Characterising pre-MBT nucleus and transcriptional dynamics in the developing zebrafish embryo

by

Haseeb Khalid Qureshi

A thesis submitted to

The University of Birmingham

for the degree of

Doctor of Philosophy

April 2021

UNIVERSITY OF
BIRMINGHAM

University of Birmingham Research Archive

e-theses repository

This unpublished thesis/dissertation is copyright of the author and/or third parties. The intellectual property rights of the author or third parties in respect of this work are as defined by The Copyright Designs and Patents Act 1988 or as modified by any successor legislation.

Any use made of information contained in this thesis/dissertation must be in accordance with that legislation and must be properly acknowledged. Further distribution or reproduction in any format is prohibited without the permission of the copyright holder.

Contents

Acknowledgments **8**

Abbreviations **9**

1 Abstract **10**

2 General Introduction **11**

2.1 Preamble . . . 11

2.2 Setting the scene: The transcriptionally silent environment of the early embryo . . . 12

2.2.1 Maternal zygotic transition and Mid blastula transition . . . 12

2.2.2 The early embryo is in a state of transcriptional repression . . . 13

2.2.3 Zygotic genome activation describes the release of transcriptional repression during MZT . . . 14

2.2.4 The MZT initiates a minor, early wave of transcription . . . 15

2.2.5 Initiation of transcription in the context of genome activation . . . 16

2.2.6 Phase separated bodies compartmentalise nucleus functions . . . 16

2.3 Single cell to blastula: Cell divisions are a focal point in understanding the maturation of the embryo . . . 18

2.3.1 Pre-MBT cell cycles are rapid . . . 18

2.3.2 The elongation of the cell cycle and gap phases is a major outcome of MBT . . . 19

2.3.3 Chromatin conformation changes with cell cycle phases . . . 20

2.3.4 Rapid cell cycles present challenges to the cooordination of transcription and DNA synthesis . . . 21

2.3.5 Inherited variations in dividing cells: Epigenetic inheritance . . . 21

2.4 How to investigate development in real time: An overview of microscopic methods and their considerations . . . 22

2.4.1 Light scattering . . . 22

2.4.2 Light sheet fluorescence microscopy . . . 23

2.4.3 Airyscan confocal microscopy . . . 24

2.4.4 Data Handling: Current methods and issues . . . 26

2.5 *Danio rerio*: a fine laboratory model for studying developmental phenomena . . . 27

2.5.1 Zebrafish model . . . 27

2.5.2 Zebrafish development . . . 27
2.6 Relevance to this thesis . . . 28

3 Methods **29**
3.1 Zebrafish . . . 29
3.1.1 Husbandry . . . 29
3.1.2 Protein preparation . . . 29
3.1.3 Injections . . . 30
3.1.4 Transcription block . . . 30
3.1.5 Cell cycle elongation . . . 30
3.1.6 Generation of tetraploid embryos . . . 30
3.1.7 Lines . . . 30
3.2 Imaging . . . 31
3.2.1 Light sheet . . . 31
3.2.2 Airyscan . . . 31
3.3 Statistical analysis . . . 31
3.4 Software . . . 31
3.4.1 2D and 3D processing . . . 31
3.4.2 Software Development . . . 32
3.4.3 Scripts . . . 33

4 *In vivo* detection and characterisation of transcription and cell cycle during the early transcriptional wave **34**
4.1 Introduction . . . 34
4.1.1 The Midblastula transition in the zebrafish embryo . . . 34
4.1.2 Early wave of transcription . . . 35
4.1.3 *In vivo* imaging of transcription . . . 35
4.1.4 Transformation of 4D lightsheet data . . . 36
4.1.5 How are cell cycle and transcription organised during the early wave of genomic activation, prior to MBT? . . . 37
4.1.6 General aims . . . 37
4.2 Results . . . 38
4.2.1 Establishing the intensity topography and implementing manual tracking of nuclei . . . 38
4.2.2 Semi-manual tracking using Icy Bioimage spot detection and custom scripts . . . 41

4.2.3 Developing a 3D tracking workflow . . . 43
4.2.4 Initial observations: Transcription of *miR430* manifests as dual hyper intense foci . . . 45
4.2.5 Transcription expression is switched on asynchronously across the embryo, over multiple embryonic stages . . . 47
4.2.6 Characterisation of stereotypical features of cell cycle and transcription . . . 48
4.2.7 Characterising variations in transcription persistence and cell cycle duration . . . 51
4.2.8 Cell cycle length is not significantly affected by histone and morpholino injection . . . 53
4.2.9 Cell cycle and transcription progressively elongate with successive embryonic stages 55
4.2.10 Transcription focus persistence and cell cycle length are correlated . . . 56
4.2.11 Characterising the size of the transcription focus . . . 58
4.2.12 The transcription focus occupies a histone deficient region within the S-phase nucleus 59
4.2.13 Multiple early wave transcripts are identified in transcription foci . . . 61
4.2.14 The number of transcription foci is increased in tetraploid embryos . . . 62
4.2.15 Early expressing nuclei are more likely to produce transcription expressing nuclei . . . 66
4.2.16 Cells with longer cell cycle are more likely to produce daughters with longer cell cycles 68
4.3 Discussion . . . 69
4.3.1 Transcription of select early wave genes manifests as chromosomally associated foci 69
4.3.2 Transcription foci appear in a metasynchronous manner across the pre-MBT embryo 69
4.3.3 The transcription focus bears the hallmarks of being a large phase separated body . . . 71
4.3.4 Transcriptional states may be inherited . . . 71
4.3.5 Cell cycle inheritance may influence transcription . . . 72
4.3.6 Technical considerations and next steps . . . 73

5 A lightweight segmentation and lineage tracking tool for noisy, low frame rate 4D fluorescence microscopy data 75

5.1 Introduction . . . 75
5.1.1 The compromise of light scattering in *in vivo* imaging . . . 75
5.1.2 Multidimensional imaging data is computationally unwieldy . . . 75
5.1.3 Watershed algorithm . . . 76
5.1.4 Feature tracking and aims . . . 76
5.2 Results . . . 78
5.2.1 Software architecture, displaying how separate functional modules integrate in the workflow . . . 78

5.2.2 Graphical User Interface design . . . 79
5.2.3 One-dimensional peak detection locates area of interest . . . 80
5.2.4 Isolating intensity peaks in 1 dimension forms the basis of feature detection . . . 82
5.2.5 Two-Dimensional feature analysis allows for creating watershed seed candidates . . . 84
5.2.6 A user defined exclusion zone sets up watershed basin . . . 84
5.2.7 Local environment sensing on detected watersheds removes unwanted features . . . 86
5.2.8 3D segmentation reveals the expected number of features at specific embryonic stages 87
5.2.9 Nuclei are tracked using KD Trees and a user defined detection radius . . . 89
5.2.10 Lineages can be visualised in 3D . . . 91
5.2.11 Detected features are validated by comparing detected coordinates with the raw image 92
5.2.12 Resource utilisation . . . 93
5.3 Discussion . . . 94
5.3.1 Overview of software design . . . 95
5.3.2 The segmentation and tracking tool is a dedicated, lightweight solution to segmenting and tracking 3D datasets . . . 95
5.3.3 Advantages and limitations . . . 96

6 Investigating heritable cell cycle features within pre-MBT embryonic cell lineages 98

6.1 Introduction . . . 98
6.1.1 Epigenetic memory . . . 98
6.1.2 Coordination of mitosis with cell cycle . . . 99
6.1.3 Nucleocytoplasmic ratio . . . 99
6.1.4 Direction and aims . . . 100
6.2 Results . . . 101
6.2.1 Parameter selection for tracking algorithm . . . 101
6.2.2 Parent daughter cell cycle lengths show correlation in preMBT embryonic stages, within a single embryo . . . 102
6.2.3 Correlations lost between parent daughter cell cycle lengths in preMBT embryonic stages, across multiple embryos . . . 103
6.2.4 Timing of cell division, and cell cycle length are correlated with position in the embryo 105
6.2.5 Mitosis times are staggered across an axis in the embryo . . . 106
6.2.6 Can we discern a role for nucleocytoplasmic ratio in the spatial aspects of memory? . 108
6.2.7 Nucleus diameter correlates with position in embryo . . . 111

6.3 Discussion 113
6.3.1 Cell cycle memory 113
6.3.2 Mitosis wave 114
6.3.3 Nucleocytoplasmic ratio 115

7 Summary and conclusions **117**

8 Appendix **119**

9 References **124**

List of Figures

2.1 Genome activation across species . . . 13
2.2 Diagram of transcription factory . . . 17
2.3 Phases within interphase . . . 18
2.4 The development of gap phases . . . 19
2.5 Images of cell cycle phases in the zebrafish embryo . . . 20
2.6 Zeiss Z.1 light sheet microscope . . . 24
2.7 Airyscan hexagonal sensor array . . . 25
2.8 Cleavage and blastula stage of zebrafish embryo . . . 28
4.1 Milestones in zebrafish development . . . 34
4.2 Activation of zebrafish genome . . . 34
4.3 Schematic of transcription imaging morpholino . . . 36
4.4 Establishing the intensity topography and implementing manual tracking of nuclei . . . 39
4.5 Semi-manual tracking using Icy Bioimage spot detection and custom scripts . . . 42
4.6 Developing a 3D tracking workflow . . . 44
4.7 Transcription of miR430 manifests as dual hyper intense foci . . . 46
4.8 Transcription expression is switched on asynchronously across the embryo, over multiple embryonic stages . . . 48
4.9 Characterisation of stereotypical features of cell cycle and transcription . . . 49
4.10 Characterising variations in transcription persistence and cell cycle duration . . . 51
4.11 Cell cycle length is not significantly affected by histone and MO injection . . . 54
4.12 Cell cycle and transcription elongate with successive embryonic stages . . . 55
4.13 Transcription focus persistence and cell cycle length are correlated . . . 57
4.14 Characterising the size of the transcription focus . . . 58
4.15 The transcription focus occupies a histone deficient region within the S-phase nucleus . . . 60
4.16 Multiple early wave transcripts are identified in transcription foci . . . 62
4.17 The number of transcription foci is increased in tetraploid embryos . . . 63
4.18 Early expressing nuclei are more likely to produce transcription expressing nuclei . . . 67
4.19 Cells with longer cell cycle are more likely to produce daughters with longer cell cycles . . . 68
5.1 Overview of software architecture of tool . . . 79
5.2 GUI design . . . 79
5.3 1 dimensional peak detection locates area of interest . . . 81

5.4 Isolating intensity peaks in 1 dimension forms the basis of feature detection 83
5.5 2 Dimensional feature analysis allows for creating watershed seed candidates 85
5.6 A user defined exclusion zone sets up watershed basin 86
5.7 Local environment sensing on detected watersheds removes unwanted features 86
5.8 3D segmentation process of nuclei 88
5.9 Tracking segmented nuclei: 89
5.10 Lineages can be visualised in 3D 91
5.11 Detected features are validated by comparing detected coordinates with the raw image . . . 92
5.12 Performance benchmarks for 3 datasets 94
6.1 Parameter selection for tracking algorithm 102
6.2 Are variations in cell cycle length correlated across generations? 103
6.3 Parent daughter cell cycle lengths show correlation at in preMBT embryonic stages, within a single embryo 104
6.4 Multi-Dimensional views of mitosis in the embryo 106
6.5 Illustration of plotting mitosis locations from a time lapse showing multiple mitosis events . . 106
6.6 Mitosis times are staggered across a particular axis in the embryo 108
6.7 Nucleus diameter declines with successive cell cycles 110
6.8 Nucleus diameter correlates with position in embryo 112
6.9 Map of cell family clusters 113
8.1 Appendix 1a 120
8.2 Appendix 1b 121
8.3 Appendix 2 122
8.4 Appendix 3 123

Acknowledgements

The last five years (MRes/PhD) as part of the Mueller lab have been both challenging and enjoyable. I am grateful for the support,guidance and mentorship of both of my supervisors over this five year period, Ferenc Mueller and Attila Sik. I feel like I have learned much about how to inquire about the universe, and have grown as a scientist.

My colleagues and friends from the Mueller lab, Ada, Lucy, Ryan, Fabio, Kacper have been an invaluable source of support over the years through their friendship, patience, knowledge and advice (sorry Kacper and Yavor for my incessant requests for MORE data).

And my parents, for believing in me.

I could not have done this alone.

Thank you.

Abbreviations

- 2D - Two Dimensional
- 3D - Three Dimensional
- ACM - Airyscan Confocal Microscopy
- CCD - Charged-Coupled Device
- CLSM - Confocal Laser Scanning Microscopy
- CTD - C Terminus Domain
- GUI - Graphical User Interface
- klf17 - Krüppel Like Factor 17
- LSFM - Light Sheet Fluorescence Microscopy
- MBT - Midblastula Transition
- MIP - Maximum Intensity Projection
- miR430 - MicroRNA-430
- MO - Morpholino
- MOVIE - Morpholino Visualisation of Expression
- MZT - Maternal Zygotic Transition
- NCR - Nucleocytoplasmic Ratio
- PCNA - Proliferating Nuclear Antigen
- PCR - Polymerase Chain Reaction
- PIC - Preinitiation Complex
- pol ii - RNA Polymerase II
- PSF - Point Spread Function
- RBP - RNA Binding Protein
- sCMOS - Scientific Complementary Metal-Oxide-Semiconductor
- TBP - TATA Binding Protein
- TF - Transcription Factor
- TxnF - Transcription Factory
- YSL - Yolk Syncytial Layer
- ZGA - Zygotic Genome Activation

1 Abstract

The early embryo begins life in a transcriptionally silent state, and undergoes a process of genomic activation as it is reprogrammed from an egg cell into a mass of stem cells. In the zebrafish, this process of genome awakening, the zygotic genome activation, occurs over several embryonic stages and consists of rapid cell cycles which exclusively switch between S-phase and mitosis. This raises the question of how genome awakening is able to occur concurrently with a genome occupied by DNA replication events. We investigated how transcription is coordinated with the cell cycle. Through developing image processing and analysis tools, I was able to parse complex 4D microscopy datasets to characterise the progressive loss of mitotic synchrony and increase in cell cycle lengths in the phases prior to mid blastula transition (MBT). 3D visualisation and analysis of high resolution imaging of microRNA-430 morpholino visualised transcription revealed large localised transcription compartments within nuclei, termed here as 'foci', which contain transcripts from multiple genes. The presence and duration of transcription occurs a coordinated manner, showing heritable relationships between parent and daughter cells, and the cell cycle duration showed a similar relationship. This provoked a question regarding how cell cycle influences transcription as a heritable feature. These analyses also revealed the limitations of conventional image processing methods, so I designed a 4D microscopy image analysis tool which utilised a novel iteration of the watershed algorithm to segment nuclei in highly noisy deep tissue environments. This tool was also able to trace nuclei over generations to provide lineages, with data about cell cycles and nuclear geometries. Based off this, I was able to make early characterisations of potential cell cycle memory effects between parent and daughter cells in pre-MBT cell cycles. I also analysed these variations in cell cycle lengths in the context of nucleocytoplasmic ratio, and position in the embryo, shedding light on how cell cycle variation and asynchrony can influence a mitotic wave, which spreads across the embryo.

2 General Introduction

2.1 Preamble

Transcription and cell division are processes fundamental to the existence of life, thus their regulation and coordination are core paradigms to survival and growth. The coordination of these two processes in time and space defines the orderly transition from fertilised egg to fully grown organism.

The early developmental stages of the zebrafish embryo represent a versatile model for studying systems which undergo rapid growth and differentiation, as well as elucidating the primordial steps of embryonic organisation. This embryonic environment offers a sandbox for probing the coordination of genetic events in a rapidly dividing system, and their implications and consequences for downstream developmental states.

The zebrafish model is useful for *in vivo* imaging studies due to the accessibility to numerous transparent eggs per clutch, relative accessibility for genetical manipulation to produce bespoke strains, and resilience of embryos to the handling involved in preparing live imaging samples.

The rise of light sheet fluorescence microscopy (LSFM) has presented exciting opportunities for *in vivo* biologists interested in the inner mechanics of such specimens in real time. This technology has made it possible to conduct long term imaging in 3 dimensions, giving additional insights into embryogenesis, tumorigenesis and neural function. The additional depth of data offered by such imaging modalities provides the potential for discerning organism level implications of cellular function, but also provides challenges in how questions are asked and how data is handled.

This thesis aims to shed light on spatiotemporal aspects of the organisation of genomic activation across the embryo; specifically addressing the overarching biological questions of:

- Where/when does the early wave of transcription manifest across the embryo, and what it is its relationship with cell cycle regulation
- Are cell cycle features inherited within pre-MBT embryonic cell lineages?

This work follows the handling of LSFM data in trying to understand the timing and manifestation of genomic activation in the early zebrafish embryo and addressing the subsequent technical challenges and biological questions spawned by this inquiry.

2.2 Setting the scene: The transcriptionally silent environment of the early embryo

The role of transcription is a pivotal concept to understand when investigating nuclear phenomena in early embryonic cells. The emergence of this fundamental life process is a hallmark of early embryonic development. The transitional environment of the very early embryo presents a unique biological state which helps inform when and where to look when trying to characterise the presence of transcriptional activation in the embryo, as well as informing experimental design when probing the onset of zygotic transcription in model systems.

2.2.1 Maternal zygotic transition and Mid blastula transition

The early embryo is transcriptionally silent with maternal proteins and mRNAs responsible for driving the majority of developmental processes; a phenomenon thought to be universal across the animal kingdom (Vastenhouw et al. 2019). This period of transcriptional silence marks the duration in which the differentiated egg and sperm cell are reprogrammed back into a totipotent cell. This totipotent cell will go on to divide into the myriad cells which will construct a new organism. These maternally proteins and mRNAs are maternally deposited in the egg pre-fertilisation and deprogram the fertilised egg by reverting it from a differentiated germ cell to stem cells and perform metabolic and developmental duties in the zygote, whilst also repressing zygotic transcription (K. N. Schulz and Harrison 2019). Very early development is thus primarily translationally regulated, with translation of maternal transcripts being coordinated to perform the metabolic and life support functions of the primordial embryo.

The maternal zygotic transition (MZT) is a coordinated period of development during which molecular and developmental processes are shifted from being driven by maternally deposited gene products to zygotic controlled processes (Vastenhouw et al. 2019). Activation of endogenous transcription gradually activates across the embryo and causing changes in the organisation and duration of the cell cycle (K. N. Schulz and Harrison 2019). In slow developing organisms (e.g. mammals: humans and mice), this process occurs until shortly after the first cell division following fertilisation, yet in fast developing organisms (e.g. poikilothermic egg laying organisms such as *Drosophila*, *Xenopus* and zebrafish) this process occurs over several synchronous cell divisions following fertilisation (Tadros and Lipshitz 2009). This is illustrated in figure 2.1.

The cell cycle lengths of slow developing animals are also much longer than those of fast developing animals, with the time until the first mitosis in mice occuring after 22hrs. In zebrafish, the first cell division occurs after 40mins, with subsequent cell cycles until MBT ranging between 15 and 30mins (figure 2.1). The embryos of slow developing animals such as mice and humans develop in the relative safety of their mothers' womb, and as such developmental processes may be delayed in favour of developing physiology

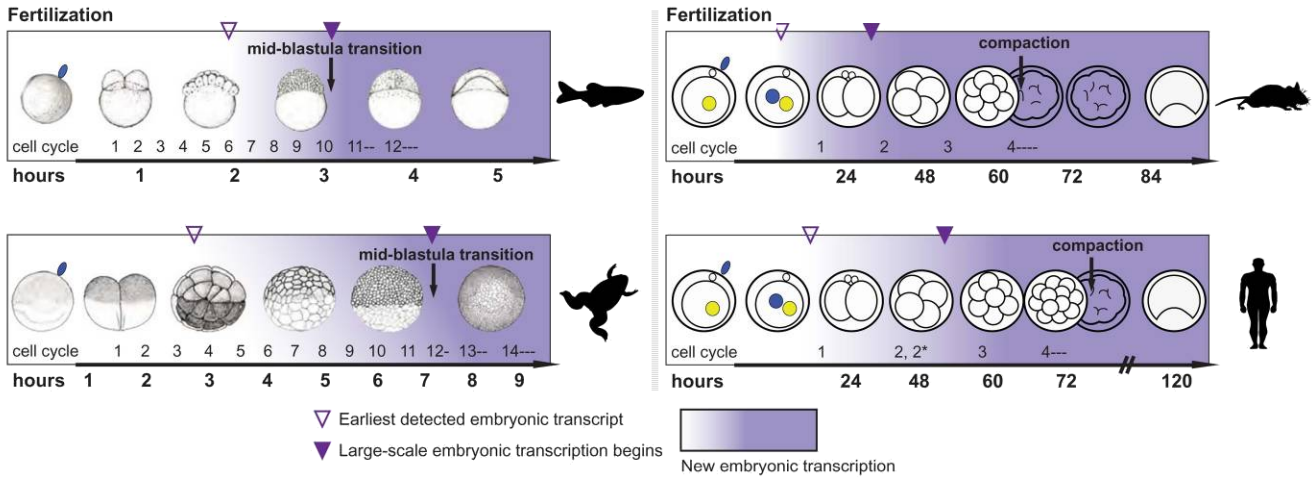


Figure 2.1: Organisation of genome activation in animals models representing fast developing animals (left) and slow developing animals (right). Clear triangular arrow denoting earliest detected transcripts represents the minor early wave of ZGA and the filled triangular arrow indicates the onset of the major wave of transcription. Diagram adapted from Jukam et al. 2017.

to utilise resources provided by the mother (O'Farrell et al. 2004). The embryos of zebrafish or *Xenopus* are within eggs, released into the wild for fertilisation. Development into larvae must be rapid to achieve motility and increase chances of survival (O'Farrell 2015, K. N. Schulz and Harrison 2019, Yuan et al. 2016).

The Mid Blastula Transition (MBT) is the point at which the body plan of the organism begins to organise (ibid.). Increased cell cycle lengths permits time for more transcription as well as time for cytoskeletal remodelling so cells may migrate and drive shape changes during epiboly and gastrulation (O'Farrell et al. 2004). MBT is a process absent from mammals, but is prevalent in fast developing animals (Yasuda 1992, Jukam et al. 2017). The MBT defines the point of global cell cycle desynchronisation and elongation (Tadros and Lipshitz 2009). Whereas the maternal-zygotic transition (MZT) and the mid blastula transition share overlap in their basic functions, they are distinct phases in embryonic development. Contrary to MZT, which encompasses a prolonged period of transition, the MBT describes a singular switching point between developmental stages, marking the onset of the major wave of zygotic genome activation (ZGA), and the introduction of fully formed gap phases in nuclei (Yuan et al. 2016, Tadros and Lipshitz 2009, K. N. Schulz and Harrison 2019).

2.2.2 The early embryo is in a state of transcriptional repression

The transcriptional silence of the early embryo is the result of genomic repression. The cellular environment is non conducive to active transcription, either by inhibition or by not providing the conditions for it to occur.

Maternally deposited histones are thought to be involved in transcriptional repression in the zygote.

The binding of histones to DNA may form heterochromatin that is inaccessible to transcription machinery. Additionally, unbound histones may repress signalling pathways involved in permitting transcription. The tail region of unbound histone 3 has been implicated in inhibiting the onset of gap phases, ordinarily during which transcription is supposed to occur (Shindo and Amodeo 2020). Unbound histone 3 levels have been shown to be elevated in the early embryo, prior to the conclusion of the MZT. Replication factors such as treslin, RecQ4, Drf1 and Cut5 also contribute to repression of transcription (C. Collart et al. 2013, Langley et al. 2014). Elevated levels of these factors drive increased levels of replication initiation events in the rapidly dividing cells of the preMBT embryo, allowing complete genome replication in the short cell cycles. In fast developing animals, the extremely short cell cycles also serve to repress transcription as there is not enough time to transcribe longer genes (C. Collart et al. 2013).

The lack of essential transcription factors in the very early embryo is another reason for genomic repression. A key component of transcription initiation, TATA binding protein (TBP), is scarce in the pre MBT embryo (Ferg et al. 2007, Jallow et al. 2004). Experiments expressing TBP prior to MBT in *Xenopus* show it has the capacity to bring the onset of MBT earlier (Veenstra et al. 1999).

2.2.3 Zygotic genome activation describes the release of transcriptional repression during MZT

The process of zygotic genome activation is not a discrete step, but encompasses multiple waves of gene activation across the zygote, beginning with a minor early wave (figure 2.1)(discussed in more detail in sections 2.2.4 and 4.1.2) of activation of a limited subset of genes, and then followed by one or more major waves of activation in which genes are activated on a larger scale (Clara Collart et al. 2014, Harvey et al. 2013). The timing and organisation of these waves varies between species, yet this multistep process of zygotic genome activation is conserved across animal species [Lee, Bonneau, and Giraldez 2014, K. N. Schulz and Harrison 2019, Hwang et al. 2018].

A leading model regarding how zygotic transcription globally activates posits that the concentrations of repressive maternal factors are titrated out by successive cell divisions. Termed Nucleocytoplasmic Ratio model (NCR), cytoplasmic volume reduces drastically with early cell divisions and the concentration of non-renewable maternally deposited factors drops concurrently (Newport and Kirschner 1982). Support for this model primarily comes from experiments involving ploidy manipulation. Halving the nucleus content of pre MBT cells in *Xenopus* and *Drosophila* by generating haploid embryos delayed MBT (Edgar et al. 1986, Jevtić and Levy 2017). The concentration of maternally deposited transcriptionally repressive histones decreases with cell divisions in the pre-MZT embryo. The titration of these proteins correlating with the onset of transcription further supports the NCR model. TBP has been shown to be involved in this process through its interaction with putatively repressive histones (Veenstra et al. 1999). The early expression of TBP

in preMBT *Xenopus* egg extract only hastens the onset of transcription if chromatin assembly is disrupted, suggesting that the titration of chromatin through successive divisions may be permissive for TBP to exert its effect (Veenstra et al. 1999). One potential candidate repressive histone is the aforementioned histone 3 tail, whose titration over successive divisions has been implicated in the progression of MZT (Shindo and Amodeo 2020).

Another model to explain the timing of the onset of MZT is the maternal clock model, in which a cascade of biochemical reactions initiated by fertilisation function as a timer for transcriptional activation. Once particular points in the cascade are reached, certain factors may be activated to begin clearance of maternal factors and allow the onset of transcriptional activation. The accumulation of transcriptional activators, and of mRNA degrading proteins, typically RNA binding proteins (RBP) is also pivotal in establishing the shift to zygotic control transcription. The RBP Smaug is translated from maternal RNAs and its main role during MZT is to bind to and target maternal RNAs for degradation (Benoit et al. 2009), which may reduce the translation of repressive maternal gene product. The TBP levels in the early *Xenopus* embryo evidences this model. TBP concentrations remain very low in the early embryo but translation of maternal TBP mRNA lead to accumulation up until a peak at MBT (Veenstra et al. 1999). At MBT, TBP can facilitate ZGA through enabling transcription initiation and aiding the clearance of maternal RNAs (Ferg et al. 2007).

Multiple models for the shift from maternal to zygotic may not be mutually exclusive occurrences, with separate lines of evidence supporting either model, and TBP plays a key role in this process.

Small non coding RNAs are also involved in targeting maternal factors for degradation, which include various microRNAs (miRNA), PIWI-interacting RNAs (piRNA) and small interfering RNAs (siRNA) (Pauli et al. 2011). miRNAs are well characterised in this regard, having been shown to facilitate deadenylation and repression of maternal transcripts (Giraldez 2010). There is further discussion of miRNA in section 4.1.2.

2.2.4 The MZT initiates a minor, early wave of transcription

A defining feature of the MZT is the early wave of transcription, which heralds the onset of zygotic genome activation (ZGA). This is the gradual activation of the genome in cells across the embryo, presenting as a low number of genes being expressed (Heyn et al. 2014, Tadros and Lipshitz 2009). These genes produce products which facilitate the transition from reliance on maternal gene products. In mice, where ZGA begins prior to the first cell division, many genes show a low level of activation although most of the resulting mRNA is non functional due to inefficient processing, representing another layer of maternal repression (Abe et al. 2015). In zebrafish, over 500 genes are active in the early wave. These genes are short, suggesting quick transcription times, seemingly limited by the repressive environment of the early embryo (Heyn et al. 2014) (further discussion in section 4.1.2). Across species, the early wave increases in gene production

and intensity until the level of maternal products has been reduced and the subsequent major wave of transcription may commence. Clearance of maternal gene product occurs in multiple stages beginning during the MZT.

2.2.5 Initiation of transcription in the context of genome activation

There are numerous transcription factors which have been implicated in ZGA. Chiefly, Zelda (Zld) has been shown to be instrumental in the activation of transcription in *Drosophila* embryos. Zld mRNA is maternally deposited in the *Drosophila* embryo, and has been shown to associate with open chromatin regions, to the extent that it is deemed a predictor of imminent transcription occurrences (Harrison et al. 2011). Zld is associated with chromatin accessibility to transcription machinery, and is important in advancing ZGA via pioneer factor like activities in enabling transcription (Katharine N Schulz et al. 2015).

