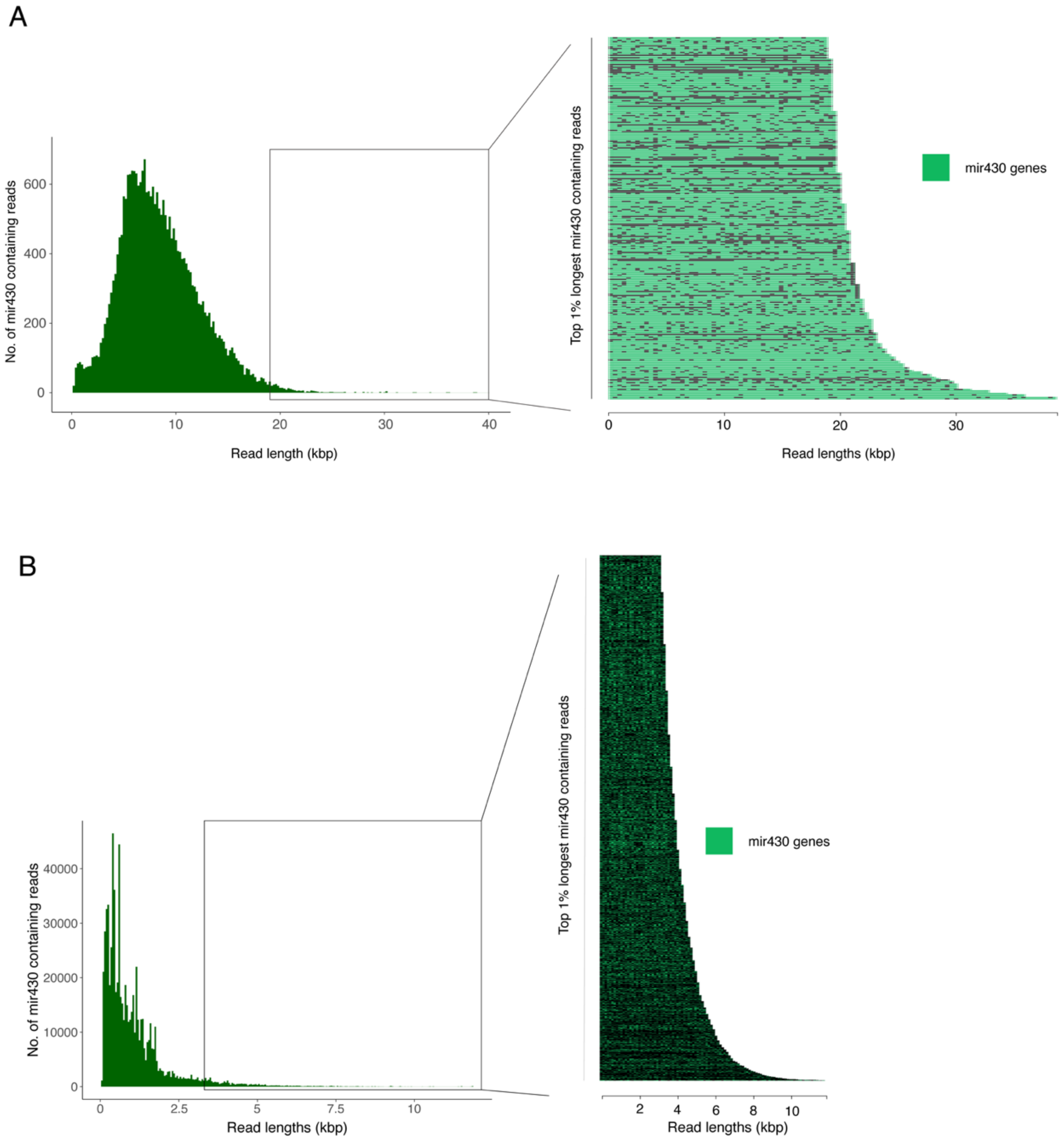

Figure 2.1.5: mir430-containing reads length distribution before and after SACRA chimera splitting.

A: Left - Distribution of mir430-containing CCS read lengths prior to SACRA chimera splitting. Right – Top 1% of mir430 containing reads ranked according to read length. Green regions represent mir430 genes while black regions represent non-mir430-containing DNA. B: Left - Distribution of mir430-containing CCS read lengths after SACRA chimera splitting. Right – Top 1% of mir430 containing reads ranked according to read length. Green regions represent mir430 genes while black regions represent non-mir430-containing DNA.

2.1.4 De novo assembly of the mir430 locus

The de novo contig assembly gave a total of 5479 primary contigs with an N50 of 12365 bp. While 146 of these contigs contained *mir430* genes, a large fraction of these *mir430*-containing contigs (119/146) were below 10 kbp in size. Contigs above 10 kbp fell into 3 groups: 1) partially *mir430*-containing (contigs with a single end terminating within the mir430 locus), 2) completely *mir430*-containing (contigs with both ends terminating within the *mir430* locus), or 3) end-to-end assemblies (contigs with both ends terminating outside of the *mir430* locus). Most of the contigs above 10 kbp were either partially *mir430*-containing or completely *mir430*-containing – 14/27(52%) and 11/27(41%), respectively. However, we were able to assemble 2 end-to-end assemblies. One of the 2 end-to-end contigs was the longest contig assembled (ptg000002l) with a length of 224297 bp. On closer inspection, we saw that this contig harboured a *mir430* gene cluster approximately 150 kbp in size (Fig. 2.1.6). This *mir430* gene cluster had a total of 71 mir430 promoters and 424 *mir430* genes of a, b, and c isoforms. The more than 10-fold difference in size of the *mir430* locus assembled here compared to the GRCz11 assembly confirms suspicions that the locus was previously collapsed into a 16 kb region. The second end-to-end contig, ptg000265l, was 143539 bp in size and contained a smaller *mir430* locus spanning 20 kb with 20 *mir430* genes and 8 *mir430* promoters (Fig. 2.1.6). This is likely a duplication event that resulted in a satellite *mir430* locus. Overall, we were able to use Xdrop targeted long read sequencing to generate end-to-end assemblies of the *mir430* locus. This demonstrates the utility of targeted sequencing approaches in resolving complex genomic loci. However, a clear problem observed in our dataset was the lack of contiguity amongst contigs - as evidenced by the large number of partial contigs. This could be the result of extensive splitting of chimeric reads and/or the inability of Hifiasm to resolve shorter reads containing only repeats.

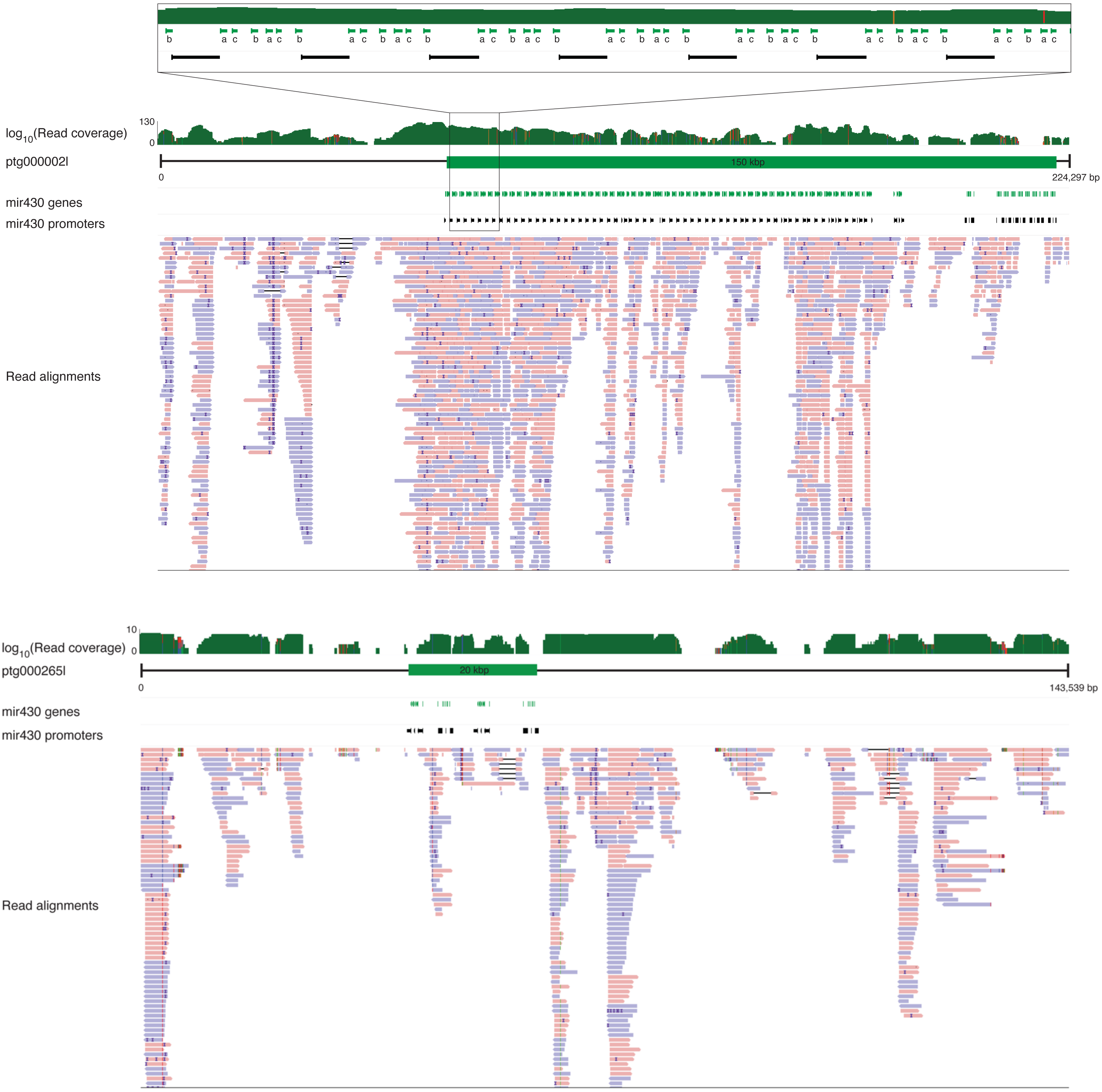

Figure 2.1.6: End-to-end assemblies of the mir430 locus

Top panel: Contig ptg000002l 224 kbp long containing a mir430 locus ~150 kbp in size. Bottom panel: Contig ptg000265l 143 kbp long containing a smaller mir430 locus ~20 kbp in size. In both cases, read alignments show best aligned reads to the contigs, not including multi-mapping reads.

We next looked deeper into the sub-structure within the assembled *mir430* locus in ptg000002l. We found that the organization of the *mir430* repeats closely resembled that described for the GRCz11. Each repeating unit was made up of a *mir430* promoter followed by 6 *mir430* genes (2 X a-c-b). The computed pairwise distance of the consensus repeat sequence from the GRCz11 and our assembled *mir430* locus revealed that the sequence of the repeat units are at least 98% identical. Together, these findings show that contrary to what is shown in the GRCz11 reference assembly, the *mir430* locus is at least 150 kbp large with 71 mir430 repeat copies. This is at least an order of magnitude larger than what has been previously reported. The identified differences occur mainly at the structural level with different repeat copy numbers. However, the underlying repeat consensus sequence remains largely identical to the sequence reported in the GRCz11. Furthermore, using this approach, we were also able to identify a satellite cluster of *mir430* repeats distinct from the 150 kbp large *mir430* locus.

2.2 Chapter 2: Transcriptional outputs from a mega-repetitive locus

In the previous chapter, we established that the *mir430* locus is a mega-repetitive locus containing at least 71 *mir430* repeat copies. These repeats are typically made up of a promoter followed by 6 *mir430* genes (2 X a-c-b). The tandemly repeated nature of this locus raises the question of how this locus is transcribed, and what the resultant transcriptional output from this locus would be. In particularly, does each *mir430* repeat represent a discrete transcriptional unit or does RNA Pol II traverse multiple repeat units – resulting in a mRNA transcript potentially tens of kilobases large? If indeed larger *miR430* transcript species exist and transcription begins at the promoters, what defines the end point of *mir430* transcription? To address these questions, I set out to characterise the transcripts produced from the mega-repetitive *mir430* locus.

2.2.1 Northern blots identify non-random populations of miR430 transcripts

To address the above uncertainties, I set out to determine the size of primary transcripts generated from the *mir430* locus using a northern blot. Compared to modern sequencing approaches that rely on potentially erroneous read mapping to the *mir430* locus, northern blots provide an unambiguous snapshot of the size distribution of primary miR430 transcripts. Total RNA was isolated from embryonic stages when *mir430* is transcribed – 64-cells, 512-cells, High stage, Sphere stage and 30% epiboly. The resultant RNA was ran on a gel, transferred onto a nitrocellulose membrane, and blotted for miR430 transcripts using a 700 bp anti-sense digoxigenin (DIG) labelled miR430 probe. This miR430 probe hybridizes to the microRNA encoding region of each *mir430* repeat unit, and the hybridised RNA can be visualized on a blot using chemiluminescence.

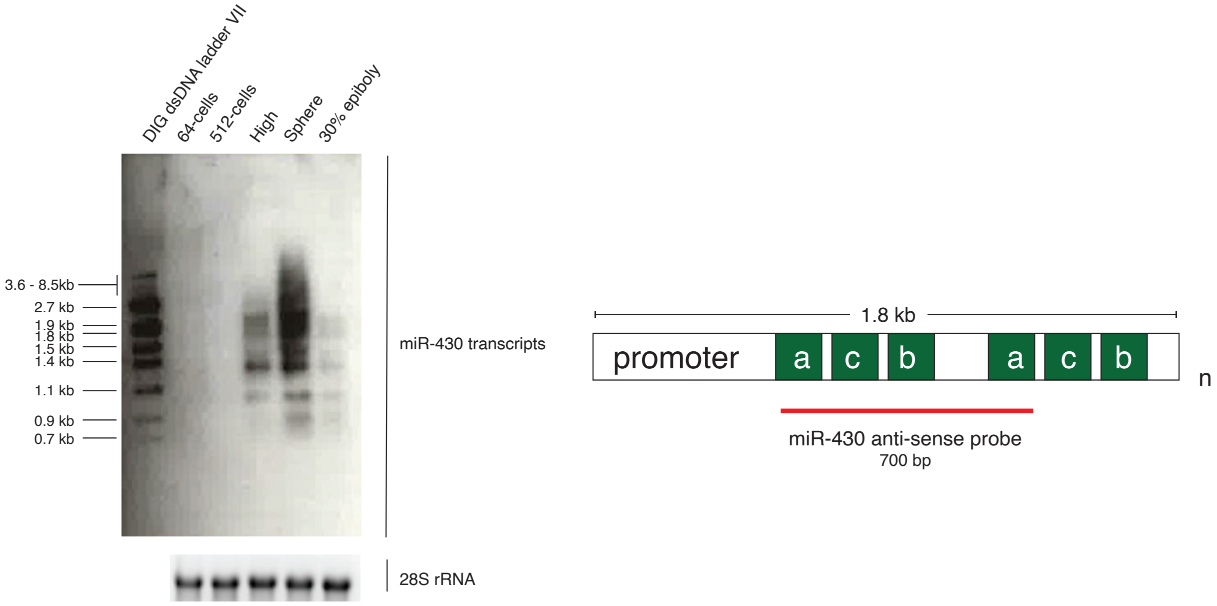

Figure 2.2.1 Northern blot for miR430 transcripts over development.

Left panel: Representative northern blot showing miR430 transcripts detected starting from the High stage to 30% epiboly. miR430 transcripts detected show a highly banded patter. N=2 Right panel: Design of a 700 bp miR430 anti-sense probe to detect primary miR430 transcripts.

Throughout the embryonic stages tested, miR430 transcripts were first detected on the northern blot at High stage (Fig. 2.2.1). Subsequently at Sphere, overall levels of miR430 transcripts increased dramatically and, eventually reduce at 30% epiboly (Fig. 2.2.1). The delay in timing of observation of miR430 transcripts here compared to its known activation time at the 64-cells stage is likely due to the higher detection threshold of the northern blot technique. The reduction in overall transcript levels at 30% epiboly is also in-line with RNA-seq studies showing that *mir430* is silenced during epiboly (Fischer et al., 2019; Giraldez et al., 2005). Thus, using the northern blot approach, I was able to observe miR430 transcript changes at post-MBT stages.

2.2.2 miR430 transcript sizes correlate with repeat unit multiplicities

The *miR430* transcripts that appeared throughout the stages when they were visible on the northern blot showed a very consistent pattern of bands, indicating that miR430 transcription is non-random. Rather, they may have defined start and end sites. One possibility was that each *mir430* repeat represented a single transcription unit and, transcripts could span different multiplicities of these *mir430* repeats. In support of each repeat representing a defined transcriptional unit, I was able to detect a polyadenylation signal (PAS) on the 5' end of *mir430* promoters, adjacent to the 3' end of the upstream repeating unit (Fig. 2.1.1 B).

This suggests that a single *mir430* transcriptional unit may begin at the miR430 promoter TSS and ends in the downstream PAS.

To further confirm this idea, I set out to determine what these different sized miR430 transcripts could represent. I reasoned that if these bands represented different multiplicities of *mir430* repeats, their sizes would correspond accordingly. To do so, however, I needed a ssRNA ladder that could also be detected using chemiluminescence. Commercially available options for ssRNA ladders detectable by chemiluminescence such as the DIG-labelled RNA Molecular Weight Marker (Roche) or the RNA Century Marker templates (Thermo-Fisher Scientific) were either discontinued or did not give clear ladders. However, the DIG-labelled dsDNA ladder VII (Roche) gave promising results on the northern blot. To obtain size equivalents of ssRNA bands using the DIG-labelled dsDNA ladder, I compared the ssRNA ladder with the DIG-labelled dsDNA ladder on a gel. In general, I observed that ssRNA travels approximately 1.5-2 times faster on a 1% TAE gel than DIG-lablled dsDNA. For example, 6 kb ssRNAs migrate approximately the same distance as 2.7 kb DIG-labelled dsDNA (Fig. 2.2.2). In this way, I was able to obtain approximate sizes of miR430 RNA transcripts by comparing against a DIG-labelled dsDNA ladder on the northern blot.

.

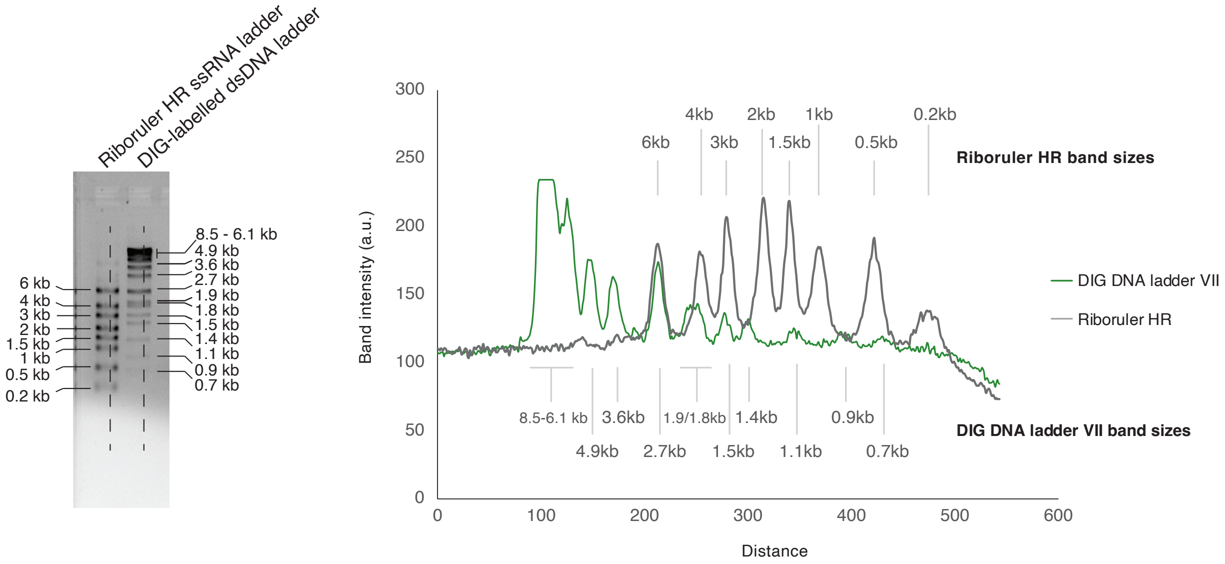

Figure 2.2.2 Size comparisons between a ssRNA ladder and a DIG-labelled dsDNA ladder

Left panel: Gel showing size comparisons between bands for the ssRNA Riboruler HR (ThermoFisher Scientific) ladder and the DIG-labelled dsDNA VII (Roche) ladder. Right panel: Line profiles of band intensities representing the size equivalents of ssRNA to DIG-labelled dsDNA ladders

By comparing alongside the DIG-labelled dsDNA ladder, I observed that specific bands of miR430 transcripts that first appear at High were 1 kbp and 1.4 kbp on the DIG dsDNA ladder (Fig. 2.2.3). This equates to ssRNA sizes of 1-1.5 kbp and 2-3 kbp, respectively (Fig. 2.2.3). While these are broad estimates, the transcript sizes corresponded well to singlets and doublets of *mir430* repeats. For instance, the 1-1.5 kbp band corresponds well with a transcript beginning at the known *mir430* TSS (Hadzhiev et al., 2023), and ending on the downstream PAS. Whereas the 2-3 kbp band would correspond to transcripts containing 2 tandem *mir430* repeats. Overall, these results show that each *mir430* repeat is a transcriptional unit, and that transcripts can span multiple repeat units. The occurrence of multiple repeats in a transcript may be a result of the skipping of PAS sites on the nascent mRNA. Primary *miR430* transcripts are known to be bound by RNA-binding proteins (RBPs), such as hnrnpa1 (Despic et al., 2017). Binding of RBPs may occlude the PAS, resulting in longer miR430 transcripts spanning multiple repeats.

In addition to the singlets and doublets, less defined transcript bands were also detected between 3-6 kbp. These could represent higher multiplicities of repeat unit transcripts. The reason for a lack of a defined band, however, is unclear. Potentially, transcription could terminate more stochastically beyond a certain transcript length.

Subsequently, at Sphere, I also observed smaller bands of approximately 750 bp. These are likely miR430 transcripts that have been further processed, perhaps via splicing. I reasoned that they are less likely to be direct miR430 transcriptional products given that they do not show up in earlier stages.

2.2.3 High transcription rates could form mega-*miR430* transcripts

Previous long read RNA-seq studies on zebrafish embryos have suggested that transcription of *mir430* results in the production of a 9 kbp long 'mega-miR430' transcript (Nudelman et al., 2018). To determine if I could detect these mega-miR430 transcripts using northern blots, I focused on miR430 transcripts at Sphere when the largest transcripts were present. Compared to High, Sphere stage embryos had a marked increase in large miR430 species, and this was visible as a smear between 1.5 – 8.5 kbp on the DIG-labelled dsDNA ladder (Fig. 2.2.3). Given that this range falls beyond the largest band on the ssRNA ladder, size estimation could only be done using the DIG-labelled dsDNA ladder. Assuming a 1.5-2 fold difference in migration rates between the DIG-labelled dsDNA and ssRNA ladder, this would suggest that the largest miR430 transcripts produced at Sphere could range between 12 - 17 kbp. While these are preliminary size estimates, they fall within the same range, or potentially even larger, as the previously described mega-miR430 transcripts.

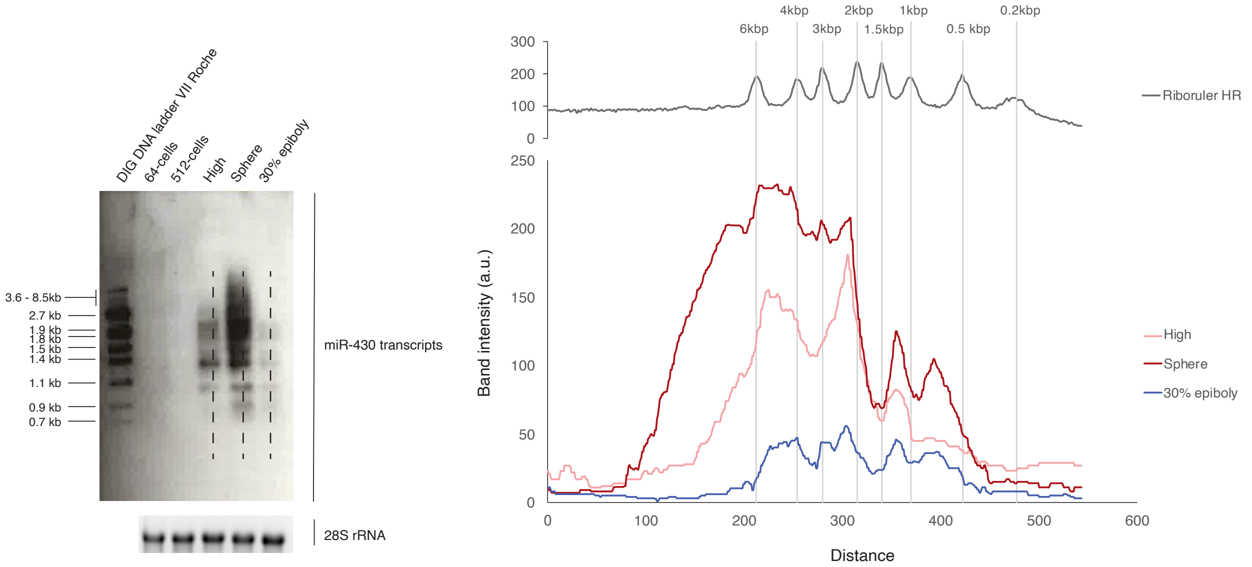

Figure 2.2.3: Size distributions of miR430 transcripts generated

Northern blot for miR430 transcripts (same of Fig. 2.1) and the corresponding sizes as determined by comparisons of the bands with the DIG-labelled dsDNA ladder and the ssRNA ladder. Line profiles of the miR430 transcript bands show that transcripts from High stage range from 1 kbp to 6 kbp in size with specific bands at ~1.4 kbp, ~2.3 kbp and 4-6 kbp. At Sphere, miR430 transcripts range from 500 bp to an estimated 17 kbp with specific bands at ~1.4 kbp and ~2.3 kbp. Finally at 30% epiboly, miR430 transcript levels are reduced with the ~1.4 kbp, ~2.3 kbp and 4-6 kbp bands looking noticeably weaker.

Overall, these findings provide a snapshot of the miR430 transcripts present at post-ZGA stages. They indicate that each *mir430* repeat represents a transcriptional unit, whereby transcription begins at the promoter and terminates at the end of the repeat unit via a PAS. Transcription through tandem *mir430* repeats can produce singlets, doublets and potentially transcripts with higher multiplicities of *mir430* repeats. These multi-repeat spanning transcripts could occur because of PAS skipping. In addition, I have also identified miR430 transcript species that may range between 12-17 kbp, consistent with the existence of a mega-miR430 transcript (Nudelman et al., 2018).

2.3 Inter-strain variations of the *mir430* locus and insights into its activation

In the previous chapters, I characterised the *mir430* locus by looking at its structure and its transcriptional output. I found that the *mir430* locus is much larger than the reported size in the GRCz11 reference assembly, with other published reporting similar findings (Hadzhiev et al., 2023; Pownall et al., 2023). Each of the *mir430* repeating units in the *mir430* locus likely represents the predominant *mir430* transcriptional unit. However, this mega-repetitive locus can also be transcribed over multiple *mir430* repeating units to produce mega-miR430 transcripts (Nudelman et al., 2018). These findings are consistent with and, could explain the enormous levels of miR430 transcripts detected during development (Heyn et al., 2014; Thatcher et al., 2008; White et al., 2017). Firstly, due to the sheer numbers of *mir430* genes present and secondly, due to the numerous promoters found across the locus – each of which represent a landing site for transcription to begin. A question that has remained elusive to the field, however, is what drives the early activation of *mir430* during ZGA? Addressing this question is of particular interest because it provides insights into the regulatory logics that govern gene activation during ZGA.

Previous work from our lab has shown that competition between histones and TFs can regulate ZGA onset (Joseph et al., 2017). In zebrafish, embryos inherit massive amounts of histones from oocytes (Joseph et al., 2017). These high levels of soluble histones present in the embryo pre-ZGA out-compete TFs for occupation of TF binding sites, thus preventing the onset of ZGA. Over developmental time, concentrations of nuclear histones gradually reduce (Joseph et al., 2017). This reduction in nuclear histone concentrations is accompanied by the accumulation of TFs over time. When sufficient TFs have been accumulated, and nuclear histones reduced, TFs would out-compete histones to gain access to DNA for initiating transcription. Consistent with this model, injections of embryos with a cocktail of all four core histones (H2A, H2B, H3 and H4) resulted in delays in ZGA while overexpression of TFs (Pou5f3 and Sox19b) resulted in advanced ZGA (Joseph et al., 2017). While this competition model can help to explain ZGA on broader time scales, they do not fully explain why different genes turn on at different times. Specifically, why is *mir430* activated earlier than other genes if the dynamics of histone-TF competition act upon the whole genome? One possibility is that the TF composition present in the early zygote determines which genes are transcribed during early ZGA. Indeed, specific TFs have been shown to activate zygotic genes early during development (De Iaco et al., 2017; Duan et al., 2021; Gaskill et al., 2021; Gassler et al., 2022; Ji et al., 2023; Lee et al., 2013; Leichsenring et al., 2013; Liang et al., 2008). However, TF specificity alone is unlikely to be the sole underlying cause since TF motifs can occur spuriously throughout the genome. Thus, the most logical explanation for differences in timing of activation would be that certain genomic features unique to the *mir430* locus drive its early activation.