Analogous to *Zld* in zebrafish are the transcription factors SoxB1, Nanog and Pou5f3. These transcription factors activate transcription of the earliest genes during MZT. These factors are maternally deposited and are highly translated prior to MZT where their activity comes into effect (Lee, Bonneau, Takacs, et al. 2013). Additionally, TBP, as already discussed, is important to ZGA and MBT in zebrafish and *Xenopus*. TBP, the levels of which rise up until MBT, forms part of the TFIID complex, which sets up the pre initiation complex to which RNA polymerase ii (pol ii) is recruited (Margottin et al. 1991, Klages and Strubin 1995, Veenstra et al. 1999).

Central to the production of RNA are the RNA polymerases. The main polymerase of interest during MZT is pol ii, which is responsible for transcription of most mRNAs, microRNAs and sRNAs. Pol ii elongates transcripts upon being recruited to the preinitiation complex (PIC) at the gene promoter (Petrenko et al. 2019). The PIC consists of transcription factors (TFs) which prepare the gene and pol ii for transcription. Pol ii transcribes a small number of nucleotides, encounters a pause and then reinitiates transcription following phosphorylation in serine 5 on the main subunit, to continue transcribing in this stop-start manner (Bartman et al. 2019). The process of transcription initiates and progresses with the involvement of numerous TFs, painting a picture of a dynamic machine with many moving parts needing to be in proximity with pol ii.

2.2.6 Phase separated bodies compartmentalise nucleus functions

Transcription factories (TxnFs) are foci of transcription initiation where regions of DNA requiring transcription are localised (Iborra et al. 1996). These factories will contain all the relevant RNA polymerase ii (pol ii) and TFs required to initiate and maintain transcription. TxnFs appear to be small foci, smaller than 100nm in size, just large enough for the clustering of the key components of transcription machinery (figure 2.2) (Rieder et al. 2012). Understanding the relationship of TxnFs with nuclear and chromatin organisation may provide

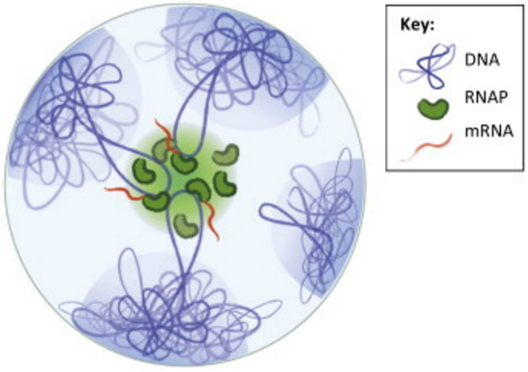


Figure 2.2: Diagram of transcription factory showing genes being transcribed from central location adapted from Weng and Xiao 2014

important navigational cues when trying to visualise transcriptional processes.

TxnFs are sub-compartments within the nucleus, akin to other nuclear bodies such as the nucleolus or Cajal body. These nuclear bodies are discrete, membrane-less compartments which contain the necessary transcription factors to access DNA and produce RNA. They are thought to exist as phase separated liquid regions with separate physical properties to the nucleoplasm surrounding them (Hnisz et al. 2017). High concentrations of TFs and pol ii may facilitate this phase transition. The C terminus domain (CTD) of pol ii is an important regulatory site, involved in numerous post translational modifications, and interactions with TFs and chromatin modifiers. Additionally, it has been shown that the heptapeptide repeat sequence nature of pol ii CTD may facilitate liquid-liquid phase separation through the formation of a hydrogel environment in regions rich in pol ii (i.e. putative TxnFs). Clustering of TFs around a focus such as those that occur around promoter and enhancer sites may also contribute to liquid phase separation, due to myriad non-covalent linkages between the high concentration of TFs (ibid., Harlen and Churchman 2017).

How TxnFs are formed is not fully understood, although current evidence suggests TxnFs consist of a protein core in which high concentrations of TFs, notably pol ii, are tethered in place (Iborra et al. 1996, Hari-Gupta et al. 2020). The persistence of pol ii rich foci following inhibition of transcription initiation and elongation suggest regions where TxnFs form are organised prior to the initiation of transcription, and exist as discrete subnuclear compartments (Mitchell and Fraser 2008). Current models posit that genes are brought in proximity to a preformed TxnF via chromatin remodelling (A. Kim and Dean 2012, Mitchell and Fraser 2008, Sutherland and Bickmore 2009). This remodelling may be the result of histone acetylation (Canals-Hamann et al. 2013). The notion of TxnFs as discrete nuclear structures existing in the absence active transcription is not without controversy, with an alternative model suggesting TxnFs emerge at sites of high transcriptional activity as a result of high local concentrations of pol ii and other TFs producing a

highly proteinaceous microenvironment around a gene cluster, causing the induction of a phase separated droplet (Sutherland and Bickmore 2009, Yao et al. 2007).

2.3 Single cell to blastula: Cell divisions are a focal point in understanding the maturation of the embryo

2.3.1 Pre-MBT cell cycles are rapid

The cell cycle is a highly coordinated process, consisting of compartmentalisation of processes to discrete phases. Of these functions are DNA synthesis and transcription. In the pre-MBT embryo within fast developing organisms such as the zebrafish, the rapid cell cycles present an environment which contrasts with the conventional understanding of how cell cycle events are coordinated.

The embryonic cell cycle is divided into discrete phases, describing a progression of growth, synthesis and division. The basic divisions of the cell cycle are interphase, prophase, metaphase, anaphase and telophase (figure 2.3). Interphase is the phase during which DNA is synthesised and genes are transcribed, and the other phases are the mitotic phases (M-phase) where the nuclear contents are reorganised in preparation of cell division.

In organisms where MZT occurs later in development, early embryonic cell cycles are typically faster than post MZT cell cycles, and these cell cycles are highly synchronous with each other. During and following MZT, cell cycles drastically elongate, resulting in cell cycle lengths more typical of differentiated cells (Tadros and Lipshitz 2009).

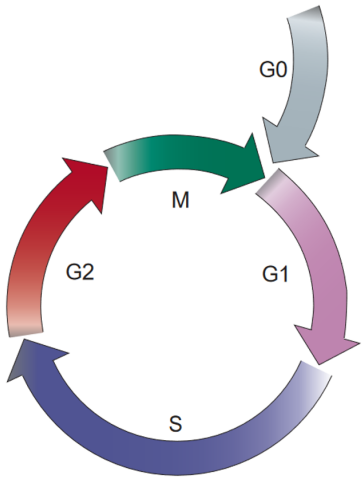


Figure 2.3: Phases within interphase (G1, S, G2), followed by mitosis during which mitotic phases (prophase, metaphase, anaphase, telophase) occur. Diagram adapted from Coleman et al. 2004.

2.3.2 The elongation of the cell cycle and gap phases is a major outcome of MBT

Interphase is further divided into subphases. The general structure of interphase follows a progression of G1-phase -> S-phase -> G2-phase (figure 2.3). The G-phases are 'gap phases' in which transcription predominately occurs. S-phase is 'synthesis phase', which is the phase for DNA replication in a non-quiescent, or a senescent cell (Sankar et al. 2016). The presentation of these early embryonic gap phases varies between organisms. In humans and mice, they are present but are relatively brief compared to gap phases beyond MZT, yet in *Drosophila*, *Xenopus* and zebrafish (our system of interest), gap phases are absent until after MZT. In these organisms, the early embryonic interphase is synonymous with S-phase (figure 2.4) (Newport and Kirschner 1982, Edgar et al. 1986, Kane and Charles B Kimmel 1993).

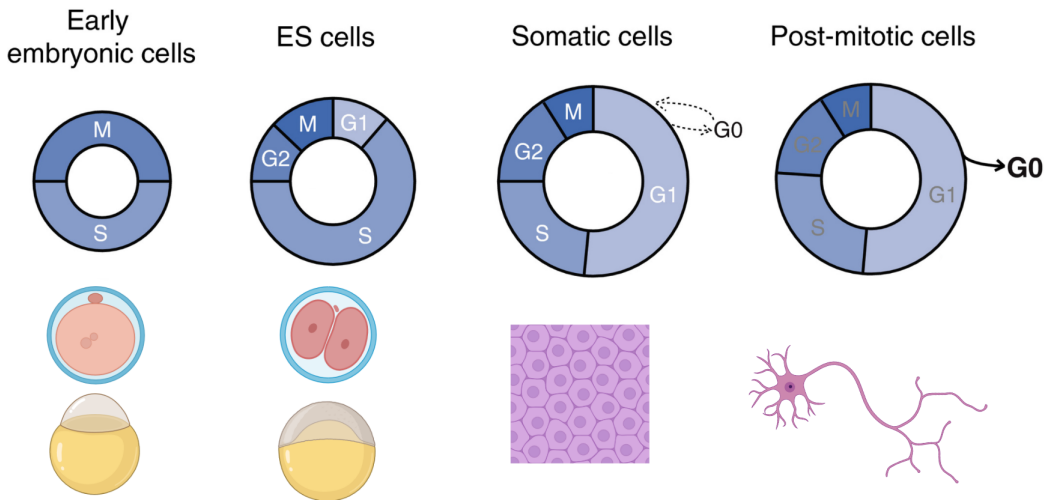


Figure 2.4: The organisation of the cell cycle changes as development progresses, with simple mitosis to S-phase making up the earliest cell cycle program, followed by the onset of gap phases (G1 and G2) later in development, which together with S-phase form the interphase, which emerges in the phase denoted by embryonic stem cells (ES cells). The embryonic stages shows depictions of both mouse and zebrafish embryos, representing the difference in onset of gap phases between slow and fast developing animals. Diagram adapted from Padgett and Santos 2020 with modifications added using resources from BioRender.com.

DNA content must be doubled via replication prior to mitosis. S-phase is the dedicated phase in which this process occurs. Transcriptional processes are significantly reduced during this period. Transcriptional repressors are elevated during S-phase, as both DNA synthesis and transcription both require access to the DNA and there is a risk of clashes between the two processes which could potentially lead to cell cycle arrest (discussed below). Other key components to advancing S-phase are replication factors. Replication factors which have relevance to MZT include treslin, RecQ4, Drf1 and Cut5. These four replication factors are maternally deposited and have been shown to repress not only zygotic transcription, but also the onset of gap phases in early embryonic cells (C. Collart et al. 2013, Langley et al. 2014).

The development and/or elongation of gap phases is a key driver of cell cycle lengthening, via increasing

the duration of interphase. The progression from G phases to S phase or mitosis is regulated by checkpoint signalling pathways (Kastan and Bartek 2004). Checkpoint mechanisms help shift the cell through the different phases of the cell cycle, by sensing if conditions are favourable to enter mitotic programs, or programs permissive to mitosis. Notable proteins in checkpoint pathways include Chk1, cyclins, and various Cyclin dependent kinases (CDK's). CDK1 has numerous functions in cell cycle progression, but has a key role in advancing S-phase. Furthermore, Chk1 is important in advancing the cell from G2 phase into M-phase (Tang et al. 2006), through its inhibition. The fast cell cycles of the early embryo exhibit Chk1 inhibition by way of excess histone 3, preventing prolongation of interphase and advancing rapid cell divisions (Shindo and Amodeo 2020). Disinhibition of chk1 due to the titrating of unbound histones contributes to gap phase development during MZT. In *Drosophila*, titration of cyclins with successive nuclear divisions is associated with cell cycle elongation. Cyclins, important for the transition from interphase into mitosis, are reduced thus permitting elongation of interphase (Stiffler et al. 1999).

2.3.3 Chromatin conformation changes with cell cycle phases

The process of M-phase prepares the nucleus to be able to divide and form two daughter cells. The visually indistinct, amorphous conformation of chromatin during interphase begins to take on order in the earliest stages of M-phase. During prophase, chromatin, whilst still diffuse begins condensing, forming small local clumps of compact chromatin. In metaphase, chromosomes are tightly wound into condensed discrete structures, ready for segregation into the two poles of the dividing cell. In anaphase, the chromosomes are pulled into their separate poles, and in telophase, chromatin begins to unpack as the new nuclear membrane reforms in the newly divided cell (these phases shown in figure 2.5) (Alberts 2008). These phases can be characterised in microscopy, though the examination of imaging parameters (discussed in section 4.2.6).

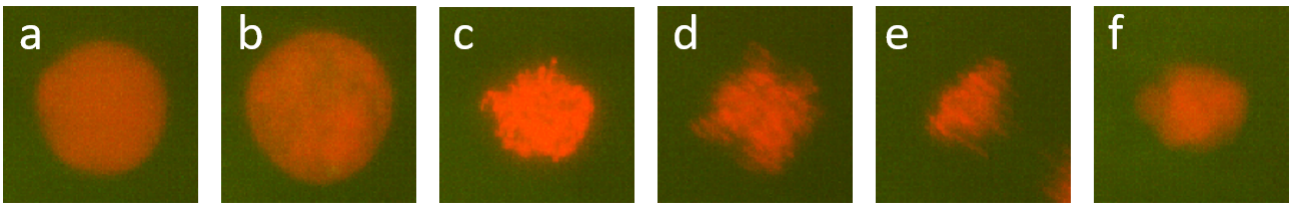


Figure 2.5: Phases of early (256 cell stage) zebrafish embryo cell cycle showing the various states of chromatin. a) S phase b) prophase c) metaphase d) anaphase e) telophase. Chromatin is visualised with H2B mRuby. Images are maximum intensity projection views of light sheet imaging by Yavor Hadzhiev.

It was long assumed that transcription ceased during M-phase, largely due to the compacted chromatin rendering its component nucleosomes inaccessible to transcription machinery (Parsons and Spencer 1997, D. Prescott and Bender 1962). Chromatin conformation is permissive to transcription as chromatin remodelling

has been shown to precede transcriptionally active states in embryonic cells during interphase (Eckersley-Maslin et al. 2018). Emerging evidence suggests a low level of transcription can be sustained into the mitotic phases, through transcription machinery which remained bound from interphase. This phenomenon was shown in cell culture on HeLa cells and is yet to be observed *in vivo*, but it represents the possibility of low levels of transcription persisting into M-phase (Palozola et al. 2017).

2.3.4 Rapid cell cycles present challenges to the cooordination of transcription and DNA synthesis

Transcription and DNA replication machinery require access to the same DNA template. Concurrent transcription and DNA replication risks collisions between machinery if both processes happen to occur at the same place at the same time. This can result in arrest of DNA replication fork, termination of transcription, or breakage of the DNA strand (Alzu et al. 2012, Brewer and Fangman 1988, Wang et al. 2021).

Typically, this conflict is avoided by temporally compartmentalising transcription and replication into gap phases and S-phase respectively. Checkpoints would also be involved in regulating the progression of cell cycle by ensuring DNA integrity is preserved, and arresting cell cycle if something is awry (Kermi et al. 2019). The early wave of transcription during MZT seems counter-intuitive, given the knowledge that early embryonic gap phases are extremely fast, or non-existent; hinting at potential conflict with how replication and transcription can coexist in this abnormally truncated cell cycle.

How replication and transcription are reconciled in the rapid cell cycles of fast developing organisms is unclear. Cell cycle length, and thus interphase/S-phase duration in fast developing organisms has been shown to influence the quantity and length of transcripts (Vastenhouw et al. 2019, Dalle Nogare et al. 2009). This hints at a juggling act between replication and transcription, and there may be limited separation between the two processes. Analysis of transcribed mRNAs in *Drosophila*, which lacks gap phases during ZGA, reveals that mostly short intronless genes are transcribed and much of the transcribed product is truncated and ineffective (Kwasnieski et al. 2019). This suggests that in organisms where gap phases are absent, the transcription-replication conflict may be unresolved and replication occurs at the expense of transcription. Additionally some truncated transcripts may be the product of incomplete transcription of longer genes which did not have the time to complete transcription during the brief pre-MBT cell cycle. This suggests that cell cycle may be a regulating factor for pre-MBT transcription, by being too brief to allow high levels of transcription to occur.

2.3.5 Inherited variations in dividing cells: Epigenetic inheritance

In addition to passing on copies of their genomes to their daughter cells, dividing cells also pass on idiosyncratic features developed over the course of their own cell cycle. This epigenetic inheritance is mainly achieved

through post translational modifications of histones within nucleosomes, and via methyl marks on DNA itself. Inherited characteristics which may affect gene regulation or cellular processes may influence cell cycle and transcriptional dynamics in embryos undergoing MZT [Cheedipudi et al. 2014, Probst et al. 2009, Xu et al. 2011].

During MZT, chromatin is remodelled to open various chromatin sites for transcription (Orsztynowicz et al. 2017). Chromatin states have been implicated in epigenetic inheritance through multiple mechanisms. The conformation of chromatin is thought to be passed on to the next generation via methylation and acetylation modifications to the histones which comprise it. Nucleosomes have been shown to 'remember' their positions within the nucleus following division, and this has been suggested to have implications for transcriptional regulation (Reinberg and Vales 2018).

Heterochromatin is often marked with Histone 3 Lysine 9 Trimethylation (H3K9me3), a ubiquitous histone modification, which can be inherited by daughter cells in the early embryo. In zebrafish H3K9me3 appears during MZT, but not prior. This is reflected in the lack of heterochromatin prior to the completion of MZT (Laue et al. 2019). The appearance of H3K9me3 opens the possibility of transmitting chromatin states, and potentially influencing gene regulation in embryonic cell lineages.

An example of inheritance of active transcriptional states is from experiments in Dictyostelium which showed through the use of live imaging of transcription that Histone 3 Lysine 4 methylation (H3K4me) was associated with correlated transcriptional activity between parent and daughter cells. Intefering with H3K4me abolished such correlated activity (Muramoto et al. 2010). Inheritance of transcription promoting chromatin states may be involved in driving differentiation and cell fate.

In this vein, transcriptional memory was identified in the early *Drosophila* embryo, through visualisation of early transcribed genes during MZT. The asynchronous activation of transcription across the embryo shows that only a small number of nuclei show transcription activation at the beginning of MZT. These experiments showed that parent nuclei which transcribed earlier during MZT were more likely to produce daughter cells with active transcription (Ferraro et al. 2016).

2.4 How to investigate development in real time: An overview of microscopic methods and their considerations

2.4.1 Light scattering

Prior to acquiring and analysing light microscopy data, it is contextually important to discuss the nature of the diffractive properties of light. An understanding of diffraction phenomena as a useful orientational primer

before exploring microscopic data in a meaningful way, as diffraction is an unavoidable phenomenon which will define the clarity of an acquired image. When designing imaging experiments, an understanding of diffraction is implicit when selecting numerical aperture (NA) and imaging media.

Light emitted from a circular point source is diffracted when it enters the objective lens of the microscope. This scattering of light causes interference with overlapping diffracted signals, producing diffraction pattern. The 2 dimensional cross sectional pattern of diffracted light in the microscope focal plane is known as the point spread function (Cole, Jinadasa and Brown, 2011).

The light emitted from the central source point spreads in the x,y directions, perpendicular to the direction of emission. The point spread function (PSF) of a circular point source and numerical aperture consists of a central beam of focussed emitted light, and a repeated concentric pattern surrounding it are scattered light. The intensity, of scattered light decreases in magnitude with more distant rings (Cole et al. 2011). The central focus of light is known as the airy disk, the diameter of which is the Airy Unit. The peripheral concentric rings are known as airy rings (Korobchevskaya et al. 2017). Together, the airy disk and rings produce the airy pattern. In terms of applying this knowledge to imaging, airy rings carry photonic information from the light source, but their quality and intensity progressively degrades with each ring (Karman et al. 1998).

This is informative for data analysis because the way light is scattered and is collected before reaching the sensor determines the clarity of the image, the introduction of artifacts and noise, the ability to distinguish between adjacent light sources (resolution) and the ability to determine the approximate size of a light source in an image. It is also representative of the notion that the image seen on the computer screen following a round of imaging is only ever an approximate representation of the light source. Some information is invariably lost in transmission, via interference and computational reconstruction, from light source to image output.

An important and unavoidable factor which will affect image quality regardless of imaging modality is depth. Image acquisition from deep within a sample will be of lower quality as light emitted from fluorophores will be scattered by the sample around them. In the case of imaging a whole zebrafish embryo, fluorescence emitted from the centre of the embryo will undergo more scattering and signal degradation than fluorescence emitted from the surface of the embryo.

2.4.2 Light sheet fluorescence microscopy

Light sheet fluorescence microscopy (LSFM) relies on passing a sheet of laser light through a transparent sample, and exciting fluorophores. Although the fundamental principles of lightsheet microscopy are constant across machines, the light sheet set up described here will be based off the workings of the Zeiss Z.1, as that is the instrument relevant to this report.

The specimen is suspended in a column of low melting point agarose. A beam of laser light is expanded into a sheet by passing it through a series of lenses, mirrors and prisms. This sheet of light is passed through the specimen, with the plane of light oriented perpendicular to the objective lens (figure 2.6). Fluorophores introduced in/endogenous to the specimen are excited, emitting fluorescence which is collected by the objective lens and sent to the sensor. The light is not passed through a pinhole, as the field of view is large and is expected to cover numerous light sources. The light is focussed via lenses onto a single sCMOS sensor.

A key issue with this with regards to producing images for analysis is the optical anisotropy across the field of view. In images covering a large area, fluorophores from the periphery will be of lower clarity. The PSF properties from fluorophores distant to the focal centre will be different to the PSF of the central fluorophores. This can lead to an 'out of focus' effect for peripheral regions, making them harder to interrogate during image analysis (Wu et al. 2016).

The specimen is moved in the z axis to image in 3D. LSFM is useful as it can quickly perform a 3D scan of a specimen, while minimising photo-toxicity and photo-bleaching. Adverse effects of laser light exposure are minimised as laser light sheet illumination only occurs in a single plane at a time, rather than illuminating the whole specimen, preserving fluorophores in the rest of the specimen until the light sheet passes through (Pampaloni et al. 2013). This method of imaging can relatively quickly produce an entire embryo scan, as the sheet of light illuminates an entire Z plane at once.

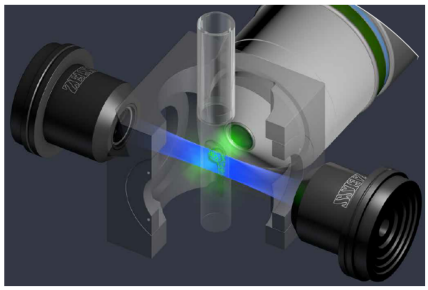


Figure 2.6: Zeiss Z.1 lightsheet microscope render, showing embryo (located at intersection between blue and green) mounting, sheet of laser light (blue) and fluorescence (green). Adapted from (Zeiss, 2014).

2.4.3 Airyscan confocal microscopy

Airyscan confocal microscopy (ACM) is a variant of confocal laser scanning microscopy (CLSM) able to utilise more photonic information from airy rings emitted from the laser excited fluorophore, to construct more clear, detailed images. Generally, CLSM is favourable over widefield microscopy as a single laser beam illuminates a single point on the specimen, rather than bathing the whole specimen in light. This can

reduce phototoxicity and photobleaching (Korobchevskaya et al. 2017).

In conventional CLSM, the laser beam stimulates fluorescence from fluorophores in the specimen, which is collected by the objective lens, and passed through a pinhole, which allows only the airy disk to reach the sensor while blocking the less focussed airy rings.

The airyscan module for a confocal microscope uses the same laser illumination and objective lens, but redirects the collected light down a different path where the light is not passed through a pinhole but is focussed through lenses and the entirety of the emitted fluorescence is detected. ACM does not discard the airy rings but collects scattered light from the source and reconstructs the image using this extra information, via it's array of 32 tessellated hexagonal sensors (figure 2.7)(Huff 2015). This can lead to a higher quality image, as the PSF is collected not from a single sensor, but discrete portions of it are detected by individual sensors, and the original image is reconstructed by digitally re-stitching the scattered airy rings together to form the original image, prior to diffraction at the NA. In this way, it is expected to reconstruct some of the information obscured by diffraction, and usually lost in conventional CLSM (Korobchevskaya et al. 2017).

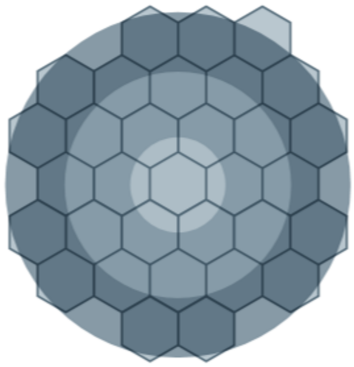


Figure 2.7: The spread of the airy pattern on the hexagonal sensor array of the airyscan microscope

For both described methods of confocal microscopy, it is important to note that a single laser beam is used for illumination, which needs to scan in the X and Y dimensions, per Z plane. This can make obtaining a single 3D image (per time point) relatively time consuming, making ACM suboptimal for whole embryo scanning, but is best used for focussing on a region of tissue, for maximal image quality and clarity.

Spinning disk technology substantially decreases the image acquisition time of CLSM. This has potential for increasing the field of view while maintaining reasonable temporal resolutions for 4D imaging. Spinning disk modules utilise an array of pinholes covering the entire sample, and tracking over the sample for simultaneous excitation and detection of fluorescence in a sample, and in doing so, eliminate the need for x,y scanning.

2.4.4 Data Handling: Current methods and issues

4D imaging methods output large volumes of data, particularly in the case of whole embryo LSM which outputs data sizes into the terabyte range. Large data sets, while powerful in terms of the data they can contain, have large computational overheads. There are many technical aspects of imaging technology to be aware of when querying data for reliability and analytical purposes. The structure of an output image from 4D microscopy will determine how to approach analysis, and which manipulations to make at the fundamental level for computational tasks to be viable, while maintaining integrity of the image quality.

Image file size increases in magnitude with the addition of more dimensionality. This dimensionality can be broken down into bit depth, and resolution in x, y, z and t. Bit depth and resolution arise from the physical capabilities of the camera sensor (sCMOS, or CCD) in the microscope. Dimensionality relies on the mechanical capabilities of the microscope.

The first factor which will set the size of the output image file is the bit depth. Bit depth represents the amount of bits of data that are encoded per pixel. The level of bit depth determines the fidelity of the image, as it determines the levels of light (termed grey scales) the sensor is able to discriminate between (Wallace et al. 2015). A higher bit depth increases the number of bits per pixel, increasing the data per pixel, thus using more memory per pixel, and increasing the final file size. A 16 bit pixel carries twice as much data as an 8 bit pixel. In geometrical terms, this can be seen as data representation in 1 dimensional space. An image processing consideration arising from this could be to reduce the bit depth of a high bit image, though this is dependent on what level of quality is required for performing a reliable analysis.

The image resolution in 2 dimensions (x,y) is the next determinant of file size. Modern sCMOS chips can support resolutions up to 2560 x 2160 pixels, increasing image resolution and file size (Fowler et al. 2010). A 3D image is a stack of 2D images (the z plane). The number of stacks used to obtain a 3D image will affect the size of the data file, making the axial resolution an important consideration when obtaining data; better axial resolution will produce higher fidelity images for analysis but will inflate data sizes, making computation difficult. With regards to fidelity, the z-stack parameters will determine how the pixels for each 2D sheet are converted to voxels (display unit of 3D image, with an x, y and z coordinate and a colour/gray value) in 3D rendering. This can affect how volumes are represented, and reduce 3D geometrical accuracy. A 4D output is a series of 3D stacks, making temporal resolution (t) the focus of data size compromise.

With this in mind, it is not unreasonable that a high fidelity 4D image will have a file size in the terabyte range. Parsing this kind of data in its raw form can be problematic for all but the most powerful computers, due to system memory limitations. Image processing on large data sets thus becomes prohibitively time consuming or simply not possible. As a result large datasets need to be manipulated before processing can

begin. Understanding how to store data, and what tools are available to process and analyse it follows from understanding how the imaging data is constructed.

2.5 *Danio rerio*: a fine laboratory model for studying developmental phenomena

2.5.1 Zebrafish model

Zebrafish are useful laboratory animals due to their high fecundity, ease of manipulating their breeding cycles, rapid development from fertilisation to larva, and relatively short maturation period from egg to sexual maturity of approx. 3 months (C. B. Kimmel et al. 1995). They function well as a model vertebrate system as development occurs in transparent eggs, and the larvae remain transparent during organogenesis, allowing researchers to observe visceral phenotypes that would ordinarily be hidden in other developing vertebrates. These attributes make phenotypes easy to screen very early on bypassing the need for complex, risky invasive procedures thus making *in vivo* intracellular imaging feasible.

Zebrafish are useful in genetic studies as their genome is sequenced and well-studied. There is a large degree of genetic similarity, with 71.4% of human genes having at least 1 orthologue in zebrafish, and 69% of zebrafish genes having at least 1 orthologue in humans (Howe et al. 2013). Processes critical to life, such as early embryonic transcription, show high levels of genetic and functional conservation among vertebrates (Duret et al. 1993). In addition to the homology between humans, this conservation suggests that studies performed in zebrafish may be applicable to human biology also. As a result, zebrafish are a useful model organism to study transcription, particularly before translating research into mammalian systems.

2.5.2 Zebrafish development

The cell cycle divisions immediately following fertilisation are highly synchronous and rapid. MZT begins at the 6th cleavage, or 64 cell stage, which occurs at approximately two hours after fertilisation (Kane and Charles B Kimmel 1993). MZT concludes and ZGA is achieved at approximately 1000 cell stage, midway through the blastula period approximately one hour after MZT onset. During this time period, blastomeres form, the yolk syncytial layer forms and global cell cycles are elongated and asynchronous. This is the primary time period of interest for this thesis.

The dividing cells of the embryo sit atop the yolk, from where they draw nutrients for growth. The yolk is a prime target for delivery of substances into the very early embryo as the cells sitting atop the yolk cell are still in partial syncytium with the yolk and substances introduced into the yolk may be actively transmitted into the cells of the cleavage stage embryo (ibid.). After cleavage stage, the syncytium with the animal cop

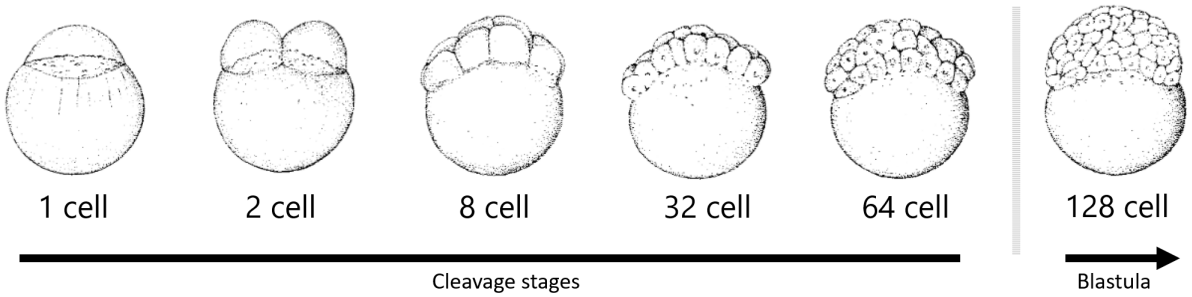


Figure 2.8: The progression from cleavage to blastula stage, showing the arrangement of cells of the animal cap in relation to the yolk. Adapted from C. B. Kimmel et al. 1995

begins to decline as cells cluster atop one another, as shown in figure 2.8 (Kane and Charles B Kimmel 1993). This poses a challenge for late stage injections into yolk as it would result in unpredictably uneven distribution of substance across embryonic cells, assuming this was intention of injection.

The single large cell divides into smaller cells, conserving the overall geometry of the animal cap: resulting in drastic reductions in cytoplasmic volume relative to nuclear volume. Nuclear volume also decreases with these early divisions, although not nearly as drastically. This physiology is consistent with the NCR model of MZT.

2.6 Relevance to this thesis

In this introductory chapter, I have discussed the period of maternal zygotic transition, and how cell cycles and transcription fit into this environment. This primer lays the groundwork for the following chapters, where in the context of MZT in the zebrafish embryo, I will discuss the activation of transcription, microscopy processing and analysis methods, and inheritance of cell cycle features and transcription.

Our initial inquiry is about the activation of transcription in the early zebrafish embryo. We will attempt to understand global spatiotemporal aspects of transcription activation across the embryo. Secondly, I attempt to address technical challenges brought forth from handling 4D data sets. I design and create a software tool to analyse inherited features in early embryonic lineages. Thirdly, I implement said tool on light sheet imaging datasets to identify the presence of inherited cell cycle characteristics.

3 Methods

3.1 Zebrafish

All zebrafish handling and imaging was conducted by Yavor Hadzhiev, Kacper Wozniak, Aleksandra Jasiulewicz.

3.1.1 Husbandry

Zebrafish were kept in a recirculating system (ZebTEC, Tecniplast) in an aquatics facility. Fish were kept at 26^{o}C in a set 10hr dark/14hr light photoperiod, and were fed three times per day. Breeding was set up in a pair wise manner. Males and females pairs were put in breeding tanks with a clear breeding barrier between the pair. The following morning, within an hour of the light cycle, the barriers in each breeding tank were removed to allow males and females to mate. Eggs were harvested and sent immediately for injection.

3.1.2 Protein preparation

All protein preparation methods outlined in this section were carried out by Divyasree Poovathumkadavil of the Gambus Lab

mRuby3:H2B was amplified via PCR from the *pKanCMV-mRuby3-10aa-H2B* (Addgene, 74258) plasmid, and mApple:PCNA was PCR amplified from *mApple-PCNA-NLS-4* (Addgene, 54937). Using an In-Fusion HD Cloning Plus Kit (Clontech, 638910), mRuby3:H2B and mApple:PCNA were each sub-cloned into a NdeI/BamHI linearised *pET-28a+* plasmid, keeping in frame with the N-terminal 6xHis-Tag.

The resultant plasmids underwent transformation into *E. coli* BL21 (DE3) and the subsequent proteins was expressed in LB media following IPTG induction at 20^{o}C.

Cell lysates containing mApple:PCNA were resuspended in 50mM Tris HCL (pH8), 250mM NaCl, 10mM imidazole, 2mM $MgCl_2$ and 10% glycerol.

Cell lysates containing mRuby3:H2B were resuspended in in 50mM Tris HCl (pH8), 500mM NaCl, 10mM imidazole, 2mM $MgCl_2$, 5% glycerol and 0.3% Brij.

Buffers were supplemented with BitNuclease, PMSF and protease inhibitors.

Sonication, followed by centrifugation of lysates clarified cell extracts, and proteins were purified using Nickel affinity chromatography (Generon). mApple:PCNA and mRuby3:H2B were dialysed against 500mM KCl, 20mM PIPES, 100μ M EGTA at pH6.8 using 12kDa columns (Pur-A-Lyzer, Sigma).

3.1.3 Injections

One-cell stage wild type zebrafish embryos were injected with 2nL solutions of 7μM morpholino (Gene Tools LLC) in the yolk in nascent transcription visualisation experiments.

400ng/uL of both mApple:PCNA protein, and mRuby3:H2B were injected at one-cell stage for their respective experiments. mApple:PCNA was used to visualise S-phase *in vivo*, and mRuby3:H2B was used as a general *in vivo* nuclear dye.

Following injections, embryos were incubated at 28^{o}C in E3 fish media until imaging.

3.1.4 Transcription block

Transcription block was applied at one cell stage zebrafish embryos by microinjection with 200pg α -amanitin (Sigma A2263), or treatment from one cell stage until imaging with 1μ M triptolide (Sigma T3652).

3.1.5 Cell cycle elongation

Wild type embryos were injected at one cell stage with a solution mixture containing 200pg of *Xenopus chk1* mRNA, 200pg miR430 MO and mRuby3:H2B. Controls were injected with 200pg *mCherry* mRNA, 200pg miR430 MO and mRuby3:H2B.

3.1.6 Generation of tetraploid embryos

Tetraploid embryos were generated by applying heat shock of 42^{o}C at 20mins following fertilisation. Embryos were transferred from 28^o C incubator into a dish in a 42^o C water bath for 2mins. Embryos were then returned to 28^o C to continue development. Heat shock treated embryos were selected based on failure to undergo cytokinesis at the timing of the first cleavage, and were used to imaging once they reached the appropriate developmental stage.

3.1.7 Lines

In addition to using wild type embryos injected with mRuby3:H2B for visualising the nucleus, a transgenic line with fluorescent Histone H2B was initially used. This was line Tg(Xla.Eef1a1:h2b-mrfp1) (http://zfin.org/ZDB-TGCONSTRUCT-130711-1 26), provided from the University of Sheffield, UK. Embryos from this line were also used in transcription imaging experiments.

3.2 Imaging

3.2.1 Light sheet

Light sheet microscopy was conducted on Zeiss Light Sheet Z.1 using a 20x objective lens. Embryos were mounted in a column composed of 1% low melting point agarose, and imaged at 28^{o} C. Images in the light sheet microscope were taken every 30-90s, with between 200-300 z-slices acquired for different datasets (datasets and accompanying metadata in appendix). Post imaging processing was performed in Zeiss Zen Black software.

3.2.2 Airyscan

Airyscan microscopy (used for 3D analysis in chapter 4) was conducted on Zeiss 880 confocal microscope with a fast scanning airyscan module. Embryos were mounted in 0.5% low melting point agarose, inside a glass bottom imaging dish maintained at 28^{o}C during imaging. 100-200 Z-slices were taken every 30-60s using a 40x objective lens. Post processing of imaging outputs was performed using Zeiss Zen Black software.

3.3 Statistical analysis

Statistical analyses were performed using the R programming language, specifically R 3.5.1 (R Core Team 2018). R was used to perform Pearson regressions and to create boxplots. No special statistical packages were downloaded for this use.

3.4 Software

All computational work was conducted by myself, with a contribution from Andor Magony in setting up the base for the 3D segmentation tool.

3.4.1 2D and 3D processing

4D (3D + t) microscopy data from light sheet imaging was converted to maximum intensity projection using Zeiss Zen Black Software (Carl Zeiss GmbH, Germany). Zeiss Zen Blue software was used to subset data for small scale 3D analyses in chapter 5.

2D Image analysis of maximum intensity projections (for chapter 4) were principally conducted using Icy Bioimage (versions 1.9.5.0-1.9.7.0, Quantitative Image Analysis Unit, Pasteur Institute, France). A script was written in the Icy script writing tool, using the Icy JavaScript libraries. 3D image analysis of Airyscan data was also conducted via the same versions of Icy Bioimage as used for 2D analysis. Github URLs for scripts linked in section 3.4.3.

ROI's were generated around features automatically using the spot detector plugin (further elaboration on development and implementation of this method in section 4.2.2, with size criteria for feature area or volume being created individually for each dataset. Spot detection was used in both 2D and 3D.

Fiji ImageJ (Schindelin et al. 2012) was used to convert .czi files as .tiff series for processing in my custom segmentation software.

3.4.2 Software Development

Python scripts were constructed using Python 3.8.5 within the Linux Ubuntu operating system, version 20.04 LTS (Canonical, USA). Key libraries used in the construction of the 2D segmentation software are as follows: Watershed seed creation was built using find_peaks function from scipy library (Virtanen et al. 2020). Peak clustering was performed using DBSCAN clustering algorithm from scikit-learn library (Pedregosa et al. 2011). OpenCV2 library was used for dilation image transforms, creating exclusion zones around watershed seeds, gaussian blurring and creating the watersheds from the seeds.

The 3D segmentation part of the tool was constructed in MATLAB (Mathworks, MATLAB ver. 2019b). A segmentation base was provided by Andor Magony, which used the in built watershed library to convert raw images into 3D segmented outputs. This basic version of the tool operates on raw microscopy slices via implementing a median filter to remove noise. A threshold is applied to create a binary, and a small object filter to remove debris. 8-connected components (bwconncomp) labelling between binary image stacks joins binarized features per slice into 3D objects. Imfill function is used to close holes in segmented objects to give solid volumes, the measurements for which can be exported as metadata. My modifications bypass the raw image processing and feed 2D segmented stacks into the 3D segmentation portion. Multithreading was also added, using the parfor function.

The transformation module part of the segmentation/tracking tool was written in Python to link the 3D segmentation output with the tracking tool. Centroids of 3D features are stored in a csv file, with metadata containing their volumes and x,y,z and t dimensions. All values used in processing to this point have been in pixels, so the transformation converts this to microns, based on user input provided from the graphical user interface. The tracking module was also built using Python. The tracking functionality was based on the KDTree function from the scikit-learn library. Github URLs to scripts linked in section 3.4.3.

More detailed discussion regarding how this tool came together is in chapter 5.

3.4.3 Scripts

The scripts created for this thesis can be found at:

- Icy javascript for 2D tracking analysis - `https://github.com/Subject-Infinite/icy_roi_group`
- Segmentation and tracking software tool - `https://github.com/Subject-Infinite/SME`
- File and folder arranging before segmentation/tracking - `https://github.com/Subject-Infinite/SME/tree/main/utility_scripts`
- Validation of tracks using 4D max projection - `https://github.com/Subject-Infinite/SME/commit/d9be7597180b515d4c3ba9054ad2b8b727c4fb99`
- Script to compare cell cycle lengths between parent and daughter cells - `https://github.com/Subject-Infinite/SME_output_lineage_plot`
- Create 2D plots of mitosis locations in the transition between pre-MBT embryonic stage - `https://github.com/Subject-Infinite/Mitosis_point_detect`

4 *In vivo* detection and characterisation of transcription and cell cycle during the early transcriptional wave

4.1 Introduction

4.1.1 The Midblastula transition in the zebrafish embryo

The zebrafish embryo undergoes rapid growth in the first 5 days post fertilisation, developing from a single cell to a free swimming, free feeding larva with a mostly developed set of internal organs (C. B. Kimmel et al. 1995). The first 3-4hrs of development post fertilisation see near exponential cellular growth (figure 4.1), a time during which the developmental physiology of the embryo shifts from maternal control to endogenous control.

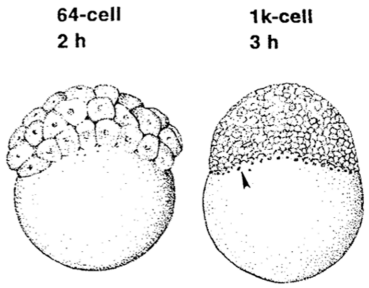


Figure 4.1: Milestones in zebrafish development, showing the extent to which zebrafish develop in the first 3hrs hours post fertilisation adapted from C. B. Kimmel et al. 1995. At 64 cell (2hrs post fertilisation) stage the first vestiges of transcription are detected, and by 1000 cells (approx. 3hrs post fertilisation), full genome activation has occurred. This hour of development will be the subject of discussion for this thesis.

The earliest developmental stages in animals are controlled by maternal RNAs and proteins, deposited in the oocyte during its formation (Tadros and Lipshitz 2009). The switch from maternal control of development to embryonic/zygotic control is a critical milestone in the development of an animal and involves the degradation of maternally deposited factors and the activation of the embryonic/zygotic genome. In zebrafish, MBT occurs at cell cycle 10, or 1000 cell stage figures 4.1, 4.2(C. B. Kimmel et al. 1995, Vastenhouw et al. 2019).

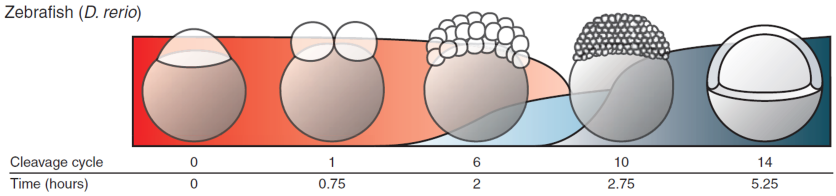


Figure 4.2: Diagram depicting activation of zebrafish genome. The red section denotes maternally driven transcription. Light blue denotes the early wave of transcription which occurs at approximately 64 cell stage when ‘early’ genes are transcribed, and dark blue denotes full embryonic genome activation when the rest of the genome awakens. Image adapted from Tadros and Lipshitz, 2009 (Tadros and Lipshitz 2009).

In the pre-MBT zebrafish embryo, initial cell divisions are very rapid, and synchronous with cells undergoing mitosis at approximately the same time (Tadros and Lipshitz 2009). From the earliest division up until 32 cell stage, cell cycle lengths decrease in length. Following this, cell cycles progressively elongate and cells begin to lose synchrony. Initial loss of synchrony is slight, but the level of asynchrony amplifies with each cell division. Post MBT, cell cycles become much longer to accommodate for gap phases during which high levels of transcription occur (Olivier et al. 2010).

4.1.2 Early wave of transcription

Maternal RNA and proteins are used for much of the cellular functions of the early embryo, and the pre MBT genome is mostly silent (Kane and Charles B Kimmel 1993). An early, minor wave of transcription is a characteristic feature of MZT in many organisms (figure 4.2). This early wave consists of the activation of a small number of genes. In the zebrafish, such genes include miRNAs, DNA binding proteins and histones (Heyn et al. 2014).

microRNA-430 (miR430) is a notable transcription product in the pre-MBT zebrafish embryo as it is involved in clearing maternally deposited transcripts (Giraldez 2010, Giraldez et al. 2005). *miR430* transcription is first detected at the 64 cell stage (Heyn et al. 2014). miR430 has been shown to facilitate clearance of maternal mRNAs, establishing a functional role in MBT onset. Transcription of krüppel-like factor (*klf*) transcription factors (TFs) and histones are other notable early wave transcripts (ibid.). *miR430* and *klf17* are located on different chromosomes in the zebrafish. They are located on chromosomes 4 and 2 respectively (Pei and Grishin 2013, Thatcher et al. 2008). These were selected as model transcripts to use for investigating ZGA, in the imaging used herein.

4.1.3 *In vivo* imaging of transcription

Current established methods of imaging transcription *in vivo* rely on using transgenic MS2 stem loop structures; the main implementation of which is the MCP-MS2 system, used to image nascent RNA transcripts (Tutucci et al. 2017). This technology relies on genetically modifying the target organism with a reporter gene preceding the gene of interest. The reporter gene will carry sequences which become stem loops. Applying a GFP tagged viral coat protein (MCP) from the MS2 bacteriophage, with a high affinity to stem loop structures will cause the MCP to bind the stem loops, producing a bright fluorescent focus at the point of transcription (Larson et al. 2011). The advantages of this approach are that it can be used for *in vivo* real time imaging of transcription. Notable shortcomings include: the need to produce a transgenic organism beforehand, optimising the reporter gene structure and length to get the best signal to noise ratio (SNR) while not disrupting the chromatin organisation (too long a reporter gene could interfere with the heterochromatin)

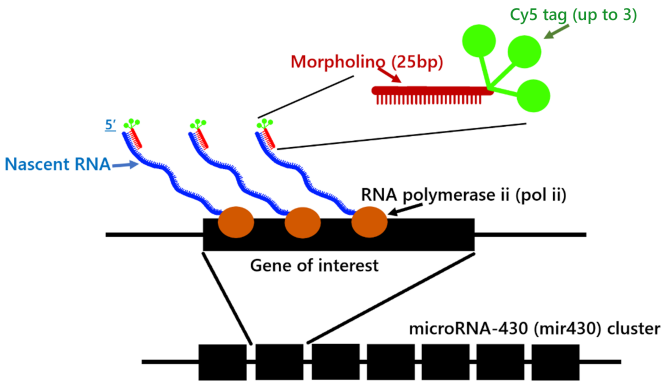


Figure 4.3: Schematic of how the transcription imaging MO works, using *miR430* as an example gene.

(Coleman et al. 2004, Ferraro et al. 2016). Our laboratory engineered a morpholino (MO) based system to address these shortcomings. This technology, named MOVIE (**MO**rpholino **VI**sualisation of **E**xpression) involves applying an external MO compound to the system of interest, and allowing the hybridisation of MO with nascent RNA transcripts (Hadzhiev et al. 2019).

Morpholino (MO) oligonucleotides are DNA analogs with an altered sugar/phosphate backbone. Normal nucleic acid bases can be attached to this backbone, and they bind with complementary nucleic acid strands with the same affinity as ordinary nucleic acids. The atypical backbone structure of MOs prevents them from being digested by nucleases. MO oligomers need to be at least 25 base pairs long in order to stably bind to complementary sequences (Corey and Abrams 2001).

The MO used to label nascent transcripts in the experiments discussed herein are tagged on the 3' end with up to 3 Cy5 moieties. The Cy5 tagged MO binds at the 5' end of nascent transcripts, as this is the first region to emerge from the transcriptosome (figure 4.3).

The MO is injected into the early embryo prior to the first cleavage. The MO will be transported to the nucleus where MO molecules will bind to emerging transcripts. The MO will accumulate at sites of transcription, specific to the gene whos transcript products are being targeted by the MO.

4.1.4 Transformation of 4D lightsheet data

With the smallest whole embryo lightsheet data sets taking up over 500GB space, the data needed to be manipulated to make it feasible for timely computational analysis. Maximum intensity projections (MIP's) were constructed from whole embryo lightsheet data, using Zen Black. A maximum intensity projection is a 2D reconstruction of 3D data, where the brightest pixel in the z dimension is selected as the representative brightness value for its corresponding x,y position and plotted onto a static image. In this image, each pixel corresponds to the brightest pixel in the z plane for that x,y coordinate, in the original 3D image stack.

4.1.5 How are cell cycle and transcription organised during the early wave of genomic activation, prior to MBT?

The fast dividing cells of the pre MBT zebrafish embryo present a compelling case study into the organisation of nuclear functions. How and where does transcription fit into the pre-MBT cell cycle, in the absence of gap phases?

Using MOVIE in conjunction with LSFM, researchers in our laboratory were able to generate insightful images showcasing the dynamics of the nucleus over the course of the cell cycle, and where transcription may fit into this picture.

4.1.6 General aims

The aims of this first set of inquiries were to address the spatiotemporal aspects of the relationship between cell cycle and transcription, and how the two processes are coordinated. My role in this arm of the investigation was principally to design workflows to interrogate obtained microscopy data, and to implement such analyses to derive biological insights from this data. I aimed to measure and quantify the cell cycle and find when transcription occurred in relation to it. To achieve this, I aimed to:

- Create a workflow to be able to process light sheet imaging data
- Measure when and where endogenous transcription is first detected
- Measure cell cycle lengths, and length of time transcription is detected
- Determine how transcription is localised within the nucleus

---

4.2 Results

All laboratory and microscopy work used for the following analyses was conducted by Yavor Hadzhiev. My contribution to this chapter was image processing and image analysis.

Developing a 2D analysis workflow The primary goal of the analyses in this chapter was to try to understand how the cell cycle and transcription are spatiotemporally coordinated in the rapid pre-MBT cell cycles of the zebrafish embryo. Temporal analysis of cell cycle and transcription from whole wmbryo data required 2D analysis due to the sheer size and complexity of the raw data output by the light sheet microscope. The first step to performing temporal analysis in this way, was to transform large scale light sheet and confocal microscopy data into a computationally useful format: the maximum intensity projection. Following this, I needed to set up workflows to detect MO to address transcription, and to detect nuclei to address the cell cycle.

4.2.1 Establishing the intensity topography and implementing manual tracking of nuclei

The premiere challenge to overcome before drawing biological conclusions was to create a workflow which enabled the analysis of large scale 4D datasets. The workflow development consisted of two principle parts, one was developing a workflow for 2D analysis, and one was for spatial analysis of smaller scale 3D data. The design of these workflows dictated the course of analysis, and determined which biological conclusions we could draw.

The imaging data processed for temporal analysis, and basic quantification of features was received in 2D maximum intensity projection format. The image analysis workflow designed to parse incoming data underwent iterations as I became familiar with the data.

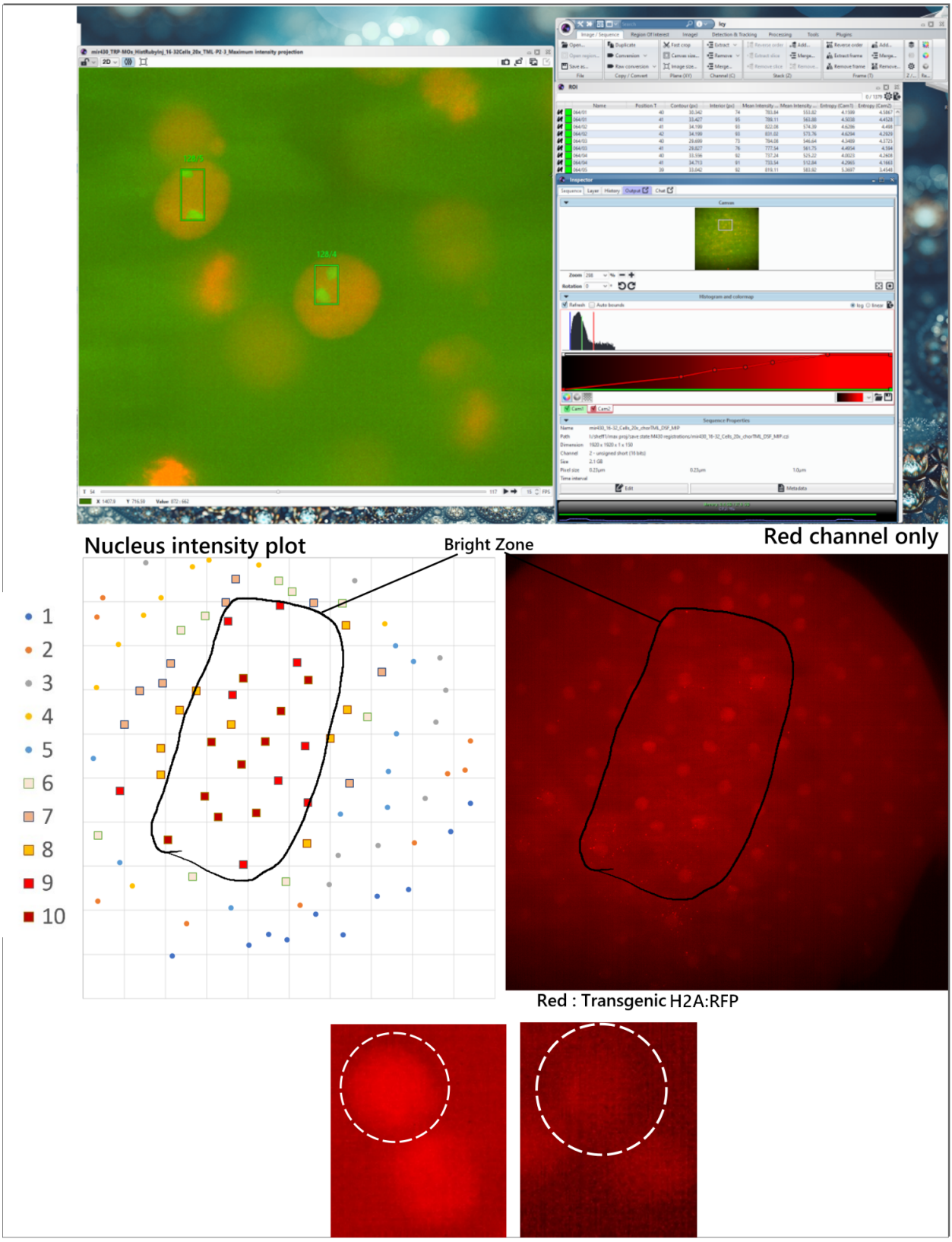

Nucleus intensity plot
Bright Zone
Red channel only
1
2
3
4
5
6
7
8
9
10
Red : Transgenic H2A:RFP

Figure 4.4

Figure 4.4: **Top.** Screenshot of Icy bioimage workspace, showing manual registrations drawn over 2D maximum intensity projection data. Nucleus in red. *miR430* nascent transcription signal highlighted in green. **Middle.** Characterisation of brightness intensity topography across embryo. Left: Scatter plot showing positions of nuclei and symbols representing the average intensity of nucleus ROI, as percentile figure. Squares represent nuclei in top 50% of average brightness (in 10% increments, with 10 the brightest), and circles the lower 50% of brightness. The percentiles of different symbols are described in the legend on the left. Right: Corresponding maximum intensity projection image from which nucleus intensity data was derived. Bright zone is highlighted in both with black outline and label. **Bottom.** Contrasting visibility of nuclei (red, highlighted with dotted white circle) in bright (left) and dark (right) zones of embryo.

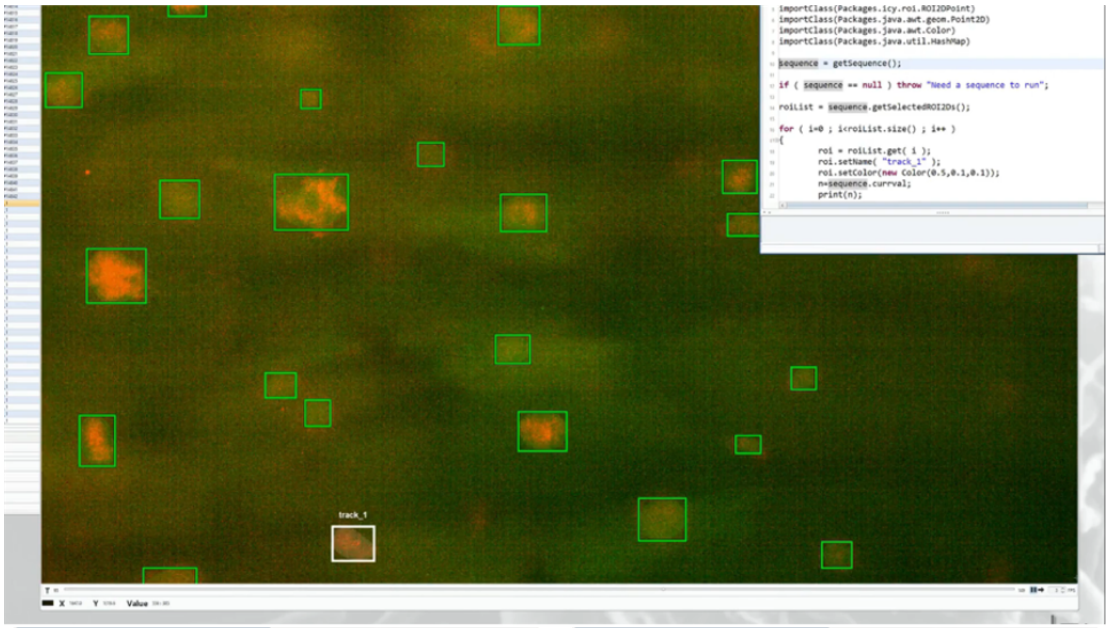
Initial forays into exploration were defined by manual frame by frame registration of nuclei and transcription signals using Icy Bioimage ROI drawing tool (figure 4.4, 4.5 top). ROI's were taken from across the embryo, and their average intensities were compared, allowing me to produce a map of intensity variation of nuclei across the embryo (figure 4.4). The peak brightness of a nucleus was dictated by its background (figure 4.4), and the unevenness of illumination across the embryo led to poorer quality of imaging in the periphery of the image. This notion rendered a basic thresholding approach for 2D analysis less feasible, unless one focussed analysis on regions with similar intensity. Manual registrations of nuclei and transcription was a good way to familiarise oneself with the data, but it was time consuming, and increased the scope for human error and judgement. Quantification of bright/dark zones is shown here calculated from a single embryo, but this method was applied to all analysed MIP datasets prior to analysis.