In this chapter, I attempt to identify these genomic features that drive early activation of *mir430*. Past studies in different model organisms have shown that early transcribed genes, such as the Dux locus in mice and the histone locus in *Drosophila*, tend to be highly repetitive. Here, I hypothesised that the high repetitiveness at the *mir430* locus could underlie its early activation during ZGA. To test this, I took a comparative approach by comparing *mir430* activity between WT zebrafish strains. The rationale behind this approach was that highly repetitive loci, such as the *mir430* locus, may have copy number variations between strains. By correlating these potential variations with the timing of *mir430* activation, I hoped to identify the relationship between *mir430* repeat numbers and *mir430* activation.

2.3.1 Timing of *mir430* activation is distinct amongst WT zebrafish strains

To investigate the relationship between *mir430* repeat copy numbers and the timing of *mir430* activation using inter-strain variation, I first set out to determine if the timing of *mir430* activation varies between WT strains. I identified a panel of WT strains – AB, TÜ, TL and NHGRI-1. These WT strains historically originated from different sources and have been genetically isolated as distinct WT strains commonly used in the zebrafish community. To determine the timing of *mir430* activation, I turned to imaging as it provided the highest sensitivity for *miR430* mRNA detection at single nucleus resolution.

Transcription of *mir430* results in the formation of two discrete and long-lived transcription bodies enriched with nascent *miR430* transcripts, RNA pol II and other transcriptional machinery (Chan et al., 2019; Hadzhiev et al., 2019; Hilbert et al., 2021, n.d.; Kuznetsova et al., 2023; Sato et al., 2019; Ugolini et al., 2023). For imaging, these *mir430* transcription bodies provide an excellent readout for probing *mir430* activity in the WT strains. Visualisation of miR430 transcripts was done using a previously established method, MOVIE, where embryos were injected with fluorescently labelled morpholinos (MOs) complementary to nascent miR430 transcripts (Hadzhiev et al., 2019). To follow the nucleus throughout the embryonic cell cycles, I co-injected the MOs with fluorescently labelled antigen-binding fragments (Fabs) which recognise H3K27Ac marks (Sato et al., 2019). These combined approaches allowed the live visualisation of *mir430* activation in the nucleus throughout development in the panel of WT strains.

Whole-mount embryos from the 4 WT strains were imaged from the 64-cells stage to the 512-cells stage. In general, I saw that the *mir430* transcription bodies began to appear in a fraction of 64-cells stage nuclei for all WT strains. The number of nuclei where *mir430* transcription bodies were detected increased progressively in the following cell cycles, reaching full activation by the 512-cells stage (Fig. 3.2). In all strains, transcription at the 64-cells stage typically only begins in 1 allele and both alleles are eventually activated in subsequent cell cycles (Fig. 3.2). Thus, *mir430* activation in all WT strains begin in a fraction of 64-cells stage nuclei, typically in single alleles, and progressively increases thereafter. These findings confirm the general temporal profile of *mir430* activation shown from previously published work (Chan et al., 2019; Hadzhiev et al., 2019; Heyn et al., 2014; Kuznetsova et al., 2023; White et al., 2017). However, it also shows the stochastic nature of *mir430* activation at the 64-cells stage since not all nuclei synchronously activate *mir430* and, both alleles do not typically turn on together. Stochasticity between cells in acquiring transcriptional competence is a previously described phenomenon during development and in general (Kærn et al., 2005; Stapel et al., 2017).

Next, I set out to test if different strains have different likelihoods of activating *mir430* in 64-cells stage nuclei. While the general timing of *mir430* activation was at the 64-cells stage for all strains, I reasoned that differences in the percentages of nuclei that were active would be representative of differences in timing of *mir430* activation at a single-nucleus level. For this comparison, I scored nuclei as 'active' based on the detection of either one or both *mir430* transcription bodies, and 'inactive' when no *mir430* transcription bodies were detected. This was done for all stages and all WT strains. This comparison showed that the percentage of active nuclei at the 64-cells stage was varied in the different WT strains. For TÜ, TL and NHGR-1 strains, 30%, 30% and 35% of nuclei, respectively, were active. In contrast, 69% of AB nuclei were active – approximately 2-fold higher compared to the other strains (Fig. 2.3.1 C). In the subsequent cell cycles, the differences between AB, TÜ, TL and NHGRI-1 are gradually lost, and by the 256-cells stage, the percentages of active nuclei equalises across all strains. These findings suggest that AB nuclei typically have an earlier timing of activation than TÜ, TL and NHGRI-1 nuclei. Furthermore, it shows that the WT strains are in fact, not all identical, but rather have distinct *mir430* activation dynamics.

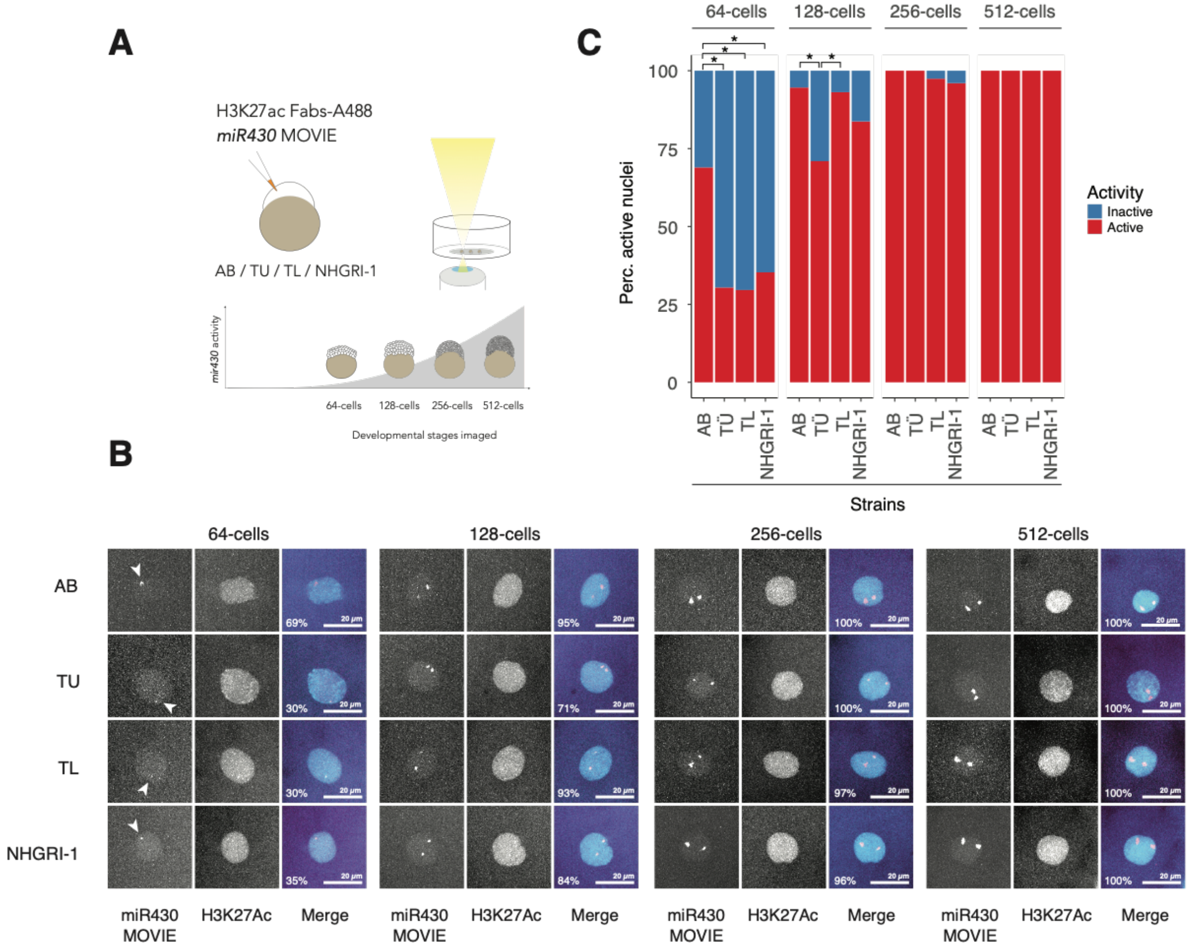

Figure 2.3.1 WT strains have distinct mir430 activation profiles.

*A. Schematic of experimental setup for imaging mir430 activation across WT strains. B. Images of active nuclei from 64-cells to 512-cells stage in WT strains. Typically, only a single allele is active at 64-cells whereas in later stages, both alleles are active. C. A comparison of percentages of active nuclei from 64-cells to 512-cells across WT strains. Fisher's exact test was used to compare inter-strain differences: * - p-value < 0.05.*

2.3.2 Timing of mir430 activation positively correlates with mir430 repeat numbers

I next wondered if differences in timing of activation observed above correlated with *mir430* repeat numbers in the different strains. To test this, I compared *mir430* repeat copy numbers in our lab WT strains using qPCRs on genomic DNA (gDNA) targeting *mir430* promoters and, *mir430a, b* and *c* genes. These 4 qPCR comparisons provide independent validations of *mir430* repeat number differences between strains. Quantification of *mir430* repeat numbers were done with respect to a single copy gene, Sox19a.

Using this approach, relative quantifications of *mir430* repeat copy numbers using the promoter, *mir430a, b* and *c* genes gave results with high standard deviations. Furthermore, the calculated absolute numbers of *mir430a, b and c* genes did not fall within the same range. This is surprising given that the expected ratios for isoforms a:b:c based on the consensus *mir430* repeating unit is 1:1:1 (Fig 2.1.1 B). To account for these issues, I compared fold-differences from AB per experiment for each qPCR target. Comparing the fold-difference of promoters and *mir430a* genes to AB per experiment revealed that TÜ, TL and NHGRI-1 had significantly lower promoters and *mir430a* genes by 0.8-fold on average (Fig. 2.3.2). Differences between TÜ, TL and NHGRI-1 were not significant for both promoters and *mir430a* genes. In contrast, *mir430b* genes were similar between all strains, and *mir430c* genes were lower in TÜ and NHGRI-1 by 0.8-fold, but not in TL. Overall, these findings identify a general trend of *mir430* repeat numbers being higher in AB compared to TÜ, TL and NHGRI-1. The fact that not all *mir430* gene isoforms follow this trend, however, reiterates the case that the stereotypic *mir430* repeat unit of 2 X a-c-b (Fig 2.1.1 B) is variable. As such, *mir430* gene copies may not always follow repeat number ratios.

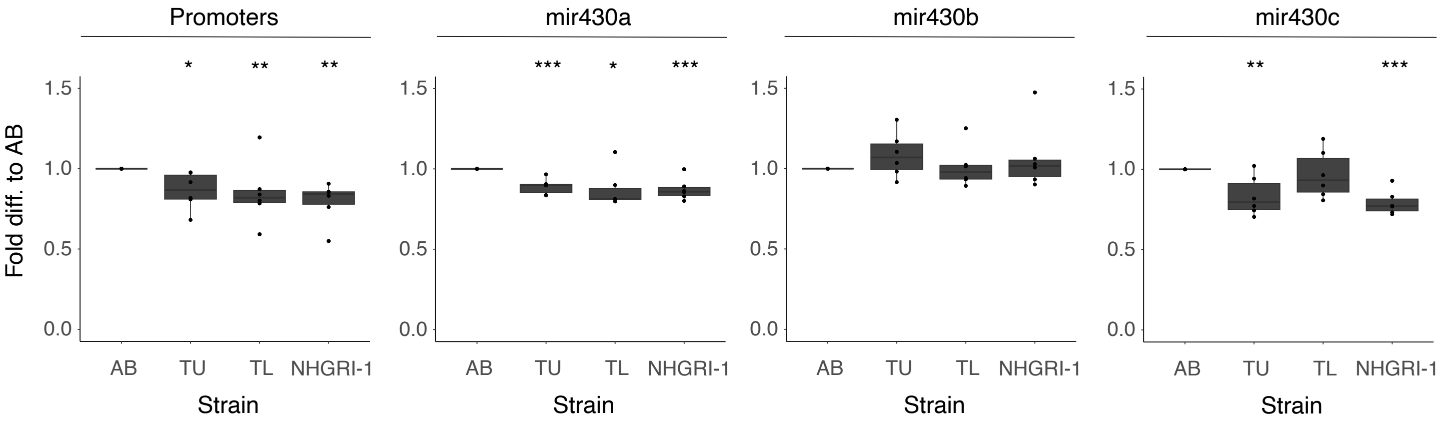

Figure 2.3.2: Inter-strain differences in mir430 repeat copy number.

*qPCR quantification of copy number differences of mir430 promoters and mir430a, b and c genes between WT strains. For each qPCR target, fold difference to AB was calculated per experiment. Students t-tests with 95% confidence interval was done to compare differences: df = 10, * - p-value < 0.02, ** - p-value < 0.005, *** - p-value < 0.0006, t-statistic (promoters): AB vs TÜ = 2.9213, AB vs TL = 1.9066, AB vs NHGRI-1 = 3.944. t-statistic (mir430a): AB vs TÜ = 5.5112, AB vs TL = 2.6127, AB vs NHGRI-1 = 4.5587. t-statistic (mir430c): AB vs TÜ = 3.3161, AB vs NHGRI-1 = 6.6081.*

This data indicates that *mir430* repeat copy numbers could be 0.8-fold lower in TÜ, TL and NHGRI-1 compared to AB. While a difference of 0.8-fold seems small, this difference can make up 50 – 60 more *mir430* repeats. In DNA terms, this is a *mir430* locus potentially 100 kb longer in AB compared to TÜ, TL and NHGRI-1 – a difference that could have biological significance. Indeed, the higher *mir430* repeat numbers in AB positively correlate with the higher fraction of active nuclei at the 64-cells stage.

2.3.3 Regulatory information within *mir430* repeats

The findings so far suggest that *mir430* repeat numbers could play an important role in its activation during ZGA. *Mir430* repeats contain valuable *cis*-regulatory information. Past work has shown that *mir430* activation is dependent on the TFs Nanog, Pou5f3 and Sox19b (Lee et al., 2013; Leichsenring et al., 2013) Motif analyses show that Nanog, Pou5f3 and Sox19b motifs can be found within each *mir430* repeating unit (Fig. 2.3.3). These are presumably functionally relevant TF binding sites as ChIP-seq for these TFs showed cognate TF occupancy and, the independent loss of each of these factors results in downregulation of *mir430* (Lee et al., 2013). Thus, it is possible that the collective effect of many Nanog, Pou5f3 and Sox19b TF binding sites clustered within the *mir430* locus results in a robust activation signal early during ZGA. Following this idea, higher numbers of *mir430* repeats would result in earlier *mir430* activation, and conversely, lower repeats numbers would result in later activation. This hypothesis will be explored in further detail in the next chapter.

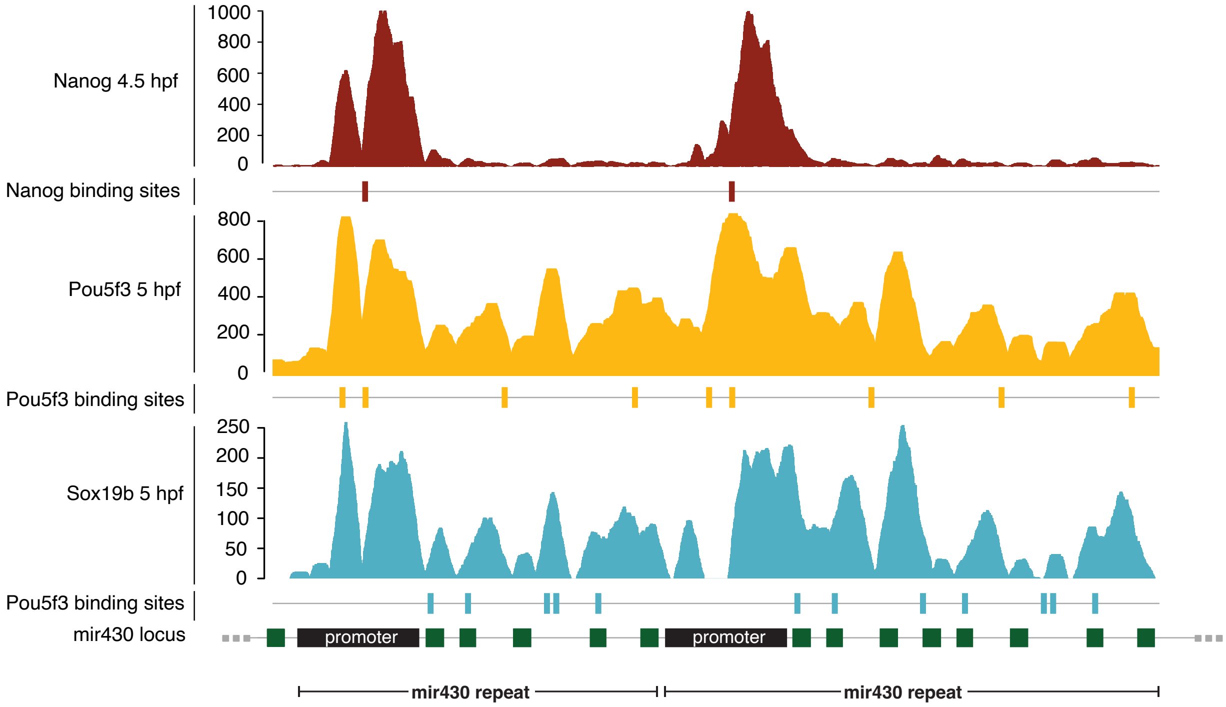

Figure 2.3.3: Example region of mir430 locus with Nanog, Pou5f3 and Sox19b TF binding sites.

ChIP-seq profiles for Nanog (Xu et al., 2012), Pou5f3 and Sox19b (Leichsenring et al., 2013b) are also shown.

2.3.4 Distinct cell cycle transcriptional dynamics between WT strains

The observation that higher *mir430* repeat numbers correlate with earlier detection of *mir430* transcription bodies in AB embryos suggest 2 possibilities: 1) *mir430* activates later in TÜ, TL and NHGRI-1 compared to AB, or 2) Higher numbers of *mir430* repeats result in higher rates of miR430 transcript production and therefore, easier detection of the miR430 transcript foci. This would result in miR430 transcript foci showing up more often in AB nuclei compared to the other strains, despite similar activation times. To distinguish between these two possibilities, I imaged the miR430 transcript foci with shorter time intervals to capture transcript foci growth rates. I expected if rate of transcript production determined at which developmental stage miR430 transcript foci showed up, I would find close correlation between the rates of miR430 transcript production and percentages of active nuclei. That is, high transcript foci growth rates would correlate with a high a percentage of active nuclei while low transcript foci growth rates would correlate with a low percentage of active nuclei. Due to the short lifetimes of miR430 transcript foci at the 64-cells stage, I followed transcript foci growth rates for all WT strains between 128-cells to 512-cells stages.

Following the miR430 transcript foci throughout the cell cycle, I observed the foci generally have a linear "growth phase" following initial detection (Fig. 2.3.4 A). The growth phase is succeeded by a sharp drop in total intensity – associated with the dissolution of the foci at the end of the cell cycle (Fig. 2.3.4 A and Fig. 5.4.1; Hadzhiev et al., 2019). The foci's growth and eventual dissolution equates to lifetimes of at least 4 mins in all cell cycles and WT strains (Fig. 2.3.4 A).

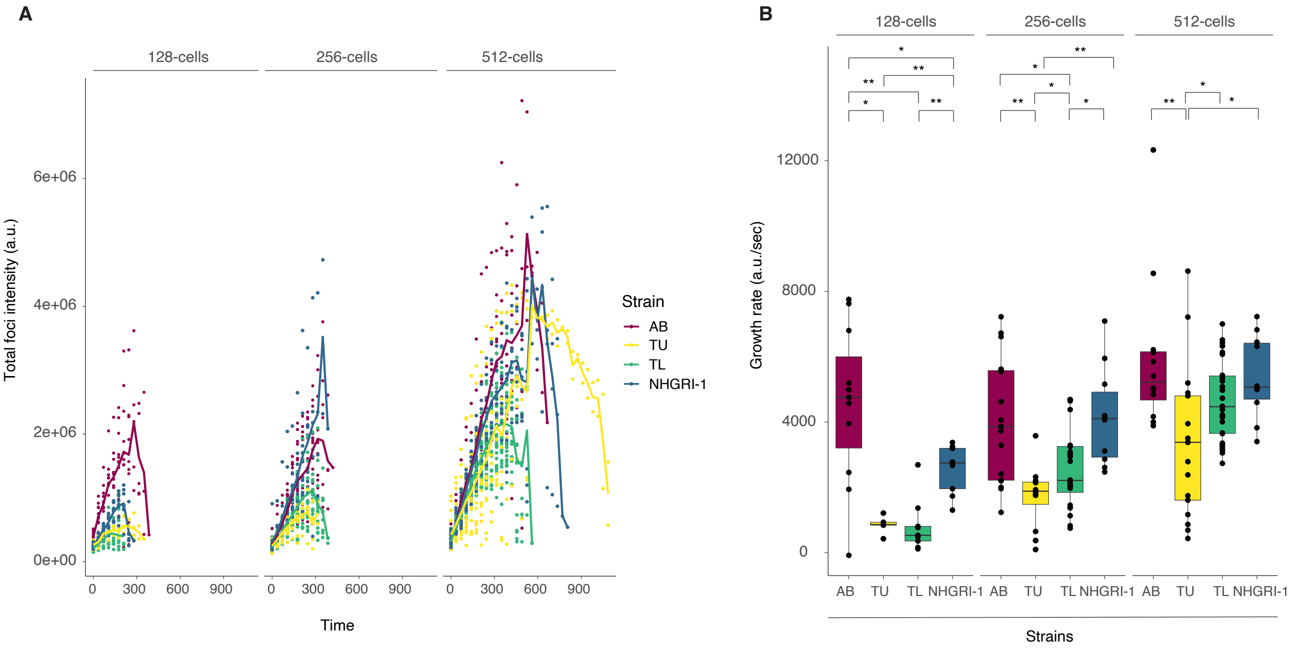

Figure 2.3.4: WT strains have distinct rates of miR430 transcript production that do not correlate with timing of activation

*A: Sum intensities of miR430 transcript foci over time for WT strains. B: Growth rates of miR430 transcript foci over the first 3 mins of miR430 transcript foci lifetimes. Comparisons of growth rates between strains was done using a Mann-Whitney-Wilcoxon (MW) test: Medians(128-cells) = 4757.8, 861.2, 532.9 and, 2749.2 a.u. for AB, TÜ, TL and NHGRI-1, respectively. Medians(256-cells) = 3860.2, 1884.8, 2212.7 and, 4102.4 a.u. for AB, TÜ, TL and NHGRI-1, respectively. Medians(512-cells) = 5223.2, 3379.1, 4470.5 and, 5067.4 a.u for AB, TÜ, TL and NHGRI-1, respectively. Mann-Whitney-Wilcoxon U-statistic (MWU; 128-cells): AB vs TÜ = 50, AB vs TL = 88, AB vs NHGRI-1 = 77, TÜ vs NHGRI-1 = 45, TL vs NHGRI-1 = 76. MWU(256-cells): AB vs TÜ = 180, AB vs TL = 297, TÜ vs TL = 203, TÜ vs NHGRI-1 = 116, TL vs NHGRI-1 = 189. MWU(512-cells): AB vs TÜ = 169, TÜ vs TL = 363, TÜ vs NHGRI-1 = 136. * - p-value < 0.05, ** - p-value < 0.01. N=2.*

To compare rates of *miR430* transcript production between strains, I obtained foci growth rates by calculating the change in total foci intensity over the first 3 mins of the foci's lifetime. I reasoned that this was most representative of the foci in its linear growth phase.

Here, I saw that transcript production rates followed very different dynamics in different strains. In AB, foci growth rates were already high at 128-cells stage and remained transcribing at a similar rate throughout the subsequent stages. In TÜ, TL and NHGRI-1, foci growth rates grew steadily going from 128-cells to 512-cells stages. Thus, *mir430* transcription in AB seems to reach its maximum rate earlier while the other strains increase *mir430* transcription more gradually. These results show that different strains have distinct *mir430* transcription dynamics.

Next, I wanted to relate the miR430 transcript foci growth rates to the percentages of active nuclei observed as *mir430* is activated. If percentages of active nuclei were determined by *mir430* transcription rates, I expected that high foci growth rates would correspond to a high percentage of active nuclei while low foci growth rates would correspond to a low percentage of active nuclei. Above a certain rate of transcript production, *miR430* transcript foci would always be visible, and this would likely be the case at the 256-cells and 512-cells stages. For this reason, I focused on the 128-cells stage for these comparisons. At the 128-cells stage, I saw that *mir430* transcription rates correlated poorly with percentage of active nuclei detected (Fig. 2.3.4 B). For example: at 128-cells stage, AB and TL strains both had above 90% active nuclei. However, transcript production rate was more than 5-fold lower in TL than AB (mean 809.4 a.u./sec vs 4542.2 a.u./sec). Similar findings were observed when comparing TÜ and NHGRI-1. At 128-cells, both strains had at above 70% active nuclei. However, transcript production rate was 3-fold lower in TÜ than NHGRI-1 (mean 856.6 a.u./sec vs 2557.4 a.u./sec). The lack of concordance between rate of miR430 transcript foci growth rates and percentages of active nuclei at the 128-cells suggest that these two metrics are independent of one another. I conclude that the observed differences in percentages of active nuclei at 64-cells between WT strains is unlikely to be resultant from differences in *mir430* transcription rates. However, it should also be noted that despite the general lack of correlation between transcription rate and percentage of active nuclei, AB does have the highest *mir430* transcription rates at 128-cells and, highest percentage of nuclei that activate at 64-cells. To reconcile these findings, I propose that the high *mir430* repeat numbers in AB may determine both the timing of activation and the rate of transcription, but independently. Timing of activation may be defined by TF binding sites within the repeat units while rate of transcription may be defined by a separate yet unknown mechanism.

2.3.5 Strain-specific maternal backgrounds define activation times

The findings presented so far show different *mir430* loci activating in the context of their respective WT strain backgrounds. However, variations in the *mir430* locus are unlikely to be the only existing inter-strain differences. Past studies have already reported transcriptomic variation between different WT strains (Holden & Brown, 2018). This raises the possibility that the above-described differences in timing of *mir430* activation may be resultant from *trans*-regulatory effects, such as differences in the maternally loaded background. To disentangle the *cis*-regulatory effects of copy number variation from differences in maternal background, I set out to probe the timing of *mir430* activation of the different WT strain *mir430* alleles given a constant maternal background. If the differences in timing of activation between strains were resultant from *trans*-regulatory effects, I expected these differences to equalise given the same background. Any remaining differences would thus be solely derived from *cis*-regulatory effects.

For this experiment, I took advantage of a *mir430* mutant previously generated in our lab (Kuznetsova et al., 2023; Ugolini et al., 2023), and the fact that in zebrafish, early embryogenesis is completely supported by maternally loaded gene products, with the father providing only a haploid genome. I crossed female *mir430* mutants with male fish from each of the WT strains (Fig. 2.4.5; left). The results in embryos carrying only one strain-specific *mir430* allele, and they all have the same maternal background. These hybrid embryos will from hereon be referred to as AB^{mirX}, TÜ^{mirX}, TL^{mirX} and NHGRI-1^{mirX}. I imaged *mir430* transcription body formation in embryos from these crosses from the 64-cells to 512-cells stage. These experiments showed that, like wildtypes, heterozygous *mir430* fish first activate *mir430* at the 64-cells stage and by the 512-cells stage, all nuclei are actively transcribing *mir430*.

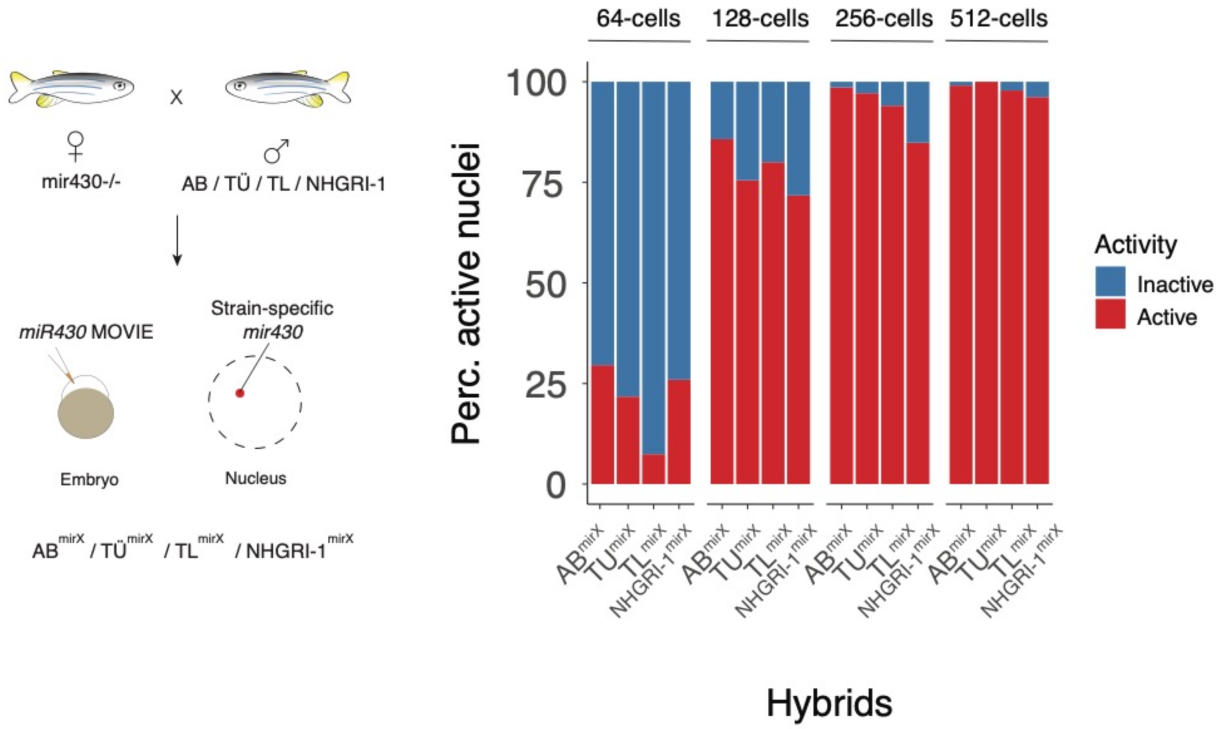

Figure 2.3.5: Maternal background accounts for inter-strain differences

Left panel: Schematic of mir430-/- X strains to get AB^{mirX}, TÜ^{mirX,} TL^{mirX} and, NHGRI-1^{mirX} embryos. Embryos were injected with H3K27Ac Fabs and miR430 MOVIE for imaging. Right panel: Percentage of active nuclei across cell cycle stages show that accounting for maternal contribution results in a loss of inter-strain differences, except in TL.