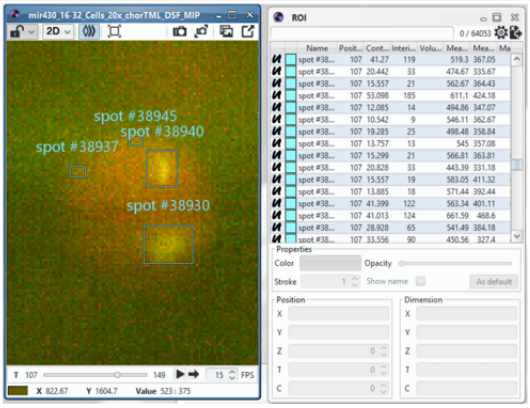
4.2.2 Semi-manual tracking using Icy Bioimage spot detection and custom scripts

To refine on nucleus and transcription detection, and improve the efficiency of nucleus and transcription detection in maximum intensity projection data, the workflow took on a semi-manual design as 2D feature registrations were carried out automatically by the spot detection tool in Icy Bioimage (figure 4.5). The spot detection plugin Icy was used to detect both nuclei and transcription signal, to allow transcription signal to be tied to specific nucleus tracks and lineages.

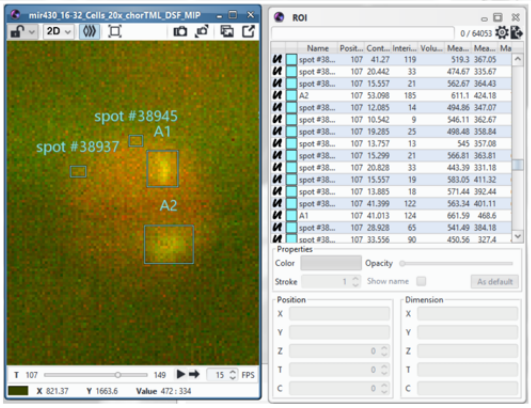
To complement automated spot detection, I designed a script in JavaScript to speed up the process of grouping tracked features together (Github URL for Icy 2D tracking in section 3.4.3). Spot detection can only detect features per frame, but is unable to perform tracking on detected features. The script I designed would colour code ROI's of features in a track and give them a unique family ID, so they could be extracted from the exported metadata and analysed together as a single track. In this iteration, spot detection automatically detected features in bright regions within pre-defined size constraints. This was followed by manual selection of positive hits, as the tool was highly susceptible to the noise in the data, as visible by the non selected/renamed hits (figure 4.5 bottom panels).

This became the standard method for segmentation and tracking for light sheet maximum intensity data for the rest of this chapter, and was the foundation for analysis of cell cycle and transcription duration.

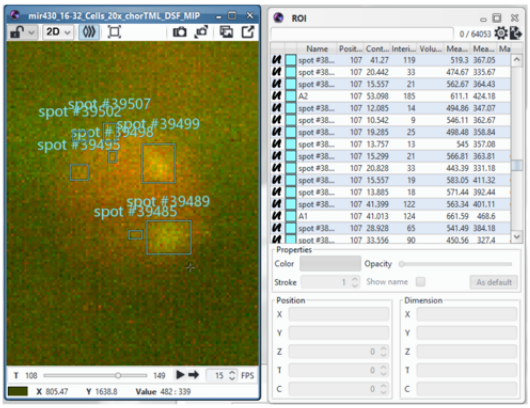




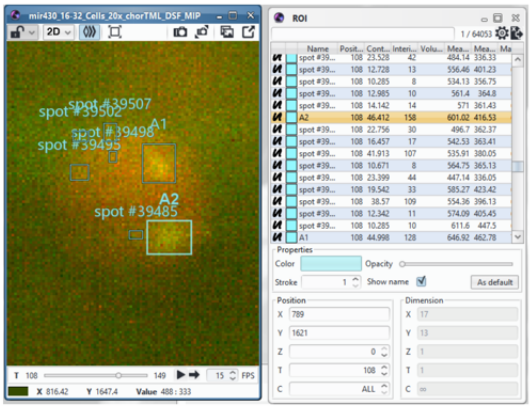
feature detection detects RNA bodies



Positive candidates renamed



Move to next frame



rename RNA bodies appropriately
to match with previous frame

Figure 4.5

Figure 4.5: **Upper, main panel.** Screenshot of Icy bioimage workspace, showing manual registrations drawn over 2D maximum intensity projection data. Nucleus (mRuby3:H2B) in red. *miR430* nascent transcription signal in green. **Lower 4 panels,to be read left to right for chronology**. Screenshots from Icy Bioimage workspace showing preview screen and ROI metadata box (right hand side of each panel). Step by step process of automatic detection and manual refinement of *miR430* signal (green) ROI's. Positive candidates labelled as A1 and A2. Top left: Spot detection on frame n. Top right: Manual selection of positive hits and autorename using custom grouping script. Bottom left: Move to next frame. Bottom right: Repeat process as described for Top right image. Additional detected spots, not renamed (keeping their original name of 'spot #'), are noise.

4.2.3 Developing a 3D tracking workflow

Another question we asked was about spatial coordination of transcription within the nucleus; more specifically we were interested in how much transcription (by volume) was occurring in the rapid cell cycles, and whether it was actually possible to gauge this using the MO visualsation technology.

To address this, I used 3D imaging data from confocal airy scan microscope to take granular looks within the nucleus, to gauge the size of the transcription signal.

The method of 3D image processing relied on creating subsets of airyscan data to focus on specific regions of the embryo, and importing these 3D subsets into Icy Bioimage. This method of 3D segmentation consumed a lot of RAM, as it would load the entire dataset to RAM. Segmentations would offer overhead on top of this, so data was subset to specific regions of the 4D datasets, limiting the scope of tracking. The 2 fluorophores were visualised in 2 different channels. *miR430* MO signal occupied the green channel, and histone mRuby:H2B occupied the red channel. Removing a channel (figure 4.6 top right) reduced the memory footprint of the dataset, allowing a larger subset to be used, providing a larger field of view to work with.

3D Tracking utilised similar principles of tracking to the 2D data. Spot detection automatically segmented bright regions, which needed to be manually grouped together in a track. To streamline this process, ROI's from multiple time points were overlayed in the same viewing frame and features visually forming tracks were grouped using the grouping JavaScript script that was written for figure 4.6. This method made it possible to inspect and track nuclei and transcription signals in small subsets of data, RAM permitting, in a semi manual process of automated feature detection and manual refining and tracking.

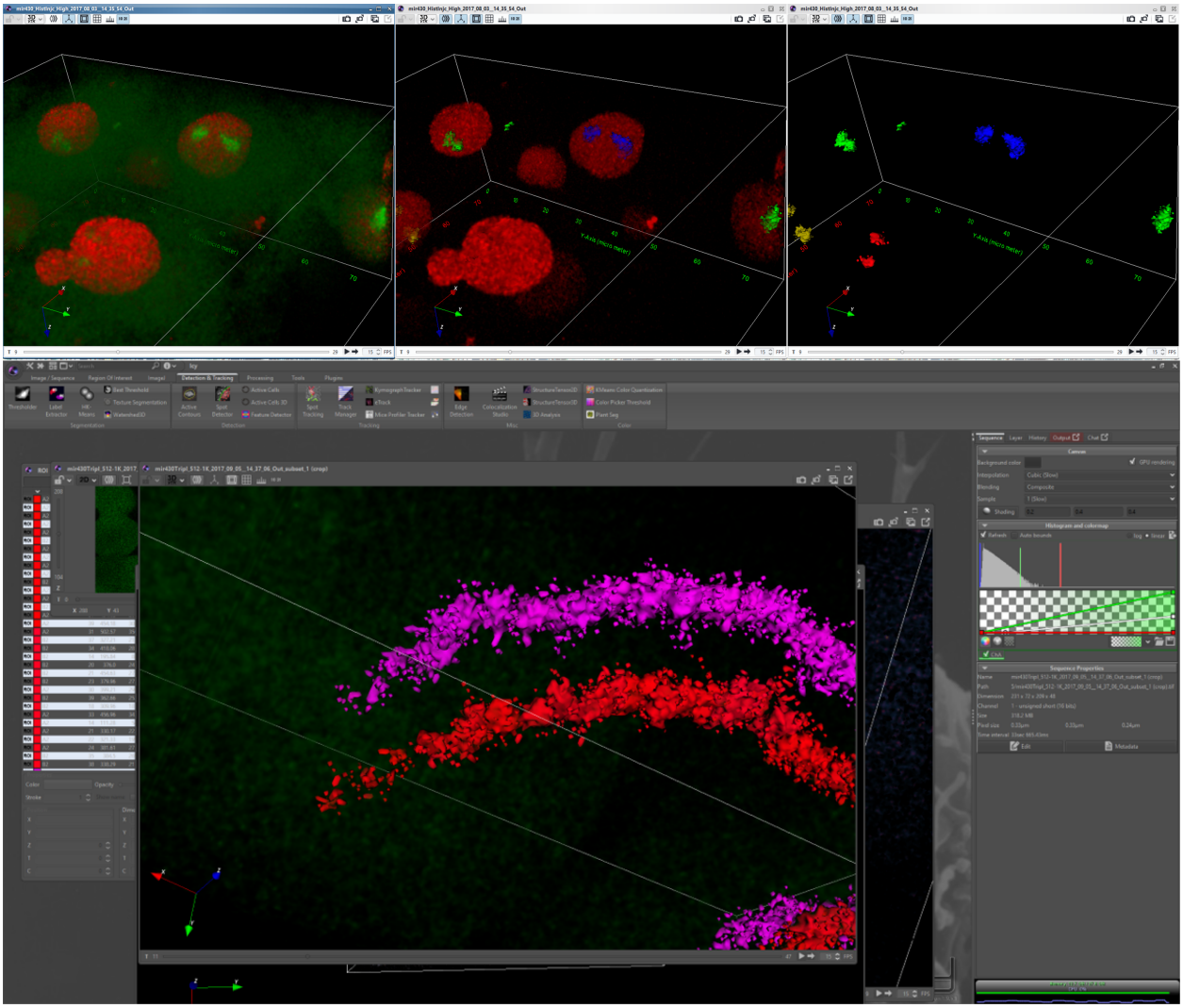


Figure 4.6: **a.** Using 3D spot detection and thresholding to isolate bright regions of miR430 localisation. Visualisations in Icy Bioimage 3D VTK viewer. **Top left**: Raw image of 3D airyscan output being visualised in Icy VTK 3D viewer. *miR430* MO Transcription signal in green. Red is histone, nucleus signal. Thresholding has been applied to only show nuclei and transcription signal in field of view. **Top middle:** using more aggressive thresholding on green channel to remove background green fluorescence. Icy spot detection has been run to create 3D ROI's, in green and blue. **Top right**. isolated transcription signal sites (multicolours). Raw image data has been hidden to only show 3D segmented regions of interest. **Bottom** Screenshot of Icy Bioimage workspace showing tracking of dual miR430 transcription signals (Red and purple) from within the same nucleus. Nucleus signal is hidden to expose *miR430* MO channel only. ROI's of segmented transcription signals were visualised in 4D space, showing the progression of growth and decay of the signal.

4.2.4 Initial observations: Transcription of *miR430* manifests as dual hyper intense foci

The very first inquiry I made of the light sheet imaged data was how does the transcription signal present in imaging? The first step of analysing MO signals in embryos was to perform a basic characterisation of how the active signal presents in imaging data. The aim of this analysis was to gain an overview of how MO signals appear in nuclei, and how this is visible across the embryo.

Global observations of the *miR430* MO signal in embryos show a transcription signal which either appears as dual signals, or evolves into 2 signals during a cell cycle. The proportion of nuclei showing dual dots increases with successive embryonic stages. These hyper intense regions of nascent transcripts will be termed **transcription foci**.

The basic organisation of transcription foci is 2 foci per nucleus, with each one possibly corresponding to a single allele. The quantification of the proportion of double focus nuclei showed an initial low level of double foci, which increased with successive cell divisions. This gradual increase in transcription detection over successive embryonic stages opened questions regarding how transcription was activated over time across the embryo (4.7). Although seldom observed, nuclei bearing single foci were observed in the timeframes analysed in this figure. These nuclei were not counted, and as the occurence was so rare it is not deemed to have had an impact on analysis here.

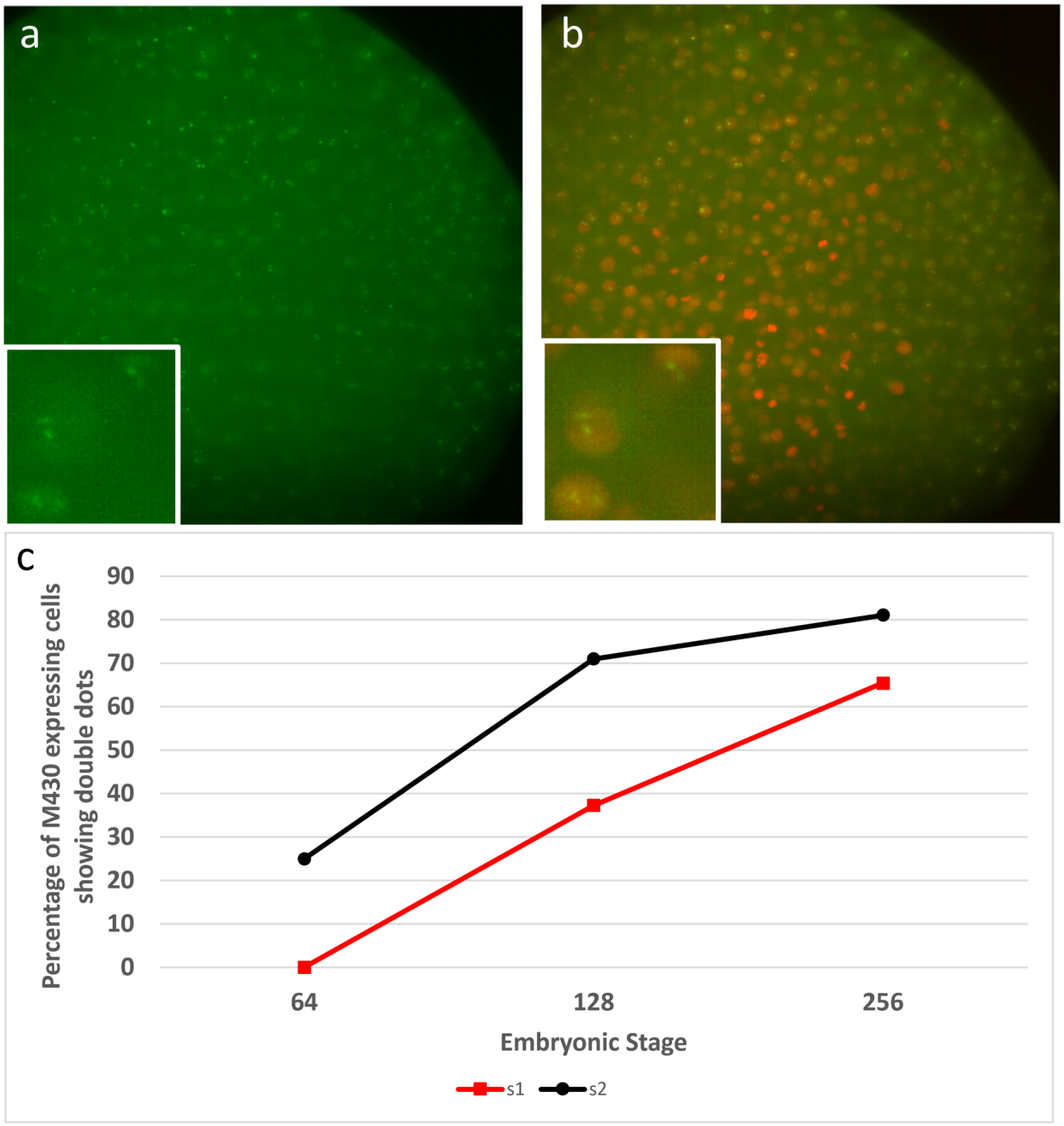


Figure 4.7: **Top left.** light sheet image (maximum intensity projection), whole embryo view, 512 cell stage. Green (*miR430* MO) channel only, showing signal foci. Inset shows zoomed in dual signal per nucleus. **Top right.** same light sheet frame as in top left panel, whole embryo view. Green and red (histone, H2B:mRuby) channels, showing signal foci and interphase nuclei. Inset shows zoomed in nuclei, from same region as inset for top left panel. **Bottom.** Percentage of *miR430* signal expressing nuclei which exhibit a dual dot signal at any point during their cell cycle, for 3 cell cycle stages. 2 lines represent data from 2 different embryos.

Temporal analysis of transcription focus

4.2.5 Transcription expression is switched on asynchronously across the embryo, over multiple embryonic stages

Having established the dual dot nature of the transcription focus, I was able to probe when and where these signals arose across the embryo. To address the question of how early transcription is detected, and how the activity increases over the course of MZT, and prior to MBT, I quantified the levels of transcription activity in an embryo. In this way we could probe the growing wave of ZGA.

Global temporal analysis shows discrete waves of signal in the pre-MBT embryonic stages, with a growing expression time per successive stage and a progressive loss of synchrony for time of signal onset. The number and proportion of expressing nuclei increases with each successive embryonic stage (figure 4.8).

miR430 transcription occurs in waves, with the time of transcription onset showing synchrony. This is in line with the synchronous cell cycles and divisions of the pre-MBT embryo. The length of time transcription foci persist extends with each embryonic stage, and synchrony is steadily lost. At 256 cell stage, some nuclei are still showing *miR340* signal, while new 512 cell stage nuclei are beginning expression. By 512 cell stage the embryo is approaching MBT where cell cycle synchrony is mostly lost, and the timing of transcription signals reflect this. At 512 cell stage we begin to see lengthening of cell cycle duration (Figure 4.12).

The proportion of visible nuclei expressing mir430 transcription never reaches 100% (figure 4.8). In the earliest stages showing transcription (64 and 128), transcription may not yet have been activated across the embryo, but as MBT is approached the proportion of expressing nuclei at a given time point does not breach 80%. At 512 cell stage, the maximum proportion of expressing nuclei at a given time decreases.

These data show a metasynchronous activation of transcription across the embryo, beginning with only a small number of nuclei showing transcription and progressively growing with each stage.

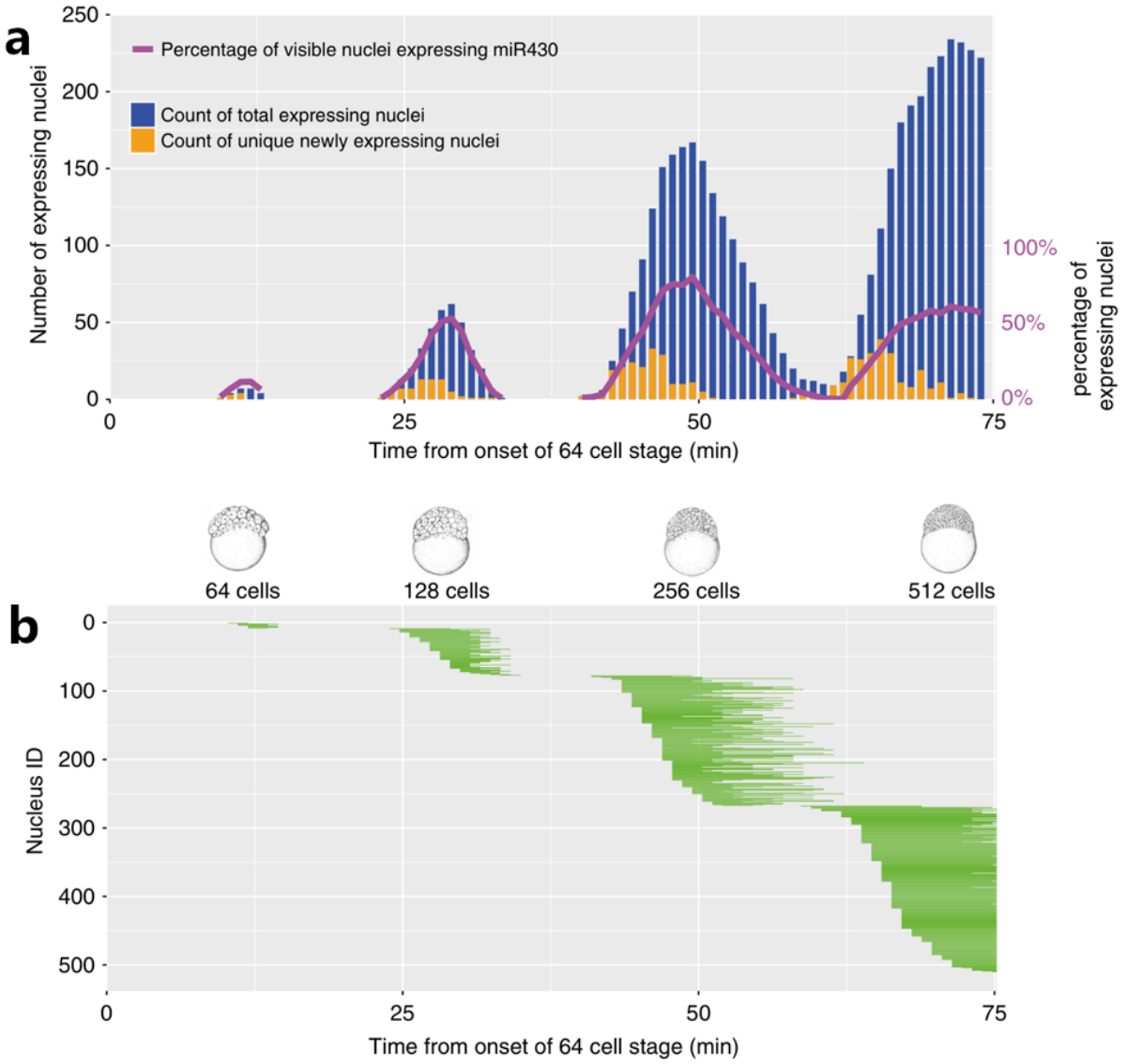


Figure 4.8: **a.** Timing of the onset of gene expression (orange) and number of nuclei (green) showing expression activity, total and new miR430 expressing nuclei per frame. Stages of development are indicated at the bottom of the panels, purple line represent percentage of nuclei expressing miR430 at the time and stages indicated. **b.**Expression span of individual nuclei over time. Each horizontal bar represents the miR430 expression time of a single nucleus. The frame rate (51s) was used as reference for the elapsed time in (b). Adapted from Hadzhiev et al. 2019.

Identification of cell cycle phases and transcription foci duration

4.2.6 Characterisation of stereotypical features of cell cycle and transcription

Identifying when transcription is occurring in relation to the cell cycle required identifying different cell cycle phases from the light sheet data. Using this information, I could find when the transcription signal begins and ceases in relation to different phases.

To build a quantitative reference of cell cycle phase characteristics (in 2D), nuclei were specifically

selected for being exceptionally clear, as based on my analysis in figure 4.4, and matching literature references for cell cycle phase milestones.

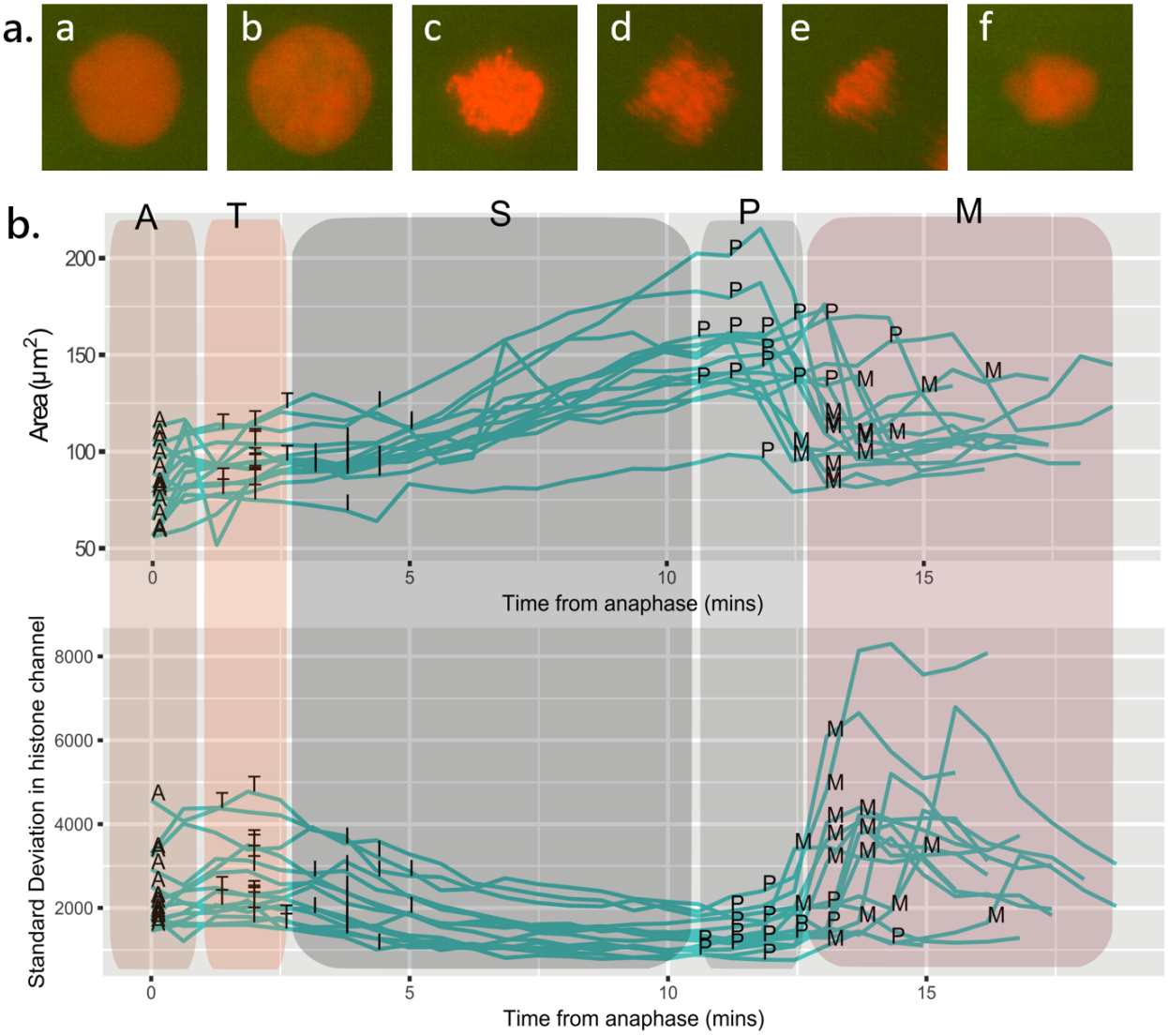


Figure 4.9: **a.** example of reference images used for quantification of cell cycle features. Images of mRuby3:H2B visualised nuclei from 256 cell stage. **b.** Cell cycle reference chart constructed from cell cycle analysis showing characteristic trends and features explored to help inform cell cycle segmentation. Individual lines represent imaging parameters from 2D ROI's generated around 256-cell stage nuclei from lightsheet maximum intensity projection data. Annotations on lines denote manually labelled cell cycle phase transition points, based off comparisons with reference library. Overlaid coloured boxes show overall segmentation into cell cycle phases based off common features in the line graph.
A=anaphase T=telophase S = S-phase P=prophase M = metaphase Figure adapted from Hadzhiev et al. 2019.

Condensed chromatin states are easiest to see, as shown by the spike in average signal intensity. Surface level nuclei tend to provide the clearest signals, particularly when trying to discern prophase where the difference from interphase is subtle. The consistency of the chromatin becomes less smooth as condensation begins. This phase can be difficult to characterise in deeper nuclei due to scattering and due to the very

short amount of time spent in prophase. In the imaging data used to construct this reference, prophase would only last for up to 3 frames. This highlighted a frame rate issue in our data, but with large numbers of nuclei to choose from in the embryonic stages running up to MBT, I was able to determine prophase characteristics.

This reference shows how the area of the nucleus changes with different phases, and how the intensity varies over the course of the cell cycle via the ‘Standard deviation in histone channel’ chart in figure 4.9b. The standard deviation parameter would calculate the standard deviation of the fluorescence intensity in a region of interest. In this case, regions of interest were manually drawn around candidate nuclei over multiple frames, and the specific cell cycle phases were highlighted per frame. The measure of intensity gives an idea of chromatin density in the nucleus, and was selected for its consistent trends across nuclei sampled.