To quantify the activity of the different strain-specific alleles under the constant maternal background, I scored ‘active’ nuclei by the presence of a single *mir430* transcription body and ‘inactive’ by the absence of *mir430* transcription bodies. Surprisingly, under the same maternal background, the fraction of active nuclei at 64-cells stage equalises between AB, TÜ and NHGR-I (Fig. 2.3.5; right). I found that 29.5% of AB^{mirX} nuclei, 21.7% of TÜ^{mirX} nuclei and, 25.9% of NHGRI-1^{mirX} nuclei were active at the 64-cells stage. In contrast, only 4.8% of TL^{mirX} nuclei were active. In the subsequent stages, percentages of active nuclei equalise amongst different strain hybrids. The dramatic loss of fold-differences between AB^{mirX}, TÜ^{mirX} and NHGRI-1^{mirX} suggest that the high levels of activation previously observed in AB were resultant from *trans*-effects, such as levels of TF mRNAs encoded in maternally loaded transcriptome. Interestingly, TL^{mirX} embryos had even lower percentages of active nuclei compared to others in the same maternal background. This suggests that certain features, potentially enhancers, present in the AB, TÜ,and NHGRI-1 genomes but absent in TL may serve to boost *mir430* transcription.

Overall, I have shown that different WT strains have distinct *mir430* loci and transcriptional dynamics. Across WT strains, the percentages of nuclei where *mir430* is active at the 64-cells stage varies. These differences, while correlated with the number of *mir430* repeats, are, in fact, resultant from differences in the maternal background of different strains. Accounting for these inter-strain differences of maternal background, I saw that timing of *mir430* activation at 64-cells stage became largely equalised. Interestingly, I also found that rate of *mir430* transcription varies widely amongst WT strains. These patterns of rate do not correlate with the number of *mir430* repeats and could be regulated by a yet unknown mechanism. One possibility could be that differences in rate of transcription are also influenced by the maternal background. Finally, these findings show that the maternal background plays an important role in timing ZGA.

2.4 *mir430* repeat copy number defines timing of activation

2.4.1 Introduction

In the previous chapter, I have shown that the maternal background has an important role in determining timing of *mir430* activation. While I found in the end that *mir430* repeat numbers do not underlie inter-strain differences in timing of activation, the mega-repetitive nature of the *mir430* locus warranted further studies into how repeat copy number might regulate *mir430*. Moreover, the small fold-differences in *mir430* repeat numbers between strains may only go so far to reveal the effects of less or more *mir430* repeats. Differences in the number of promoters between the *mir430* locus and other early zygotic genes can be more than a 100-fold, and correlate with differences in timing of activation spanning 2-3 cell cycles. Thus, comparisons of the *mir430* locus with a smaller *mir430* locus, where copy numbers are more representative of other zygotic genes, may be more useful for studying the role of *mir430* repeat numbers on influencing timing of activation.

To this end, I set out to create a miniature transgenic (Tg) *mir430* locus at a known genomic position. This approach allows the comparison of the effects of a low copy number (Tg *mir430* locus) versus a high copy number (endogenous *mir430* locus) on the timing of activation within the same nuclear environment. If the timing of *mir430* activation is dependent on repeat copy numbers, I expected that a low copy number locus would activate later than the endogenous high copy number locus, despite the two loci having the same promoters. The ideal comparison in this case would have been replacement of the endogenous *mir430* with a miniature *mir430* at the same genomic position on chr4q – a feat that is not easily achievable. To be able to compare the effects of *mir430* repeat numbers on activation from the same genomic position, I aimed to create 3 Tg lines carrying 1x, 3x or 5x of *mir430* repeats inserted into the same genomic position. These Tg lines would provide a system to test both genomic position and *mir430* repeat numbers in determining timing of *mir430* activation. In this chapter, I will discuss the creation of these Tg lines and the insights gained from comparing timing of activation of the endogenous *mir430* locus and Tg *mir430.*

2.4.2 Generating a 3x *mir430* transgenic line

To create the Tg *mir430* loci, I set out to insert a plasmid construct containing either 1 copy, 3 copies or 5 copies of the 1.7 kb *mir430* repeat (Fig. 2.4.1 A). The single *mir430* repeat was first ordered as a gene-block based on the sequence of the *mir430* repeat consensus. Thus, it contains a *mir430* promoter, followed by 6 *mir430* genes in the 2 X a-c-b conformation. This repeat unit also contains all the TF binding sites native to the endogenous *mir430* repeat. Using this gene-block, I sequentially subcloned *mir430* repeats in the same orientation into an insertion backbone to get the 1x, 3x and 5x *mir430* insertion constructs.

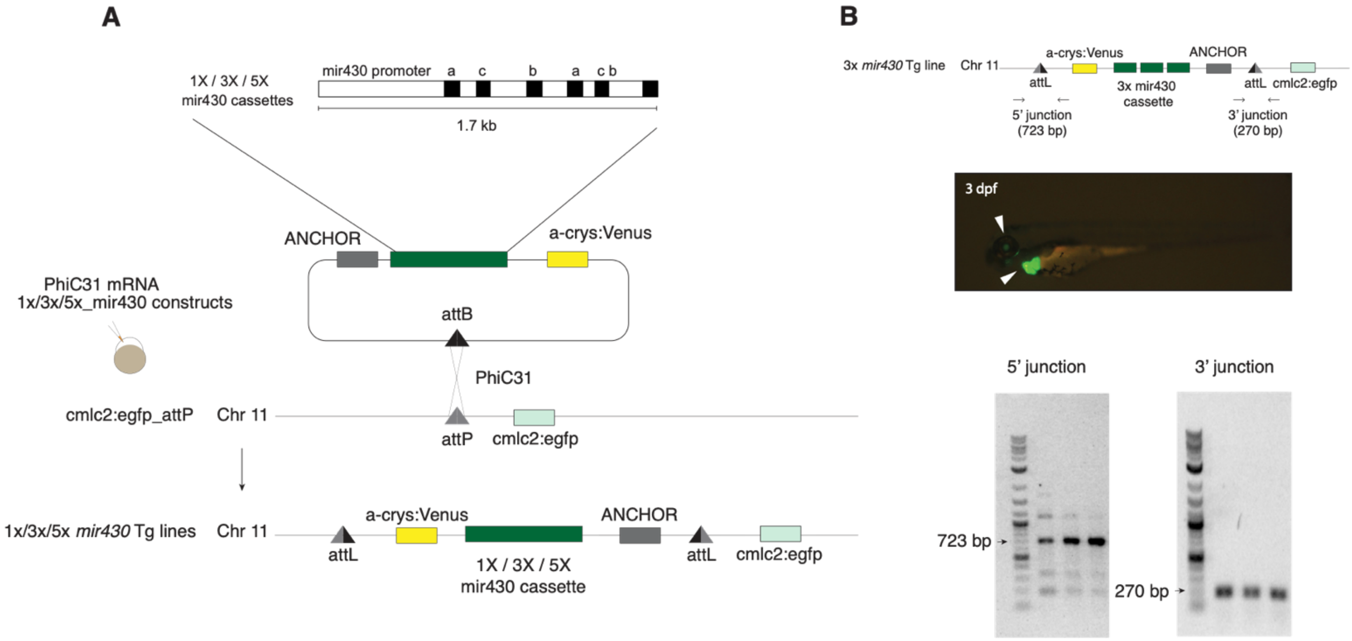

Figure 2.4.1: Transgenic insertion designs

A: Schematic of the design for inserted the 1x, 3x and 5x mir430 *constructs via site-directed transgenesis. The insertion constructs also contain a Tg marker,* α-crystallin:Venus, and an ANCHOR site for live-DNA labelling. B: Insertions of the 3x mir430 *plasmid. Confirmation of the presence of both Tg markers* (α-crystallin:Venus and cmlc2:egfp; white arrows) and presence of the attL and attR junctions were done.

Importantly, the Tg *mir430* loci needed to be visible via microscopy to allow comparisons with endogenous *mir430* activation. To this end, I adapted the ANCHOR DNA-labelling approach for zebrafish. The ANCHOR system is a DNA-labelling approach originally derived from the bacterial ParABS chromosome segregation partitioning system (Saad et al., 2014). In this system, a 1 kb long ANCHOR sequence is recognised by a cognate ParB1 protein which can be fluorescently labelled. Upon recognition of the ANCHOR sequence, fluorescently labelled ParB1 strongly binds on the ANCHOR site and oligomerises through weak ParB1-ParB1 interactions and, non-specific ParB1-DNA interactions in the surrounding region (Sanchez et al., 2015). The result of this is local signal amplification from fluorescent molecules labelling the ANCHOR Tg. This approach was chosen over other currently available live DNA-labelling approaches for several reasons: 1) Many current approaches such as the TetO/TetR or the LacI/LacO system require the insertion of high copy number repeats which are recognised by fluorescent-labelled cognate proteins (FPs) in a stoichiometric ratio. These arrays of repeats could be recombination prone and are typically large, which could negatively impact transgenesis efficiency. In contrast, the ANCHOR approach utilises a compact 1 kb long ANCHOR sequence which is not repetitive. Rather than stoichiometric recruitment of FPs, the ANCHOR system relies on oligomerisation to improve signal detection. 2) CRISPR-based approaches, such as dCas9-GFP or dCas9 coupled with single-guide RNAs (sgRNAs) with MS2 stem loops that bind to MCP-GFP proteins, still struggle to effectively label single-copy loci. This approach ultimately also depends on the number of sgRNA recognitions sites present. The utility of the ANCHOR approach over the above-described approaches has been shown previously in yeast, human cell lines, plants and, Drosophila tissues (Delker et al., 2022; Germier et al., 2018; Meschichi et al., 2021; Saad et al., 2014). Therefore, to adapt this to the *mir430* transgenic lines in zebrafish, the 1 kb ANCHOR sequence was included in the 1x, 3x and 5x *mir430* insertion constructs.

The insertions of the plasmid constructs into the zebrafish genome were done in a site-directed manner on Chr11 via gateway recombination. Specifically, I used a fish line containing a transgenic attP site present on Chr 11 previously characterised by Mosimann *et al* (2013; Fig. 2.4.1 A). This attP landing site was shown to be resistant to silencing – a common problem with transgenes. Integration of the Tg *mir430* plasmids occurs via a corresponding attB site cloned into the insertion construct. The insertion plasmids were co-injected with mRNA encoding the PhiC31 recombinase into 1-cell stage embryos from the attP landing site line to create the 1x, 3x and 5x insertions on chr11 (Fig. 2.4.1 A).

Transgenic markers were present in the attP landing site line and, the Tg *mir430* constructs. The attP landing site line contains a *cmlc2:egfp* Tg marker which results in green fluorescence detectable in the larval cardiac tissue. The Tg *mir430* constructs contain an *α-crystallin:Venus* Tg marker which results in yellow fluorescence detectable in the lens. For all 3 Tg lines, I determined successful integration of the plasmids into the genome by observing larvae that show both transgenic markers. However, in the 1x and 5x *mir430* lines, I saw that the *cmcl2:egfp* Tg marker and the *α-crystallin:Venus* Tg marker were not linked *in-cis.* Rather, I observed instances where the two Tg markers were inherited separately. This suggests that the 1x and 5x *mir430* plasmids were randomly inserted elsewhere in the genome, as the supposed attP insertion site and *cmlc2:egfp* Tg marker are closely positioned on the genome. This 'incorrect' insertion was confirmed using low coverage whole genome sequencing which revealed that both plasmids were inserted into the exact same location on Chr8 and, were concatenated several times to varying degrees.

In contrast, the Tg phenotypes co-segregated amongst 3x mir430 larvae. To confirm the site-specific insertion of the *3x mir430* plasmid, I was also able to detect the 5' and 3' junctions resultant from attP/attB recombination (Fig. 2.4.1 B). Overall, I set out to create Tg mir430 lines containing *1x, 3x and 5x mir430* repeats inserted into the same genomic position. Of the 3 Tg lines, the *1x and 5x mir430* lines were likely integrated elsewhere in the genome via homologous end joining of a unique seed sequence within the plasmid. These findings highlight the need for rigorous tests that confirm site-specific insertion events when performing targeted transgenesis. They also identify an unexpected recombination-prone site found on Chr8 of the zebrafish genome, potentially when using the *α-crystallin:Venus* Tg marker. The *3x mir430* Tg was, however, correctly inserted into the attP site. The resultant *3x mir430* Tg also contains an ANCHOR sequence which can be used for live fluorescent imaging of the Tg locus.

2.4.3 Improved ANCHOR DNA-labelling of the 3x mir430 Tg

Due to the uncertainty surrounding the nature of the 1x *mir430* and 5x *mir430* insertions, I decided to focus on the *3x mir430* line where the insertion was well characterised. Upon establishing the stable *3x mir430* Tg line, I first confirmed the functionality of the ANCHOR approach in live DNA labelling during zebrafish embryo development. I injected *3x mir430* 1-cell stage embryos with mRNA encoding ParB1-mNeongreen and, miR430 MOVIE to label nascent miR430 transcripts (Fig. 2.4.2). These injected *3x mir430* embryos were then imaged at post-ZGA stages, as they were the stages when the *3x mir430* Tg was most likely to be active. At the 1k-cells stage, I observed that ParB1-mNeongreen was expressed, localised mainly in the cytoplasm, and was largely excluded from the nucleus (Fig. 2.4.2 C). This pattern of subcellular localisation mirrors previously published work in eukaryotic cells (Germier et al., 2018; Saad et al., 2014) and, is likely due to the absence of a nuclear localisation signal (NLS) on the ParB1-mNeongreen protein. On closer inspection, I saw that ParB1-mNeongreen formed discrete foci in the nucleus (Fig. 2.4.2 C; top panel). These are presumably the labelled Tg *3x mir430* foci because I observed foci of nascent miR430 transcript signal colocalising with the ParB1-mNeongreen foci. These findings prove that the nuclear ParB1-mNeongreen foci detected represent the successful labelling of the *3x mir430* Tg. Furthermore, they provide an initial indication that the *3x mir430* Tg is actively transcribed during development.

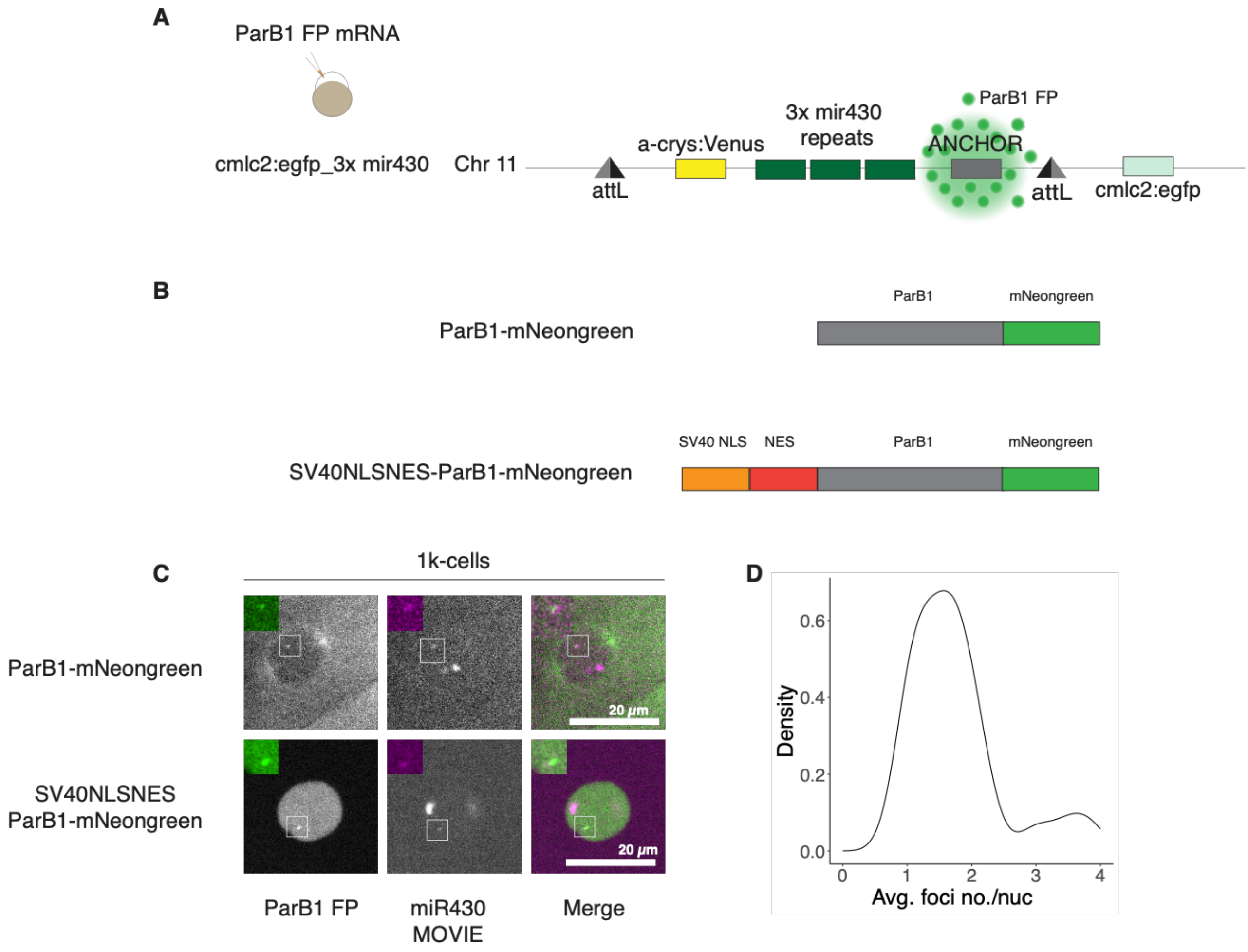

Figure 2.4.2: Addition of the NLS-NES to the N-terminus of ParB1-mNeongreen improves 3x mir430 Tg detection by ANCHOR

A: To visualise the 3x mir430 Tg, 1-cell stage embryos from the 3x mir430 line are injected with mRNA encoding the ParB1 fluorescent protein (FP). The resultant translated ParB1 FP recognises the ANCHOR sequence and oligomerises around the ANCHOR site and non-specifically around the neighbouring DNA. B: ParB1-mNeongreen with and without the NLSNES on the N-terminus. C: Comparison of ParB1-mNeongreen Tg labelling signal using ParB1-mNeongreen alone or ParB1-mNeongreen with the SV40 NLS-NES on the N-terminus. D: Distribution of the average number of ParB1-mNeongreen foci detected in nuclei throughout the cell cycle at 1k-cells stage shows that on average, only 2 NLSNES-ParB1-mNeongreen foci are detected in homozygotes (no. of nuclei=26).

Despite the functionality of the ANCHOR system in the developing zebrafish embryo, I found that the signal-to-noise ratio was often poor, impeding easy foci detection. I reasoned that this was due to poor localisation of the ParB1-mNeongreen protein to the nucleus. Rather than active import into the nucleus, cytoplasmic ParB1-mNeongreen protein was likely incorporated into the nucleus during mitosis when the nuclear envelope disintegrates and reforms. Considering this, I generated a ParB1-mNeongreen encoding construct with an SV40 NLS and a nuclear export signal (NES) on the N-terminus (Fig 2.4.2 B). I expected that the presence of the SV40NLS-NES would facilitate exchange of ParB1-mNeongreen protein in and out of the nucleus, potentially facilitating oligomerisation on the ANCHOR sequence. From imaging of *3x mir430* embryos injected with SV40NLSNES-ParB1-mNeongreen mRNA and miR430 MOVIE, I saw that adding the SV40NLSNES to the ParB1-mNeongreen resulted in clear enrichment of the FP in the nucleus compared to the cytoplasm. Compared to ParB1-mNeongreen without the SV40NLSNES, I observed a marked improvement in Tg foci detection. As before, I was able to visualise nascent miR430 transcript signal colocalising with the SV40NLSNES-ParB1-mNeongreen foci (Fig 2.4.2 C).

Previous work with the ANCHOR system has suggested that addition of an NLS to the ParB1 FP would negatively impact the signal-to-noise ratio and may cause the formation of non-specific aggregates (Germier et al., 2018). To determine if this was the case, I looked at the number of ParB1-mNeongreen spots in homozygous nuclei from early to later developmental stages. In general, only 2 SV40NLSNES-ParB1-mNeongreen foci were detected per nuclei (Fig. 4.4 D).

In summary, I showed that the ANCHOR DNA labelling system can be used to label the *3x mir430* Tg during early embryogenesis. In addition to establishing this technique for live imaging of genomic loci during zebrafish embryogenesis, I have also improved the signal-to-noise ratio of ANCHOR labelling by replacing the ParB1-mNeongreen with SV40NLSNES-ParB1-mNeongreen. The initial findings from these experiments indicate that the *3x mir430* Tg is in fact, transcriptionally active and forms a focus of nascent miR430 transcripts. Thus, live imaging of the 3x *mir430* Tg is a tractable system for studying the role of *mir430* repeat numbers on timing of activation. From hereon, I will refer to SV40NLSNES-ParB1-mNeongreen as ParB1-mNG.

2.4.4 The 3x mir430 Tg activates later during development

Having the initial indications that the *3x mir430* Tg was transcriptionally active during embryogenesis, I next set out to determine the timing of activation of the Tg with respect to the endogenous *mir430* locus. To do this, I imaged *3x mir430* embryos injected with ParB1-mNG mRNA and miR430 MOVIE from the 64-cells to 1k-cells stages. As expected, the endogenous *mir430* transcription bodies were visible in nuclei of all cell cycle stages imaged (Fig. 2.4.3 A). In contrast, transcriptional activity from the *3x mir430* Tg was only detected from the 256-cells stage onwards (Fig. 2.4.3 A).

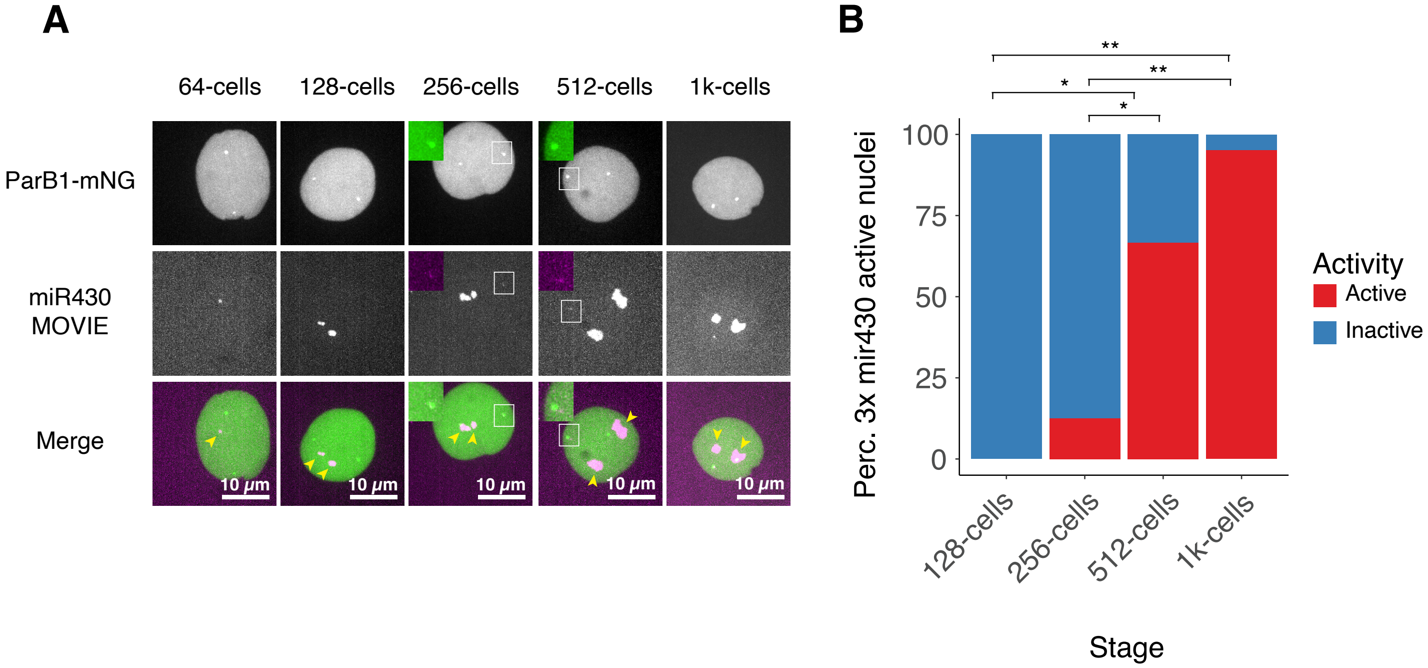

Figure 2.4.3: 3x mir430 activates later during development.

*Left panel: Representative images showing 3x mir430 activation at 64-cells, 128-cells,256-cells, 512-cells and 1k-cells stage. miR430 transcripts are shown in magenta and NLSNES-ParB1-mNeongreen is shown in green. Yellow arrows indicate endogenous mir430 transcription bodies while white box insets indicate 3x mir430 transcription. Right panel: Fraction of active nuclei for the 3x mir430 Tg from 128-cells to 1k-cells stages. Fisher's exact test was used to compare fraction of active nuclei between stages. * - p-value < 0.007, ** - p-value < 0.0001.*

To characterise *3x mir430* activity throughout development, I quantified the percentage of nuclei where the *3x mir430* Tg was transcriptionally active, which I define as nuclei where the ParB1-mNG foci and miR430 MOVIE foci colocalise in at least one frame. Due to the lack of activity observed prior to 256-cells, I quantified active nuclei only for the stages between 128-cells and 1k-cells (Fig. 2.4.3 B). This analysis showed that the *3x mir430* Tg is activated in a small fraction of nuclei at 256-cells stage (7.7%) and increases progressively in the following cell cycle stages – 69.9% at 512-cells, and 95.2% at 1k-cells. These findings reveal that the *3x mir430* Tg turns on 2 cell cycle stages later than the endogenous *mir430*, despite the same maternal background and promoter type. The observed activation profile of the *3x mir430* Tg recapitulates findings in a recent study where transcription from a single *mir430* promoter upstream of a reporter gene was detected around the 256/512-cells stage (Hadzhiev et al., 2023). I conclude that the *3x mir430* Tg is activated later during development than the endogenous *mir430*. This proves that *mir430* repeat numbers do define the timing of *mir430* activation.

Interestingly, during this analysis, I also observed that a fraction of ParB1-mNG foci across all stages co-localised with the endogenous *mir430* transcription body (Fig. 5.4.2). As I am unable to distinguish between transcripts deriving from the endogenous *mir430* locus and *3x mir430* Tg, I assume that both loci are transcriptionally active within the same nuclear compartment. This finding was particularly interesting as it suggested that similar promoters may tend to co-transcribe together in nuclear space, despite being present on different chromosomes. In fact, 25.9% of ParB1-mNG foci at 512-cells and 1k-cells stages had an instance of colocalization with the endogenous *mir430* transcription body, suggesting that these might not be random events (Fig. 5.4.2). In the future, more work will need to be done to understand if similar promoter types may define spatial co-transcription of genes.

2.4.5 The *3x mir430* Tg activates at the same time as other zygotic genes

So far, I have showed that *mir430* repeat copy numbers can define the timing of *mir430* activation during ZGA. Recently, it was also proposed that *mir430* is regulated by similar mechanisms as other early ZGA genes (Haberle et al., 2014; Hadzhiev et al., 2023). Specifically, it was proposed that early activating genes (including *mir430*) typically had promoters that contained a TATA-box and sharp TSSs. In contrast, later activating gene promoters lacked a TATA-box and had broad TSSs (Hadzhiev et al., 2023). While these findings suggest a broad regulatory logic for early transcription during ZGA, they do not explain the differences between the timing of activation of *mir430* and the other early genes which turn on 2-3 cell cycles later (Bhat et al., 2023; Hadzhiev et al., 2023; Heyn et al., 2014). I hypothesised that in addition to there being a promoter code for early transcription, higher copy numbers of these early promoters may distinguish *mir430* from the other early zygotic genes. This may explain why *mir430* activation precedes that of other early zygotic genes. To test this, I compared the expression of endogenous *mir430*, *3x mir430* and other early zygotic genes throughout development. If the high number of promoters distinguished the endogenous *mir430* from other early zygotic genes, I expected that the *3x mir430* Tg would then activate at the same time as other zygotic genes due to its lower promoter copy number.