Some subjective oversight was necessary to create such a reference, in the use of judgment to match the light sheet image with reference information to determine cell cycle phases (figure 4.9a), however, this tool provides a set, consistent reference point for producing cell cycle related analysis.

4.2.7 Characterising variations in transcription persistence and cell cycle duration

Using my cell cycle phase reference, I asked where transcription signal placed over the course of the cell cycle. Initial observations showed that some foci would still persist into metaphase, when transcription is thought to be mostly silent. I examined the extent of the duration of the transcription focus across multiple cell cycles to see if it related to the duration of cell cycle phases (figure 4.10).

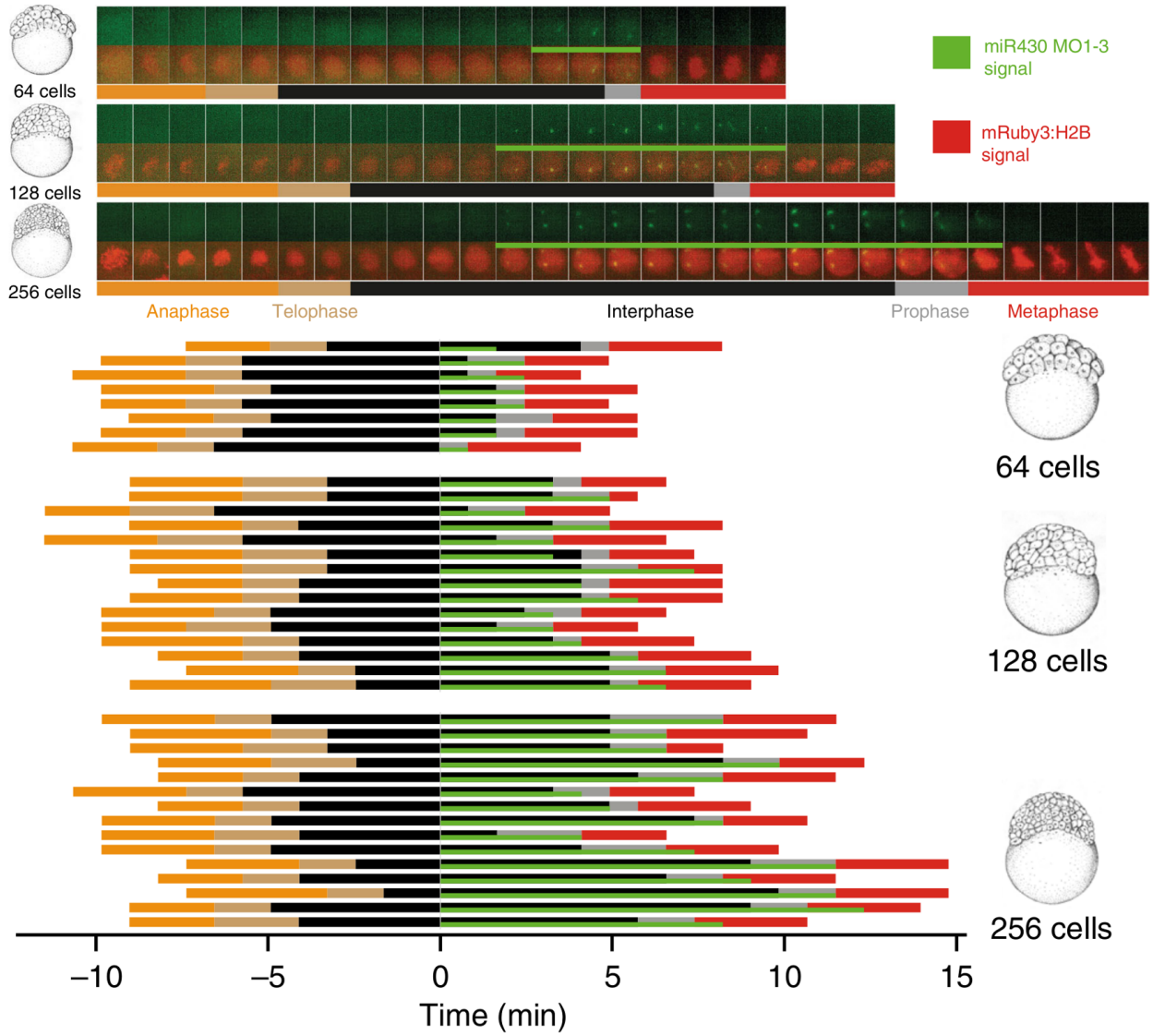


Figure 4.10: **Top.** Timelapses of cell cycle per embryonic stage, showing how cell cycle phases are segmented, based on the reference constructed in figure 4.9. Red channel is histone H2B, denoting nucleus. Green represents *miR430* MO. **Bottom.** Cell cycle phase segmentation as shown in top panel. Horizontal bars represent a nucleus each, they are aligned to onset of detection of miR430 expression and ordered according to length of miR430 expression and cell cycle length. x-axis shows time relative to initial detection of *miR430* signal in minutes
Colour code beneath cell cycle timelapse.
Figure adapted from Hadzhiev et al. 2019.

Using the reference constructed in figure 4.9, I probed cell cycle progression and transcription occurrence

in granular detail. In 128 and 256 cell stages, *miR430* transcription focus was first detected midway during S-phase, and continue into prophase. In 64 cell stage, the transcription focus appeared during the end of S-phase. This may be due to detection limitations. Cell cycle and *miR430* transcription lengths were also seen to increase.

Analysis of when transcription is detected relative during the cell cycle shows that transcription signal is detected part way through S-phase. The timing relative to onset of S-phase gets earlier with later cell cycles. There is variability in when the mir430 transcription focus is first detected within embryonic stages. Additionally, the transcription signal would often extend into the mitotic phases. The duration of transcription appeared to follow the length of the cell cycle, and this offered us initial evidence of a relationship between cell cycle and transcription (figure 4.10).

Cell cycle phase analysis showed transcription focus persistence into prophase, and sometimes into metaphase. Transcription is known to stop prior to the mitotic phases of the cell cycle as chromatin becomes condensed and closed off to transcription machinery. This suggests that the transcription signal being detected is an accumulation of RNA, rather than a direct measure of the process of transcription. Emerging literature suggests a low level of transcription may be maintained during mitosis (Palozola et al. 2017), and this may be a contributing factor to the persistent transcription signal seen during mitosis.

Elongation of cell cycle and transcription focus presence

4.2.8 Cell cycle length is not significantly affected by histone and morpholino injection

As this investigation used a mixture of transgenic histone lines and wild types with histone injected to be able to visualise nuclei, we were concerned that adding additional histone may influence MZT and disrupt cell cycle. Although we were injecting Histone H2B into the yolk for cellular uptake, and the primary candidate for genomic repression is H3, we deemed it cautious to check the effects on cell cycle. We measured whether injected histone interfered with baseline cell cycle length, measured using transgenic fluorescent histone line (see methods). Furthermore, we also tested whether MO uptake would also interfere with cell cycle by comparing cell cycle lengths from embryos injected with MO, and embryos injected with water (figure 4.11).

There was no significant difference between histone injected and transgenic only cell cycle lengths across 4 embryonic stages. Furthermore, inspecting the effect of introducing MO into the embryo showed no significant differences between the cell cycle lengths in embryos injected either with water or MO. This outcome suggested MO or histone injections were not likely to cause changes to cell cycle.

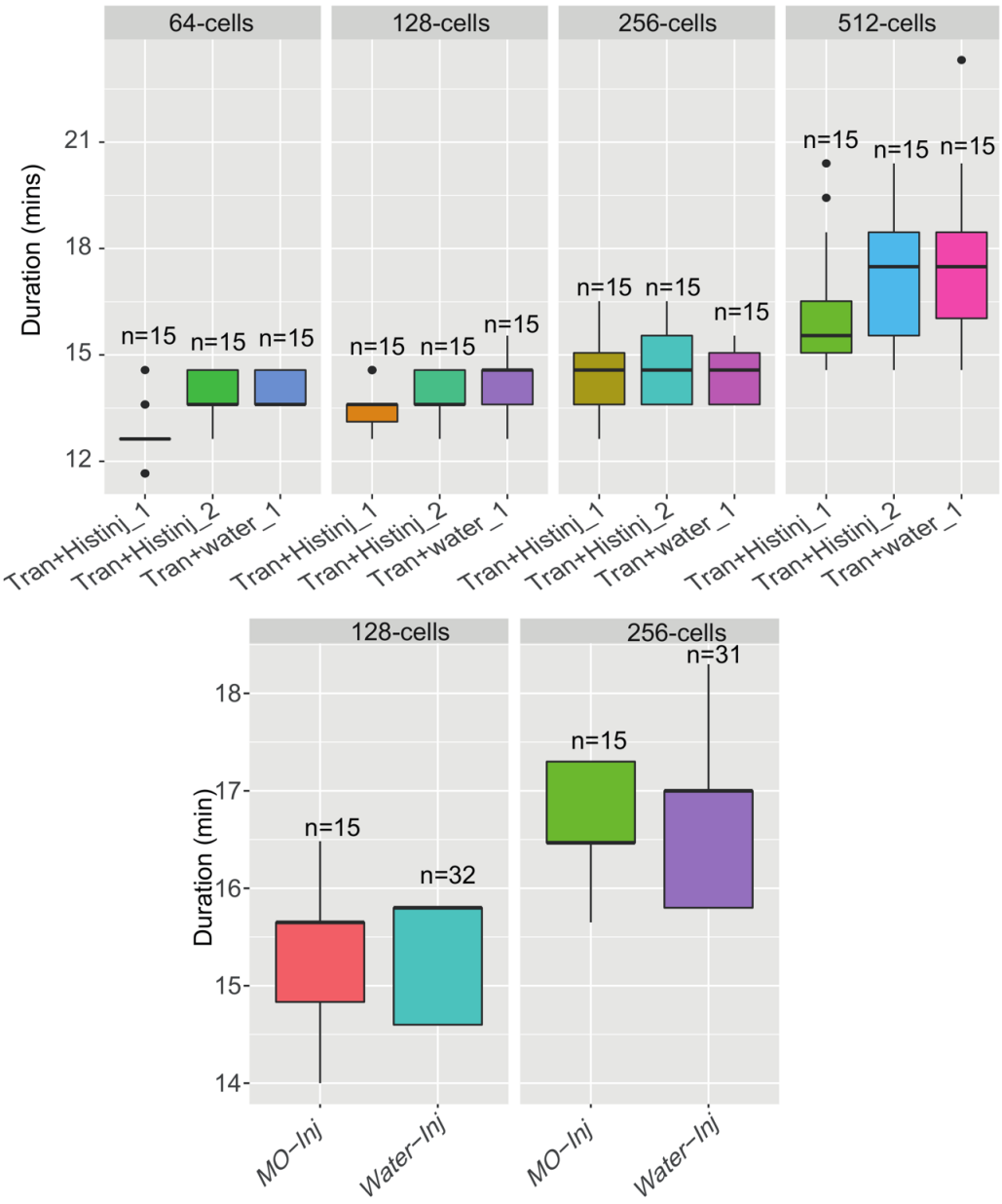


Figure 4.11: **a.** Cell cycle lengths measured from anaphase to anaphase, based on LSFM MIP data. 'Tran' denotes transgenic, 'Histinj' denotes histone injection group. Each box shows duration data from different cells within a single embryo. Embryonic stages denoted in grey headers. **b.** Box plot comparison of cell cycle durations between MO injected and water injected embryos to assess the impact of MO on the cell cycle. Each box shows duration data from different cells within a single embryo. Embryonic stages denoted in grey headers.
Figure adapted from Hadzhiev et al. 2019.

4.2.9 Cell cycle and transcription progressively elongate with successive embryonic stages

The aim of the following analysis was to establish the relationship between cell cycle length and transcription duration by measuring the lengths of both across embryonic stages. It is already established that cell cycles begin to elongate during MZT, and become fully elongated past the MBT (Vastenhouw et al. 2019, K. N. Schulz and Harrison 2019). Consistent with this knowledge, data analysis on cell cycle and transcription focus signal duration (figure 4.12) also showed successive pre-MBT embryonic stages followed this trend of progressive elongation. We noted variation within embryonic stages, showing a non uniform increase in cell cycle length. We also noted transcription duration appeared to increase in duration also. This data showed concomitant progression of transcription and cell cycle duration with successive stages, furthering support for the linked nature between transcription and the cell cycle.

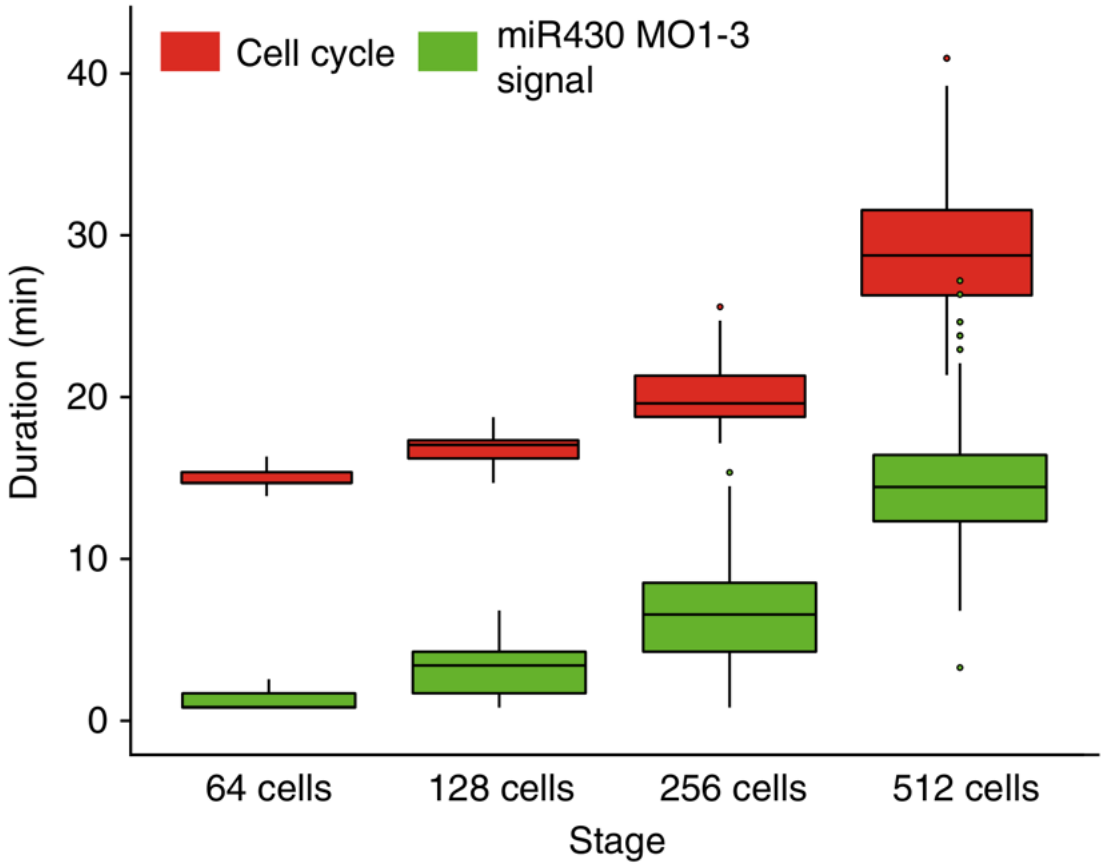


Figure 4.12

Figure 4.12: Box plot quantification of miR430 signal (green) alongside cell cycle length (red). Data from 2 embryos used. Total number of nuclei (cell cycle): 64-cell=17, 128-cell=60, 256-cell=62. All groups were significantly different from one another ($p<0.001$) (One-way ANOVA with multiple comparisons). Middle line=median. whiskers=nearest hinge±1.5 * interquartile range. Cell cycles were measured for nuclei across 4 embryonic stages, across 2 embryos (Cell cycle: 64 cell n=26 nuclei, 128 cell: n=60 nuclei, 256 cell: n=62 nuclei; mir430 length: 64 cell n=17 nuclei, 128 cell: n=120 nuclei, 256 cell: n=335 nuclei). Cell cycle length was measured from onset of anaphase to end of metaphase. Progressive cell cycles show cell cycle lengthening (Median Cell cycle length: 64 cell: 14.7mins, 128 cell: 17.05mins, 256 cell: 19.62mins), and concurrent M430 transcription signal lengthening (Median transcription focus length: 64 cell: 0.85mins, 128 cell: 3.42mins, 256 cell: 6.57mins). Figure adapted from Hadzhiev et al. 2019.

4.2.10 Transcription focus persistence and cell cycle length are correlated

I next asked where elongation was occurring. The aim of this question was to investigate how cell cycle elongation may be increasing transcription duration. The rapid cell cycle of the pre-MBT embryo suggests S-phase would be limiting to transcription as DNA is being rapidly replicated during the short cell cycle, reducing capacity for transcription machinery access to the DNA. I hypothesised that prolonged S-phases would facilitate longer transcription duration as there is increased time for transcription to occur (figure 4.13).

Longer cell cycle lengths were correlated with longer miR430 expression times. Cell cycle length increase appears to be mostly due to S-phase elongation, but time for mitosis is not correlated with cell cycle length. This suggests that cell cycle elongation is principally due to S-phase elongation. This supports the hypothesis that S-phase elongation may be responsible for an increase in transcription duration (figure 4.13).

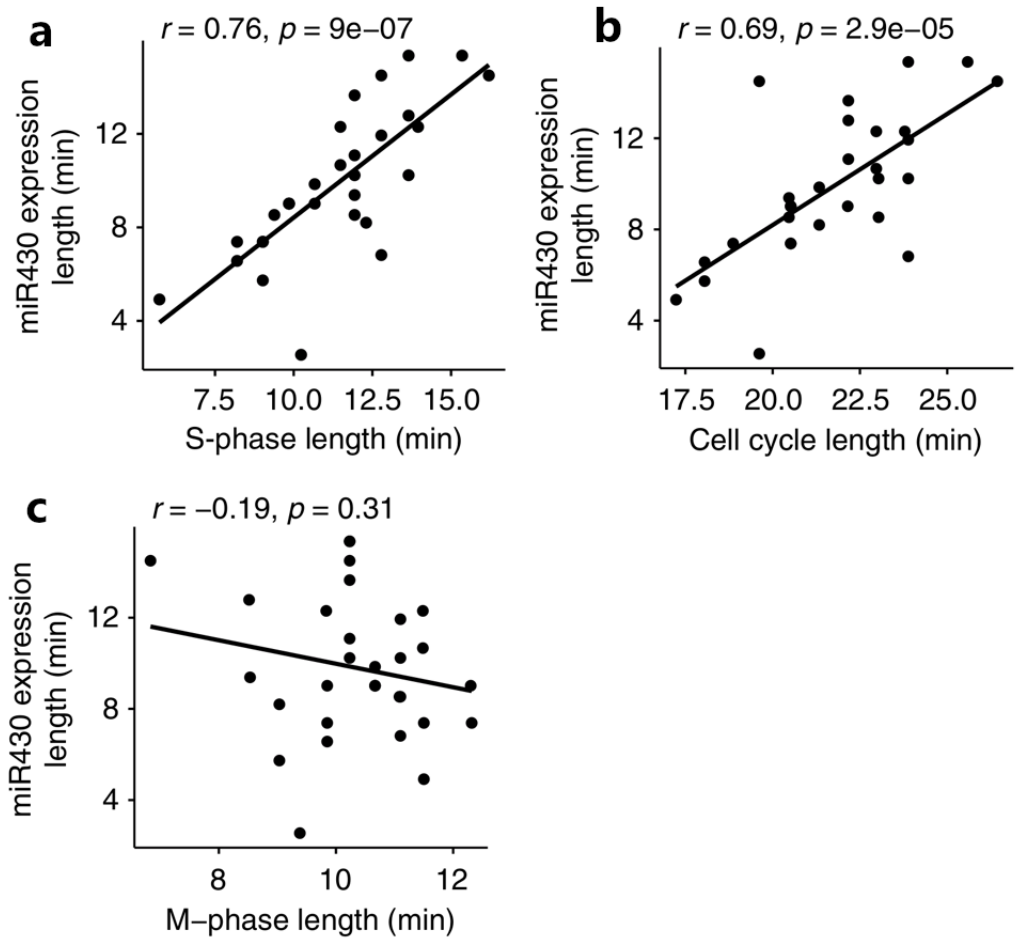


Figure 4.13: Correlations of cell cycle, S-phase and M-phase length with miR430 expression length, at 256-cell stage (2 embryos, 15 nuclei each). The Pearson correlation coefficient (r) and the corresponding p-value (p) are shown on each plot.
Figure adapted from Hadzhiev et al. 2019.

Volumetric analysis of the transcription focus: The transcription focus appears proportionally very large in relation to the nucleus, when viewed in the LSFM MIPs. The notion of a focus of dedicated transcriptional activity within the nucleus is not a novel concept, this idea is posited by TxnFs. The notable difference with established transcription factory theory is the size of the transcription focus that we were detecting; the current model suggests TxnFs have a diameter in the order of 10's of nanometers (Iborra et al. 1996). In contrast, our measurements yield transcription foci with volumes ranging in the 10's of cubic microns (figure 4.14). To validate the appearance of transcription foci in 2D analysis, and to gain a spatial understanding of this feature, I used airyscan confocal microscopy data to perform 3D analysis on the size and life cycle of the transcription focus.

4.2.11 Characterising the size of the transcription focus

The size dynamics of the transcription focus were probed in time series to gain an understanding of how much transcription may be occurring in the focus. This analysis would provide an understanding of how the focus develops, and which sizes it has the potential to reach. I also viewed the 3D shapes of foci over time to observe how it grows and decays (figure 4.14).

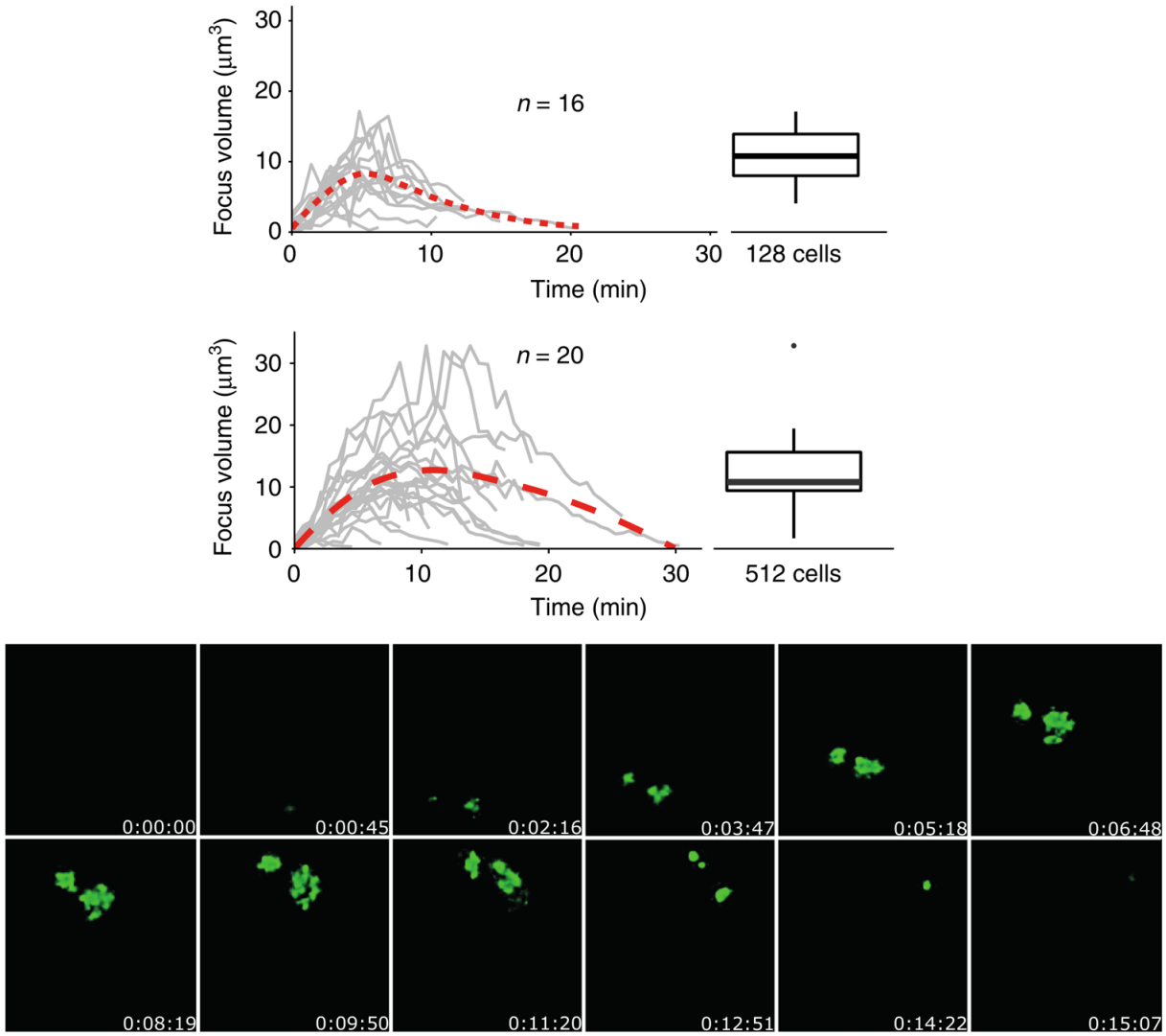


Figure 4.14: **Top.** Charts representing the transcription focus volume over time based on miR430 MO signal, measured using spot detection plugin Icy Bioimage software. Plots aligned at onset of signal. top panel: 128-cell stage, bottom: 512-cell stage. Boxplots on the right show variation in peak volume. Middle line=median. whiskers=nearest hinge±1.5 * interquartile range. **Bottom.** Timelapse showing growth and decay of isolated transcription focus (green), from a single cell during 512 cell stage, imaged using Airyscan microscope. Timestamps show time from onset of signal detection. Focus isolated computationally using Icy Bioimage.
Figure adapted from Hadzhiev et al. 2019.

Computationally isolating the transcription foci from high resolution, high magnification 3D datasets

allowed me to probe the life cycle of the transcription focus. To explore the aspects of the spatial dynamics of the transcription focus, volume data was extracted from 3D ROI's segmented around transcription foci from subsetted datasets. 3D analysis was very RAM intensive, and required individual nuclei or small nuclei clusters to be isolated to allow for analysis, presenting computational challenges.

Focus pairs which merged to form a single focus were excluded from analysis. Foci showed steeper growth and decay phases at 128 cell stage in comparison to 512 cell stage. The plateau phase was also extended in 512 cell stage. An abnormal terminal phase in a single nucleus produced an aberrant tail on the terminal phase for the best fit curve.

The visual observations of focus growth and decay align with the numerical analyses, showing a rapid ballooning phase of growth, followed by plateau and then a decline in size. Foci would often crumble and fall apart as the cell cycle entered mitosis (figure 4.14 bottom). (Figure 4.14 bottom) is supported by a video I generated using Icy biomage. This video was uploaded as supplementary material for Hadzhiev et al. 2019. The supporting video is the source from which the images for figure 4.14 bottom were taken. This video can be found at: `https://static-content.springer.com/esm/art%3A10.1038%2Fs41467-019-08487-5/MediaObjects/41467_2019_8487_MOESM6_ESM.mp4`.

The extent of symmetry between dual foci is unclear. Data from 128 cell stage and upwards consistently shows dual *miR430* transcription foci per nucleus, although volume measurements and visually inspecting 3D segmentation show that some nuclei have a disparity between focus sizes within them, while other nuclei show 2 similarly sized foci.

4.2.12 The transcription focus occupies a histone deficient region within the S-phase nucleus

Further to characterising the size and life cycle of the transcription focus, I next asked how this large feature exists alongside chromatin, which occupies the majority of the nuclear volume. Was the transcription focus a diffuse but heavily concentrated cloud of RNA, or did it occupy exclusive space within the nucleus? I investigated this by close examination of high resolution, high magnification 3D LSFM data of nuclei visualised using a Proliferating Cell Nuclear Antigen fusion protein (mAPple:PCNA). I also examined dual channel 3D confocal airyscan images of embryos with histone and MO visualised (figure 4.15).

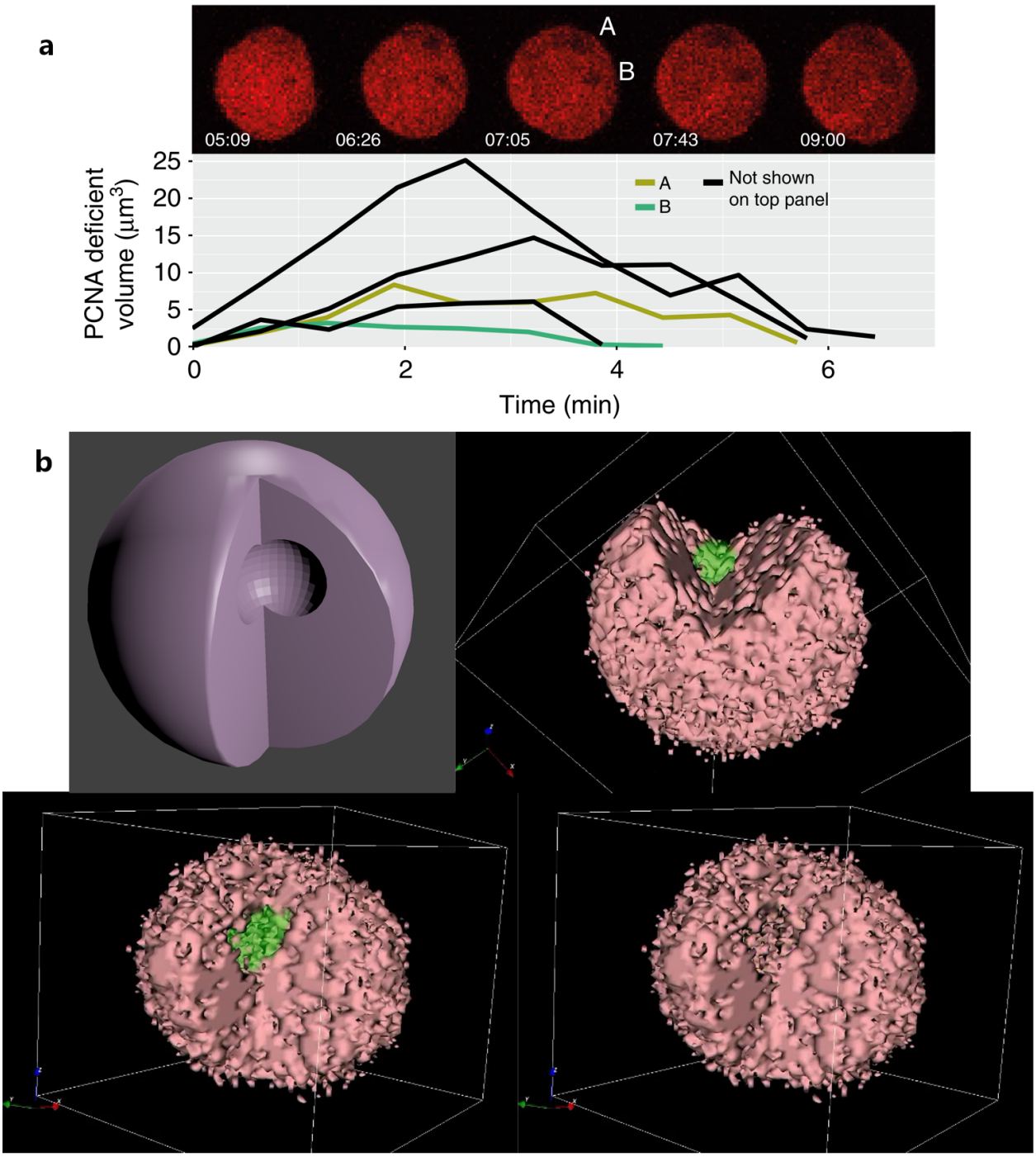

a
A
B
05:09
06:26
07:05
07:43
09:00
PCNA deficient volume (μm³)
25
20
15
10
5
0
0
2
4
6
Time (min)
A
B
Not shown on top panel
b

Figure 4.15

Figure 4.15: **a.** Time-lapse of 512-cell stage nucleus, injected with mApple:PCNA (red). Image shows individual z-slices from within the nucleus exhibiting two PCNA deficient regions. Graph below shows the evolution of the individual depicted PCNA deficient regions over time (A & B), in addition to deficient regions from other 512-cell stage nuclei from the same embryo, detected using spot detection plugin in Icy Bioimage. Time indicates duration from first detection of PCNA signal within the cell cycle in seconds and minutes. **b.** 3D rendered image showing nucleus segmented for histone channel (rendered in pink), with a 90^o spherical wedge cut from the histone ROI to reveal the interior. MO signal (green) occupies a region in the ROI, which was occupied by low intensity histone signal, and thus was not segmented with the rest of the nucleus. **Top left:** 3D schematic diagram showing how the nucleus was computationally sliced for this visualisation, revealing the cavity (white arrow) within. **Top right:** view of the sliced nucleus ROI, showing the MO within. **Bottom left:** Front view of the slice, showing the position of the transcription focus. **Bottom right:** Histone ROI only, with MO removed to show cavity

The role of PCNA imaging is to visualise S-phase. PCNA is a protein that enhances the binding of DNA polymerase to DNA (J. E. Celis and A. Celis 1985, Schönenberger et al. 2015). Accumulation of fluorescent PCNA indicates that the nucleus is undergoing S-phase, a period in which the entire genome is replicated. A nucleus in the midst of S-phase will have its volume visualised by fluorescent PCNA (evident from the left-most panel of figure 4.15a) (ibid.). Analysing nuclei visualised via PCNA showed histone deficient regions where the transcription focus occurred, and the volume of these deficient zones mirror the size, and trends of growth and decay of the transcription focus. Closer examination of the chromatin structure shows histone deficient regions where the transcription signal is detected. These cavities persist into prophase and eventually collapse in metaphase. The peak volume of the *mir430* transcription signal filling these cavities varies between nuclei, but individual signals regularly reach up to 10μm^3 in volume. This represents a sizeable volume within the nucleus. These data show that the transcription focus occupies a chromatin deficient space, suggesting a form of compartmentalisation of transcripts.

Quantifying the presence of mRNA from other genes within the transcription focus: Due to the size of the transcription focus, we questioned whether this contained other early transcripts. I tested the hypothesis that multiple genes were outputting their products into the transcription focus. To address this, I received dual channel imaging visualising the transcripts of 2 genes, *miR430* and *klf17*, and measured whether they occupy the same transcription focus and to what extent this occurs (figure 4.16).

4.2.13 Multiple early wave transcripts are identified in transcription foci

My preliminary analysis was to establish that multiple transcripts occupy the same space in the nucleus, and thus the same focus.

In imaging experiments in which a different MO designed to visualise *klf17* was used alongside miR430 MO, transcription foci of *klf17* MOs were detected, which also colocalised with *miR430*. The *klf17* MO showed similar behaviour to the *miR430*, highlighting distinct, large hyperintense foci within the nucleus.

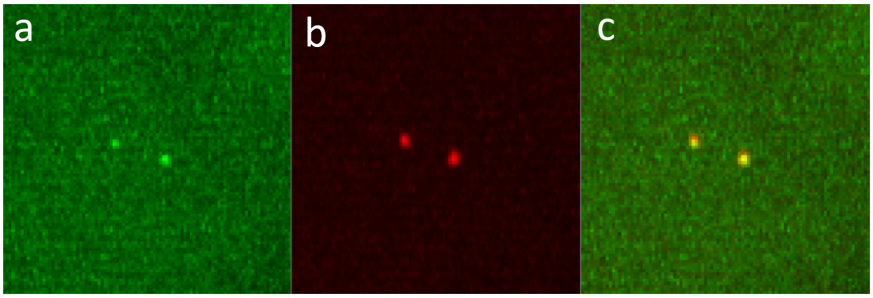


Figure 4.16: Panels are all same frame with different channel settings, from diploid embryo.**a.** Example of klf17 transcription signal, represented by green hyper intense spots. Green channel only. **b.** Example of miR430 transcription signal, represented by red hyper intense spots. Red channel only. **c.** Example of overlay of klf17 and miR430 transcription signals, represented by green/red hyper intense spots, or yellow where signals overlap. Dual channel image, Green and red channels.

4.2.14 The number of transcription foci is increased in tetraploid embryos

Having established that multiple transcripts can occupy the same focus, the next step was to test whether this was a universal phenomenon and quantify levels of colocalization in transcription foci (figure 4.16). Tetraploid embryos were generated to see if the transcription focus was related to allele number, and if the symmetry of transcripts occupying dual foci held in multiple foci.

Dual channel maximum intensity projection data, with red channel representing *miR430* MO and green channel; representing *klf17* MO was analysed in Icy Bioimage. ROI's were generated for both channels, and if transcription signals were seen to overlap within a vicinity equivalent to a nucleus diameter, then that nucleus would be noted as exhibiting colocalization (figure 4.17 a).

Colocalisation was examined on 2 levels; on the nuclear level, and on the individual signal level. Nuclear colocalization sought to quantify how many discrete transcription foci were present per nucleus, and if these bodies showed colocalised transcription signals in them. Colocalisation states were defined prior to analysis.

Individual colocalization sought to quantify how much colocalization was occurring, across the whole dataset. This was a straightforward binary question applied to each focus; If there was a signal, was it colocalised with another gene transcription signal? This would provide a global colocalization rate.

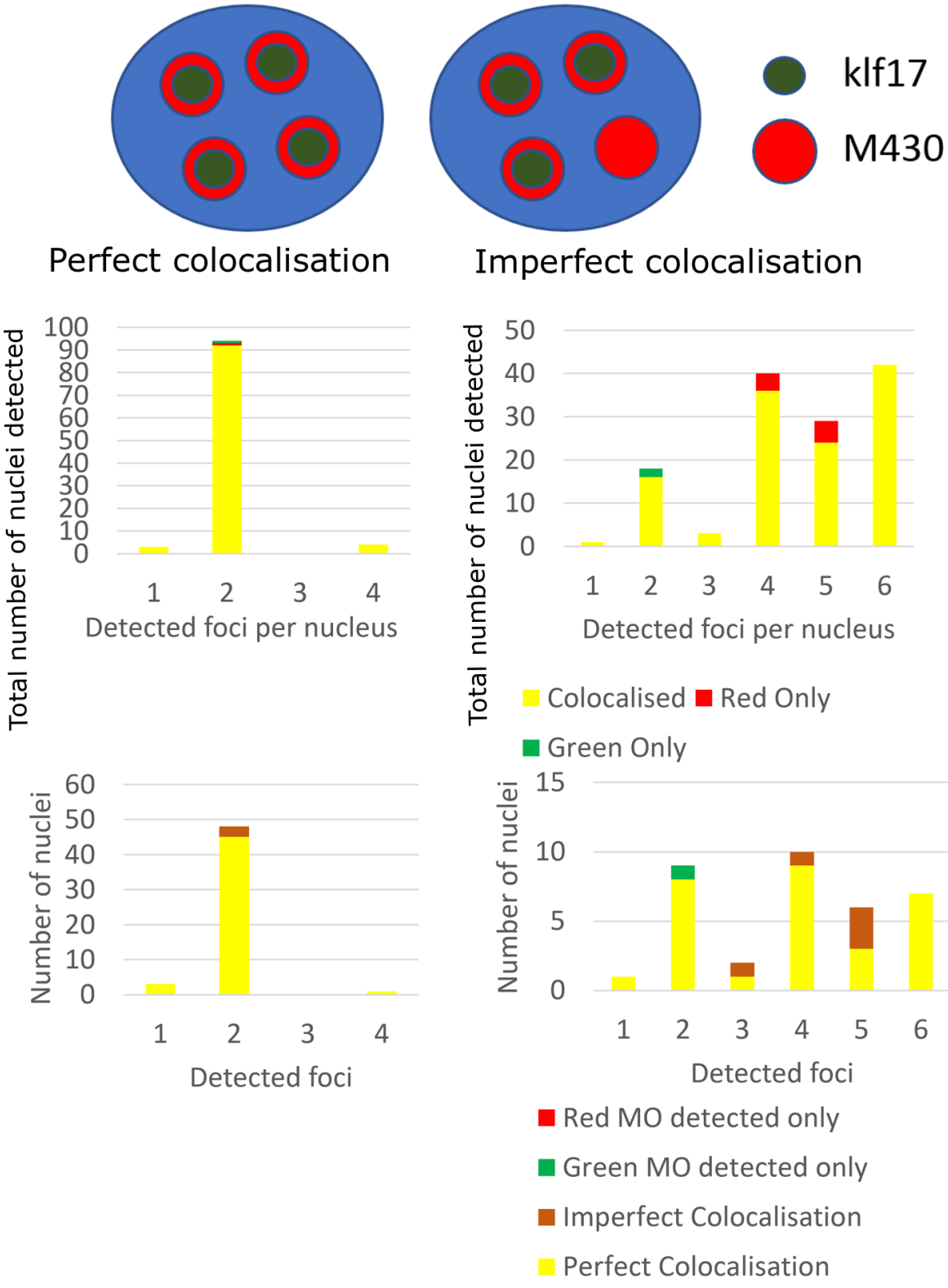

klf17
M430
Perfect colocalisation
Imperfect colocalisation
Total number of nuclei detected
Detected foci per nucleus
Colocalised
Red Only
Green Only
Number of nuclei
Detected foci
Red MO detected only
Green MO detected only
Imperfect Colocalisation
Perfect Colocalisation

Figure 4.17

Figure 4.17: **a.** Determining colocalization states. Individual transcription foci were deemed to be nuclear body-like structures. If 2 transcription signals from different genes colocalised, they were deemed to inhabit the same transcription focus. If a nucleus showed perfect colocalization, then all differential signals showed colocalization. A nucleus showing imperfect colocalization would have a disparity between the number of foci per gene, leaving one lone focus. **b.** Data from 2 diploid embryos. Nuclei grouped into the maximum number of foci they showed in analysis data set (x axis). Chart quantifies the number of transcription foci detected per nucleus group, and the colocalization status of each individual focus. **c.** Data from 2 tetraploid embryos. Nuclei grouped into the maximum number of foci they showed in analysis data set (x axis). Chart quantifies the number of transcription foci detected per nucleus group, and the colocalization status of each individual focus. **d.** Data 2 from diploid embryos. Nuclei grouped into the maximum number of foci they showed in analysis data set (x axis). Chart quantifies the number of nuclei per nucleus group, and the level of colocalization being shown per nucleus. **e.** Data from 2 tetraploid embryos. Nuclei grouped into the maximum number of foci they showed in analysis data set (x axis). Chart quantifies the number of nuclei per nucleus group, and the level of colocalization being shown per nucleus.

For the diploid group, 98% of foci in showed colocalization (total foci counted = 99), and 92% of nuclei showed 2 hubs (total number of nuclei counted = 52), in line with the ploidy of the embryo. 92% of nuclei showed perfect colocalization.

In the tetraploid group, 29% of nuclei showed 4 hubs of which 90% showed perfect colocalization, 26% showed 2 hubs and 20% showed 6 hubs (total number of nuclei counted = 35). In total, 83% of nuclei showed perfect colocalization. 92% of all individual foci (total number of foci counted = 133) counted showed colocalisation.

The presence of other transcripts from other chromosomes colocalising with mir430 suggests the transcription foci may be a storage area for transcripts. *Klf17* mRNA showed a high degree of colocalisation with *mir430*. Manipulating the ploidy of the embryo increased the number of foci, and the high level of colocalisation was maintained. Additionally, *hist2A* transcripts have also been detected colocalising with *mir430* (data not shown). *Hist2A* and *klf17* genes are located on chromosome 5 and 2 respectively.

It is notable that colocalisation is not always perfect, and some nuclei showed only partial colocalisation. This would manifest when 2 foci out of 4 would show colocalisation, yet the other two foci would only exhibit a single type of transcript.

These data from figures 4.17 and 4.16 suggest that the number of alleles may influence the number of foci present, with one focus serving as a repository for multiple gene products.

Analysing the lineages of the earliest transcription expression nuclei: Having explored how transcription takes place during the cell cycle, and documented the relationship with the cell cycle I sought to probe this relationship more deeply. We noted in figure 4.8 that transcription is turned on gradually across the embryo, beginning with only a small number of nuclei, which increases with each division. I hypothesised that the transcription state of the mother may influence the transcriptional state of the daughter cell, leading to mothers with active transcription state early in the MZT being more likely to produce early transcribing

progeny. Based on the findings from Ferraro et al. 2016, transcriptional states appear to be transmitted between mother and daughter cells, shown in *Drosophila*. I sought to investigate this phenomenon in a vertebrate model.

4.2.15 Early expressing nuclei are more likely to produce transcription expressing nuclei

With the temporal analysis of asynchronous transcription activation from figures 4.8 and 4.10 characterising the variation in cell cycle and transcription lengths, and the cell cycle/transcription relationships observed from the analysis shown in figure 4.13, I probed whether there was a pattern to how transcription activates across the embryo. To do this, I took a samples of nuclei from 128 cell stage and followed their progeny for 2 cell cycles up to 512 cell stage to examine the frequency and duration of transcription signal across lineages (figure 4.18).

Cells expressing transcription at 128 cell stage were more likely to produce transcription expressing daughter cells than those that weren't, as shown above where significantly more 256 cell stage cells from expressing parents expressed transcription signal than those from non transcription expressing parents. offspring than ($P<0.05$). By the next generation, granddaughter cells of either group of nuclei had similar chances of expressing transcription with 512 cell stage cells showing no significant difference between 512 cell stage cells from expressing 128 cell stage grandparents and non expressing 128 cell stage grandparents.

Further to the analysis into frequency of transcription presence in descendants from early expressing ancestors, I analysed the signal strength from early expressing lineages, using expression time as a proxy for expression strength. Cell lineages which showed miR430 expression at 128 cell stage produced cells which expressed transcription for longer in comparison to expressing 256 and 512 cell stage ancestors ($P<0.05$ and $P<0.001$ respectively) which did not show transcription expression at 128 cell stage.

Collectively, these data support the hypothesis for the presence of transcription memory in the early zebrafish embryo.

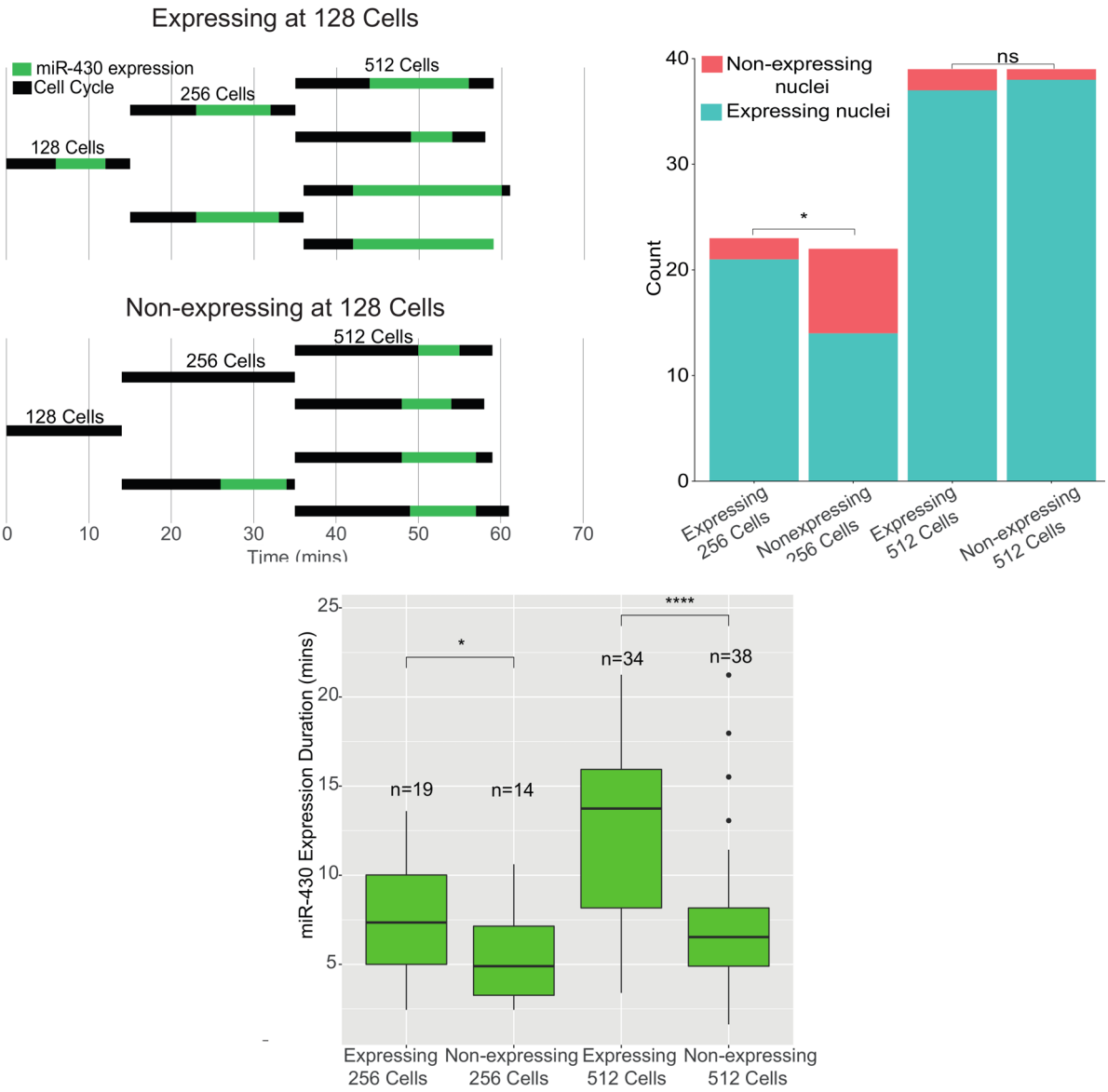


Figure 4.18: **a.** Lineage trees showing a lineage where the parent nucleus showed miR-430 expression (green section of bar) at 128 cell stage (top) and the other where the parent nucleus was not expressing at 128 cell stage (bottom); each bar represents an individual nucleus from the lineage, and how long its cell cycle persisted for. **b.** Counts of expressing and non-expressing nuclei which showed miR-430 expression, per cell cycle stage from each lineage group (lineage groups are based on whether the parent cell showed miR430 expression at 128 cell stage). **c.** Comparison of miR430 expression duration at different embryonic stages from different lineage groups. Groups designated based on their cell cycle and whether they are expressing miR430 at that stage. Figure adapted from Hadzhiev et al. 2019.

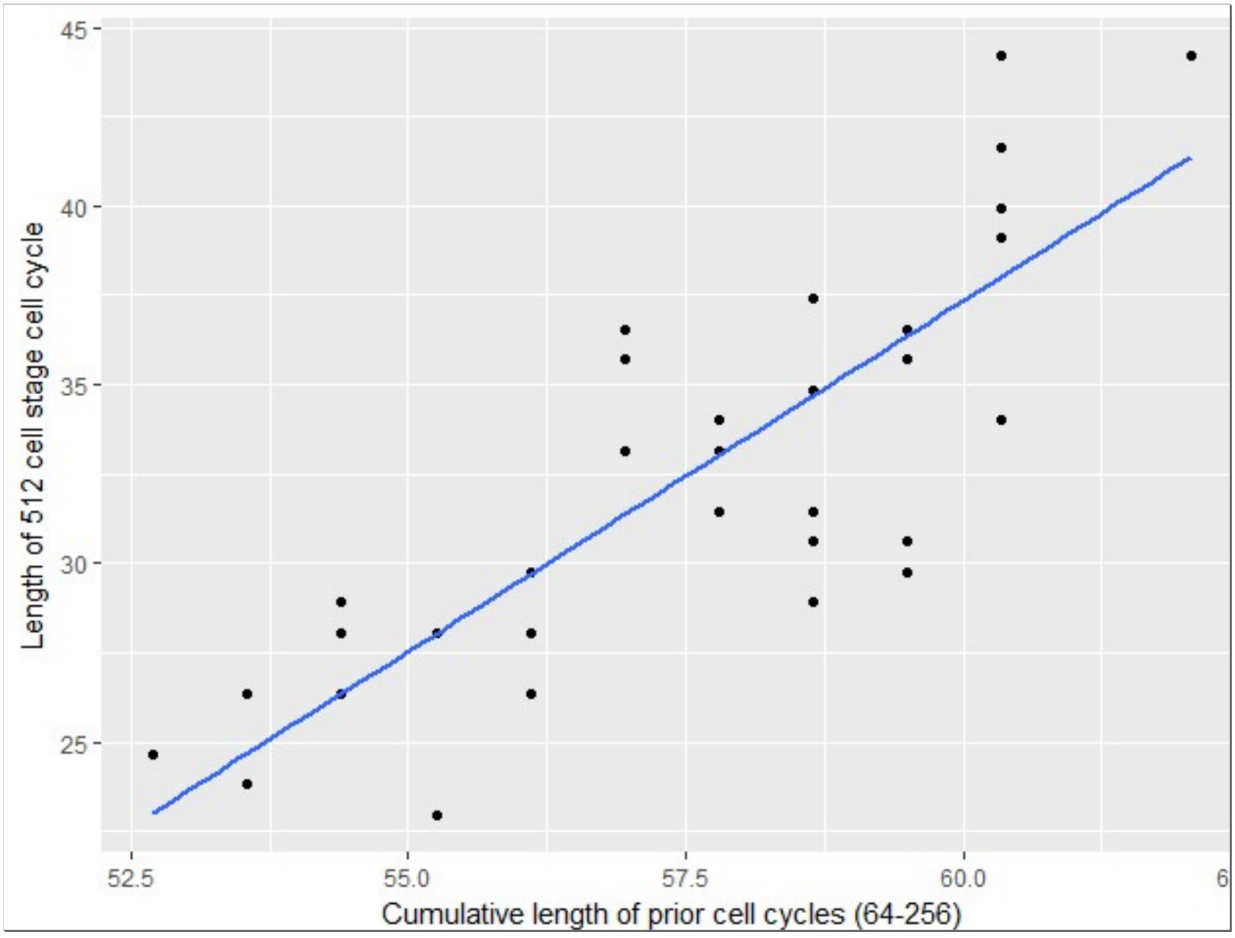


Figure 4.19: Correlation between 512 cell stage nuclei cell cycle lengths and their respective cumulative ancestral cell cycle lengths. 39 nuclei from single embryo, r^2=0.81

4.2.16 Cells with longer cell cycle are more likely to produce daughters with longer cell cycles

The finding of transcriptional memory like behaviour regarding inheritance of transcription duration, coupled with the relationship between cell cycle length and transcription duration identified in figure 4.18 led me to investigate whether the cell cycle also exhibited memory.

I measured the length of cell cycles at 512 cell stage, and compared them to the cumulative cell cycle time of ancestors up to and including 64 cell stage (3 prior cycles). These preliminary findings from a single embryo showed correlation between ancestral cell cycle length and descendent cell cycle length (figure 4.19).

4.3 Discussion

The analysis and observations made regarding the nucleus dynamics prior to the MBT present a dichotomy between the activation of transcription and the desynchronisation and elongation of cell cycles. At first glance there appears to be a link between cell cycle and transcriptional dynamics, prompting the question of which factor is driving this link. How is transcription occurring during rapid cell cycles alongside DNA synthesis, and how does this relate to the events of the cell cycle?

4.3.1 Transcription of select early wave genes manifests as chromosomally associated foci

In experiments and analysis performed by other members of the laboratory, the transcription focus was shown to be abundant in phosphorylated RNA polymerase II, suggesting it as a site of active transcription (Hadzhiev et al. 2019). The appearance of double transcription foci per nucleus is consistent with the diploid nature of the zebrafish. A single focus may represent a gene locus on a single chromosome, in the case of *miR430*, chromosome 4. This is supported by the generation of polyploid embryos, where the number of transcription foci is increased. These embryos did not end up being perfectly tetraploid (confirmed with FACS in an independent experiment by Joseph Wragg, (ibid.)), and this is reflected in the abnormal numbers of foci. Examining of the transcription focus in tetraploid embryos helped establish the consistency of the existence of the transcription focus, as modifications to ploidy were mirrored by changes to the number of transcription foci. Through this we were able to observe that the transcription focus is a consistent product of early genomic function prior to MBT.

4.3.2 Transcription foci appear in a metasynchronous manner across the pre-MBT embryo

Analysis of real time *in vivo* temporal dynamics of transcription activation shows that transcription foci appear asynchronously across the embryo between 64 cell stage and 1000 cell stage. This is consistent with early data from fixed *Drosophila* specimens, showing a gradual increase in early wave transcription expression (Pritchard and Schubiger 1996). The length of time the signal persists for suggests a low level of transcription may be occurring throughout S-phase, indicating concurrent transcription alongside DNA replication. The length of time the transcription focus persists increases with cell cycle length suggesting a relationship between transcription time and cell cycle length, indicating either cell cycle length is being elongated by the presence of transcription, or that cell cycle length is driver making it permissible for transcription to occur, or a combination of both of these ideas.

In comparison to the post-MBT embryo, a comparatively low number of genes is active during the early, minor wave of transcription of the pre-MBT embryo (Heyn et al. 2014, Tadros and Lipshitz 2009).

The entirety of the genome is not replicated simultaneously (D. Prescott and Bender 1962), indicating that transcription may occur in the regions where DNA synthesis is not occurring (D. M. Prescott and Kimball 1961). The pre-MBT embryonic cells may thus be able to fit in the lower number of genes for transcription around DNA synthesis during those short cell cycles. The early wave of transcription shows a gradual increase in the level of transcription with successive cell cycles. The variation in cell cycle lengths may reflect a different complement of genes being activated at different rates in different cells in the early embryo. Consistent with this notion, Boettiger and Levine showed stochastic activation of genes, in addition to a set of genes synchronously activated across all nuclei, in the early *Drosophila* embryo. Such stochasticity of genome activation may explain the variations in cell cycle and transcription length. Transcription interference experiments conducted within our laboratory showed that transcription itself was not driving cell cycle elongation (Hadzhiev et al. 2019), but the cell cycle elongation could be permissible of further gene activation in addition to the set of obligatory developmental genes which are synchronously activated across the early embryo (Boettiger and Levine 2009).