For this experiment, I assayed the expression of non-specific miR430 transcripts (transcripts deriving from the endogenous *mir430* locus and *3x mir430* Tg*)*, 3x miR430 *transcripts* (transcripts deriving specifically from the *3x mir430* Tg) and transcripts of early zygotic genes (*dusp6*, *grhl3* and, *mxtx2;* Bhat et al., 2023; Hadzhiev et al., 2023; Heyn et al., 2014). This was done using RT-qPCR on total RNA from the 2-cells stage to Sphere stage. An RT-qPCR was used in this case instead of live microscopy to assay expression as live-labelling approaches were not available for assaying *dusp6*, *grhl3* and *mxtx2* expression. While this results in lower mRNA detection sensitivity, it also allows the simultaneous probing of expression for multiple genes throughout development.

From this experiment, I observed that non-specific miR430 transcripts were expressed earliest at the 64-cells stage and transcript levels increased drastically in the following stages (Fig. 2.4.4; left). The non-specific miR430 transcripts observed here are likely derived mainly from the endogenous *mir430* locus given that the *3x mir430* is least 50-fold smaller than the endogenous *mir430* locus and, that the imaging results show that the *3x mir430* Tg is not active at these early stages (Fig. 2.4.3). Next, I looked at the expression of the 3x miR430 transcripts. Consistent with the imaging results, I observed that the *3x mir430* Tg is activated later than the endogenous *mir430* – around the 1k-cells/High stages (Fig. 2.4.4; right). These timings did not correspond exactly with the timing of *3x mir430* activation observed via imaging (256/512-cells stages), likely due to the lower sensitivity of the RT-qPCR approach. Finally, I compared *3x mir430* expression with expression of the other early zygotic genes *dusp6, grhl3* and *mxtx2.* I observed that *mxtx2* is expressed around the 1k-cells/High stages, while *dusp6* and *grhl3* are expressed around the High/Sphere stages (Fig. 2.4.4; right). These transcripts follow similar expression profiles as the *3x mir430* Tg suggesting that they are activated around the same time. Similar to the *3x mir430* Tg, these early genes are likely activated 2-3 cycles early but are missed due to the lower sensitivity of the RT-qPCR approach.

Overall, these findings corroborate the differences in timing of activation between the endogenous *mir430* and *3x mir430* observed via imaging. They further show that a low copy number *mir430* locus activates around the same time as other early zygotic genes (*dusp6*, *grhl3* and *mxtx2*). This suggests that high copy numbers of the proposed "early" promoter, which contains a TATA-box and a sharp TSS (Haberle et al., 2014; Hadzhiev et al., 2023), could distinguish the endogenous *mir430* locus from other early zygotic genes for activation at the 64-cells stage.

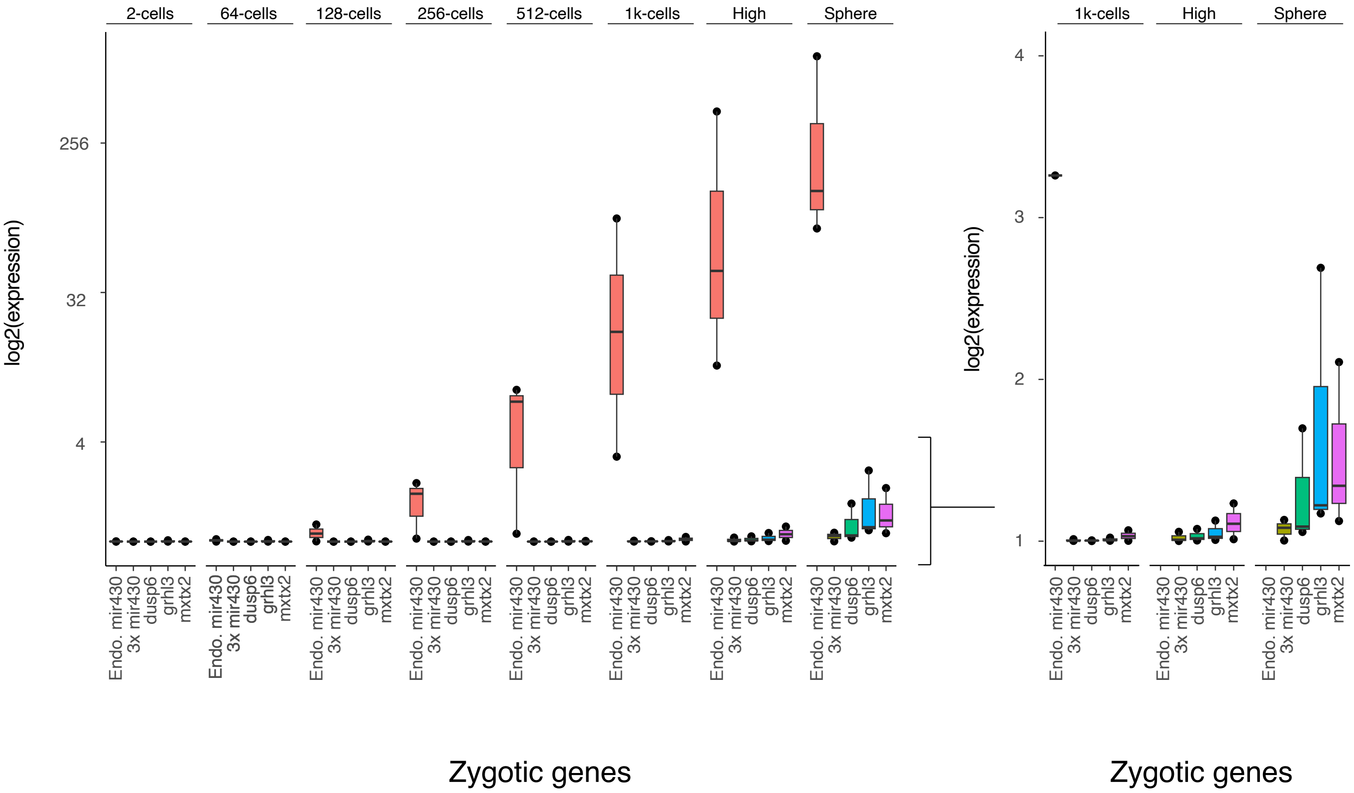

Figure 2.4.4: 3x mir430 activates around the same time as minor wave genes

RT-qPCR on total RNA from 2-cells to Sphere stage for zygotic transcripts. The y-axis scale is log2(expression) and is skewed to represent the large differences in transcripts levels. Values higher than 5 on the log2(expression) scale are represented only as 5. These results show distinct timings of activation between the endogenous mir430 and the 3x mir430 Tg. Endogenous miR430 transcripts are detected from 64/128-cells onwards while the 3x miR430 transcripts are only detected from 1k-cells/High stage onwards. A comparison of 3x mir430 Tg timing of activation to that of other minor wave zygotic genes defined in Hadzhiev et al (2023) showed that 3x mir430 activates around the same time as dusp6,grhl3 and mxtx2.

2.4.6 Nanog overexpression compensates for lower TF binding sensitivity

Throughout this thesis, I have showed using multiple lines of evidence that repeat copy numbers of the *mir430* locus can determine the timing of *mir430* activation. What exactly do high repeat numbers confer to the *mir430* locus? As shown previously, each of the *mir430* repeats contain TF binding sites for TFs known to regulate *mir430*, such as Nanog, Pou5f3 and Sox19b (Fig. 2.3.3). These binding sites are presumably functional as they are occupied by their cognate TFs and loss of these factors have been shown to result in downregulation of mir430 expression (Fig. 2.3.3; Lee et al., 2013). Of these TFs, Nanog has been shown to be indispensable for *mir430* activation and, likely sits at the top of the activation cascade for *mir430* (Kuznetsova et al., 2023; Lee et al., 2013). High numbers of Nanog binding sites (and potentially Pou5f3 and Sox19b sites) localised at the repetitive *mir430* locus could collectively confer higher sensitivity of the *mir430* locus for TF binding. In other words, many Nanog binding sites could increase the changes of Nanog binding, even when Nanog levels are limiting during early embryonic stages. In the context of the competition model, where histones compete with TFs for access to TF binding sites, a higher number of TF binding sites concentrated at a genomic locus could result in localised out-competition of histones by TFs. Thus, allowing earlier gene activation despite a repressive nuclear environment.

I aimed to test the Nanog sensitivity hypothesis for *mir430*. Specifically, if the lower repeat numbers at the *3x mir430* Tg result in lower Nanog binding sensitivity, it would be predicted that higher amounts of higher amounts of available Nanog would compensate for this, resulting in earlier *3x mir430* activation. To test this, I overexpressed Nanog in *3x mir430* Tg embryos. I injected 120 pg of Nanog-HA mRNA together with ParB1-mNG mRNA and miR430 MOVIE into *3x mir430* 1-cell stage embryos. I then imaged the injected embryos from the 256-cells to 1k-cells stage and compared whether the timings of *3x mir430* activation differed from *3x mir430* embryos without Nanog overexpression.

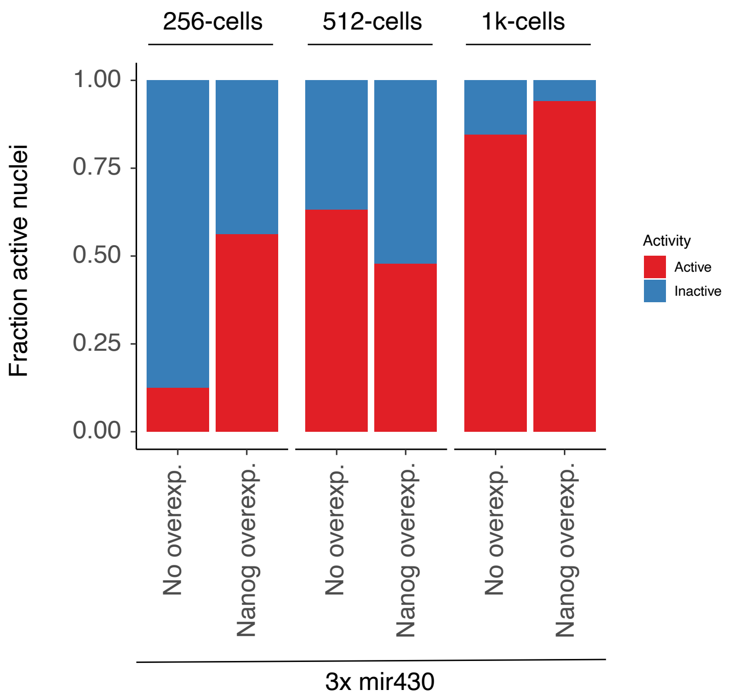

Figure 2.4.5: Nanog overexpression advances 3x mir430 activation during development.

Overexpression of Nanog results in a higher fraction of nuclei where the 3x mir430 Tg is active at the 256-cells stage. Fraction of active nuclei at 512-cells and 1k-cells is unchanged.

Without Nanog overexpression, the *3x mir430* Tg is only active in 7.7% of 256-cells stage nuclei, and this increases steadily in the subsequent cell cycle stages. With Nanog overexpression, I saw an increase in the percentage of nuclei where the *3x mir430* Tg was active at the 256-cells stage from 7.7% to 56% percent (Fig. 2.4.5). The percentage of active nuclei in the subsequent stages were comparable to conditions without Nanog overexpression. These results confirm that higher amounts of Nanog can compensate for a *mir430* locus with lower Nanog binding sensitivity. Thus, the ability of the endogenous *mir430* locus to activate so early during development depends on the high number of Nanog (and potentially Pou5f3 and Sox19b) binding sites that confer the locus with higher sensitivity for Nanog binding.

A particularly curious observation was that Nanog overexpression did not result in an increase in percentage of active nuclei at the 512-cells stage, despite the increases observed at the 256-cells stage. At the moment, the reasons for these results remain unclear.

2.4.7 3x miR430 transcript foci are observed later during the cell cycle

(This section contains collaborative work from J. Brenner and R. Purkanti)

Thus far, we have seen that the *3x mir430* activates later than the endogenous *mir430* locus due to its lower sensitivity for Nanog binding. In addition to the differences in timing of activation between the endogenous *mir430* locus and the *3x mir430* Tg observed during *development*, we also observed, on closer inspection, that the timing of activation during the *cell cycle* was distinct. In particular, the endogenous *mir430* locus often activates earlier during the cell cycle than the *3x mir430* Tg (Fig. 2.4.6 A- B). These differences account for 1-2 mins of delay between detection of endogenous *mir430* transcriptional activity and *3x mir430* transcriptional activity at both 512-cells and 1k-cells stages (Fig. 2.4.6 C). Knowing now that the *3x mir430* Tg has lower sensitivity to Nanog binding compared to the endogenous *mir430* locus, we hypothesised that later activation during the cell cycle may be resultant from a longer time required for sufficient Nanog to be recruited to the *3x mir430* locus to initiate activation. This would be consistent with a model where higher numbers of Nanog binding sites not only increase the probability of Nanog binding, but also enhances the rate of Nanog recruitment, resulting in earlier cell cycle activation times. Conversely, lower numbers of Nanog binding sites would confer a slower rate of Nanog recruitment and result in later cell cycle activation times.

Interpreting endogenous *mir430* and *3x mir430* activation times is complicated by the fact that the endogenous *mir430* and *3x mir430* Tg may synthesise *miR430* transcripts at vastly different rates given the difference in number of *mir430* genes. As such, the *3x mir430* transcript foci may require a longer time to become detectable. Here, I term differences in true activation times as "activation delays" and differences due to different rates of *miR430* transcript production as "detection delays" (Fig. 2.4.7).

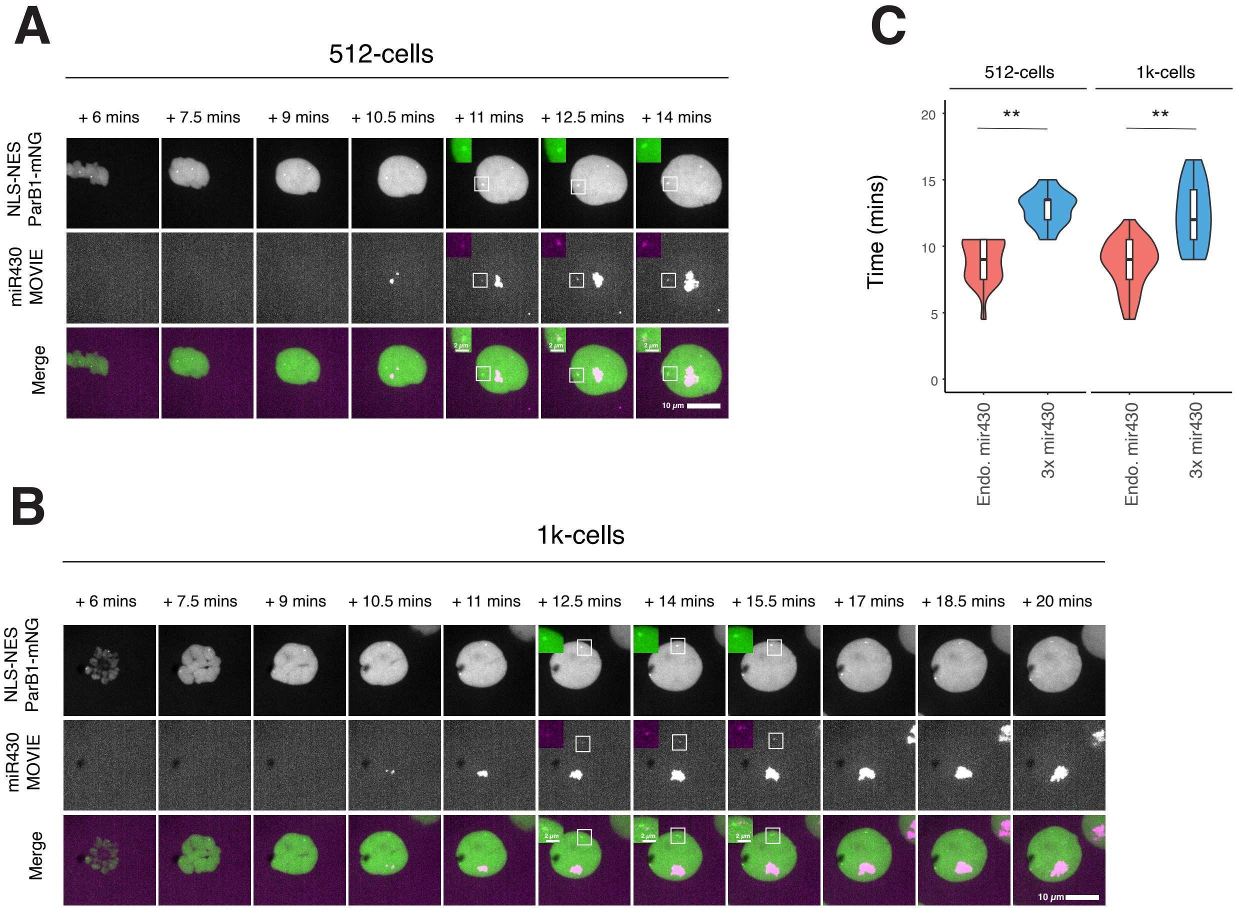

Figure 2.4.6: Timing of 3x mir430 transcript foci detection is later during the cell cycle than the endogenous mir430

*A and B: Representative images showing that the 3x miR430 transcript foci are detected later in the cell cycle than the endogenous miR430 transcript foci at both the 512-cells and 1k-cells stage. The indicated time represents time following the formation of the metaphase plate. C: Quantification of differences in timing of foci detection shows that transcript foci of the endogenous mir430 are detected later than the 3x mir430 Tg at both 512-cells and 1k-cells stage. A Welch t-test was used to compare differences in means between the endogenous miR430 and 3x miR430 transcript foci detection times. For 512-cells: df=27.3, ** - p-value < 0.0000006, t-statistic=-6.4845. For 1k-cells: df=45.3, ** - p-value < 0.0000001, t-statistic=-6.3521.*

To determine the extent to which these two factors influence the timing of foci detection, I compared rates of transcript production at the endogenous mir430 locus and the 3x mir430 locus at the 1k-cells stage. Specifically, I followed the growth of *miR430* transcript foci at these respective loci over time (see chapter 2.3.4). As expected, I observed a marked difference in rate of transcript production between the *3x mir430* Tg and the endogenous *mir430* (Fig. 2.4.7 B). To determine if these differences can fully account for the delays in cell cycle detection, I used the transcript foci growth rates over the first 1.5 mins (2 time points) to extrapolate true activation times for the *3x mir430* Tg and the endogenous *mir430* (Fig. 2.4.7 C). Based on previous analysis on *miR430* transcript production rates, I assume that transcript production following activation is largely linear over time. Therefore, I determined true activation times by doing a linear extrapolation. From this analysis, I found the true activation times for the endogenous *mir430* and *3x mir430* Tg to be 8.57 mins and 8.83 mins, respectively. This is still preliminary data as only 6 tracks were acquired for the *3x mir430* transcript foci. However, they indicate, rather surprisingly, that cell cycle activation times between the endogenous *mir430* and *3x mir430* are comparable. This means that the differences in observed timing of *3x mir430* and endogenous *mir430* transcript foci during the cell cycle are largely due to differences in rate of transcript production. In the context of Nanog recruitment to these loci, it suggests that both the endogenous *mir430* and *3x mir430* take the same amount of time to recruit sufficient Nanog to activate transcription. Thus, the effects of different numbers of TF binding sites operate at the developmental level but seem to be neutral at a cell cycle level.

In conclusion, I have shown throughout this thesis that the ability of *mir430* to activate so early during development is dependent on the number of copies of *mir430* repeats. I have shown using a transgenic *mir430* model that different numbers of repeats, and their TF binding sites therein, confer different sensitivities to Nanog binding. Higher number of repeats, result in more Nanog binding sites and, higher sensitivity of a locus for Nanog binding. In the context of competition between histones and TFs, higher sensitivity of the *mir430* locus for TF binding may allow localised out-competition of the TFs against histones for access to their cognate binding sites, despite a generally repressive nuclear environment. These genetic features, therefore, drive the early activation of *mir430.*

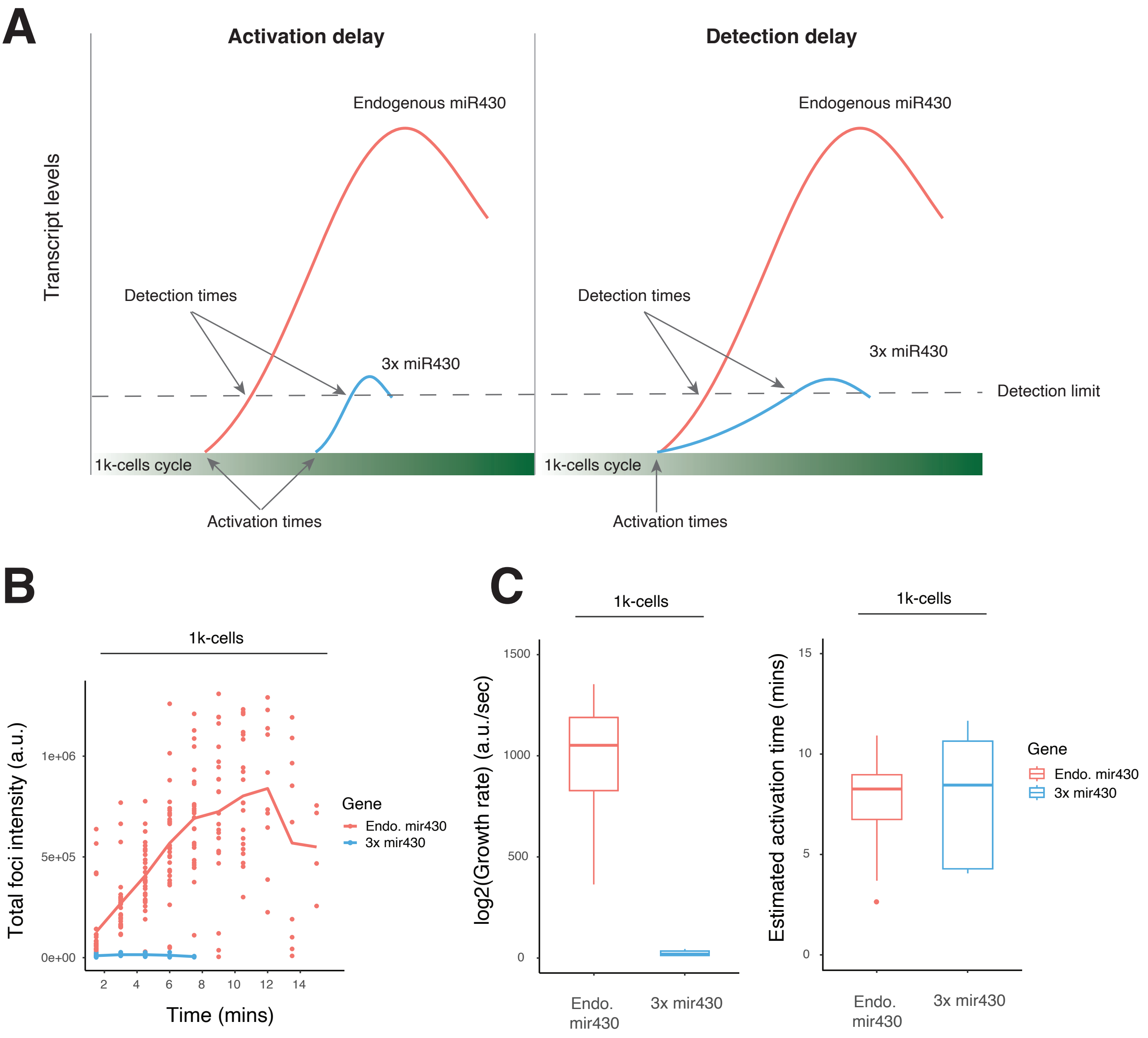

Figure 2.4.7: Detection delays explain differences in timing of transcript foci within cell cycles

A: Graphical explanations of Activation delays vs Detection delays. Activation delays are resultant from differences in timing of activation during the cell cycle, even though the endogenous mir430 and 3x mir430 transcribe at identical rates. Detection delays are resultant from differences in transcription rates even though both the endogenous mir430 and 3x mir430 are activated at the same time. B: Tracking of transcript foci for the endogenous mir430 and 3x mir430 over time at the 1k-cells stage. No. of foci for endogenous=29, No. of foci for 3x mir430=8. C: Left panel - Transcript foci growth rates over the first 3 mins from foci detection were quantified and compared for the 1k-cells stage. Y-axis is log2(Growth rate). Right panel – Estimation of activation times using linear profiles of transcript foci growth rates for the endogenous miR430 and 3x miR430. No. of foci tracked for B and C are 29, and 6 for endogenous mir430 and 3x mir430, respectively. These preliminary findings show that activation occurs largely within similar time periods.

3 Discussion

For this thesis, I had set out to gain a comprehensive understanding of the mir430 locus and how it is activated during ZGA. A major obstacle in studying mir430 is the fact that the locus is so repetitive, rendering many genetic and transcriptomic approaches difficult. To overcome this obstacle, I took advantage of both novel and traditional approaches to characterise the structure of the *mir430* locus and the transcripts it produces. Furthermore, I showed using multiple lines of evidence that *mir430* activation is dependent on its high repetitiveness, which confer to it higher sensitivities to TF binding. Together, these findings contribute towards a more holistic understanding of mir430 and its activity during ZGA. Furthermore, they reveal core principles of transcriptional regulation that may help us understand gene regulation in general.

3.1.1 The mir430 locus is mega-repetitive and hundreds of kilobases large

Using Xdrop targeted long read sequencing, I have re-assembled the *mir430* locus in an AB fish which is at least 150 kbp in size, containing 71 *mir430* repeat units and 424 *mir430* genes. This is at least 10-fold larger than the 16 kbp locus represented in the GRCz11 reference genome. This demonstrates that with modern advancements in sequencing technologies, we can better resolve complexed genomic regions such as the *mir430* locus. Concerning particularly repetitive genomic regions such as the zebrafish *mir430* locus, long read sequencing approaches and its variants will likely have a pivotal role in future research.

Of note, the *mir430* locus re-assembled here was done on an AB fish genome. The finding that different strains have distinct *mir430* repeat copy numbers calls to question how representative the *mir430* sequence assembled here is for the different strains. A proper characterisation of these inter-strain differences would require re-sequencing of the genomes of different WT strains - a feat that would be expensive, to say the least. For now, the strong conservation of the sequence of *mir430* repeat units, strongly suggests that inter-strain differences in *mir430* occur mainly at the repeat copy number level.

During this work, two research groups had also independently set out to re-assemble the *mir430* locus using long read sequencing (Hadzhiev et al., 2023; Pownall et al., 2023). Rather than taking a targeted approach, they used whole genome long read sequencing (Oxford Nanopore (ONT)-seq and PacBio HiFi) to resolve the *mir430* locus in an AB/TU/TL hybrid fish (Pownall et al., 2023) or in an AB fish (Hadzhiev et al., 2023). In both cases, they re-assembled a *mir430* locus approximately 550 kb in size with at least 300 copies of the stereotypic *mir430* repeat units. Only in one of the cases was a single end-to-end *mir430* locus-containing contig assembled (Pownall et al., 2023). In contrast, the *mir430* locus assembled in Hadzhiev et al. (2023) was split between 14 contigs and no end-to-end contig was assembled. While the results from both studies agreed with one another, by these metrics, the *mir430* locus presented in Pownall *et al.* (2023) is superior in quality. Albeit the *mir430* locus represented there is likely an average between the AB, TU and TL WT strains.

Both *mir430* loci assembled in the above-mentioned studies however, differed in size from our assembled *mir430* locus. The disparity between the results from the whole genome sequencing approaches and ours may be resultant from the limitations of the Xdrop technique. Previous publications of Xdrop usage reported up to 30 kb of extensions around the Xdrop target site, suggesting a limit on the Xdrop assay read lengths (Madsen et al., 2020). This could be further exacerbated by the extensive splitting of potentially chimeric reads (Fig. 1.3.2 B). Indeed, at the start of the project, we did not anticipate the dramatic difference in size between the reference assembly and the actual locus. Therefore, it is possible our dataset suffered the same problem as the GRCz11 assembly, where the read lengths were insufficient long to properly resolve the mega-repetitive *mir430* locus. Given these limitations, I expect that the ~550 kbp *mir430* loci assembled in the above-mentioned studies are closer representations of the actual locus.

3.1.2 Emergence of the mega-repetitive mir430 locus on chr4q

The mir430 locus resides on chr4q of the zebrafish genome, a region which is highly repetitive, and contains many duplicated genes (Howe et al., 2013). The finding that the mir430 locus is ~550 kbp in size with at least 300 mir430 repeat units attests to the duplication rates found within this chromosome arm. The path to the emergence of this mega-repetitive locus may, therefore, be a consequence of its genomic positioning. In support of this, it has been reported that 77.5% of the genes on chr4q belong to just 31 ancestral gene families (Howe et al., 2013). Interestingly, high duplication rates on chr4q may be zebrafish specific. In medaka, a distant teleost relative of the zebrafish, only 16 *mir430* genes have been identified on its chr4 (Tani et al., 2010). This number may be an underestimation given that the medaka genome was also assembled using short reads. Nevertheless, it is still lower compared to the number of mir430 genes in the short read generated GRCz11 assembly.

The zebrafish chr4q remains a particularly mysterious part of the genome. This region of the genome has been collectively associated with high recombination rates and, high numbers of repeat elements and transposons (Chang et al., 2022; Howe et al., 2013). Despite these seemingly deleterious features, chr4q clearly encodes important functions for it to be selectively retained within the zebrafish genome. This is exemplified by the *mir430* locus and its vital role during development. Further research will be required to fully understand the encoded contents of chr4q and their importance in zebrafish physiology.

3.1.3 Copy number of mir430 repeats and the maternal load define the timing of mir430 activation

In trying to understand the genetic features that underlie the early activation of mir430, I discovered that both the maternal background and the number of clustered *mir430* repeats can define the timing at which the mir430 locus is transcriptionally activated.

1) Maternal background defines the timing of *mir430* activation in WT strains

   By taking a comparative approach between WT zebrafish strains, I found that different strains have different timings of *mir430* activation (Fig. 2.3.1). These differences, while correlated with *mir430* repeat numbers, are in fact a result of differences in the maternal background between strains. By accounting for this maternal background in the *mir430*-/- X strain crosses, I showed that the timing of *mir430* activation equalises amongst WT strains (Fig. 2.3.5). This suggests that important ZGA regulators, likely loaded as maternal mRNAs, may be variable in amount or composition amongst WT strains. Indeed a recent study showed, using an elegant experimental design, that maternal background defines the timing of ZGA (Gert et al., 2021). Here, the authors discovered the gene, *Bouncer*, to be required for fertilisation in zebrafish. By expressing *Bouncer* in a closely related species, *Medaka*, they were able to produce *Medaka* eggs fertilised with zebrafish sperm (Gert et al., 2021; Herberg et al., 2018). In *Medaka*, zygotic transcription only begins around 6 hpf (Aizawa et al., 2003; Tani et al., 2010). In these *Medaka*-zebrafish hybrid genomes, genes that would normally turn on earlier in the zebrafish genome followed *Medaka* ZGA times instead (Gert et al., 2021). Thus, maternal background plays a very important role in ZGA. In the future, studying differences in maternal transcriptome between strains, or even between *Medaka* and zebrafish, could be informative of what molecular determinants inherited from the mother determines ZGA timing.

   In the past, multiple studies have reported evidence of divergence between commonly used WT zebrafish strains (Deng et al., 2022; Guryev et al., 2006; Holden and Brown, 2018b). Notably, a comparison of an AB strain genome against the GRCz11 TÜ genome showed extensive single nucleotide polymorphisms (SNPs) and structural variants (SVs) occurring throughout the zebrafish genome (Deng et al., 2022). However, most of these studies have sought only to characterise divergence between strains either at the genomic or at the behavioural level (Deng et al., 2022; Guryev et al., 2006; Holden and Brown, 2018b; Lange et al., 2013). Within the WT strains, I have found that *mir430* activates at different times. Since ZGA begins with *mir430* transcription, it is also possible that the cascade of other zygotic genes that are activated following *mir430* may also be impacted by inter-strain differences in *mir430* activation. These could have direct consequences on development of the embryo. Future work will need to be done to determine the extent of which inter-strain differences of the maternal background could impact development. These studies are also important in highlighting that inter-strain genetic differences can result in functional consequences. Better awareness, therefore, should be taken in usage and reporting of WT zebrafish strains as it could have significant consequences on data reproducibility and how results are interpreted.

2) A transgenic *3x mir430* locus activates later during development

To study the impact of *mir430* repeat copy number on its activation, I have shown that a low copy number mir430 locus containing only 3 mir430 repeats inserted into a landing site on chr11 of the genome activates later during development. This was shown using live labelling of miR430 transcript foci that form on the *3x mir430* Tg labelled using the ANCHOR approach. I found that the *3x mir430* Tg begins transcribing at the 256-cells stage and becomes more active in the following cell cycle stages. This was corroborated by RT-qPCR results which also showed that the *3x mir430* locus activates later than the endogenous *mir430* locus. This proves that lower *mir430* repeat numbers result in later activation during ZGA.

The findings here are derived from an insertion into a single site within the zebrafish genome. A possible critique could therefore be that the differences in timing of transcriptional activation observed between the *3x mir430* Tg and endogenous *mir430* is a result of the differences in local chromatin environment. In support of the findings observed here, similar insertions in 2 alternative sites of the genome have shown that a transgenic *mir430* promoter begins transcription around the 256-cells/512-cells stage (Hadzhiev et al., 2023). However, rather than complete *mir430* repeats, only a single *mir430* promoter upstream of a reporter gene was inserted (Hadzhiev et al., 2023). Despite these differences, the concordance in observed timing of activation of the transgenic *mir430*s prove that the later activation of *the 3x mir430* Tg is not a positional effect. Furthermore, a study on chromatin architecture during zebrafish embryo development has shown that during early ZGA, the genome is largely unstructured and uniform (Wike et al., 2021). Distinct chromatin architectures, such as heterochromatin, only emerge later during development around 4 hpf (Laue et al., 2019). Thus, chromatin architecture is unlikely to be the reason for the later activation of the *3x mir340* Tg.

Despite similar findings between the work presented here and that published in Hadzhiev *et al.* (2023), we arrive at different conclusions. In Hadzhiev *et al.* (2023), the authors posit that ZGA in zebrafish is broadly categorised into the earlier "minor wave" and the later "major wave". They further propose that "minor wave" genes have a characteristic promoter which has sharp TSSs and a TATA-box, whereas "major wave" genes have broad TSSs and lack a TATA-box (Haberle et al., 2014; Hadzhiev et al., 2023). To test this, the authors created a Tg line with a single insertion of the *mir430* promoter upstream of a reporter gene and found that it indeed shows transcriptional activation prior to bulk ZGA at the 1k-cells stage. However, the authors do not address the differences observed between the endogenous *mir430* activation and the transgenics. I propose that higher copy numbers of these distinct "minor wave" promoters may distinguish *mir430* from other "minor wave" genes for earlier activation (Fig. 2.4.1). This could therefore, represent an additional layer of regulation by which the zygote uses to finetune gene activation even amongst "minor wave" genes. This is supported by the fact that the low copy number *3x mir430* locus activates around the same time as other "minor wave" genes such as *dusp6*, *grhl3* and *mxtx2*.

3.1.4 Higher *mir430* repeat numbers enhance TF recruitment

How does the higher number of *mir430* repeats result in earlier activation during development? To understand the biological significance of higher repeat numbers, I studied the encoded information within each mir430 repeat unit. This showed that TF binding sites for known regulators of *mir430*, Nanog, Pou5f3 and Sox19b (N, P and S), are localised throughout each repeat unit. These factors bind to the *mir430* locus and induce chromatin opening, likely via the recruitment and passage of RNA Pol II (Pálfy et al., 2020). The independent loss of N, P or S has been shown to result in a reduction in accessibility at the *mir430* locus, in addition to thousands of other genomic loci (Miao et al., 2022; Veil et al., 2019). Nanog, specifically, is the only one of these 3 factors which is indispensable for *mir430* activation (Lee et al., 2013). Combinatorial losses further exacerbate transcription at the *mir430* locus (Lee et al., 2013) Thus, the TF binding sites found within the *mir430* repeat units represent functional *cis*-regulatory elements. This raises the question of whether just the promoter or the whole *mir430* repeat unit has a role in regulating early activation. This is a relevant question because Nanog binding sites are only found within the promoter, while the rest of the repeat unit contains many Pou5f3 and Sox19b binding sites (Fig. 2.3.3). Given past studies on cooperativity between TFs (Adams and Workman, 1995; Miao et al., 2022;

Michael et al., 2020; Mirny, 2010; Veil et al., 2019), I speculate the whole repeat unit, and the NPS sites therein, function cooperatively to drive *mir430* activation.

What then, is the molecular function of having multiple of these repeats? I had hypothesised that the high number of TF binding sites that accompany these repeats result in higher sensitivity of the locus for TF binding. In other words, when many TF binding sites are available, a lower threshold amount of nuclear TFs need to be present for TF binding to occur, or in the context of the histone-TF competition model, to out-compete histones for DNA binding. In line with this, I reasoned that a higher number of available TFs could also compensate for lower sensitivity to TFs at genes with a lower number of TF binding sites. Indeed, this was the case when I overexpressed Nanog in the *3x mir430* Tg line. I observed that following Nanog overexpression, the *3x mir430* Tg was active in 56% of nuclei at the 256-cells stage - higher than the 7.7% without Nanog overexpression (Fig. 2.4.5). Thus, supporting the TF sensitivity idea. Expanding from this work, in the future, drawing correlations between the number of TF binding sites, and the timing of gene activation could be informative of how generalisable this principle of gene regulation during early ZGA could be.

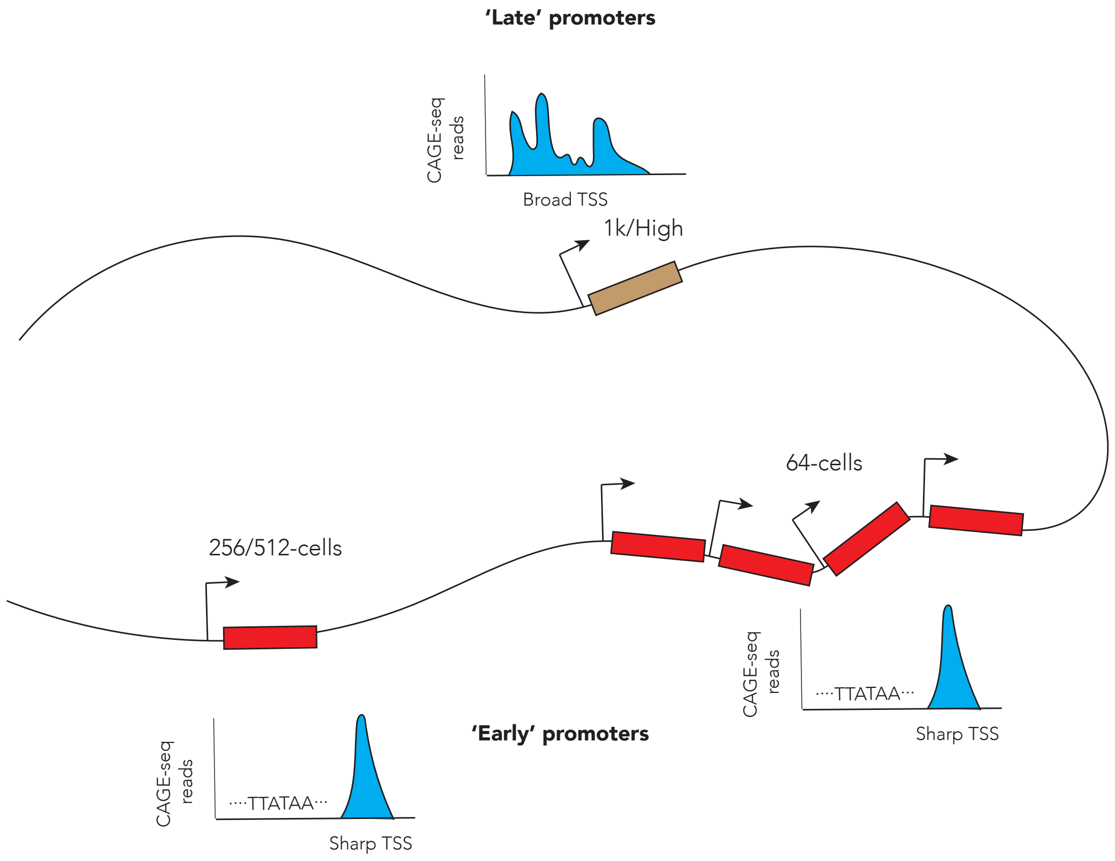

Figure 3.1.1: Model of promoter copy number effects on timing of activation

In this model, 'early' promoters have TATA-boxes and sharp TSSs as previously proposed. However, different copy numbers of these promoter types may distinguish their activation times. In addition, there are also the 'late' type promoters which do not have TATA-boxes and have broad TSSs.

The importance of TF binding sites has been explored in many contexts which continue to evolve with the introduction of novel concepts in gene regulation over the years. Perhaps the context which has been subjected to the most scrutiny is the role of TF binding sites within enhancers. The immense number of studies on enhancers have reinforced that the number, affinity, order, density, and variety of TF binding sites in enhancers are paramount to their ability to orchestrate transcriptional programs. Changes in one or more of these features could result in loss or mis-regulation in enhancer activity. Given the features observed at *mir430* so far, I speculate that *mir430* could behave similarly to an enhancer, or rather super-enhancer, given its exceptional size. In the future, it would be interesting to see if this putative enhancer could function distally to activate other zygotic genes too.

In recent years, many studies have found that TFs form discrete clusters/condensates within the nucleus. These clusters are DNA bound and have been shown to regulate transcription (Gaskill et al., 2023b; Kuznetsova et al., 2023; Mir et al., 2018, 2017; Sabari et al., 2018; Shrinivas et al., 2019a). In zebrafish, we have shown previously that Nanog and Sox19b clusters on the *mir430* locus contribute to its activation. In Drosophila, nuclear clusters of Zelda and GAF have been shown to independently activation or repress zygotic genes during development (Foo et al., 2014; Gaskill et al., 2023b; Mir et al., 2018). The formation of these subnuclear clusters of TFs may be facilitated by intrinsically disordered domains (IDRs) on the TFs (Kawasaki & Fukaya, 2023; Kuznetsova et al., 2023; Meeussen et al., 2023). However, there is no current unifying stand on a general role of IDRs in organising TF clusters as studies on different TFs have yielded contradictory findings (E. and B., 2023; Gaskill et al., 2023a; Kawasaki and Fukaya, 2023; Meeussen et al., 2023). In contrast, the role of the interaction between the DNA binding domain (DBD) of TFs and their cognate binding sites in seeding TF clusters is usually necessary. These interactions have been shown to be prevalent at repetitive regions where each repeat contains the cognate TF binding sites. A recent study in Drosophila reported that the GAGA factor (GAF) forms subnuclear clusters in early embryonic nuclei and, that these GAF clusters are seeded by AAGAG satellite repeats occurring. Formation of subnuclear GAF clusters were found to be required for the silencing of AAGAG repeats, likely to protect genome integrity (Gaskill et al., 2023a). Thus, repeat regions containing high numbers of TF binding sites may be particularly adept at seeding the formation of nuclear TF clusters to repress or activate genes in concert. In the case of *mir430*, the ~550 kbp of repeated TF binding sites could seed the formation of large TF clusters of NPS. Indeed, unpublished work from our lab has found that the *mir430*-associated subnuclear clusters of Nanog are typically the largest and brightest when imaged by fluorescence microscopy. Together, both published work and the findings presented here support a role of a high number of TF binding sites in regulating transcription, potentially by serving as a platform for recruiting a high local concentration of TFs, or promoting multivalent interactions between TFs by bringing them into close proximity. Interestingly, the absolute number of TF motifs alone is not definitive of transcriptional activity either. Studies have also shown that factors such as the density and diversity of TF motifs could also be important factors in determining transcriptional outcomes (Shrinivas et al., 2019b; Singh et al., 2021). Further work will be needed to draw a decisive link between the organisation and content of TF binding sites at repeat regions, TF clusters and transcriptional outcomes.

3.1.5 The competition model at single nucleus resolution

Previous work from our lab has shown that the competition between histones and TFs for DNA binding regulates the timing of bulk ZGA onset (Joseph et al., 2017). To show this, the authors injected a cocktail of core histones (H2A, H2B, H3 and H4) into 1-cell stage embryos and saw this this resulted in delayed ZGA. Reciprocally, overexpression of Pou5f3 and Sox19b was able to advance ZGA. In the context of this histone-TF competition, the higher sensitivity of the *mir430* locus for TF binding may facilitate localised out-competition of histones against TFs, even when histone:TF ratios are not permissive for activation on other zygotic genes. Whether this histone-TF dynamic affects *mir430* activation, however, remains unclear. The advancement of *3x mir430* activation upon Nanog overexpression suggests that competition may regulate *mir430* activation too. In the future, measuring mir430 activity (endogenous and 3x mir430 Tg) following the injections of the histone cocktail would shed more light on this.

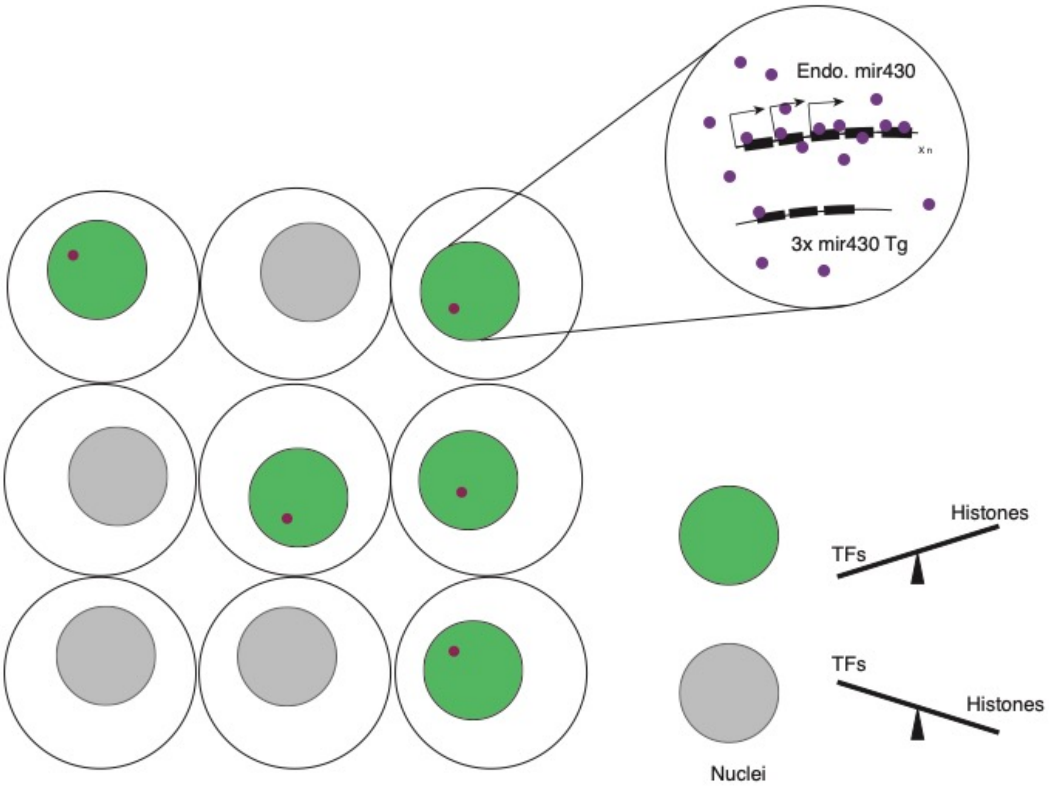

Figure 3.1.2: Competition model in the 3xmir430 line at single nucleus resolution

The variability between active nuclei at different stages may be resultant from extrinsic noise in histone-TF levels. Different genomic loci may respond in different ways to these histone TF levels. Those with higher numbers of repeats (black rectangles), and therefore higher number of TF binding sites may be better able to recruit TFs (violet circles) compared to loci with less repeats/TF binding sites.

Interestingly, mir430 activation at the 64-cells stage seems to be stochastic – it begins only in some nuclei, and often only on 1 allele. Stochastic activation of zygotic genes has been previously observed in zebrafish embryos (Stapel et al., 2017). If indeed competition between histones and TFs regulate mir430 activation, cell-to-cell variability in transcriptional outcomes would suggest that histones or TF levels are variable between nuclei (Fig. 3.1.2). Given that TFs such as Nanog have been shown to regulate *mir430* (Lee et al., 2013), it is probable variable TF levels would result in variability in *mir430* transcription between cells. Cell-to-cell variability in TF levels is a form of "extrinsic noise" which has been shown to result in cell-to-cell variation in transcription outcomes (Bar-Even et al., 2006; Blake et al., 2003; das Neves et al., 2010). Importantly, this variability would have a more evident effect under conditions of limiting resources. Over time, as TFs accumulate, nuclei in more cells reach a threshold number of TFs for *mir430* to activate. This model would be consistent with the increases in fraction of active nuclei from the 128-cells stage onwards.

3.1.6 Future directions

The work presented in this thesis provides evidence that the high numbers of TF binding sites at the mega-repetitive *mir430* locus result in higher sensitivity to TFs, allowing earlier activation of *mir430* during zebrafish development. In the future, more work will need to be done to understand to what extent the number and types of TF binding sites regulate the timing of activation of other zygotic genes, and how these factors influence the dynamic between histones and TFs in a gene-specific manner. In addition to pinning down the mechanisms of *mir430* activation, this study revealed 2 interesting findings that warrant future pursuit.

3.1.7 Regulatory rules of spatial co-transcription

Imaging of the *3x mir430* line showed that active *3x mir430* Tg tended to colocalise with the endogenous miR430 transcription bodies. Out of 58 active *3x mir430* ParB1-mNG spots, 15 (25.8%) had instances of colocalization with the endogenous miR430 transcription bodies. Given that the *3x mir430* Tg and the endogenous *mir430* locus are located on different chromosomes, this percentage of colocalization suggests a non-random occurrence. Past studies of spatial co-transcription of genes have shown gene colocalization percentages ranging from 42% to 60% (Osborne et al., 2004).

The spatial co-transcription of genes has been previously reported in different contexts. Perhaps the most well-studied one is that of the globin genes in erythroid cells, whereby globin genes were found to spatially co-transcribe within shared RNA Pol II clusters (Osborne et al., 2004). It remains unclear what defines the rules of spatial co-transcription. Previous studies have shown that similar promoters on plasmids transfected into cell lines tended to spatially co-transcribe (Xu and Cook, 2008). However, spatial co-transcription was negatively impacted by the addition of an intron, suggesting that regulation might not occur at the promoter level. Co-binding of distal elements by similar TFs may bring distal genomic elements, such as promoters or enhancers, together. Indeed, in embryonic stem cells (ESCs), pluripotency TFs such as Nanog and Oct4/Pou5f3 bind to regions of high cognate motif density and drive the formation of interchromosomal contacts (De Wit et al., 2013). However, the contrary has been reported too, whereby co-binding by similar TFs were poor predictors of long-range contacts (Friman et al., 2023). Thus, it remains largely unclear what drives spatial co-transcription of genes present on distinct chromosomes. One possibility is that a threshold concentration of TF binding sites is required for sufficient TFs to bind and bring distal promoters together. The utility of the ANCHOR DNA labelling system is exemplified here as it can be used in the *3x mir430* Tg line to address further questions on TF recruitment and spatial co-transcription of the *3x mir430* Tg with the endogenous miR430 transcription bodies. In the future, this technique could also be expanded further to test various promoters and their co-localisation potential.

3.1.8 Identifying interactors of long miR430 transcripts

The endogenous mir430 locus was discovered to generate very long ssRNA transcripts, potentially 17 kbp in size. miR430 transcripts have been shown in the past to bind to an RNA binding protein (RBP), hnrnpa1 (Despic et al., 2017). However, this represents only a limited set of known *miR430* binding RBPs. Past studies have shown that RNAs can act as architectural elements that maintain the structure of nuclear compartments such as nuclear paraspeckles (Clemson et al., 2009). An idea could be that the mega-miR430 transcript acts as a structural scaffold, partly to maintain the miR430 transcription bodies. To study this, I propose to pull out RBPs that may interact with these mega-miR430 transcripts. These may provide an initial clue as to the RBPs that interact with miR430 transcripts and what their functions may be in the context of the miR430 transcription bodies.

4 Conclusions

To conclude this thesis, I have shown that ZGA, a well-timed biological process, is regulated by differential biomolecular interactions between TFs and the DNA template.

This is important from a developmental standpoint, as it may be a generalisable principle by which temporal regulation of gene activation is achieved. From a broader perspective, these findings represent a step towards drawing a link between DNA encoded elements, high local clustering of transcriptional activators and machinery, and developmentally relevant outcomes.

5 Materials and Methods

5.1 Chapter 1: Pinning down the structure of the mir430 locus

5.1.1 Zebrafish strains and

Zebrafish were maintained and raised under standard conditions. WT AB adults were acquired from the fish facility at the Ecole Féderale Polytechnique Lausanne (EPFL). A whole adult female fish was dissected of internal organs, including gonads, and snap-frozen in liquid nitrogen.

5.1.2 High molecular weight genomic DNA extraction

Snap-frozen adult body was ground in liquid nitrogen into a fine powder. Ground tissue was lysed overnight in lysis buffer (400 mM NaCl, 20 mM Tris pH8.0, 30 mM EDTA pH 8.0, 0.5% SDS, 100 µg/mL proteinase K) at 55°C. Following overnight lysis, debris was removed by spinning at 4000 x g at room temperature for 30 mins. The resultant supernatant was transferred to a fresh tube and treated with RNase A (50 µg/mL final conc.) at 37°C for 1 hr. DNA was extracted using Phenol-chloroform, followed by 2 washes with chloroform to remove residual phenols. gDNA was precipitated by ethanol precipitation and the resultant gDNA fibres were using a glass hook and further washed in 70% ethanol. Spooled gDNA was airdried at room temperature and then eluted in Tris-EDTA buffer.

5.1.3 Xdrop targeted enrichment of the mir430 locus

Xdrop target enrichment was performed at Samplix, Denmark. The following protocol was provided as a report by Samplix following target enrichment:

5.1.3.1 Xdrop dPCR droplet generation

Droplet generation was performed using XdropTM instrument and reagents. In short, each of the DNA samples was compartmentalized into droplets with dPCR master mix and relevant dPCR primer sets. After droplet production the DNA in droplets were subjected to PCR amplification.

5.1.3.2 Sorting of Xdrop droplets

After the dPCR protocol, the droplets were collected and dyed in 1 ml 1x dPCR buffer and 10 µl droplet dye and incubated at room temperature for 5 min, protected from light. The positive droplet populations were sorted from the negative using a SONY benchtop SH800S cell sorter with a 100 µm nozzle (Sony Biotechnology). Droplets were gated on Forward

Scatter (FSC) and Side Scatter (SSC) height to separate them from debris. Then the identified droplets were gated in a new plot to identify the negative and positive green fluorescent populations of droplets. This was done by subjecting the droplets to excitation using a 488 nm laser and detect the emission in a green channel and plotting green fluorescence versus side scatter (SSC). The positive green fluorescent droplets were sorted from the negative droplets and collected into 15 µl of molecular grade H2O at the bottom of a 1.5 ml DNA LoBind collection tube. Positive droplets were broken and used for dMDA.

5.1.3.3 dMDA reaction

Isolated enriched DNA was compartmentalized into droplets with dMDA master mix and enzyme. The dMDA reactions were loaded into dMDA cartridges and collected single emulsion droplets were incubated at 30°C for 16 h followed by 65°C for 10 mins.

After the 16-hour incubation, dMDA DNA was isolated by breaking the droplets. The collected DNA was quantified by Quantus and size distribution evaluated by TapestationTM System (Agilent Technologies Inc.), using Genomic DNA ScreenTape according to the manufacturer's instructions. If dMDA DNA yields are below a microgram, an additional 20 µL of 2x dMDA mastermix is added to the harvested dMDA and incubated for additional 2 hours at 30 °C and quantified again by Quantus.

5.1.3.4 qPCR enrichment evaluation

Enrichment of Xdrop enriched DNA was evaluated by qPCR. dMDA reactions were diluted in molecular grade H2O (1:9 vol/vol) and subjected to qPCR reactions using validated qPCR assays at a site adjacent but not overlapping the Xdrop target sites. 10 ng unenriched DNA was used as a reference.

5.1.4 PacBio HiFi sequencing

PacBio HiFi library preparation was done at the DNA sequencing facility at the Max Planck Institute of Cell Biology and Genetics (MPI-CBG). DNA samples for each Xdrop target site was multiplexed

5.1.5 Analysis of Xdrop targeted long-read sequencing of the mir430 locus

5.1.5.1 Multiple sequence alignment of mir430 repeat units

Mir430 repeat units were defined as the sequence between the start of a mir430 promoter and the start of the downstream mir430 promoter. Promoter positions were defined based on the mir430 promoter described in (Hadzhiev et al., 2023). These definitions were used to extract mir430 repeat units and perform multiple sequence alignments using the R package, ggmsa (Zhou et al., 2022).

5.1.5.2 Mapping of PacBio HiFi reads to reference and contig assemblies

For Fig. 1.3.1, Pacbio HiFi reads for all 4 Xdrop target sites were pooled and mapped to the GRCz11 reference assembly using Minimap2 default settings for PacBio HiFi reads (-ax map-hifi) (Li, 2018).

For Fig. 1.3.4, SACRA output reads larger than 1kbp for all 4 Xdrop target sites were pooled and mapped to Hifiasm contigs using Minimap2 with default settings for Pacbio HiFi reads (-ax map-hifi). Data was visualised using IGV (Thorvaldsdóttir et al., 2013).

5.1.5.3 SACRA splitting of chimeric reads

Pacbio HiFi reads for all 4 Xdrop target sites were pooled and split using SACRA (Kiguchi et al., 2021b) default settings and reads larger than 1 kbp were kept for downstream analysis. Visualisation of mir430 genes in top 1% of pre- and post-SACRA splitting was done using the R package, ChromoMap (Anand and Rodriguez Lopez, 2022).

5.1.5.4 Hifiasm de novo contig assembly of SACRA split reads

Split reads larger than 1 kbp from all 4 Xdrop target sites were used for *de novo* contig assembly using Hifiasm (Cheng et al., 2021). Settings used for Hifiasm were -D 100 -n 10000 –max-kocc 1000 –hg-size 188m. In particular, genome size was defined as 188m rather than the actual genome size as recommended by Samplix in order to account for highly skewed coverage at the mir430 locus compared to the rest of the genome. Contigs were visualized on IGV.

5.1.5.5 Identification of mir430 promoters and genes

Mir430 promoters, a, b and c gene isoforms were identified using blastn against mir430 promoter, a, b and c sequences from GRCz11.

5.2 Transcriptional outputs from a mega-repetitive mir430 locus

5.2.1 Northern blot DIG-labelled probe generation

To generate the northern blot probes targeting *miR430* transcripts, I designed primers that amplified a 700 bp sequence within the *mir430* repeat unit. This product was subsequently subcloned into a pCRII backbone downstream of a T7 promoter using restriction cloning. From there, DIG-labelled RNA probes complementary to *miR430* transcripts were synthesised using the mMessage T7 *in vitro* transcription kit (Thermo Fisher cat. No.:

AM1344) with addition of a DIG labelling mix (Roche cat. No. 11277073910). Synthesised probes were quality-checked on a 1% TBE agarose gel.