Transcription focus expression lengths show variation between cells, within embryonic stages. S-phase ranges between 7-10mins at 128 cell phase. The growth phase alone of the *mir430* transcription focus volume at 128 cell stage is approximately 7mins long, followed by a plateau and then decline, suggesting that the transcription focus may persist for the full length of S-phase. This is supported by the large scale analysis of transcription expression and S-phase lengths which showed similar timings for each. It is also notable that the very earliest stages of the transcription focus may be missed by lower magnification 2D maximum intensity projection visualisation method due to thresholding and sensitivity. High magnification 3D airyscan data suggests the focus begins earlier than is suggested by whole embryo 2D analyses, based on the length of time the foci persist for in high magnification 3D airyscan data compared to the MIP whole embryo data. This would also support the notion that transcription may begin at the onset of S-phase, as the MOVIE detection we observe from our imaging in early S-phase may represent a relatively matured transcription focus. The imaging would detect the focus when it was of volume, potentially missing the earliest growth phase of the focus. As the focus would decline and degrade, transcription focus would often persist into the mitotic phase.

A potential explanation for the persistence of the transcription focus into prophase and metaphase may be the existence of a low level of transcription persisting into M phase (Palozola et al. 2017). It is unclear whether this low level of transcription would be discernible using the MOVIE tool, and whether the persistent transcripts would include the gene products of *miR430*. Palozola et al suggest that mitotically active transcription would concern genes involved in the maintenance and growth of post mitotic daughter cells, and miR430 is not involved in such function (Giraldez 2010, Giraldez et al. 2005).

4.3.3 The transcription focus bears the hallmarks of being a large phase separated body

A distinct nuclear storage compartment for transcripts is likely to be non-membrane bound as membranes would leave a trace in the imaging, which we have not detected. A phase separation model is a potential explanation for the structure of the focus, in line with current understanding regarding TxnFs. The density of transcription factors (Hadzhiev et al. 2019) and RNA may form hydrogel with different fluid dynamics to the surrounding matrix. This is consistent with the presence of sub-nuclear organelles such as nucleolus and histone locus body, and additional experimentation from this investigation suggest this compartment to be distinct to those bodies (ibid.).

To further support this theory that the transcription focus is a phase separated body, following the morphology of transcription foci during growth and decay shows that it forms as a small, low intensity amorphous cloud, and grows with irregular geometries until it plateaus in volume and then declines in size. At the end of its life during prophase or metaphase, it crumbles, falls apart and dissipates. These dynamics are contrary to the expected growth and decay dynamics of membrane bound organelles which decay by breaking into approximately spherical blebs.

The transcription foci measured in this research are orders of magnitude larger than TxnFs and seem to influence the organisation of the chromatin around them. This may be a unique transient structure in the early zebrafish embryo, related to regulating transcription in the pre-MBT nucleus. Phase separated structures have been posited to provide protective environments for transcribing genes (Hilbert et al. 2018). In line with the conclusions from (ibid.), the transcription focus could protect transcribing genes from the transcriptionally repressive environment of the pre-MBT embryo. A case strengthened by the idea that gene products such as miR430 are important in clearing maternal RNAs in preparation for MBT, and must be free to transcribe to promote ZGA. Further experiments carried out in the laboratory (Hadzhiev et al. 2019) show that the transcription focus is undetectable after 50% epiboly (5.5hrs post fertilisation) at which point the embryo is forming the basic body structure of the larval zebrafish, and nucleoli are an established feature of the nuclear architecture.

4.3.4 Transcriptional states may be inherited

Analysis of transcription activation by following cell lineages over generations, revealed patterns in how the activation metasynchrony manifests. A minority of nuclei exhibit miR430 expression at 64 and 128 cell stage, and the offspring of these nuclei had an increased tendency to show transcription in comparison to nuclei from ancestors which did not express transcription at 128 cell stage. Furthermore, lineages which show their first transcription expression at 256 cell stage had a ‘weaker’ expression, as shown by shorter average

durations of transcription focus detection. The finding that 128 cell stage nuclei expressing transcription are more likely to produce offspring expressing transcription, prior to MBT, hints at a form of memory. Phase separated bodies containing RNA have been implicated in forms of cellular memory (Lau et al. 2020), suggesting a potential mechanism by which the transcription focus may be involved, but this is untestable with the data we have access to currently. The nature of this memory is unclear from these transcriptional findings alone, but analysis of how transcription relates to the cell cycle offers further insights into how this memory may operate.

4.3.5 Cell cycle inheritance may influence transcription

Much like the variations in transcription detection durations, cell cycles became progressively elongated and less synchronised with successive generations, between 64 cell stage and MBT. Within embryonic stages, nuclei showed variation in cell cycle lengths, which was mainly attributed to S-phase elongation. The elongation of S-phase alongside activation of transcription may suggest that the rate of DNA synthesis is slowing to allow for concurrent transcription. Tests into cell cycle duration during MO injected and mRuby3:H2B injected embryos showed that the visualisation manipulations were unlikely to be behind such variation, and additional injected histone was not interfering with cell cycle elongation. mApple:PCNA dataset analysis showed that the PCNA signal does not fade outside of mitotic phases, suggesting continuous S-phase, and ruling out gap phase interference. The length of the mitotic phase did not show correlation with cell cycle length, ruling out elongation of M-phase also.

Nuclei with longer cell cycles also showed longer transcriptional durations, encouraging us to explore whether there is a link between the two. Other data obtained and analysed by Aleksandra Jasciulewicz, a researcher in our laboratory (Hadzhiev et al. 2019), showed inhibiting transcription with triptolide did not affect the progressive elongation of cell cycle lengths, yet increasing cell cycle length with synthetic Chk1 increased transcription focus duration. Taken together, these data support the notion that cell cycle is the principal driving factor for transcription presence and duration, with longer cell cycles offering permissive environments in which transcription can occur with minimal interference with DNA synthesis.

Regions where DNA replication occurs typically show local transcriptional repression, suggesting DNA synthesis can be limiting to transcription (Hamperl and Cimprich 2016, Tsirkas et al. 2022). Our data supports this, with elongated cell cycles producing longer transcription durations (Hadzhiev et al. 2019). In extremely fast cell cycles, the rapid pace of genome replication may only be permissive for a select few key genes to be transcribed. Genome replication in developed animals typically follows a temporal order, with regions of the genome being replicated in a particular sequence (Rhind and Gilbert 2013). In contrast, current evidence suggests pre-MBT DNA replication does not adhere to a strict schedule, with multiple simultaneous

replication events, potentially enabling a faster S-phase. As there is less overall transcription going on in these early cell cycles, a replicating cell can afford to invest in faster synthesis with minimal expense to transcription.

The variation in cell cycle lengths may be a heritable phenomenon due to the potential stochastic variation in genome activation shown by Boettiger and Levine 2009. Genes may be activated in cell primed for particular cell fates, or perhaps due to their position in the embryo. Differing genome activation levels of cells positioned in various regions of the embryo may contribute to coordination of early cell divisions Ding et al. 1998, or the coordination of developmental events such as YSL formation or epiboly. This coordination is explored and discussed in chapter 6.

Analysis of cell cycle lengths within pre-MBT cell lineages showed correlations of longer parental cell cycle lengths with longer descendant cell cycle lengths. Just as we showed with trends of transcriptional activation within lineages, we observed similar familial trends with cell cycle. These cell cycle trends were only from a single embryo so the data is extremely preliminary, yet they corroborated data from transcription duration and activation trends from multiple embryos. As a result, familial cell cycle trends offer a promising avenue of investigation to better understand hereditary phenomena in the pre-MBT embryo. This is explored in more detail in chapter 6.

As the putative inheritance of transcriptional states appears to be dependent on potential hereditary factors regarding the cell cycle, our next priority became deep, granular analysis of inheritance of cell cycle states, and how this relates to developmental organisation in the early embryo.

4.3.6 Technical considerations and next steps

The early conclusions regarding transcription suggest that the earliest transcriptional activity occurs in synchrony, with a steady loss of synchrony with each embryonic stage in line with the loss in cell cycle synchrony. Transcription likely begins at the onset of S-phase, with the transcripts persisting into the mitotic phases. This may be due to transcription products being stored in discrete foci, the number of which possibly match the ploidy of the cell. This focus may be a cavity within the chromatin where transcripts from multiple genes are stored.

Technical difficulties which arise from handling multiple terabytes of multidimensional imaging data made it difficult to achieve efficient analysis of familial trends. Analysis of 2D maximum intensity projection data was a necessary compromise to achieve timely analysis of early embryonic transcription and cell cycle durations, yet 2D images offer challenges in tracking lineages as nuclei move behind each other and converting to 2D only reliably allows analysis of nuclei closer to the surface of the animal cap.

Further limitations were imposed by the manual nature of much of the image analysis, limiting samples

which could be analysed in the given time. These data are no less valid, but require further elaboration from greater and varied sample sizes.

To better understand the potential role of cellular memory in cell cycle elongation, we would need to improve the image processing pipeline and use alternative solutions to the ones thus far implemented.

5 A lightweight segmentation and lineage tracking tool for noisy, low frame rate 4D fluorescence microscopy data

5.1 Introduction

The early stage data regarding correlations between ancestral and descendant cell cycles within cell lineages was promising, but was limited by how the imaging data was manipulated for analysis and by the low throughput of the semi-manual methods of image analysis. This low throughput lead to low numbers of embryos analysed, especially with regards to cell lineage tracking, owing to cell tracking being a time consuming and computationally complex task. MIPs offered a global overview of transcriptional and cell cycle events within the embryo, but the compression of 3D into 2D led to issues in tracking lineages of cells as cells frequently move behind each other during mitosis and this led to only comparatively few lineages being traced in comparison to how many cell are present in the embryo. The lineage tracing method used was also time consuming due to the semi manual nature of tracking nuclei. Existing software solutions struggled to handle the large datasets, and the speed at which nuclei moved during mitosis would lead to issues in tracking, chiefly because such software relies on comparing overlapping features between consecutive frames, which was often not the case with our data due to the frame rate of the video time lapse we found ourselves working with. In this chapter, I discuss a novel software tool I designed and created for the purpose of detecting and tracking nuclei in the developing embryo.

5.1.1 The compromise of light scattering in *in vivo* imaging

A major disadvantage of *in vivo* imaging is the light scattering which occurs deep within the specimen. Light scattering is particularly pronounced deep within the specimens, which may obscure nuclei inside developing embryos, such as the zebrafish embryos analysed for this thesis. Obtaining clearer images from the depths of the embryo would require a compromise on the temporal resolution, rendering the tracking of nuclei during development less feasible. Alternatively, increasing the laser power to greater stimulate fluorophores from deep nuclei risks phototoxic damage and will hasten bleaching (Philipp J. Keller et al. 2008, Jemielita et al. 2013, Dima et al. 2011).

5.1.2 Multidimensional imaging data is computationally unwieldy

Single channel light sheet imaging yields monocolour grayscale outputs in 4 dimensional format, carrying information in dimensions x,y,z and t (time). These outputs are often very large. In the case of the data used

in chapter 4, the datasets fell between 300 GB and 500GB. This is too much for consumer computers to natively handle for manipulations and analysis due to RAM constraints, as conventional image processing tools rely on storing datasets to the RAM for quick and efficient image transformations. Free, open source tools which are available for complex image analysis for 4D data include Fiji ImageJ and XPIWIT. Fiji is a very popular Java based tool due to the wide array of tools it offers for image processing, but relies heavily on storing data to memory for image processing rendering it unfeasible for large scale analysis in the office setting. XPIWIT shows potential for memory efficient image analysis but requires creation of customised pipelines, and the tracking function may struggle with low frame rate datasets.

5.1.3 Watershed algorithm

The process of isolating features within an image, be it 2D or 3D, is known as segmentation. Segmentation relies on detecting pixels which similar characteristics, defined by the user.

The watershed algorithm is a commonly used segmentation tool. It relies on dividing an image into a foreground, which contains the features of interest, and a background, which is to be excluded. Watershed algorithms work by placing 'seeds' in regions of the image which meet user defined criteria which would define them as interesting parts of the foreground (i.e. a feature). A catchment basin is set up around the seed, within which the shape of the feature is determined and the segmentation occurs. The catchment basin is scanned to locate pixels which are similar to the pixels in the seeding region. The similar pixels are grouped together to be defined as a feature.

The watershed algorithm relies on basic thresholding to identify watershed seeds (Preim and Botha 2014), where regions above an intensity threshold are considered features, and anything below is discarded as background. A region around the seeds is determined as the zone within which to search for similar pixels, using a flood-fill algorithm which inspects the intensities of neighbouring pixels, progressively scanning a larger radius until it meets the boundary of the catchment basin.

5.1.4 Feature tracking and aims

A caveat with the fast moving nuclei of the early zebrafish embryo is that they require high frame rate 4D imaging, which can lead to the aforementioned compromises in illumination and image quality. This interferes with standard cell tracking algorithms as cell trackers use a method which inspects overlap between features over time to determine whether they are the same feature moving through space (Xiaowei et al. 2006).

The solution I sought to design was aimed to be able to work in noisy datasets with low frame rates, whilst computationally lightweight, and accessible to use without necessitating the use of supercomputing

clusters. In this chapter, I present a novel segmentation and tracking pipline composed of an original 2D segmentation algorithm, a 3D segmentation tool that was created with the help of a fellow researcher (Dr Andor Magony), and a tracking tool. The aim of this tool is to:

- Effectively segment features in the varying fluorescence environment of the embryo
- Track features which do not spatially overlap between time frames

---

5.2 Results

All laboratory and microscopy work used for the following analyses was conducted by Yavor Hadzhiev. My role in this chapter was designing and creating a piece of software. All programming work described in this chapter was conducted by me, except for a contribution from Andor Magony, who is credited in the section which uses his work.

Software design overview:

5.2.1 Software architecture, displaying how separate functional modules integrate in the workflow

The process of segmenting and tracking light sheet microscopy images required multiple discrete steps. I needed to outline how these tasks would be organised and how data would flow between program modules in the overall pipeline. This would inform how each module output and receive data, and would also be crucial to designing a graphical user interface to bring this together into a software package (figure 5.1).

The pipeline is divided into discrete modules which each perform distinct functions. Except for 3D detection which is built in MATLAB, all modules are built in using the Python programming language. Image data is passed between modules for the individual stages of processing, 2D Segmentation, 3D segmentation, Transformation and Tracking. The pre-processed images are passed into 2D segmentation where they are segmented into binary images which are stacked into 3D features in the MATLAB based 3D feature detection program. The 3D segmentation outputs are data files containing geometries of the segmented objects, which are transformed into tracking friendly format. Quality control is user inspection of outputs, to determine whether the process has executed correctly.

The preprocessing steps are not handled by this software I developed, but are the processes by which raw files from the microscope are made available for the computer. This involves splitting the raw file into 2D images, which can be achieved using Fiji ImageJ. Each timepoint is represented by a stack of 2D slices, with each slice corresponding to a z plane in the stack. These 2D images will be segmented one by one, converted into a binary image and restitched together for 3D segmentation.

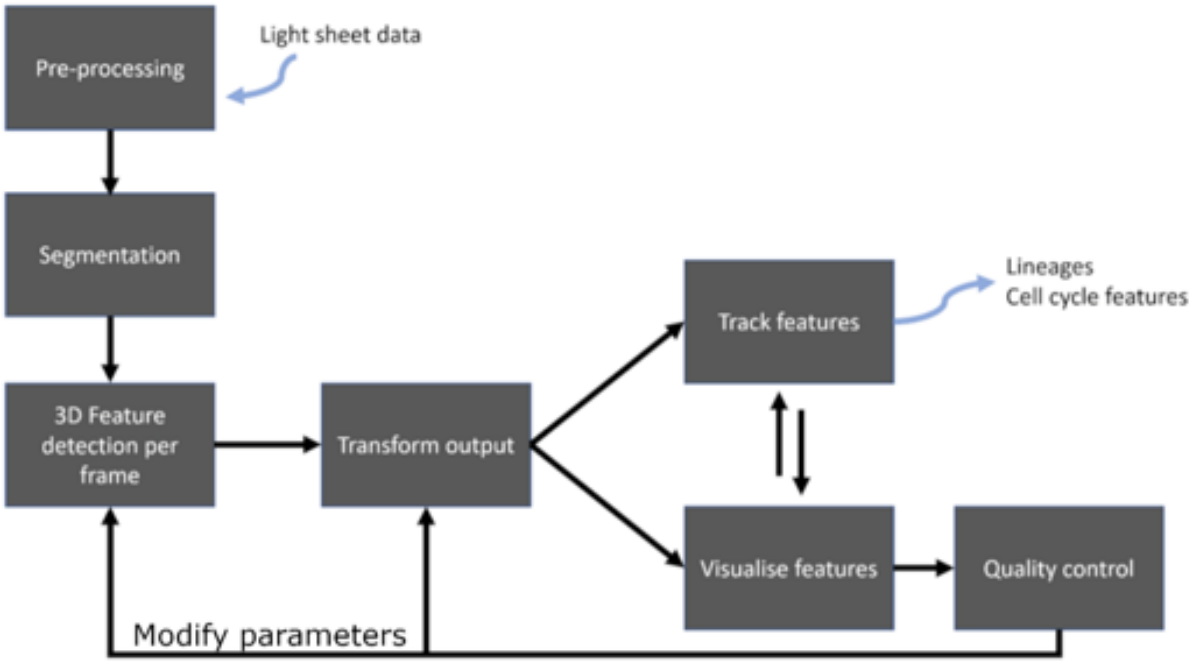


Figure 5.1: Architecture of tracking pipeline: Different modules are dedicated to different phases in processing.

5.2.2 Graphical User Interface design

A graphical user interface (GUI) was designed and implemented using the PySimpleGUIQt Python library to integrate all of the python based modules described in section 5.2.1 to streamline usage of the pipeline and to make it user friendly. Different tabs contain parameter inputs for different modules, and a console window displays the progress of the module being executed. The console is primarily for debugging, and will also display errors which occur during execution of a module (figure 5.2).

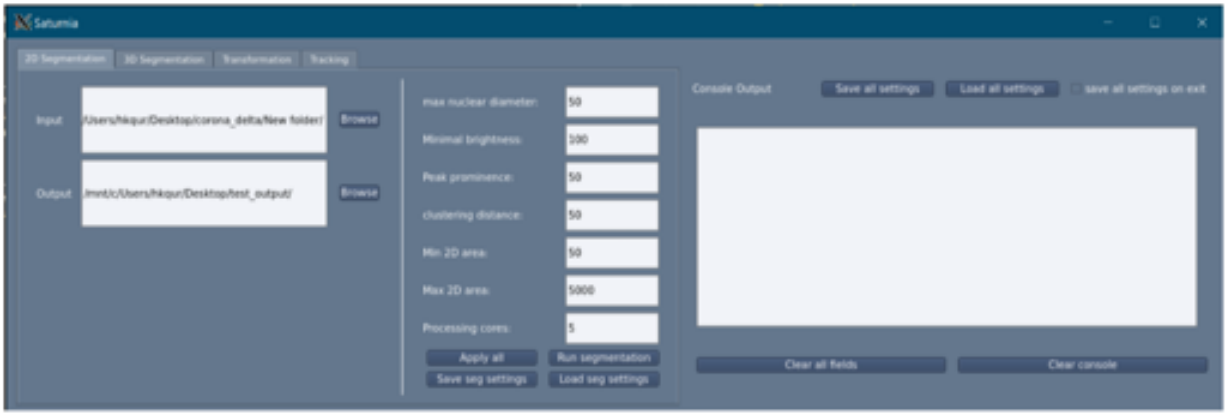


Figure 5.2: Graphical User Interface design: The GUI offers a more user friendly method for running the segmentation and tracking script.

Uneven intensity topographies in light sheet images influenced 2D Segmentation algorithm design:

5.2.3 One-dimensional peak detection locates area of interest

Reliably segmenting light sheet data in 2D, in chapter 5, was a challenging task, owing to the amount of light scattering which occurs within the embryo. I initially sought to understand how fluorescence intensity varies around the embryo, in order to understand how scattering was interfering with image properties. Based on this knowledge I could design strategies to work around it. I investigated the topography of intensity across images, and the effects of conventional thresholding on segmentation of light sheet images (figure 5.3).

Analysing the brightness histogram along a 1 dimensional line taken across a single z slice of imaging data from the embryo at various depths reveals the vast intensity variation across the plane of view (figure 5.3a). The intensity topography shows the embryo periphery is the most dim, with the fluorescence of peripheral nuclei also being dim. Deeper into the embryo, scattering of light reduces intensity and clarity of nuclei, making them more difficult to detect, with nuclei in the centre of the embryo showing similar fluorescence intensities to those on the periphery.

With ongoing embryonic development, the fluorescence intensity decreases, due to bleaching and fluorophore diffusion, presenting challenges for conventional thresholding, which is used in the foreground setup for general watershed methods .

Using conventional watershed seeding methods provided by the MATLAB programming environment revealed issues with the uneven intensity topologies in noisy light sheet data. The conventional watershed methodology in MATLAB was susceptible to noise, and would not detect all of the available nuclei.

The uneven intensity topography of the embryo proved to be an obstacle to segmentation when using the conventional method of global thresholding (figure 5.3b upper/lower). Underthresholding would occur; less nuclei would be detected when a higher thresholding value was used. Peripheral nuclei in the darker zone of the image would be excluded from segmentation as their peak fluorescence was below the set threshold value . At 128 cell stage, only 16 nuclei were detected, and 61 nuclei were detected at 256 cell stage. Overthresholding occurred when the threshold was lowered to include more features, and this modification, while capturing features from the periphery of the embryo, still under detected features. 47 nuclei were detected at 128 cell stage, and 163 nuclei were detected at 256 cell stage. This represents a limit to useful thresholding, as this level of thresholding leads to features in the brightest zone of the embryo being detected as a single feature as whole regions of the embryo are above the threshold will allows capture of peripheral nuclei.

This analysis highlighted how scattering of light made deep embryo features less detectable by lowering the overall intensity environment, and causing features to be missed by thresholding limits, or otherwise

detecting too many objects as features. Variation between embryos may also contribute, possibly with regards to quality of mRuby3:H2B injection and uptake, although this has not been measured.

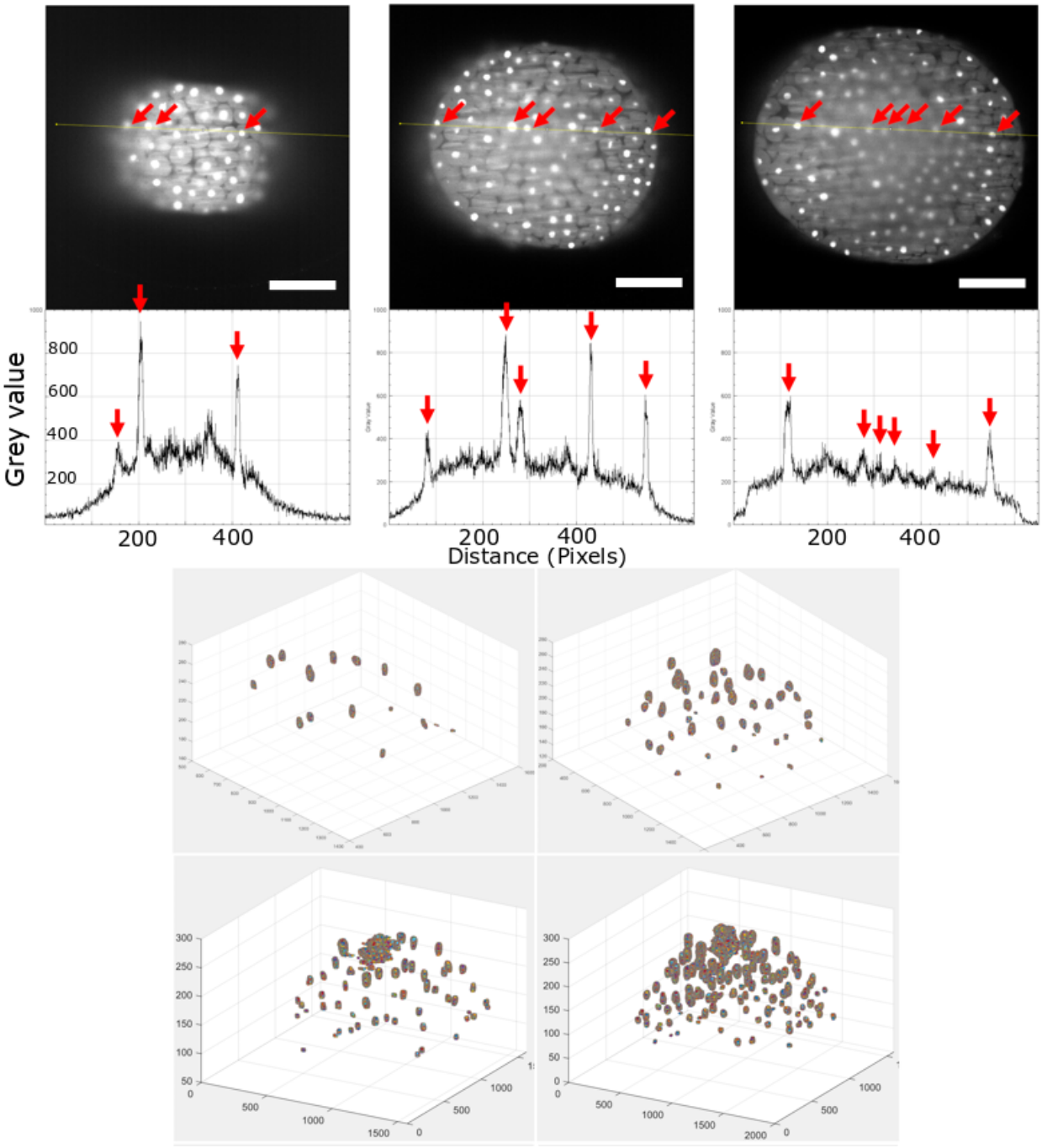


Figure 5.3: **a.** 1D intensity topography (taken along yellow line) of different z slices (upper panels) from progressively deeper (left to right) regions of embryo. Red arrows denote nuclei, and are lined up between the z slice and their graphs. Scale Bar = 50μM **b.** Consequences of 3D segmenting binary outputs of conventional watershed methodology. Left panels show 3D representation of segmented embryo, showing segmentation of nuclei from 128 cell stage and the right panels are from 256 cell stage (same embryo, same time lapse). Axis units are in pixels. **Upper:** 3D segmented feature outputs when using a high thresholding value (overthresholding). **Lower:** 3D segmented feature outputs when using a low thresholding value (underthresholding). Source data from LSFM, nuclei visualised with mRuby3:H2B.

5.2.4 Isolating intensity peaks in 1 dimension forms the basis of feature detection

Having identified how light scattering influenced the detection of objects in the embryo, I designed an algorithm which would be able to work around the vast disparities in brightness, and detect objects with a high brightness relative to their background. This algorithm is part of the overall segmentation and tracking pipeline, the code for which can be found in section 3.4.3. The purpose of this detection was to perform a preliminary sweep for feature candidates, which could produce seeds for watershed algorithm. These candidates may then be segmented in 2D (figure 5.4).

The software reads 2D images from an image stack, and the first step to detecting features involves applying a gaussian blur to enhance contrast and smooth noisy fluctuations in intensity. This highlights peaks related to features and reduces the complexity of peak detection calculations performed by the peak prominences package. The minimum height of a prominence is a user defined input. This figure shows how gaussian blur highlights features against gaussian noise in an image, and how dim and bright nuclei alike are selected against the varying intensity topography of the embryo. This algorithm has been designed to process gaussian blurred images in a row-by-row manner, detecting peaks horizontally per row to generate a 2D map of peaks across the whole image (figure 5.4).