5.2.2 Total RNA extractions and RNA preparations

Total RNA was extracted from 30-50 embryos at the desired stages using the MinElute RNeasy kit and quality of extracted RNA was checked on a 1% TBE gel. In total 8 μg of total RNA was used for each northern blot condition. To ensure that the same volume was loaded into each well, I used a speedvac to concentrate total RNA samples and made-up volumes to 10 μL for all wells.

5.2.3 ssRNA ladder to DIG-labelled dsDNA size comparisons

To determine *miR430* transcript sizes on the northern blot using the DIG-labelled dsDNA ladder VII (Roche cat. No. 11669940910), I made size comparisons of ssRNA to DIG-labelled dsDNA on a separate 1% ethidum bromide TBE agarose gel. By drawing a line profile across the adjacent ladders and plotting the intensity values, I was able to acquire ssRNA size equivalents for DIG-labelled dsDNA bands. This gave DIG-labelled dsDNA to ssRNA size conversion rate of 1.5-2x (1 kbp DIG dsDNA ~ 1.5-2 kbp ssRNA).

For DIG-labelled dsDNA ladder on a northern blot, line profiles were drawn to determine sizes of *miR430* transcripts in DIG dsDNA sizes. Approximate conversions was then done based on known size conversions rates.

5.2.4 Northern blot

8 μg of total RNA was used for each northern blot condition. 8 μg of total RNA was made-up to 10 μL and incubated at 85°C for 2 mins to denature RNA secondary structures. Tubes containing the RNAs were subsequently spun briefly and placed on ice prior to running on a 1% ethidium bromide TBE agarose gel approximately 0.6 cm in thickness. Gels were ran at 40 V/100 Amp for 2 hrs and presence of 28S and 18S rRNAs were used for confirmation of RNA quality.

RNA on gels were then transferred onto Nylon membranes using overnight capillary transfer in 1X TBE buffer. The following day, RNA was crosslinked on the mebrane using a Strata UV crosslinker.

Probe hybridisation was done in DIG Easy hyb buffer using 10 mL per 100 cm^2 of membrane. DIG Easy hyb buffer was prewarmed in a long glass bottle in an oven rotator at 68°C. To block against non-specific nucleic acid binding, Salmon sperm DNA was added during the prehybridization step. For this, Salmon sperm DNA was denatured at 85°C for 2 mins and immediately placed on ice. After buffer was prewarmed, Salmon sperm DNA was added to a final concentration of 50 μg/mL. Crosslinked membrane was placed into prewarmed DIG Easy hyb buffer with salmon sperm DNA anmd allowed to hybridise in a 68°C oven with rotation for 30 mins. Alongside, 7 mL of DIG Easy hyb buffer was prewarmed for probe hybridisation. In this prewarmed buffer for probe hybridisation DIG-labelled probe was added to a final concentration of 100ng/mL alongside heat denatured Salmon sperm DNA at a final concentration of 50 μg/mL. after the 30 min pre-hybridisation time was over, the prehybridization mix was discarded and replaced with the hybridisation mix contained the DIG-labelled probe. Probes were allowed to hybridise on the membrane in a 68°C oven overnight. The following day, the membrane was washed twice for 5 mins at 68°C with 15 mL of Low stringency buffer (2X SSC containing 0.1% SDS) pre-warmed to 68°C. After which, the membrane was washed twice for 15 mins at 68°C with 15 mL of pre-warmed High stringency buffer (0.1X SSC containing 0.1% SDS).

For probe detection, washed membranes were shaken in wash buffer (Maleic acid buffer with 0.3% Tween 20) for 2mins. After which, wash buffer was discarded and blocking buffer (1% milk with Maleic acid buffer) was added and membrane was shaken for 30 mins at rom temperature. After blocking, blocking buffer was discarded and 20 mL of antibody solution (Anti-DIG-AP antibody diluted 1:10000 in blocking buffer) was added to the membrane and shaken for 30 mins. Subsequently, membrane was washed twice for 15 mins with wash buffer and then equilibrated in 20 mL of detection buffer for 3 mins with shaking. Next, the membrane was placed faced-down onto 500 μL of CDP-Star (Sigma-Aldrich cat. No. C0712) on saran-wrap to allow even spread. Membrane was subsequently wrapped up with saran-wrap and developed on film, for 5 mins.

5.3 Inter-strain variations of the mir430 locus and insights into its activation

5.3.1 WT zebrafish strains

WT zebrafish strains for AB, TÜ, TL and NHGRI-1 were obtained from the fish facility at the Ecole Féderale Polytechnique Lausanne (EPFL). All strains were maintained and raised under standard conditions. Embryos from strain in-crosses were dechorionated with Pronase immediately upon fertilisation, injected, and allowed to develop to the desired stage at 28°C.

5.3.2 Embryo injections for imaging

1-cell stage embryos were injected at room temperature with an injection mix containing H3K27Ac Fabs and miR430 MOVIE (as described in Hadzhiev et al., 2019). The injection mix was injected into the cell.

5.3.3 Mounting of embryos for live microscopy

At the 8-cells stage, embryos were mounted in 0.7% UltraPure low melting point agarose (ThermoFisher 16520050) dissolved in Danieau's media supplemented with iodixanol (OptiPrep, STEMCELL Technologies 07820) to match a refractive index of 1.3615. Embryos were mounted in Ibidi glass-bottom dishes (μ-Dish 35 mm, high Glass Bottom, 81158) and the agarose was allowed to set with the dish upside-down to ensure minimal distance between the embryos and the coverslip. Embryos were imaged when the agarose solidified.

5.3.4 Spinning disk confocal imaging of mir430 activation

Whole-mount embryos were imaged on a Nikon Ti2 microscope with a Yokogawa CSU-W1 spinning disk unit and 2x Photometrics Prime 95B sCMOS Grayscale cameras. A 60x/1.27 Plan Apochromat VC water objective was used. Dual-color imaging was done to simultaneously capture fluorescence from miR430 MOVIE (591 nm) and A488 (488 nm) fluorophores. Nuclei were acquired either in Z-stacks of 60 optical slices of 0.5 μm thickness, over the course of 1.5 hrs, with a time-interval of 1.5 mins (Fig. 3.2) or with a time-interval of 35 secs (Fig. 3.4.1).

5.3.5 Scoring for active nuclei

Active nuclei were defined as nuclei with either 1 or 2 detectable miR430 MOVIE foci that show up during the cell cycle, which was defined as the time frames between the formation of the previous metaphase plate and the formation of the next metaphase plate. Inactive nuclei were nuclei with no miR430 MOVIE foci showing up during the cell cycle. miR430 MOVIE foci were only considered if they had signal intensities at least 10% higher than background.

5.3.6 Tracking of miR430 MOVIE foci over the cell cycle and calculation of growth rate

Nuclei over the cell cycle were manually segmented in Fiji. Subsequently, miR430 MOVIE foci were tracked over time using the software, Icy, with the Spot Detector and Spot Tracking functions . Tracks were exported and plotted over time, with T0 being the first time point of foci detection, regardless of absolute time following the start of the cell cycle stage. For growth rate calculations, the signal growth over the first 5 time points (~3 mins) were used to calculate *miR430* transcript production rates for the different strains.

5.3.7 Mir430 promoter and genes qPCR quantification

Total gDNA was extracted from 40 24 hpf embryos by proteinase K lysis followed by phenol:chloroform cleanup and ethanol precipitation. Extracted gDNA concentrations were measured using the Qubit dsDNA broad range kit (Thermo Fisher cat. No. Q32850). Alongside this, concentrations of serial 1:10 dilutions of gDNA was measured to ensure concentration accuracy.

For relative quantification of *mir430* promoters, and mir430a, b and c genes, qPCR on 0.5 ng of gDNA from each strain was done. A single copy gene, Sox19a, was used as a single copy gene reference.

For the qPCR, 4 uL of gDNA (making up 0.5 ng) was added to 5 µL of Power SYBR Green PCR Master mix (Thermo Fisher cat. No. 4368577), 0.4 µL of forward and reverse primers (final concentration of 0.2 µM). The reaction was made up to 10 µL with nuclease-free water. Reactions were cycled at 95°C for 10 mins, and then 40 cycles of 95°C for 30 sec and then 60°C for 1 min. Melting curve analysis was also done. Initial tests were done on the primer pairs and they showed single melting curve peaks and 90-110% calculated efficiencies.

For relative quantification of *mir430* promoters and *mir430a, b* and *c* genes, dCt was calculated using Sox19a as reference gene and ddCt was calculated with respect to AB for each biological replicate.

5.3.8 Nanog, Pou5f3 and Sox19b motif and ChIP-seq analysis

Motif analysis was done one the GRCz11 chr4, position weight matrices (PWM) for Nanog, POU5F1 (JASPAR MA1115.1) and Sox2 (JASPAR MA0143.1). Nanog motifs were not available on JASPAR. For this, I acquired sequences of called ChIP-seq peaks from previously published Nanog ChIP-seq data (Xu et al., 2012) and used the software MEME from MEME suite to generate a Nanog motif PWM. Motif detection on chr4 was done using FIMO from MEME suite.

ChIP-seq reads for Nanog (Xu et al., 2012), Pou5f3 and Sox19b (Leichsenring et al., 2013b) was done using bowtie2 using the –very-fast setting.

5.4 Mir430 repeat copy numbers defines timing of activation

5.4.1 Generation of the 1x, 3x and 5x insertion constructs

A single *mir430* repeat unit was ordered as a geneblock based on the sequence of the *mir430* consensus sequence. Restriction enzyme cut sites for SalI and XhoI-NotI were added to the arms of the repeat unit geneblock and used for sequential subcloning of *mir430* repeats into the insertion construct in the same orientation. Sequential cloning of repeats was possible because SalI and XhoI give the same sticky ends when digested, but upon re-ligating with a non-self sticky end, would result in the destruction of the original cut-site. Thus, SalI restriction sites would be lost after each round of re-ligating to a XhoI sticky end. This then allows subsequent reusing of SalI and XhoI for repeat subcloning. Correct sizes for 1x, 3x and 5x constructs were confirmed on a gel following cloning.

5.4.2 Generating 1x, 3x and 5x *mir430* lines

25 pg of the 1x, 3x and 5x *mir430* insertion constructs were injected together with 25 pg of mRNA encoding PhiC31 recbombinase into embryos from the cmlc2:egfp_attP2B line (Mosimann et al., 2013). Embryos were allowed to grow to 3 days post fertilisation (dpf) before they were screened for the α-crystallin:Venus Tg marker. Only larvae with both α-crystallin:Venus and cmlc2:egfp markers were put to grow into adults and screened for founders.

5.4.3 ANCHOR DNA-labelling components

The plasmid constructs for ANCHOR sequences were ordered from AddGene. The ParS sequence was derived from the plasmid pFG2 (Addgene #87250) and the ParB1 sequence from derived from the plasmid pCM189-ParB1::mCherry (Addgene #87253). The sequence of ParS was identified and used to design primers that amplified the ParS sequence and used for homology arm cloning. The ParB1 was restriction cloned into an empty pCS2+_mNeongreen backbone by PCR addition and usage of restriction sites for FseI and AscI on the ParB1 cDNA. This resulted in a pCS2+_ParB1-mNeongreen plasmid. Further cloning of SV40NLS-NES was done by Gibson cloning with an empty pCS2+ backbone containing an in-frame NLS-NES upstream of the homology site. This resulted in a pCS2+_NLSNES_ParB1-mNeongreen encoding plasmid.

5.4.4 Embryo injections for imaging

3xmir430 embryos were dechorionated immediate after they were laid and then injected with 100-120 pg of NLSNES-ParB-mNG encoding mRNA along with miR430 MOVIE as described inHadzhiev et al., 2019. Importantly, mRNAs and *miR430 MOVIE* were injected in separate needles to prevent precipitation of MOVIE. Following injections, embryos were grown to the desired stage and mount for imaging as described in 4.3.3.

For Nanog overexpression experiments, dechorionated *3x mir430* embryos were injected with 120 pg of Nanog-HA mRNA, 120 pg of ParB1-mNG mRNA and, in a separate needle, *miR430* MOVIE.

5.4.5 Imaging of *3x mir430* line injected embryos

Injected embryos were imaged on the same system as described in 4.3.4. Sequential imaging of fluorophores was done instead of simultaneous dual-colour imaging. This was done to ensure that *miR430* MOVIE signal that co-localised with ParB1-mNG signal was not due to bleed-through.

5.4.6 *3x mir430* image analysis

For determining the percentage of nuclei where the *3x mir430* Tg was active, I segmented nuclei manually and did a colocalization analysis in 3D using the Fiji plugin Comdet. Prior to using comdet, I first processed signals in green (ParB1-mNG). In the green channel, I subtracted the background using a rolling circle of radius=1. Both channels in green (ParB1-mNG) and red (miR430 MOVIE) were then passed through the "Smooth" function. Subsequently. The Z-stacks were used for COMDET colocalization analysis using the an expected size of 4 pixels and standard deviation threshold of 7 for ParB1-mNG, and an expected size of 8 pixels and standard deviation of 4 for *miR430* MOVIE. Nuclei with at least 1 instance of colocalised spots were categorised as 'active' while those with no instances of colocalization were categorised as 'inactive'.

For the rate analysis (Chapter 2.4.7), segmented nuclei were used for spot segmentation and tracking on Imaris. Manual thresholding was used to segment spots for tracking. The segmented Here, both endogenous *mir430* and *3x mir430* transcript foci could be segmented and distinguished via their colocalization with the ParB1-mNG foci. For calculating growth rates, the shorter lifetime of the *3x mir430* Tg transcript foci meant that only a shorter timeframe could be used for growth rate calculations. We used increase in sum intensity between the first 2 time points following transcript foci detection (1.5 mins time span) to calculate growth rates for the *3x mir430* and endogenous *mir430* transcript foci. The data was then plotted on a log scale.

For determining true activation times, a linear equation was calculated for each *3x mir430* and endogenous *mir430* transcript foci track. Extrapolation of true activation times were then then by taking the x value when y=0. This gave a distribution of estimated true activation times for each track.

5.4.7 RT-qPCR against non-specific *miR430, 3x miR430* and other early zygotic gene transcripts

Total RNA was collected for the stages 2-cells, 64-cells, 128-cells, 256-cells, 512-cells, 1k-cells, High and Sphere stages. The staging here was done using absolute time starting with 45 mins post fertilisation for 2-cells stage, 2 hpf for 64-cells stage, and every 15 min timepoint after with the exception of High and Sphere, which were collected at 3.3 hpf and 4 hpf respectively. Total RNA for these samples was extracted using Qiazol and isopropanol precipitation after. RNA was then treated with TURBO DNase (Thermo Fisher cat. No. AM2238) at 37°C for 30 mins and subsequently cleaned up with the RNeasy MinElute Cleanup kit (QIAGEN cat. No. 74004).

Resultant RNA was quality checked on a 1% TBE ethidium bromide agarose gel. 1 μg of each RNA sample was used for reverse transcription using the Superscript II RT kit (Invitrogen cat. No. 12574026) with random hexamers. The resultant cDNA was diluted with a factor of 1:40. 4 μL of diluted cDNA was used in a qPCR reaction together with 5 μL of Power SYBR Green PCR Master mix (Thermo Fisher cat. No. 4368577), 0.4 μL of forward and reverse primers (final concentration 0.2 μM) and water making the reaction up to 10 μL.

6 Supplementary

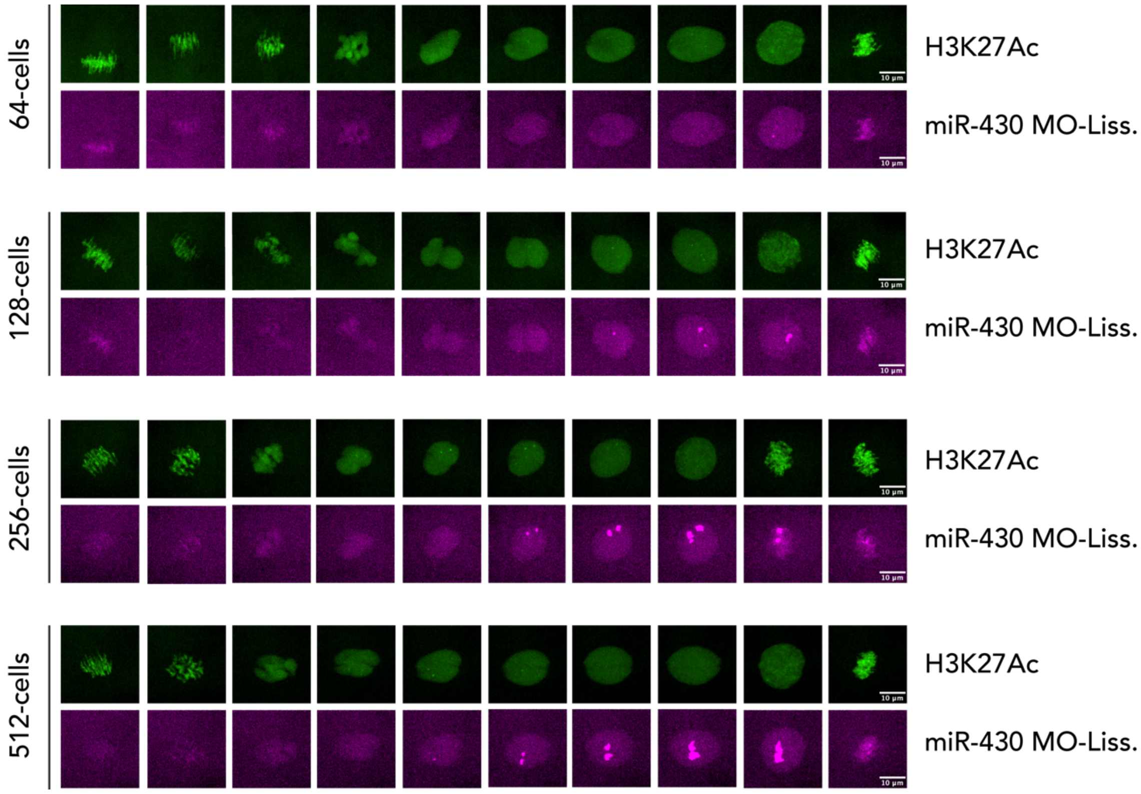

Figure 5.4.1: mir430 transcription body formation in an AB embryo throughout the cell cycle.

H3K27 Ac is shown in green and miR430 MOVIE is shown in magenta. miR430 transcription bodies are discrete and long-lasting. They typically form and only dissolve at the end of the cell cycle.

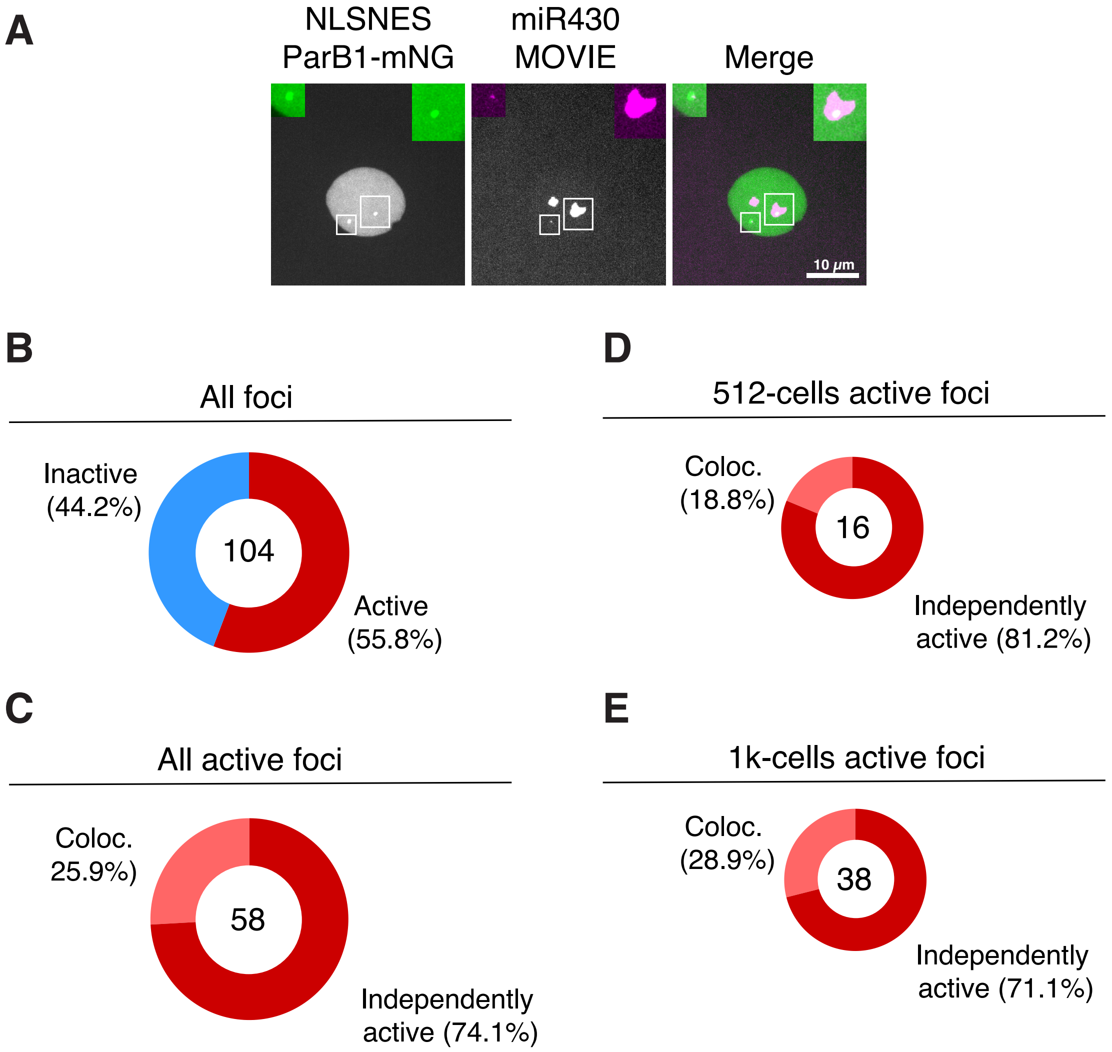

Figure 5.4.2: The 3x mir430 Tg transcribes either together with the endogenous mir430 or independently

A: Representative images of independently active and colocalising 3x mir430 Tg in a single nucleus. B: Pie chart showing distribution of active and inactive 3x miR430 NLSNES-ParB1-mNG foci (104). C-E: Pie chart showing distribution of active 3x mir430 NLSNES-ParB1-mNG foci that either have instances of colocalisation with the endogenous miR430 transcription body, or are independently active. D-E show the same but for the 512-cells and 1k-cells stage.

7 References

Aanes, H., Winata, C.L., Lin, C.H., Chen, J.P., Srinivasan, K.G., Lee, S.G.P., Lim, A.Y.M., Hajan, H.S., Collas, P., Bourque, G., Gong, Z., Korzh, V., Aleström, P., Mathavan, S., 2011. Zebrafish mRNA sequencing deciphers novelties in transcriptome dynamics during maternal to zygotic transition. Genome Res 21, 1328–1338. https://doi.org/10.1101/GR.116012.110

Adams, C.C., Workman, J.L., 1995. Binding of disparate transcriptional activators to nucleosomal DNA is inherently cooperative. Mol Cell Biol 15, 1405–1421. https://doi.org/10.1128/MCB.15.3.1405

Adamson, E.D., Woodland, H.R., 1974. Histone synthesis in early amphibian development: Histone and DNA syntheses are not co-ordinated. J Mol Biol 88, 263–285. https://doi.org/10.1016/0022-2836(74)90481-1

Aizawa, K., Shimada, A., Naruse, K., Mitani, H., Shima, A., 2003. The medaka midblastula transition as revealed by the expression of the paternal genome. Gene Expression Patterns 3, 43–47. https://doi.org/10.1016/S1567-133X(02)00075-3

Akkers, R.C., van Heeringen, S.J., Jacobi, U.G., Janssen-Megens, E.M., Françoijs, K.J., Stunnenberg, H.G., Veenstra, G.J.C., 2009. A hierarchy of H3K4me3 and H3K27me3 acquisition in spatial gene regulation in Xenopus embryos. Dev Cell 17, 425. https://doi.org/10.1016/J.DEVCEL.2009.08.005

Almouzni, G., Wolffe, A.P., 1995. Constraints on transcriptional activator function contribute to transcriptional quiescence during early Xenopus embryogenesis. EMBO J 14, 1752–1765. https://doi.org/10.1002/J.1460-2075.1995.TB07164.X

Anand, L., Rodriguez Lopez, C.M., 2022. ChromoMap: an R package for interactive visualization of multi-omics data and annotation of chromosomes. BMC Bioinformatics 23, 1–9. https://doi.org/10.1186/S12859-021-04556-Z/FIGURES/5

Anderson, K. V., Lengyel, J.A., 1980. Changing rates of histone mRNA synthesis and turnover in Drosophila embryos. Cell 21, 717–727. https://doi.org/10.1016/0092-8674(80)90435-3

Bar-Even, A., Paulsson, J., Maheshri, N., Carmi, M., O'Shea, E., Pilpel, Y., Barkai, N., 2006. Noise in protein expression scales with natural protein abundance. Nat Genet 38, 636–643. https://doi.org/10.1038/NG1807

Bhat, P., Cabrera-Quio, L.E., Herzog, V.A., Fasching, N., Pauli, A., Ameres, S.L., 2023. SLAMseq resolves the kinetics of maternal and zygotic gene expression during early zebrafish embryogenesis. Cell Rep 42, 112070. https://doi.org/10.1016/J.CELREP.2023.112070

Blake, W.J., Kærn, M., Cantor, C.R., Collins, J.J., 2003. Noise in eukaryotic gene expression. Nature 2003 422:6932 422, 633–637. https://doi.org/10.1038/nature01546

Blythe, S.A., Wieschaus, E.F., 2015. Zygotic Genome Activation Triggers the DNA Replication Checkpoint at the Midblastula Transition. Cell 160, 1169–1181. https://doi.org/10.1016/J.CELL.2015.01.050

Bogdanović, O., Fernandez-Miñán, A., Tena, J.J., De La Calle-Mustienes, E., Hidalgo, C., Van Kruysbergen, I., Van Heeringen, S.J., Veenstra, G.J.C., Gómez-Skarmeta, J.L., 2012. Dynamics of enhancer chromatin signatures mark the transition from pluripotency to cell specification during embryogenesis. Genome Res 22, 2043. https://doi.org/10.1101/GR.134833.111

BRAVO, R., KNOWLAND, J., 1979. Classes of Proteins Synthesized in Oocytes, Eggs, Embryos, and Differentiated Tissues of Xenopus Zaevis. Differentiation 13, 101–108. https://doi.org/10.1111/J.1432-0436.1979.TB01572.X

Brown, D.D., Littna, E., 1964. RNA synthesis during the development of Xenopus laevis, the South African clawed toad. J Mol Biol 8, 669–687. https://doi.org/10.1016/S0022-2836(64)80116-9

Brown, J.L., Wu, C., 1993. Repression of Drosophila pair-rule segmentation genes by ectopic expression of tramtrack. Development 117, 45–58. https://doi.org/10.1242/DEV.117.1.45

Bushati, N., Stark, A., Brennecke, J., Cohen, S.M., 2008. Temporal Reciprocity of miRNAs and Their Targets during the Maternal-to-Zygotic Transition in Drosophila. Current Biology 18, 501–506. https://doi.org/10.1016/J.CUB.2008.02.081

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., Moreno-Mateos, M.A., Giraldez, A.J., 2019. Brd4 and P300 confer transcriptional competency during zygotic genome activation. Dev Cell 49, 867. https://doi.org/10.1016/J.DEVCEL.2019.05.037

Chang, H., Yeo, J., Kim, J. gyun, Kim, H., Lim, J., Lee, M., Kim, H.H., Ohk, J., Jeon, H.Y., Lee, H., Jung, H., Kim, K.W., Kim, V.N., 2018. Terminal Uridylyltransferases Execute Programmed Clearance of Maternal Transcriptome in Vertebrate Embryos. Mol Cell 70, 72-82.e7. https://doi.org/10.1016/J.MOLCEL.2018.03.004

Chang, N.C., Rovira, Q., Wells, J., Feschotte, C., Vaquerizas, J.M., 2022. Zebrafish transposable elements show extensive diversification in age, genomic distribution, and developmental expression. Genome Res 32, 1408–1423. https://doi.org/10.1101/GR.275655.121/-/DC1

Chari, S., Wilky, H., Govindan, J., Amodeo, A.A., 2019. Histone concentration regulates the cell cycle and transcription in early development. Development 146. https://doi.org/10.1242/DEV.177402

Charney, R.M., Forouzmand, E., Cho, J.S., Cheung, J., Paraiso, K.D., Yasuoka, Y., Takahashi, S., Taira, M., Blitz, I.L., Xie, X., Cho, K.W.Y., 2017. Foxh1 Occupies cis-Regulatory Modules Prior to Dynamic Transcription Factor Interactions Controlling the Mesendoderm Gene Program. Dev Cell 40, 595-607.e4. https://doi.org/10.1016/J.DEVCEL.2017.02.017

Chen, H., Einstein, L.C., Little, S.C., Good, M.C., 2019. Spatiotemporal Patterning of Zygotic Genome Activation In A Model Vertebrate Embryo. Dev Cell 49, 852. https://doi.org/10.1016/J.DEVCEL.2019.05.036

Chen, H., Good, M.C., 2022. Nascent transcriptome reveals orchestration of zygotic genome activation in early embryogenesis. Current Biology 32, 4314-4324.e7. https://doi.org/10.1016/J.CUB.2022.07.078

Chen, K., Johnston, J., Shao, W., Meier, S., Staber, C., Zeitlinger, J., 2013a. A global change in RNA polymerase II pausing during the Drosophila midblastula transition. Elife 2013. https://doi.org/10.7554/ELIFE.00861

Chen, K., Johnston, J., Shao, W., Meier, S., Staber, C., Zeitlinger, J., 2013b. A global change in RNA polymerase II pausing during the Drosophila midblastula transition. Elife 2013. https://doi.org/10.7554/ELIFE.00861

Cheng, H., Concepcion, G.T., Feng, X., Zhang, H., Li, H., 2021. Haplotype-resolved de novo assembly using phased assembly graphs with hifiasm. Nature Methods 2021 18:2 18, 170–175. https://doi.org/10.1038/s41592-020-01056-5

Cho, C.-Y., O'Farrell, P.H., 2023. Stepwise modifications of transcriptional hubs link pioneer factor activity to a burst of transcription. Nat Commun 14, 4848. https://doi.org/10.1038/S41467-023-40485-6

Cho, W.K., Jayanth, N., English, B.P., Inoue, T., Andrews, J.O., Conway, W., Grimm, J.B., Spille, J.H., Lavis, L.D., Lionnet, T., Cisse, I.I., 2016. RNA Polymerase II cluster dynamics predict mRNA output in living cells. Elife 5. https://doi.org/10.7554/ELIFE.13617

Chong, S., Dugast-Darzacq, C., Liu, Z., Dong, P., Dailey, G.M., Cattoglio, C., Heckert, A., Banala, S., Lavis, L., Darzacq, X., Tjian, R., 2018. Imaging dynamic and selective low-complexity domain interactions that control gene transcription. Science (1979) 361. https://doi.org/10.1126/SCIENCE.AAR2555/SUPPL_FILE/AAR2555_CHONG_SM.PDF

Cisse, I.I., Izeddin, I., Causse, S.Z., Boudarene, L., Senecal, A., Muresan, L., Dugast-Darzacq, C., Hajj, B., Dahan, M., Darzacq, X., 2013. Real-time dynamics of RNA polymerase II clustering in live human cells. Science (1979) 341, 664–667. https://doi.org/10.1126/SCIENCE.1239053/SUPPL_FILE/CISSE.SM.PDF

Clemson, C.M., Hutchinson, J.N., Sara, S.A., Ensminger, A.W., Fox, A.H., Chess, A., Lawrence, J.B., 2009. An Architectural Role for a Nuclear Non-coding RNA: NEAT1 RNA is Essential for the Structure of Paraspeckles. Mol Cell 33, 717. https://doi.org/10.1016/J.MOLCEL.2009.01.026

Collart, C., Allen, G.E., Bradshaw, C.R., Smith, J.C., Zegerman, P., 2013. Titration of four replication factors is essential for the Xenopus laevis midblastula transition. Science 341, 893–896. https://doi.org/10.1126/SCIENCE.1241530

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

Colonnetta, M.M., Abrahante, J.E., Schedl, P., Gohl, D.M., Deshpande, G., 2021. CLAMP regulates zygotic genome activation in Drosophila embryos. Genetics 219. https://doi.org/10.1093/GENETICS/IYAB107

Core, L., Adelman, K., 2019. Promoter-proximal pausing of RNA polymerase II: a nexus of gene regulation. Genes Dev 33, 960. https://doi.org/10.1101/GAD.325142.119

Cramer, P., 2019. Organization and regulation of gene transcription. Nature 2019 573:7772 573, 45–54. https://doi.org/10.1038/s41586-019-1517-4

Crippa, M., Davidson, E.H., Mirsky, A.E., 1967. Persistence in early amphibian embryos of informational RNA's from the lampbrush chromosome stage of oögenesis. Proc Natl Acad Sci U S A 57, 885–892. https://doi.org/10.1073/PNAS.57.4.885/ASSET/46DE63F2-CA09-4B04-90C2-F9C19DC66BE8/ASSETS/PNAS.57.4.885.FP.PNG

das Neves, R.P., Jones, N.S., Andreu, L., Gupta, R., Enver, T., Iborra, F.J., 2010. Connecting variability in global transcription rate to mitochondrial variability. PLoS Biol 8. https://doi.org/10.1371/JOURNAL.PBIO.1000560

De Iaco, A., Planet, E., Coluccio, A., Verp, S., Duc, J., Trono, D., 2017. DUX-family transcription factors regulate zygotic genome activation in placental mammals. Nature Genetics 2017 49:6 49, 941–945. https://doi.org/10.1038/ng.3858

De Renzis, S., Elemento, O., Tavazoie, S., Wieschaus, E.F., 2007. Unmasking Activation of the Zygotic Genome Using Chromosomal Deletions in the Drosophila Embryo. PLoS Biol 5, e117. https://doi.org/10.1371/JOURNAL.PBIO.0050117

De Wit, E., Bouwman, B.A.M., Zhu, Y., Klous, P., Splinter, E., Verstegen, M.J.A.M., Krijger, P.H.L., Festuccia, N., Nora, E.P., Welling, M., Heard, E., Geijsen, N., Poot, R.A., Chambers, I., De Laat, W., 2013. The pluripotent genome in three dimensions is shaped around pluripotency factors. Nature 2013 501:7466 501, 227–231. https://doi.org/10.1038/nature12420

Delker, R.K., Munce, R.H., Hu, M., Mann, R.S., 2022. Fluorescent labeling of genomic loci in Drosophila imaginal discs with heterologous DNA-binding proteins. Cell Reports Methods 2, 100175. https://doi.org/10.1016/J.CRMETH.2022.100175

Deng, Y., Qian, Y., Meng, M., Jiang, H., Dong, Y., Fang, C., He, S., Yang, L., 2022. Extensive sequence divergence between the reference genomes of two zebrafish strains, Tuebingen and AB. Mol Ecol Resour 22, 2148–2157. https://doi.org/10.1111/1755-0998.13602

Despic, V., Dejung, M., Gu, M., Krishnan, J., Zhang, J., Herzel, L., Straube, K., Gerstein, M.B., Butter, F., Neugebauer, K.M., 2017. Dynamic RNA-protein interactions underlie the zebrafish maternal-to-zygotic transition. Genome Res 27, 1184–1194. https://doi.org/10.1101/GR.215954.116/-/DC1

Du, Z., Zheng, H., Huang, B., Ma, R., Wu, J., Zhang, Xianglin, He, J., Xiang, Y., Wang, Q., Li, Y., Ma, J., Zhang, Xu, Zhang, K., Wang, Y., Zhang, M.Q., Gao, J., Dixon, J.R., Wang, X., Zeng, J., Xie, W., 2017. Allelic reprogramming of 3D chromatin architecture during early mammalian development. Nature 547, 232–235. https://doi.org/10.1038/NATURE23263

Duan, J.E., Rieder, L.E., Colonnetta, M.M., Huang, A., McKenney, M., Watters, S., Deshpande, G., Jordan, W.T., Fawzi, N.L., Larschan, E.N., 2021. Clamp and zelda function together to promote drosophila zygotic genome activation. Elife 10. https://doi.org/10.7554/ELIFE.69937

E., H.C., B., E.M., 2023. Intrinsic protein disorder is insufficient to drive subnuclear clustering in embryonic transcription factors. Elife 12. https://doi.org/10.7554/ELIFE.88221

Eckersley-Maslin, M., Alda-Catalinas, C., Blotenburg, M., Kreibich, E., Krueger, C., Reik, W., 2019. Dppa2 and Dppa4 directly regulate the Dux-driven zygotic transcriptional program. Genes Dev 33, 194–208. https://doi.org/10.1101/GAD.321174.118

Edgar, B.A., Kiehle, C.P., Schubiger, G., 1986. Cell cycle control by the nucleo-cytoplasmic ratio in early Drosophila development. Cell 44, 365–372. https://doi.org/10.1016/0092-8674(86)90771-3

Edgar, B.A., Schubiger, G., 1986a. Parameters controlling transcriptional activation during early Drosophila development. Cell 44, 871–877. https://doi.org/10.1016/0092-8674(86)90009-7

Edgar, B.A., Schubiger, G., 1986b. Parameters controlling transcriptional activation during early drosophila development. Cell 44, 871–877. https://doi.org/10.1016/0092-8674(86)90009-7

Eichhorn, S.W., Subtelny, A.O., Kronja, I., Kwasnieski, J.C., Orr-Weaver, T.L., Bartel, D.P., 2016. mRNA poly(A)-tail changes specified by deadenylation broadly reshape translation in Drosophila oocytes and early embryos. Elife 5. https://doi.org/10.7554/ELIFE.16955

Emerson, C.P., Humphreys, T., 1970. Regulation of DNA-like RNA and the apparent activation of ribosomal RNA synthesis in sea urchin embryos: quantitative measurements of newly synthesized RNA. Dev Biol 23, 86–112. https://doi.org/10.1016/S0012-1606(70)80008-2

Erickson, J.W., Cline, T.W., 1998. Key aspects of the primary sex determination mechanism are conserved across the genus Drosophila. Development 125, 3259–3268. https://doi.org/10.1242/DEV.125.16.3259

Erickson, J.W., Cline, T.W., 1993. A bZIP protein, sisterless-a, collaborates with bHLH transcription factors early in Drosophila development to determine sex. Genes Dev 7, 1688–1702. https://doi.org/10.1101/GAD.7.9.1688

Farrell, J.A., O'Farrell, P.H., 2013. Mechanism and Regulation of Twine (Cdc25) Protein Destruction in Embryonic Cell Cycle Remodeling. Curr Biol 23, 118. https://doi.org/10.1016/J.CUB.2012.11.036

Fischer, P., Chen, H., Pacho, F., Rieder, D., Kimmel, R.A., Meyer, D., 2019. FoxH1 represses miR-430 during early embryonic development of zebrafish via non-canonical regulation. BMC Biol 17, 1–17. https://doi.org/10.1186/S12915-019-0683-Z/FIGURES/6

Foo, S.M., Sun, Y., Lim, B., Ziukaite, R., O'Brien, K., Nien, C.Y., Kirov, N., Shvartsman, S.Y., Rushlow, C.A., 2014. Zelda potentiates morphogen activity by increasing chromatin accessibility. Curr Biol 24, 1341–1346. https://doi.org/10.1016/J.CUB.2014.04.032

Friman, E.T., Flyamer, I.M., Marenduzzo, D., Boyle, S., Bickmore, W.A., 2023. Ultra-long-range interactions between active regulatory elements. Genome Res gr.277567.122. https://doi.org/10.1101/GR.277567.122

Fujinaga, K., Huang, F., Peterlin, B.M., 2023. P-TEFb: The master regulator of transcription elongation. Mol Cell 83, 393–403. https://doi.org/10.1016/J.MOLCEL.2022.12.006

Gao, M., Veil, M., Rosenblatt, M., Riesle, A.J., Gebhard, A., Hass, H., Buryanova, L., Yampolsky, L.Y., Grüning, B., Ulianov, S. V., Timmer, J., Onichtchouk, D., 2022. Pluripotency factors determine gene expression repertoire at zygotic genome activation. Nat Commun 13. https://doi.org/10.1038/S41467-022-28434-1

Gaskill, M.M., Gibson, T.J., Larson, E.D., Harrison, M.M., 2021. GAF is essential for zygotic genome activation and chromatin accessibility in the early Drosophila embryo. Elife 10. https://doi.org/10.7554/ELIFE.66668

Gaskill, M.M., Soluri, I. V., Branks, A.E., Boka, A.P., Stadler, M.R., Vietor, K., Huang, H.-Y.S., Gibson, T.J., Mukherjee, A., Mir, M., Blythe, S.A., Harrison, M.M., 2023a. Localization of the Drosophila pioneer factor GAF to subnuclear foci is driven by DNA binding and required to silence satellite repeat expression. Dev Cell 58, 1610-1624.e8. https://doi.org/10.1016/J.DEVCEL.2023.06.010

Gaskill, M.M., Soluri, I. V., Branks, A.E., Boka, A.P., Stadler, M.R., Vietor, K., Huang, H.-Y.S., Gibson, T.J., Mukherjee, A., Mir, M., Blythe, S.A., Harrison, M.M., 2023b. Localization of the Drosophila pioneer factor GAF to subnuclear foci is driven by DNA binding and required to silence satellite repeat expression. Dev Cell 58, 1610-1624.e8. https://doi.org/10.1016/J.DEVCEL.2023.06.010

Gassler, J., Kobayashi, W., Gáspár, I., Ruangroengkulrith, S., Mohanan, A., Hernández, L.G., Kravchenko, P., Kümmecke, M., Lalic, A., Rifel, N., Ashburn, R.J., Zaczek, M., Vallot, A., Rico, L.C., Ladstätter, S., Tachibana, K., 2022a. Zygotic genome activation by the totipotency pioneer factor Nr5a2. Science (1979) 378, 1305–1315. https://doi.org/10.1126/SCIENCE.ABN7478/SUPPL_FILE/SCIENCE.ABN7478_MDAR_REPRODUCIBILITY_CHECKLIST.PDF

Gassler, J., Kobayashi, W., Gáspár, I., Ruangroengkulrith, S., Mohanan, A., Hernández, L.G., Kravchenko, P., Kümmecke, M., Lalic, A., Rifel, N., Ashburn, R.J., Zaczek, M., Vallot, A., Rico, L.C., Ladstätter, S., Tachibana, K., 2022b. Zygotic genome activation by the totipotency pioneer factor Nr5a2. Science 378, 1305–1315. https://doi.org/10.1126/SCIENCE.ABN7478

Germier, T., Sylvain, A., Silvia, K., David, L., Kerstin, B., 2018. Real-time imaging of specific genomic loci in eukaryotic cells using the ANCHOR DNA labelling system. Methods 142, 16–23. https://doi.org/10.1016/J.YMETH.2018.04.008

Gert, K.R., Cabrera Quio, L.E., Novatchkova, M., Guo, Y., Cairns, B.R., Pauli, A., Biocenter, V., Program, P., 2021. Reciprocal zebrafish-medaka hybrids reveal maternal control of zygotic genome activation timing. bioRxiv 2021.11.03.467109. https://doi.org/10.1101/2021.11.03.467109

Giraldez, A.J., Cinalli, R.M., Glasner, M.E., Enright, A.J., Thomson, J.M., Baskerville, S., Hammond, S.M., Bartel, D.P., Schier, A.F., 2005. MicroRNAs regulate brain morphogenesis in zebrafish. Science (1979) 308, 833–838. https://doi.org/10.1126/SCIENCE.1109020/SUPPL_FILE/PAPV2.PDF

Giraldez, A.J., Mishima, Y., Rihel, J., Grocock, R.J., Van Dongen, S., Inoue, K., Enright, A.J., Schier, A.F., 2006. Zebrafish MiR-430 promotes deadenylation and clearance of maternal mRNAs. Science (1979) 312, 75–79. https://doi.org/10.1126/SCIENCE.1122689

Gottesfeld, J.M., 2019. Milestones in transcription and chromatin published in the journal of biological chemistry. Journal of Biological Chemistry 294, 1652–1660. https://doi.org/10.1074/JBC.TM118.004162

Graham, C.F., Morgan, R.W., 1966. Changes in the cell cycle during early amphibian development. Dev Biol 14, 439–460. https://doi.org/10.1016/0012-1606(66)90024-8

Grow, E.J., Weaver, B.D., Smith, C.M., Guo, J., Stein, P., Shadle, S.C., Hendrickson, P.G., Johnson, N.E., Butterfield, R.J., Menafra, R., Kloet, S.L., van der Maarel, S.M., Williams, C.J., Cairns, B.R., 2021. p53 convergently activates Dux/DUX4 in embryonic stem cells and in facioscapulohumeral muscular dystrophy cell models. Nature Genetics 2021 53:8 53, 1207–1220. https://doi.org/10.1038/s41588-021-00893-0

Guryev, V., Koudijs, M.J., Berezikov, E., Johnson, S.L., Plasterk, R.H.A., Van Eeden, F.J.M., Cuppen, E., 2006. Genetic variation in the zebrafish. Genome Res 16, 491. https://doi.org/10.1101/GR.4791006

Haberle, V., Li, N., Hadzhiev, Y., Plessy, C., Previti, C., Nepal, C., Gehrig, J., Dong, X., Akalin, A., Suzuki, A.M., Van Ijcken, W.F.J., Armant, O., Ferg, M., Strähle, U., Carninci, P., Müller, F., Lenhard, B., 2014. Two independent transcription initiation codes overlap on vertebrate core promoters. Nature 507, 381. https://doi.org/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., Sik, A., Hutvàgner, G., Tora, L., Gambus, A., Fossey, J.S., Müller, F., 2019. A cell cycle-coordinated Polymerase II transcription compartment encompasses gene expression before global genome activation. Nature Communications 2019 10:1 10, 1–14. https://doi.org/10.1038/s41467-019-08487-5

Hadzhiev, Y., Wheatley, L., Cooper, L., Ansaloni, F., Whalley, C., Chen, Z., Finaurini, S., Gustincich, S., Sanges, R., Burgess, S., Beggs, A., Müller, F., 2023. The miR-430 locus with extreme promoter density forms a transcription body during the minor wave of zygotic genome activation. Dev Cell 58, 155-170.e8. https://doi.org/10.1016/J.DEVCEL.2022.12.007

Harrison, M.M., Botchan, M.R., Cline, T.W., 2010. Grainyhead and Zelda compete for binding to the promoters of the earliest-expressed Drosophila genes. Dev Biol 345, 248–255. https://doi.org/10.1016/J.YDBIO.2010.06.026

Harrison, M.M., Li, X.Y., Kaplan, T., Botchan, M.R., 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. https://doi.org/10.1371/JOURNAL.PGEN.1002266

Hendrickson, P.G., Doráis, 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., Nix, D.A., Peterson, C.M., Tapscott, S.J., Carrell, D.T., Cairns, B.R., 2017. Conserved roles of mouse DUX and human DUX4 in activating cleavage-stage genes and MERVL/HERVL retrotransposons. Nature Genetics 2017 49:6 49, 925–934. https://doi.org/10.1038/ng.3844

Herberg, S., Gert, K.R., Schleiffer, A., Pauli, A., 2018. The Ly6/uPAR protein Bouncer is necessary and sufficient for species-specific fertilization. Science (1979) 361, 1029–1033. https://doi.org/10.1126/SCIENCE.AAT7113/SUPPL_FILE/AAT7113S2.MOV

Heyn, P., Kircher, M., Dahl, A., Kelso, J., Tomancak, P., Kalinka, A.T., Neugebauer, K.M., 2014. The Earliest Transcribed Zygotic Genes Are Short, Newly Evolved, and Different across Species. Cell Rep 6, 285–292. https://doi.org/10.1016/J.CELREP.2013.12.030

Hilbert, L., Sato, Y., Kuznetsova, K., Bianucci, T., Kimura, H., Jülicher, F., Honigmann, A., Zaburdaev, V., Vastenhouw, N.L., 2021. Transcription organizes euchromatin via microphase separation. Nature Communications 2021 12:1 12, 1–12. https://doi.org/10.1038/s41467-021-21589-3

Hilbert, L., Sato, Y., Kuznetsova, K., Bianucci, T., Kimura, H., Jülicher, F., Honigmann, A., Zaburdaev, V., Vastenhouw, N.L., n.d. Transcription organizes euchromatin via microphase separation. https://doi.org/10.1038/s41467-021-21589-3

Hinegardner, R.T., Rao, B., Feldman, D.E., 1964. The DNA synthetic period during early development of the sea urchin egg. Exp Cell Res 36, 53–61. https://doi.org/10.1016/0014-4827(64)90159-4

Holden, L.A., Brown, K.H., 2018a. Baseline mRNA expression differs widely between common laboratory strains of zebrafish. Scientific Reports 2018 8:1 8, 1–10. https://doi.org/10.1038/s41598-018-23129-4

Holden, L.A., Brown, K.H., 2018b. Baseline mRNA expression differs widely between common laboratory strains of zebrafish. Sci Rep 8, 4780. https://doi.org/10.1038/S41598-018-23129-4

Hontelez, S., Van Kruijsbergen, I., Georgiou, G., Van Heeringen, S.J., Bogdanovic, O., Lister, R., Veenstra, G.J.C., 2015. Embryonic transcription is controlled by maternally defined chromatin state. Nature Communications 2015 6:1 6, 1–13. https://doi.org/10.1038/ncomms10148

Howe, K., Clark, M.D., Torroja, C.F., Torrance, J., Berthelot, C., Muffato, M., Collins, J.E., Humphray, S., McLaren, K., Matthews, L., McLaren, S., Sealy, I., Caccamo, M., Churcher, C., Scott, C., Barrett, J.C., Koch, R., Rauch, G.J., White, S., Chow, W., Kilian, B., Quintais, L.T., Guerra-Assunção, J.A., Zhou, Y., Gu, Y., Yen, J., Vogel, J.H., Eyre, T., Redmond, S., Banerjee, R., Chi, J., Fu, B., Langley, E., Maguire, S.F., Laird, G.K., Lloyd, D., Kenyon, E., Donaldson, S., Sehra, H., Almeida-King, J., Loveland, J., Trevanion, S., Jones, M., Quail, M., Willey, D., Hunt, A., Burton, J., Sims, S., McLay, K., Plumb, B., Davis, J., Clee, C., Oliver, K., Clark, R., Riddle, C., Eliott, D., Threadgold, G., Harden, G., Ware, D., Mortimer, B., Kerry, G., Heath, P., Phillimore, B., Tracey, A., Corby, N., Dunn, M., Johnson, C., Wood, J., Clark, S., Pelan, S., Griffiths, G., Smith, M., Glithero, R., Howden, P., Barker, N., Stevens, C., Harley, J., Holt, K., Panagiotidis, G., Lovell, J., Beasley, H., Henderson, C., Gordon, D., Auger, K., Wright, D., Collins, J., Raisen, C., Dyer, L., Leung, K., Robertson, L., Ambridge, K., Leongamornlert, D., McGuire, S., Gilderthorp, R., Griffiths, C., Manthravadi, D., Nichol, S., Barker, G., Whitehead, S., Kay, M., Brown, J., Murnane, C., Gray, E., Humphries, M., Sycamore, N., Barker, D., Saunders, D., Wallis, J., Babbage, A., Hammond, S., Mashreghi-Mohammadi, M., Barr, L., Martin, S., Wray, P., Ellington, A., Matthews, N., Ellwood, M., Woodmansey, R., Clark, G., Cooper, J., Tromans, A., Grafham, D., Skuce, C., Pandian, R., Andrews, R., Harrison, E., Kimberley, A., Garnett, J., Fosker, N., Hall, R., Garner, P., Kelly, D., Bird, C., Palmer, S., Gehring, I., Berger, A., Dooley, C.M., Ersan-Ürün, Z., Eser, C., Geiger, H., Geisler, M., Karotki, L., Kirn, A., Konantz, J., Konantz, M., Oberländer, M., Rudolph-Geiger, S., Teucke, M., Osoegawa, K., Zhu, B., Rapp, A., Widaa, S., Langford, C., Yang, F., Carter, N.P., Harrow, J., Ning, Z., Herrero, J., Searle, S.M.J., Enright, A., Geisler, R., Plasterk, R.H.A., Lee, C., Westerfield, M., De Jong, P.J., Zon, L.I., Postlethwait, J.H., Nüsslein-Volhard, C., Hubbard, T.J.P., Crollius, H.R., Rogers, J., Stemple, D.L., 2013. The zebrafish reference genome sequence and its relationship to the human genome. Nature 496, 498. https://doi.org/10.1038/NATURE12111

Hug, C.B., Grimaldi, A.G., Kruse, K., Vaquerizas, J.M., 2017. Chromatin Architecture Emerges during Zygotic Genome Activation Independent of Transcription. Cell 169, 216-228.e19. https://doi.org/10.1016/J.CELL.2017.03.024

Humphreys, T., 1971. Measurements of messenger RNA entering polysomes upon fertilization of sea urchin eggs. Dev Biol 26, 201–208. https://doi.org/10.1016/0012-1606(71)90122-9

Humphreys, T., 1969. Efficiency of translation of messenger-RNA before and after fertilization in sea urchins. Dev Biol 20, 435–458. https://doi.org/10.1016/0012-1606(69)90025-6

Iaco, A. De, Coudray, A., Duc, J., Trono, D., 2019. DPPA2 and DPPA4 are necessary to establish a 2C-like state in mouse embryonic stem cells. EMBO Rep 20, e47382. https://doi.org/10.15252/EMBR.201847382

Jackson, D.A., Hassan, A.B., Errington, R.J., Cook, P.R., 1993. Visualization of focal sites of transcription within human nuclei. EMBO J 12, 1059–1065. https://doi.org/10.1002/J.1460-2075.1993.TB05747.X

Jeronimo, C., Robert, F., 2014. Kin28 regulates the transient association of Mediator with core promoters. Nat Struct Mol Biol 21, 449–455. https://doi.org/10.1038/NSMB.2810

Jevtić, P., Levy, D.L., 2017. Both Nuclear Size and DNA Amount Contribute to Midblastula Transition Timing in Xenopus laevis. Sci Rep 7. https://doi.org/10.1038/S41598-017-08243-Z

Jevtić, P., Levy, D.L., 2015. Nuclear Size Scaling during Xenopus Early Development Contributes to Midblastula Transition Timing. Current Biology 25, 45–52. https://doi.org/10.1016/J.CUB.2014.10.051

Ji, S., Chen, F., Stein, P., Wang, J., Zhou, Z., Wang, L., Zhao, Q., Lin, Z., Liu, B., Xu, K., Lai, F., Xiong, Z., Hu, X., Kong, T., Kong, F., Huang, B., Wang, Q., Xu, Q., Fan, Q., Liu, L., Williams, C.J., Schultz, R.M., Xie, W., 2023a. OBOX regulates mouse zygotic genome activation and early development. Nature 2023 620:7976 620, 1047–1053. https://doi.org/10.1038/s41586-023-06428-3

Ji, S., Chen, F., Stein, P., Wang, J., Zhou, Z., Wang, L., Zhao, Q., Lin, Z., Liu, B., Xu, K., Lai, F., Xiong, Z., Hu, X., Kong, T., Kong, F., Huang, B., Wang, Q., Xu, Q., Fan, Q., Liu, L., Williams, C.J., Schultz, R.M., Xie, W., 2023b. OBOX regulates mouse zygotic genome activation and early development. Nature 2023 620:7976 620, 1047–1053. https://doi.org/10.1038/s41586-023-06428-3

Joseph, S.R., Pálfy, M., Hilbert, L., Kumar, M., Karschau, J., Zaburdaev, V., Shevchenko, A., Vastenhouw, N.L., 2017a. Competition between histone and transcription factor binding regulates the onset of transcription in zebrafish embryos. Elife 6. https://doi.org/10.7554/ELIFE.23326

Joseph, S.R., Pálfy, M., Hilbert, L., Kumar, M., Karschau, J., Zaburdaev, V., Shevchenko, A., Vastenhouw, N.L., 2017b. Competition between histone and transcription factor binding regulates the onset of transcription in zebrafish embryos. Elife 6. https://doi.org/10.7554/ELIFE.23326

Jukam, D., Kapoor, R.R., Straight, A.F., Skotheim, J.M., 2021. The DNA-to-cytoplasm ratio broadly activates zygotic gene expression in Xenopus. Curr Biol 31, 4269-4281.e8. https://doi.org/10.1016/J.CUB.2021.07.035

Kærn, M., Elston, T.C., Blake, W.J., Collins, J.J., 2005. Stochasticity in gene expression: from theories to phenotypes. Nat Rev Genet 6, 451–464. https://doi.org/10.1038/NRG1615

Kawasaki, K., Fukaya, T., 2023. Functional coordination between transcription factor clustering and gene activity. Mol Cell 83, 1605-1622.e9. https://doi.org/10.1016/j.molcel.2023.04.018

Ke, Y., Xu, Y., Chen, X., Feng, S., Liu, Z., Sun, Y., Yao, X., Li, F., Zhu, W., Gao, L., Chen, H., Du, Z., Xie, W., Xu, X., Huang, X., Liu, J., 2017. 3D Chromatin Structures of Mature Gametes and Structural Reprogramming during Mammalian Embryogenesis. Cell 170, 367-381.e20. https://doi.org/10.1016/J.CELL.2017.06.029

Kiguchi, Y., Nishijima, S., Kumar, N., Hattori, M., Suda, W., 2021a. Long-read metagenomics of multiple displacement amplified DNA of low-biomass human gut phageomes by SACRA pre-processing chimeric reads. DNA Research 28, 1–10. https://doi.org/10.1093/DNARES/DSAB019

Kiguchi, Y., Nishijima, S., Kumar, N., Hattori, M., Suda, W., 2021b. Long-read metagenomics of multiple displacement amplified DNA of low-biomass human gut phageomes by SACRA pre-processing chimeric reads. DNA Research 28, 1–10. https://doi.org/10.1093/DNARES/DSAB019

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

Koreski, K.P., Rieder, L.E., McLain, L.M., Chaubal, A., Marzluff, W.F., Duronio, R.J., 2020. Drosophila histone locus body assembly and function involves multiple interactions. Mol Biol Cell 31, 1525. https://doi.org/10.1091/MBC.E20-03-0176

Kostrewa, D., Zeller, M.E., Armache, K.J., Seizl, M., Leike, K., Thomm, M., Cramer, P., 2009. RNA polymerase II–TFIIB structure and mechanism of transcription initiation. Nature 2009 462:7271 462, 323–330. https://doi.org/10.1038/nature08548

Kuznetsova, K., Chabot, M., Ugolini, M., Kimura, H., Jug, F., Vastenhouw Correspondence, N.L., 2023. Nanog organizes transcription bodies In brief. https://doi.org/10.1016/j.cub.2022.11.015

Kwasnieski, J.C., Orr-Weaver, T.L., Bartel, D.P., 2019. Early genome activation in Drosophila is extensive with an initial tendency for aborted transcripts and retained introns. Genome Res 29, 1188–1197. https://doi.org/10.1101/GR.242164.118/-/DC1

Lange, M., Neuzeret, F., Fabreges, B., Froc, C., Bedu, S., Bally-Cuif, L., Norton, W.H.J., 2013. Inter-Individual and Inter-Strain Variations in Zebrafish Locomotor Ontogeny. PLoS One 8, 70172. https://doi.org/10.1371/JOURNAL.PONE.0070172

Larson, E.D., Komori, H., Fitzpatrick, Z.A., Krabbenhoft, S.D., Lee, C.Y., Harrison, M., 2022. Premature translation of the Drosophila zygotic genome activator Zelda is not sufficient to precociously activate gene expression. G3: Genes|Genomes|Genetics 12. https://doi.org/10.1093/G3JOURNAL/JKAC159

Laue, K., Rajshekar, S., Courtney, A.J., Lewis, Z.A., Goll, M.G., 2019. The maternal to zygotic transition regulates genome-wide heterochromatin establishment in the zebrafish embryo. Nature Communications 2019 10:1 10, 1–10. https://doi.org/10.1038/s41467-019-09582-3

Lee, M.T., Bonneau, A.R., Takacs, C.M., Bazzini, A.A., Divito, K.R., Fleming, E.S., Giraldez, A.J., 2013a. Nanog, Pou5f1 and SoxB1 activate zygotic gene expression during the maternal-to-zygotic transition. Nature 2013 503:7476 503, 360–364. https://doi.org/10.1038/nature12632

Lee, M.T., Bonneau, A.R., Takacs, C.M., Bazzini, A.A., Divito, K.R., Fleming, E.S., Giraldez, A.J., 2013b. Nanog, Pou5f1 and SoxB1 activate zygotic gene expression during the maternal-to-zygotic transition. Nature 503, 360. https://doi.org/10.1038/NATURE12632

Leichsenring, M., Maes, J., Mos�sner, R., Driever, W., Onichtchouk, D., 2013a. Pou5f1 transcription factor controls zygotic gene activation in vertebrates. Science (1979) 341, 1005–1009. https://doi.org/10.1126/SCIENCE.1242527/SUPPL_FILE/TABLES1_19SEPT2013.PDF

Leichsenring, M., Maes, J., Mos�sner, R., Driever, W., Onichtchouk, D., 2013b. Pou5f1 transcription factor controls zygotic gene activation in vertebrates. Science (1979) 341, 1005–1009. https://doi.org/10.1126/SCIENCE.1242527/SUPPL_FILE/TABLES1_19SEPT2013.PDF

Li, H., 2018. Minimap2: pairwise alignment for nucleotide sequences. Bioinformatics 34, 3094–3100. https://doi.org/10.1093/BIOINFORMATICS/BTY191

Li, X.Y., Harrison, M.M., Villalta, J.E., Kaplan, T., Eisen, M.B., 2014. Establishment of regions of genomic activity during the Drosophila maternal to zygotic transition. Elife 3, 3737. https://doi.org/10.7554/ELIFE.03737

Liang, H.L., Nien, C.Y., Liu, H.Y., Metzstein, M.M., Kirov, N., Rushlow, C., 2008a. The zinc-finger protein Zelda is a key activator of the early zygotic genome in Drosophila. Nature 2008 456:7220 456, 400–403. https://doi.org/10.1038/nature07388

Liang, H.L., Nien, C.Y., Liu, H.Y., Metzstein, M.M., Kirov, N., Rushlow, C., 2008b. The zinc-finger protein Zelda is a key activator of the early zygotic genome in Drosophila. Nature 456, 400. https://doi.org/10.1038/NATURE07388

Lindeman, L.C., Andersen, I.S., Reiner, A.H., Li, N., Aanes, H., Østrup, O., Winata, C., Mathavan, S., Müller, F., Aleström, P., Collas, P., 2011. Prepatterning of Developmental Gene Expression by Modified Histones before Zygotic Genome Activation. Dev Cell 21, 993–1004. https://doi.org/10.1016/j.devcel.2011.10.008

Ling, J., Umezawa, K.Y., Scott, T., Small, S., 2019. Bicoid-Dependent Activation of the Target Gene hunchback Requires a Two-Motif Sequence Code in a Specific Basal Promoter. Mol Cell 75, 1178-1187.e4. https://doi.org/10.1016/J.MOLCEL.2019.06.038

Lott, S.E., Villalta, J.E., Schroth, G.P., Luo, S., Tonkin, L.A., Eisen, M.B., 2011. Noncanonical Compensation of Zygotic X Transcription in Early Drosophila melanogaster Development Revealed through Single-Embryo RNA-Seq. PLoS Biol 9, e1000590. https://doi.org/10.1371/JOURNAL.PBIO.1000590

Lu, F., Liu, Y., Inoue, A., Suzuki, T., Zhao, K., Zhang, Y., 2016. Establishing chromatin regulatory landscape during mouse preimplantation development. Cell 165, 1375–1388. https://doi.org/10.1016/j.cell.2016.05.050

Lu, X., Li, J.M., Elemento, O., Tavazoie, S., Wieschaus, E.F., 2009. Coupling of zygotic transcription to mitotic control at the Drosophila mid-blastula transition. Development 136, 2101–2110. https://doi.org/10.1242/DEV.034421

Lund, E., Dahlberg, J.E., 1992. Control of 4-8S RNA transcription at the midblastula transition in Xenopus laevis embryos. Genes Dev 6, 1097–1106. https://doi.org/10.1101/GAD.6.6.1097

Lupo, O., Kumar, D.K., Livne, R., Chappleboim, M., Levy, I., Barkai, N., 2023. The architecture of binding cooperativity between densely bound transcription factors. Cell Syst 14, 732-745.e5. https://doi.org/10.1016/J.CELS.2023.06.010

Madsen, E.B., Höijer, I., Kvist, T., Ameur, A., Mikkelsen, M.J., 2020. Xdrop: Targeted sequencing of long DNA molecules from low input samples using droplet sorting. Hum Mutat 41, 1671. https://doi.org/10.1002/HUMU.24063

Mathavan, S., Lee, S.G.P., Mak, A., Miller, L.D., Murthy, K.R.K., Govindarajan, K.R., Tong, Y., Wu, Y.L., Lam, S.H., Yang, H., Ruan, Y., Korzh, V., Gong, Z., Liu, E.T., Lufkin, T., 2005. Transcriptome Analysis of Zebrafish Embryogenesis Using Microarrays. PLoS Genet 1, e29. https://doi.org/10.1371/JOURNAL.PGEN.0010029

McKnight, S.L., Miller, O.L., 1976. Ultrastructural patterns of RNA synthesis during early embryogenesis of Drosophila melanogaster. Cell 8, 305–319. https://doi.org/10.1016/0092-8674(76)90014-3

Meeussen, J.V.W., Pomp, W., Brouwer, I., de Jonge, W.J., Patel, H.P., Lenstra, T.L., 2023. Transcription factor clusters enable target search but do not contribute to target gene activation. Nucleic Acids Res 51, 5449–5468. https://doi.org/10.1093/NAR/GKAD227

Meschichi, A., Ingouff, M., Picart, C., Mirouze, M., Desset, S., Gallardo, F., Bystricky, K., Picault, N., Rosa, S., Pontvianne, F., 2021. ANCHOR: A Technical Approach to Monitor Single-Copy Locus Localization in Planta. Front Plant Sci 12, 677849. https://doi.org/10.3389/FPLS.2021.677849/BIBTEX

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., Giraldez, A.J., 2022. The landscape of pioneer factor activity reveals the mechanisms of chromatin reprogramming and genome activation. Mol Cell 82, 986-1002.e9. https://doi.org/10.1016/J.MOLCEL.2022.01.024

Michael, A.K., Grand, R.S., Isbel, L., Cavadini, S., Kozicka, Z., Kempf, G., Bunker, R.D., Schenk, A.D., Graff-Meyer, A., Pathare, G.R., Weiss, J., Matsumoto, S., Burger, L., Schübeler, D., Thomä, N.H., 2020. Mechanisms of OCT4-SOX2 motif readout on nucleosomes. Science (1979) 368, 1460–1465. https://doi.org/10.1126/SCIENCE.ABB0074/SUPPL_FILE/ABB0074_S1.MOV

Mir, M., Reimer, A., Haines, J.E., Li, X.Y., Stadler, M., Garcia, H., Eisen, M.B., Darzacq, X., 2017. Dense bicoid hubs accentuate binding along the morphogen gradient. Genes Dev 31, 1784–1794. https://doi.org/10.1101/GAD.305078.117/-/DC1

Mir, M., Stadler, M.R., Ortiz, S.A., Hannon, C.E., Harrison, M.M., Darzacq, X., Eisen, M.B., 2018. Dynamic multifactor hubs interact transiently with sites of active transcription in drosophila embryos. Elife 7. https://doi.org/10.7554/ELIFE.40497

Mirny, L.A., 2010. Nucleosome-mediated cooperativity between transcription factors. Proc Natl Acad Sci U S A 107, 22534–22539. https://doi.org/10.1073/PNAS.0913805107/SUPPL_FILE/PNAS.0913805107_SI.PDF

Mosimann, C., Puller, A.C., Lawson, K.L., Tschopp, P., Amsterdam, A., Zon, L.I., 2013. Site-directed zebrafish transgenesis into single landing sites with the phiC31 integrase system. Dev Dyn 242, 949–963. https://doi.org/10.1002/DVDY.23989

Newport, J., 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. https://doi.org/10.1016/0092-8674(82)90272-0

Newport, J., Kirschner, M., 1982b. A major developmental transition in early xenopus embryos: II. control of the onset of transcription. Cell 30, 687–696. https://doi.org/10.1016/0092-8674(82)90273-2

Nguyen, T., Costa, E.J., Deibert, T., Reyes, J., Keber, F.C., Tomschik, M., Stadlmeier, M., Gupta, M., Kumar, C.K., Cruz, E.R., Amodeo, A., Gatlin, J.C., Wühr, M., 2022. Differential nuclear import sets the timing of protein access to the embryonic genome. Nature Communications 2022 13:1 13, 1–16. https://doi.org/10.1038/s41467-022-33429-z

Nien, C.Y., Liang, H.L., Butcher, S., Sun, Y., Fu, S., Gocha, T., Kirov, N., Manak, J.R., Rushlow, C., 2011. Temporal Coordination of Gene Networks by Zelda in the Early Drosophila Embryo. PLoS Genet 7, e1002339. https://doi.org/10.1371/JOURNAL.PGEN.1002339

Nudelman, G., Frasca, A., Kent, B., Sadler, K.C., Sealfon, S.C., Walsh, M.J., Zaslavsky, E., 2018. High resolution annotation of zebrafish transcriptome using long-read sequencing. https://doi.org/10.1101/gr.223586.117

Ogiyama, Y., Schuettengruber, B., Papadopoulos, G.L., Chang, J.M., Cavalli, G., 2018. Polycomb-Dependent Chromatin Looping Contributes to Gene Silencing during Drosophila Development. Mol Cell 71, 73-88.e5. https://doi.org/10.1016/J.MOLCEL.2018.05.032

Osborne, C.S., Chakalova, L., Brown, K.E., Carter, D., Horton, A., Debrand, E., Goyenechea, B., Mitchell, J.A., Lopes, S., Reik, W., Fraser, P., 2004. Active genes dynamically colocalize to shared sites of ongoing transcription. Nature Genetics 2004 36:10 36, 1065–1071. https://doi.org/10.1038/ng1423

Owens, N.D.L., Blitz, I.L., Lane, M.A., Patrushev, I., Overton, J.D., Gilchrist, M.J., Cho, K.W.Y., Khokha, M.K., 2016. Measuring Absolute RNA Copy Numbers at High Temporal Resolution Reveals Transcriptome Kinetics in Development. Cell Rep 14, 632–647. https://doi.org/10.1016/J.CELREP.2015.12.050

Pagans, S., Ortiz-Lombardía, M., Espinás, M.L., Bernués, J., Azorín, F., 2002. The Drosophila transcription factor tramtrack (TTK) interacts with Trithorax-like (GAGA) and represses GAGA-mediated activation. Nucleic Acids Res 30, 4406–4413. https://doi.org/10.1093/NAR/GKF570

Pálfy, M., Schulze, G., Valen, E., Vastenhouw, N.L., 2020. Chromatin accessibility established by Pou5f3, Sox19b and Nanog primes genes for activity during zebrafish genome activation. PLoS Genet 16, e1008546. https://doi.org/10.1371/JOURNAL.PGEN.1008546

Paraiso, K.D., Blitz, I.L., Coley, M., Cheung, J., Sudou, N., Taira, M., Cho, K.W.Y., 2019. Endodermal Maternal Transcription Factors Establish Super-Enhancers during Zygotic Genome Activation. Cell Rep 27, 2962-2977.e5. https://doi.org/10.1016/J.CELREP.2019.05.013

Porrua, O., Libri, D., 2015. Transcription termination and the control of the transcriptome: why, where and how to stop. Nature Reviews Molecular Cell Biology 2015 16:3 16, 190–202. https://doi.org/10.1038/nrm3943

Pownall, M.E., Miao, L., Vejnar, C.E., M'Saad, O., Sherrard, A., Frederick, M.A., Benitez, M.D.J., Boswell, C.W., Zaret, K.S., Bewersdorf, J., Giraldez, A.J., 2023. Chromatin expansion microscopy reveals nanoscale organization of transcription and chromatin. Science 381, 92. https://doi.org/10.1126/SCIENCE.ADE5308

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

Pritchard, D.K., Schubiger, G., 1996. Activation of transcription in Drosophila embryos is a gradual process mediated by the nucleocytoplasmic ratio. Genes Dev 10, 1131–1142. https://doi.org/10.1101/GAD.10.9.1131

Rabinowitz, M., 1941. Studies on the cytology and early embryology of the egg of Drosophila melanogaster. J Morphol 69, 1–49. https://doi.org/10.1002/JMOR.1050690102

Rieder, L.E., Koreski, K.P., Boltz, K.A., Kuzu, G., Urban, J.A., Bowman, S.K., Zeidman, A., Jordan, W.T., Tolstorukov, M.Y., Marzluff, W.F., Duronio, R.J., Larschan, E.N., 2017a. Histone locus regulation by the Drosophila dosage compensation adaptor protein CLAMP. Genes Dev 31, 1494–1508. https://doi.org/10.1101/GAD.300855.117/-/DC1

Rieder, L.E., Koreski, K.P., Boltz, K.A., Kuzu, G., Urban, J.A., Bowman, S.K., Zeidman, A., Jordan, W.T., Tolstorukov, M.Y., Marzluff, W.F., Duronio, R.J., Larschan, E.N., 2017b. Histone locus regulation by the Drosophila dosage compensation adaptor protein CLAMP. Genes Dev 31, 1494–1508. https://doi.org/10.1101/GAD.300855.117/-/DC1

Riesle, A.J., Gao, M., Rosenblatt, M., Hermes, J., Hass, H., Gebhard, A., Veil, M., Grüning, B., Timmer, J., Onichtchouk, D., 2023. Activator-blocker model of transcriptional regulation by pioneer-like factors. Nature Communications 2023 14:1 14, 1–18. https://doi.org/10.1038/s41467-023-41507-z

Rowley, M.J., Corces, V.G., 2018. Organizational principles of 3D genome architecture. Nat Rev Genet 19, 789–800. https://doi.org/10.1038/S41576-018-0060-8

Saad, H., Gallardo, F., Dalvai, M., Tanguy-le-Gac, N., Lane, D., Bystricky, K., 2014. DNA Dynamics during Early Double-Strand Break Processing Revealed by Non-Intrusive Imaging of Living Cells. PLoS Genet 10, e1004187. https://doi.org/10.1371/JOURNAL.PGEN.1004187

Sabari, B.R., Dall'Agnese, A., Boija, A., Klein, I.A., Coffey, E.L., Shrinivas, K., Abraham, B.J., Hannett, N.M., Zamudio, A. V., Manteiga, J.C., Li, C.H., Guo, Y.E., Day, D.S., Schuijers, J., Vasile, E., Malik, S., Hnisz, D., Lee, T.I., Cisse, I.I., Roeder, R.G., Sharp, P.A., Chakraborty, A.K., Young, R.A., 2018. Coactivator condensation at super-enhancers links phase separation and gene control. Science (1979) 361. https://doi.org/10.1126/SCIENCE.AAR3958/SUPPL_FILE/AAR3958_SABARI_SM_TABLE_S3.XLSX

Sagata, N., Shiokawa, K., Yamana, K., 1980. A study on the steady-state population of poly(A)+RNA during early development of Xenopus laevis. Dev Biol 77, 431–448. https://doi.org/10.1016/0012-1606(80)90486-8

Salzler, H.R., Tatomer, D.C., Malek, P.Y., McDaniel, S.L., Orlando, A.N., Marzluff, W.F., Duronio, R.J., 2013. A sequence in the Drosophila H3-H4 promoter triggers 1 Histone Locus Body assembly and biosynthesis of replication-coupled histone mRNAs. Dev Cell 24, 623. https://doi.org/10.1016/J.DEVCEL.2013.02.014

Sanchez, A., Cattoni, D.I., Walter, J.C., Rech, J., Parmeggiani, A., Nollmann, M., Bouet, J.Y., 2015. Stochastic Self-Assembly of ParB Proteins Builds the Bacterial DNA Segregation Apparatus. Cell Syst 1, 163–173. https://doi.org/10.1016/j.cels.2015.07.013

Sandler, J.E., Stathopoulos, A., 2016. Quantitative single-embryo profile of drosophila genome activation and the dorsal–ventral patterning network. Genetics 202, 1575–1584. https://doi.org/10.1534/GENETICS.116.186783/-/DC1

Sato, Y., Hilbert, L., Oda, H., Wan, Y., Heddleston, J.M., Chew, T.L., Zaburdaev, V., Keller, P., Lionnet, T., Vastenhouw, N., Kimura, H., 2019. Histone H3K27 acetylation precedes active transcription during zebrafish zygotic genome activation as revealed by live-cell analysis. Development (Cambridge) 146. https://doi.org/10.1242/DEV.179127/VIDEO-11

Shen, W., Gong, B., Xing, C., Zhang, L., Sun, J., Chen, Y., Yang, C., Yan, L., Chen, L., Yao, L., Li, G., Deng, H., Wu, X., Meng, A., 2022. Comprehensive maturity of nuclear pore complexes regulates zygotic genome activation. Cell 185, 4954-4970.e20. https://doi.org/10.1016/J.CELL.2022.11.011

Shrinivas, K., Sabari, B.R., Coffey, E.L., Klein, I.A., Boija, A., Zamudio, A. V., Schuijers, J., Hannett, N.M., Sharp, P.A., Young, R.A., Chakraborty, A.K., 2019a. Enhancer Features that Drive Formation of Transcriptional Condensates. Mol Cell 75, 549-561.e7. https://doi.org/10.1016/J.MOLCEL.2019.07.009

Shrinivas, K., Sabari, B.R., Coffey, E.L., Klein, I.A., Boija, A., Zamudio, A. V., Schuijers, J., Hannett, N.M., Sharp, P.A., Young, R.A., Chakraborty, A.K., 2019b. Enhancer Features that Drive Formation of Transcriptional Condensates. Mol Cell 75, 549-561.e7. https://doi.org/10.1016/J.MOLCEL.2019.07.009

Singh, G., Mullany, S., Moorthy, S.D., Zhang, R., Mehdi, T., Tian, R., Duncan, A.G., Moses, A.M., Mitchell, J.A., 2021. A flexible repertoire of transcription factor binding sites and a diversity threshold determines enhancer activity in embryonic stem cells. Genome Res 31, 564–575. https://doi.org/10.1101/GR.272468.120/-/DC1

Skirkanich, J., Luxardi, G., Yang, J., Kodjabachian, L., Klein, P.S., 2011. An essential role for transcription before the MBT in Xenopus laevis. Dev Biol 357, 478–491. https://doi.org/10.1016/J.YDBIO.2011.06.010

Sönmezer, C., Kleinendorst, R., Imanci, D., Barzaghi, G., Villacorta, L., Schübeler, D., Benes, V., Molina, N., Krebs, A.R., 2021. Molecular Co-occupancy Identifies Transcription Factor Binding Cooperativity In Vivo. Mol Cell 81, 255-267.e6. https://doi.org/10.1016/J.MOLCEL.2020.11.015

Sonoda, J., Wharton, R.P., 2001. Drosophila Brain Tumor is a translational repressor. Genes Dev 15, 762–773. https://doi.org/10.1101/GAD.870801

Soutourina, J., 2017. Transcription regulation by the Mediator complex. Nature Reviews Molecular Cell Biology 2017 19:4 19, 262–274. https://doi.org/10.1038/nrm.2017.115

Spitz, F., Furlong, E.E.M., 2012. Transcription factors: from enhancer binding to developmental control. Nature Reviews Genetics 2012 13:9 13, 613–626. https://doi.org/10.1038/nrg3207

Stapel, L.C., Zechner, C., Vastenhouw, N.L., 2017. Uniform gene expression in embryos is achieved by temporal averaging of transcription noise. Genes Dev 31, 1635–1640. https://doi.org/10.1101/GAD.302935.117/-/DC1

Sun, Y., Nien, C.Y., Chen, K., Liu, H.Y., Johnston, J., Zeitlinger, J., Rushlow, C., 2015. Zelda overcomes the high intrinsic nucleosome barrier at enhancers during Drosophila zygotic genome activation. Genome Res 25, 1703–1714. https://doi.org/10.1101/GR.192542.115

Syed, S., Wilky, H., Raimundo, J., Lim, B., Amodeo, A.A., 2021. The nuclear to cytoplasmic ratio directly regulates zygotic transcription in Drosophila through multiple modalities. Proc Natl Acad Sci U S A 118, e2010210118. https://doi.org/10.1073/PNAS.2010210118/SUPPL_FILE/PNAS.2010210118.SM03.MOV

Tadros, W., Goldman, A.L., Babak, T., Menzies, F., Vardy, L., Orr-Weaver, T., Hughes, T.R., Westwood, J.T., Smibert, C.A., Lipshitz, H.D., 2007. SMAUG Is a Major Regulator of Maternal mRNA Destabilization in Drosophila and Its Translation Is Activated by the PAN GU Kinase. Dev Cell 12, 143–155. https://doi.org/10.1016/J.DEVCEL.2006.10.005

Tani, S., Kusakabe, R., Naruse, K., Sakamoto, H., Inoue, K., 2010. Genomic organization and embryonic expression of miR-430 in medaka (Oryzias latipes): Insights into the post-transcriptional gene regulation in early development. Gene 449, 41–49. https://doi.org/10.1016/J.GENE.2009.09.005

ten Bosch, J.R., Benavides, J.A., Cline, T.W., 2006. The TAGteam DNA motif controls the timing of Drosophilapre-blastoderm transcription. Development 133, 1967–1977. https://doi.org/10.1242/DEV.02373

Thatcher, E.J., Bond, J., Paydar, I., Patton, J.G., 2008. Genomic organization of zebrafish microRNAs. BMC Genomics 9, 1–9. https://doi.org/10.1186/1471-2164-9-253/TABLES/4

Thorvaldsdóttir, H., Robinson, J.T., Mesirov, J.P., 2013. Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. Brief Bioinform 14, 178–192. https://doi.org/10.1093/BIB/BBS017

Treen, N., Chavarria, E., Weaver, C.J., Brangwynne, C.P., Levine, M., 2023. An FGF timer for zygotic genome activation. Genes Dev 37, 80–85. https://doi.org/10.1101/GAD.350164.122

Ugolini, M., Kuznetsova, K., Oda, H., Kimura, H., Vastenhouw, N.L., 2023. Transcription bodies regulate gene expression by sequestering CDK9. bioRxiv 2022.11.21.517317. https://doi.org/10.1101/2022.11.21.517317

Vastenhouw, N.L., Cao, W.X., Lipshitz, H.D., 2019. The maternal-to-zygotic transition revisited. https://doi.org/10.1242/dev.161471

Vastenhouw, N.L., Zhang, Y., Woods, I.G., Imam, F., Regev, A., Liu, X.S., Rinn, J., Schier, A.F., 2010. Chromatin signature of embryonic pluripotency is established during genome activation. Nature 2010 464:7290 464, 922–926. https://doi.org/10.1038/nature08866

Veenstra, G.J.C., Destrée, O.H.J., Wolffe, A.P., 1999a. 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. https://doi.org/10.1128/MCB.19.12.7972

Veenstra, G.J.C., Destrée, O.H.J., Wolffe, A.P., 1999b. 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. https://doi.org/10.1128/MCB.19.12.7972

Veil, M., Yampolsky, L.Y., Grüning, B., Onichtchouk, D., 2019. Pou5f3, SoxB1, and Nanog remodel chromatin on high nucleosome affinity regions at zygotic genome activation. Genome Res 29, 383–395. https://doi.org/10.1101/GR.240572.118/-/DC1

Wharton, R.P., Struhl, G., 1991. RNA regulatory elements mediate control of Drosophila body pattern by the posterior morphogen nanos. Cell 67, 955–967. https://doi.org/10.1016/0092-8674(91)90368-9

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., Füllgrabe, A., Davis, M.P., Enright, A.J., Busch-Nentwich, E.M., 2017. A high-resolution mRNA expression time course of embryonic development in zebrafish. Elife 6. https://doi.org/10.7554/ELIFE.30860

Wike, C.L., Guo, Y., Tan, M., Nakamura, R., Shaw, D.K., Díaz, N., Whittaker-Tademy, A.F., Durand, N.C., Aiden, E.L., Vaquerizas, J.M., Grunwald, D., Takeda, H., Cairns, B.R., 2021a. Chromatin architecture transitions from zebrafish sperm through early embryogenesis. Genome Res 31, 981–994. https://doi.org/10.1101/GR.269860.120/-/DC1

Wike, C.L., Guo, Y., Tan, M., Nakamura, R., Shaw, D.K., Díaz, N., Whittaker-Tademy, A.F., Durand, N.C., Aiden, E.L., Vaquerizas, J.M., Grunwald, D., Takeda, H., Cairns, B.R., 2021b. Chromatin architecture transitions from zebrafish sperm through early embryogenesis. Genome Res 31, 981–994. https://doi.org/10.1101/GR.269860.120/-/DC1

Wong, K.H., Jin, Y., Struhl, K., 2014. TFIIH phosphorylation of the Pol II CTD stimulates mediator dissociation from the preinitiation complex and promoter escape. Mol Cell 54, 601–612. https://doi.org/10.1016/J.MOLCEL.2014.03.024

Xu, C., Fan, Z.P., Müller, P., Fogley, R., DiBiase, A., Trompouki, E., Unternaehrer, J., Xiong, F., Torregroza, I., Evans, T., Megason, S.G., Daley, G.Q., Schier, A.F., Young, R.A., Zon, L.I., 2012. Nanog-like Regulates Endoderm Formation through the Mxtx2-Nodal Pathway. Dev Cell 22, 625. https://doi.org/10.1016/J.DEVCEL.2012.01.003

Xu, M., Cook, P.R., 2008. Similar active genes cluster in specialized transcription factories. Journal of Cell Biology 181, 615–623. https://doi.org/10.1083/JCB.200710053

Yamada, S., Whitney, P.H., Huang, S.K., Eck, E.C., Garcia, H.G., Rushlow, C.A., 2019. The Drosophila Pioneer Factor Zelda Modulates the Nuclear Microenvironment of a Dorsal Target Enhancer to Potentiate Transcriptional Output. Current Biology 29, 1387-1393.e5. https://doi.org/10.1016/J.CUB.2019.03.019

Yáñez-Cuna, J.O., Dinh, H.Q., Kvon, E.Z., Shlyueva, D., Stark, A., 2012. Uncovering cis-regulatory sequence requirements for context-specific transcription factor binding. Genome Res 22, 2018–2030. https://doi.org/10.1101/GR.132811.111

Zalokar, M., 1976. Autoradiographic study of protein and RNA formation during early development of Drosophila eggs. Dev Biol 49, 425–437. https://doi.org/10.1016/0012-1606(76)90185-8

Zaret, K.S., 2020. Pioneer Transcription Factors Initiating Gene Network Changes. Annu Rev Genet 54, 367. https://doi.org/10.1146/ANNUREV-GENET-030220-015007

Zhang, B., Wu, X., Zhang, W., Shen, W., Sun, Q., Liu, K., Zhang, Y., Wang, Q., Li, Y., Meng, A., Xie, W., 2018. Widespread Enhancer Dememorization and Promoter Priming during Parental-to-Zygotic Transition. Mol Cell 72, 673-686.e6. https://doi.org/10.1016/j.molcel.2018.10.017

Zhang, M., Kothari, P., Mullins, M., Lampson, M.A., 2014. Regulation of zygotic genome activation and DNA damage checkpoint acquisition at the mid-blastula transition. Cell Cycle 13, 3828–3838. https://doi.org/10.4161/15384101.2014.967066

Zhang, Y., Vastenhouw, N.L., Feng, J., Fu, K., Wang, C., Ge, Y., Pauli, A., Van Hummelen, P., Schier, A.F., Liu, X.S., 2014. Canonical nucleosome organization at promoters forms during genome activation. Genome Res 24, 260. https://doi.org/10.1101/GR.157750.113

Zhou, L., Feng, T., Xu, S., Gao, F., Lam, T.T., Wang, Q., Wu, T., Huang, H., Zhan, L., Li, L., Guan, Y., Dai, Z., Yu, G., 2022. ggmsa: a visual exploration tool for multiple sequence alignment and associated data. Brief Bioinform 23. https://doi.org/10.1093/BIB/BBAC222