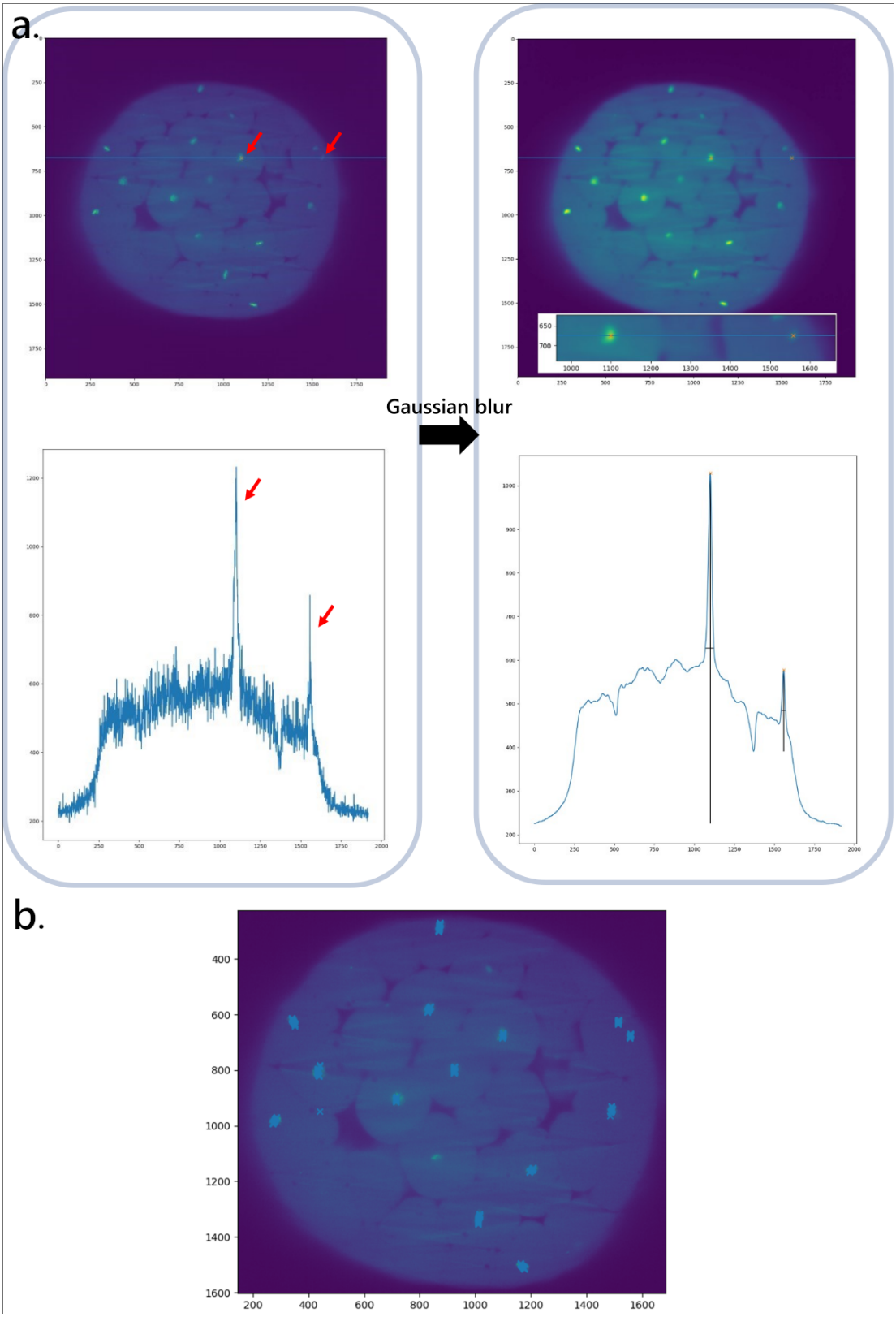


Figure 5.4

Figure 5.4: Isolating intensity peaks in 1 dimension forms the basis of feature detection. **a.** 4 panels showing the effect of gaussian blur on intensity topography in 1 dimension **Upper left:** raw unaltered Z slice. **Bottom left:** 1 dimensional intensity topography of line drawn across upper left. Arrows highlight nuclei which intersect with line. **Upper right:** Z slice post gaussian blur. **Inset:** Zoom in on nuclei of interest which intersect with line. Bottom right: Corresponding 1D intensity topography. **b.** 2D map of detected peaks (light blue crosses) from row-wise peak detection. Peaks overlaid onto original raw image.

5.2.5 Two-Dimensional feature analysis allows for creating watershed seed candidates

The row by row peak detection performed in figure 5.4 produces clusters of peaks across the 2D image, but to create a watershed seeding point, a single point from this cluster must be selected. The next step of designing the 2D segmentation algorithm was to find clusters of peaks within a defined distance, a user defined maximum feature radius value, and to group these peaks as part of single feature. These clusters would be used to generate the seeding point for creating a watershed. From this seed, the flood fill method of watershed could be used (using the standard Python OpenCV2.0 library). Once distinct clusters pertaining to individual nuclei are established, the watershed seed candidate can be extracted to offer the best starting location for producing watersheds.

Intensity peak clusters were grouped using DBSCAN module from the scikit-learn toolkit, a technique used in machine learning to group objects based on similarity between parameters (Pedregosa et al. 2011). In this case, clustering is based on proximity of 2D coordinates. Initially the seeding method was to take the middle most detected peak from the cluster and retain this as the singular seeding point, while deleting the rest of the peaks from the cluster.

The watershedding method required optimisation, as shown in figure 5.6. Initial watersheds were over flowing, suggesting the foreground/background separation was not good enough, producing vast watersheds which covered large portions of the embryo. The first step to removing the overly large watersheds was to remove watersheds which exceeded a certain size. This size was user defined.

I also modified the seed selection algorithm, selecting the brightest peak in the brightest zone of the cluster. This was determined by creating a sliding window algorithm to inspect groups of peaks within the cluster and identify the brightest group on average, and take the brightest peak from this bright cluster as the seed. These changes to the seeding algorithm reduced eliminating erroneous watersheds.

5.2.6 A user defined exclusion zone sets up watershed basin

I observed watersheds would still produce irregular morphologies inconsistent with the shape of the actual feature (figure 5.6 top panels). This stemmed from the crude thresholding method used to determine the basin which could potentially be filled by watershedding. As a result, I opted to use ‘exclusion zones’

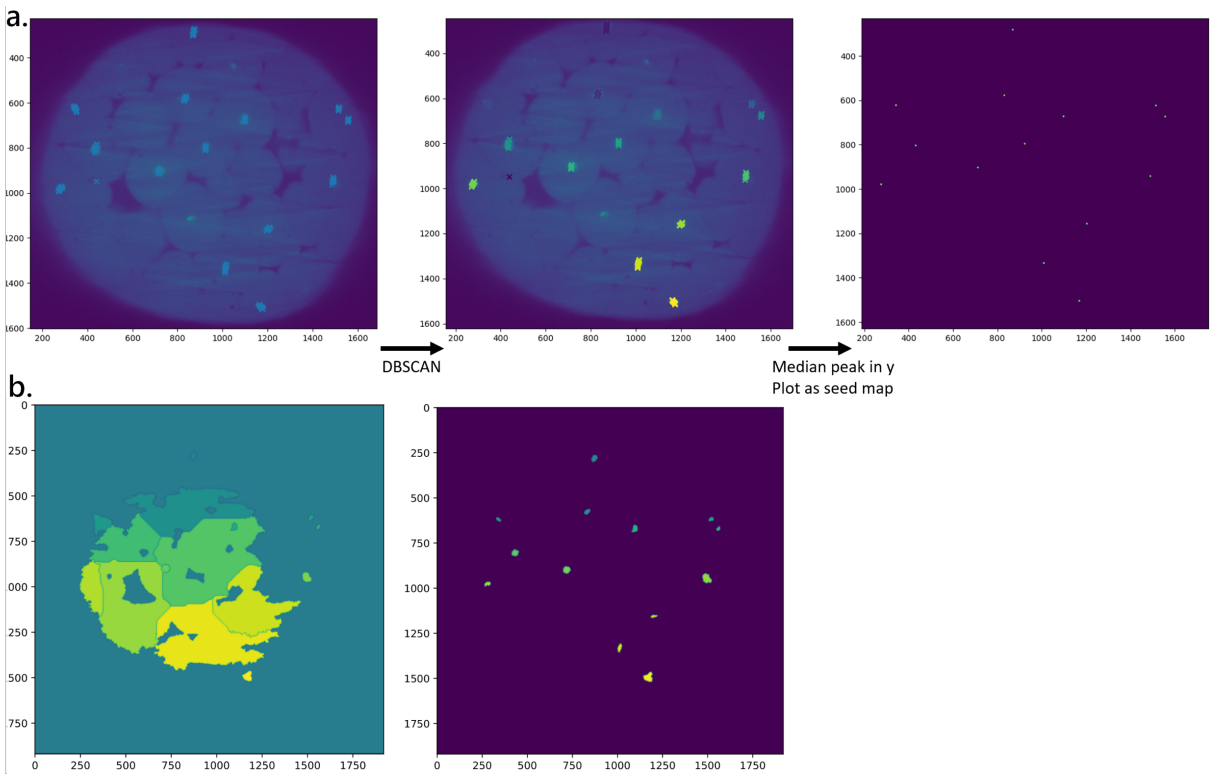


Figure 5.5: Optimising foreground detection via the process of refining watershed seeding.**a.** Process of generating watershed seeds from row by row 1D feature detection. **Left:** All detected feature candidates are annotated on Z slice. **Middle:** Clustering algorithm (DBSCAN) sorts detected feature clusters into distinct groups (visualised with different colours for different clusters). **Right:** Candidate watershed seeding points, selected from scanning for median brightest peak in cluster. **b.** Progression of filled watershed outcomes, showing progression of improved watershed seed placement. **Left:** First watershed outcome, with no watershed size filtering or refinement of watershed seed placement. **Right:** Watershed outcomes with refined watershed filtering and size filter.

(figure 5.6) around seed candidates into which a feature could grow. This would limit the size of resultant watersheds. The size of this zone would be approximately the same size as the user defined maximum feature diameter. This method produced false positive detections, which would be filtered out using local environment sensing (figure 5.7).

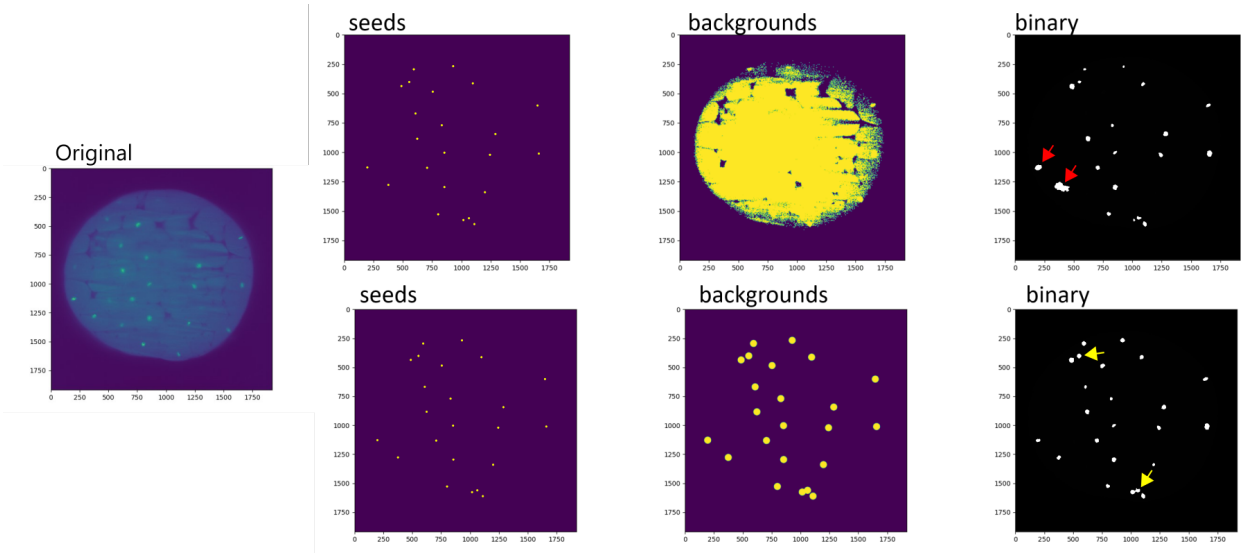


Figure 5.6: **Left.** Original z slice used for testing watershed. LSFM data, green foci are nuclei visualised using mRuby3:H2B (printed in false colour). **Top.** Standard thresholding based foreground separation method. **Left:** Watershed seeds, represented by yellow dots. **Middle:** Foreground watershed basin (yellow) set up by basic threshold of z slice image. **Right:** Resulting binary outline of detected features. Red arrows highlight improperly segmented nuclei, producing inaccurate morphologies. **Bottom.** Exclusion zone determined foreground separation method. **Left:** Watershed seeds, represented by yellow dots. **Middle:** Foreground watershed basin set up by setting up exclusion zone of user defined size around detected features. **Right:** Resulting binary outline of detected features. Yellow arrows highlight false positive detections.

5.2.7 Local environment sensing on detected watersheds removes unwanted features

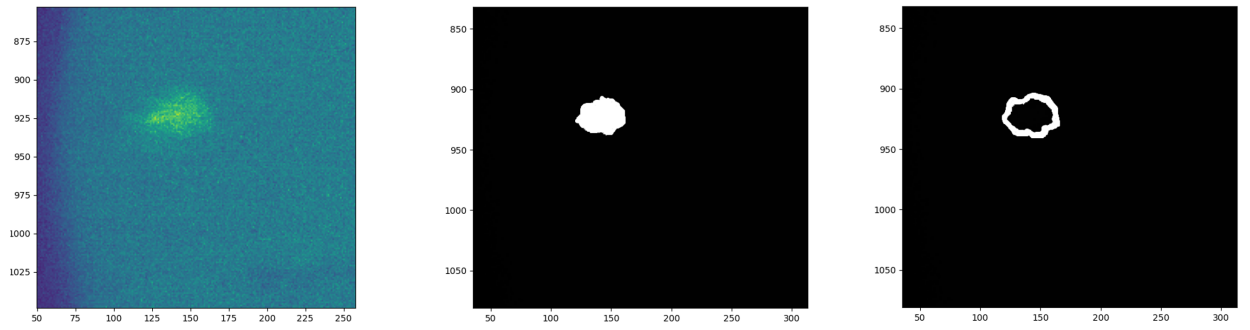


Figure 5.7: Inspecting the immediate environment around detected features to determine if it is a nucleus. **Left:** Raw image of nucleus (zoomed in). **Middle:** Binary segmentation of nucleus from left, following watershed algorithm. **Right:** Corona region created via OpenCV dilation, to inspect for brightness, for comparison with intensity within feature.

To ensure that random fluctuations in unbound fluorescence are not detected as features, a final step of feature validation was designed (figure 5.7). A corona around a feature was generated by using the dilation tool from the OpenCV2 Python library. To isolate the feature corona, the original feature is subtracted from the dilated image. The average intensity of the pixels in the corona, representing the immediate environment around a watershed, was compared with the average intensity of the pixels within the watershed. If the

difference between the 2 regions exceeds 10% intensity, then the watershed is retained, otherwise it is deleted. 10% was selected to allow for detections in darker regions with lower intensity differentials between background and foreground, but this number can be modified by the user depending on their dataset.

3D segmentation process of nuclei: The 2D segmented images are analysed as whole stacks in the MATLAB based 3D segmentation software. The 3D analysis base of this software was designed by Dr Andor Magony, but I modified and customised the base code to work with datasets like our own, through parallelising segmentation across multiple cores, reading multiple time points and to be able to read intensities from the original raw data file to increase the metadata per detected feature (Github URLs for segmentation and tracking script linked in section 3.4.3).

5.2.8 3D segmentation reveals the expected number of features at specific embryonic stages

Our novel segmentation tool follows the expected value of features for given developmental stages (C. B. Kimmel et al. 1995), and shows superior detection over established 2D detection tools, such as the ones used on MIP based data (as used in section 4.2.2). The value of comparing with the 2D detected data is that as there is human validation at every step, we have confidence that it is detecting synchronised mitotic events accurately. That our 3D tool closely follows these trends in nucleus counts is a positive validatory step which suggests that our feature detection shows accuracy in feature detection. Notably, multiple factors contribute to the number of nuclei detected by the segmentation algorithm not matching precisely with the expected values of cells at a particular embryonic stage. For example, at 256 stage, the number of detected features increases up to 256, and then increases beyond 256 features as 256 cell stage progresses. Of note is the loss of cell cycle and division synchrony, where mitosis will be occurring over a prolonged period of time. This logic may also apply to why fewer than 512 nuclei are detected at 512 cell stage, as divisions may not be complete from the previous embryonic stage. This phenomenon is also visible in manually counted nuclei from MIPs in figure 4.8 in chapter 4. Another issue is the quality of imaging, where imaging does not capture all nuclei present in the embryo, particularly nuclei deep within the embryo. Light is too scattered within the deepest parts of the embryo to allow any kind of detection.

The over detection of features, as noted from the 256 cell stage in figure 5.8, was noted to be false positives, biologically inert grains of fluorescence. A filter was designed within the transformation step (figure 5.1) to exclude these features from tracking and analysis, which works by excluding features below a specific volume (set by the user).

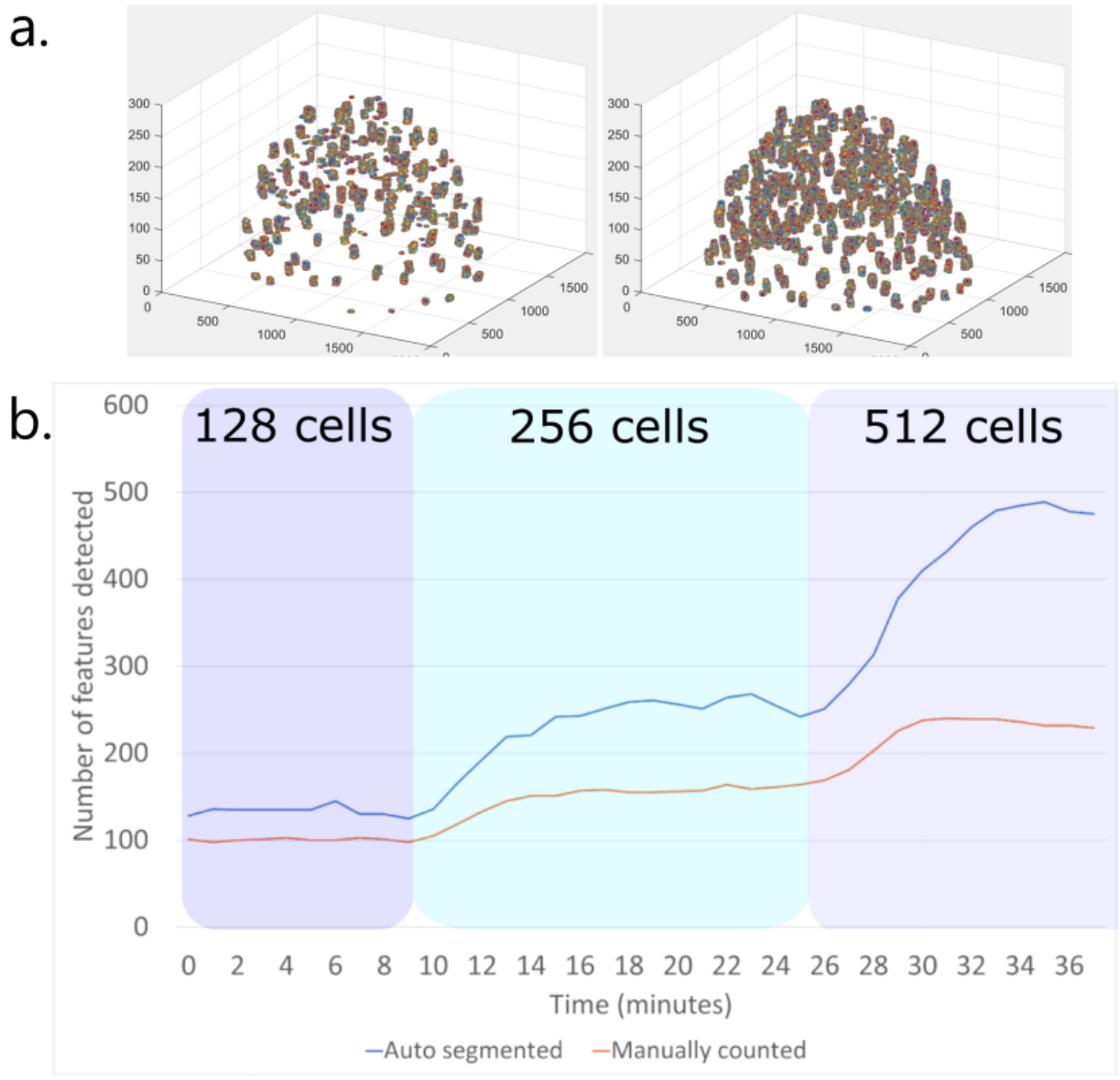


Figure 5.8: a. 3D representation of nuclei segmented by MATLAB 3D segmentation tool. Left image is segmentation taken at 128 cell stage, right image is segmentation taken at 256 cell stage. Axis units are in pixels. b.Comparison of detected features between manual counting (using 2D MIP) and 3D segmentation via our pipeline, showing the increase in features with periods of metasynchronous mitosis between 3 embryonic stages, 128, 256 and 512 cell stage - indicated by coloured boxes and labels. Data shown from single embryo.

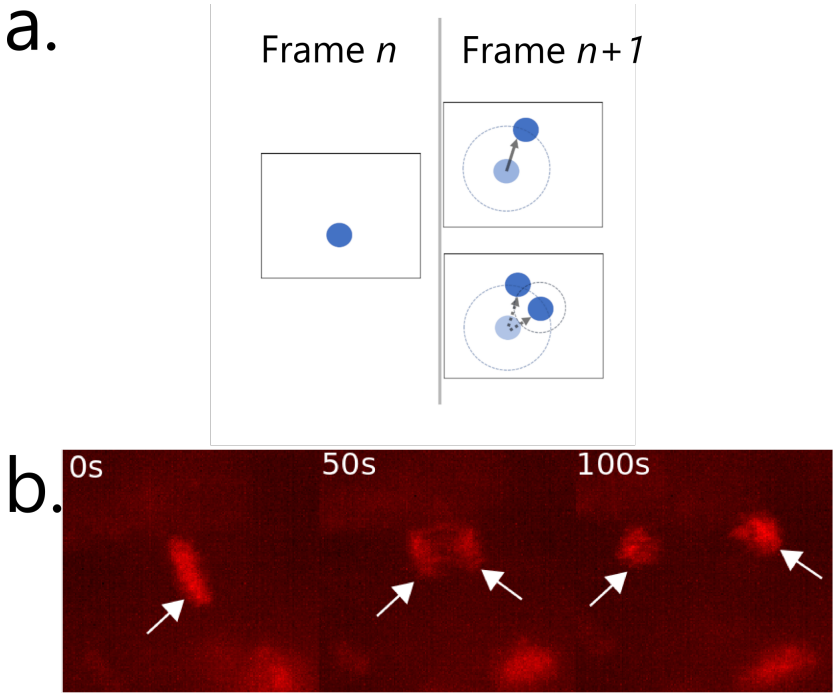


Figure 5.9: **a.** Mitosis detection is achieved by comparing the location of a feature a at time point (frame) n (an), with the number of features within a defined radius of an in the frame n+1. If there is more than one feature, conditions are checked if this is a mitotic event. Frame 1 represents nucleus in current frame, frame 2 shows 2 scenarios. The top scenario shows when there is only 1 feature in the next frame within the scan radius (large dotted circle), indicating the continuation of the track. The bottom scenario shows a mitosis event, where 2 nuclei within a minor scan radius (smaller dotted circle) indicate the current track must end, and 2 new tracks must begin. **b.** Time lapse showing the jump in nuclear movement between frames showing how the distance covered during cytokinesis does not allow for spatial overlap. Image taken from a sample dataset used in analysis. Nucleus in red, highlighted by arrows. Time in seconds in top left of each frame.

Tracking segmented nuclei: The tracking portion of the software package takes the transformed 3D segmentation data. Transformation entails reshaping the output csv files from the 3D segmentation tool, and filtering undesirable features based on geometrical parameters such as volume. Tracking was based around calculating nearest neighbours, using a KD Tree package from the scikit-learn library (Pedregosa et al. 2011).

5.2.9 Nuclei are tracked using KD Trees and a user defined detection radius

Our tracking tool sets all nuclei in the first frame of the timelapse to an origin pool, know as time point 0. These are set as generation 0 also. Any nuclei added to this pool represent the starting points of tracking lineages. The algorithm sequentially selects each nucleus and tracks it until mitosis, while recording the time point of division and the change in generation: recorded by setting the new generation value as metadata for the starting nucleus positions of the daughter nuclei. These daughter nuclei are added to the origin pool so they can be selected for tracking.

Nuclei are represented by the 3D coordinate of their centroid. Our implementation stores the centroid of a nucleus in memory and calculates the distances of nearby nuclei in the next frame. This implementation also relies on a user defined scan radius, which is a radius within which to search for the next position of the nucleus currently being tracked.

Nuclei which fall within the scan radius are candidates for the next position of the nucleus, or they may represent a mitotic division. To determine mitosis, a minor scan radius is applied to each candidate nucleus. This local radius is scanned to detect very proximal nuclei. The presence of 2 closely associated nuclei is assumed to represent a mitosis event, based on the notion that the specimen under analysis is mononuclear, and the only time multiple nuclei appear in such close proximity is following cytokinesis. If a candidate nucleus has no nearby nuclei within its minor scan radius, no mitosis is assumed to have occurred, and this nucleus is set as the next position for the currently tracked nucleus. This method also allows for the capture of subsequent steps in the track in situations where the nucleus has moved far, such as in the case of during cytokinesis.

The tracking algorithm considers the loss of cell and nuclear diameter that can occur in embryonic cell divisions by factoring a user defined decay factor, and this will reduce the scan radius by a fixed amount for a user defined number of cycles. This is notable as a fixed scan radius may become too large and detect nuclei in neighbouring cells and trigger erroneous mitotic events, which could lead to the tool running in perpetuity.

Finding the correct parameters for tracking requires some trial and error in figuring out an optimal scan radius. Choosing too small a scan radius does not yield tracks as the algorithm is not searching far enough to detect the next position of the nucleus, and selecting too wide of a radius will register mitosis events and spawn new tracks incorrectly. Tracking newly divided cells presents an issue with the scan radius method: Once the nuclei have divided, it may be the case that the 2 migrating telophase nuclei are still within each others scan radius. To work around perpetually generating mitosis events, the user can set a post mitotic buffer period which sets a number of frames during which mitotic events won't be registered. Setting this requires knowledge of how long it takes the 2 daughter nuclei to migrate sufficiently apart from one another.

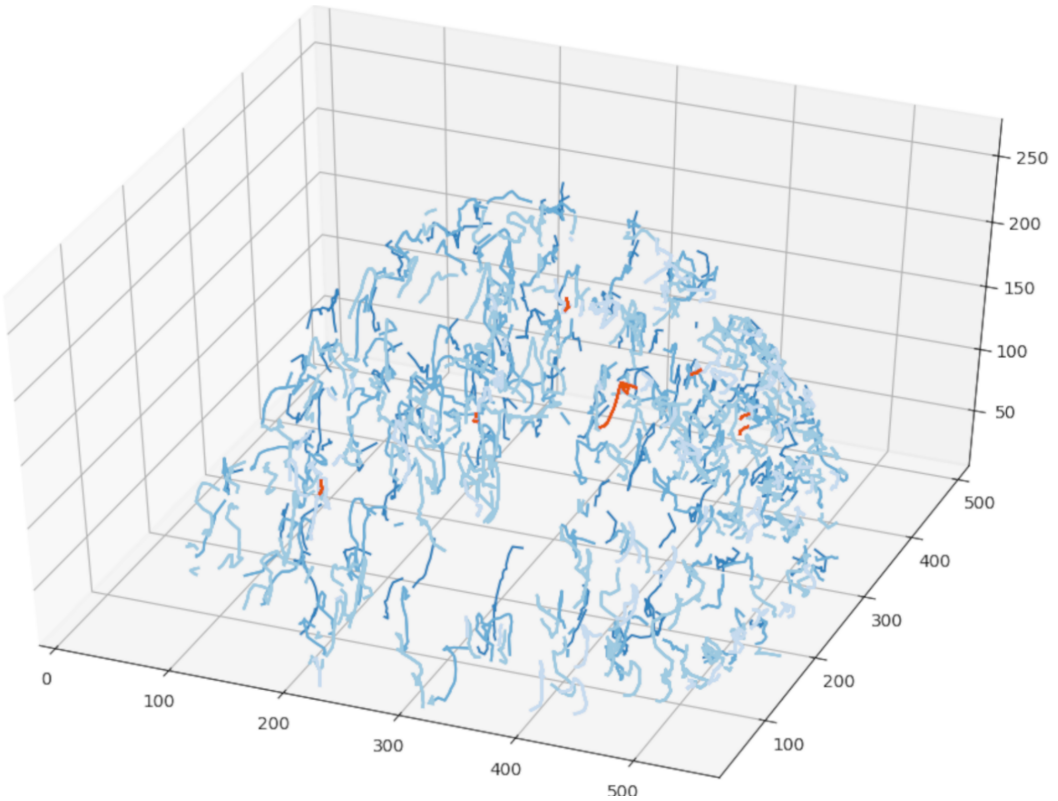


Figure 5.10: Tracking outputs shows the lineages tree in 3 Dimensional space, with lighter hues representing successive generations, and reds denoting over tracked lineages.

5.2.10 Lineages can be visualised in 3D

Once tracking is complete, track data is exported as a .csv file, which contains data regarding nuclei and their track lengths, start/end times, generation and family history. This allows the construction of family trees. A static 4D plot (a 3D plot showing trajectories over time) showing the tracks of all nuclei followed is displayed, which allows a quick visual inspection of outputs (figure 5.10).

The process of tracking outputs a 3D plot displaying the paths of all tracks and their lineages where generations within lineages are colour coded based on where they place in the family tree. In the current iteration of the module, this is represented by shades of blue, where darker blues correspond to earlier members of the family

5.2.11 Detected features are validated by comparing detected coordinates with the raw image

Tracks were validated by randomly selecting points in tracks and corroborating their positions with nuclei in the raw data. Additionally, I developed a quick visual tool using the mayavi library (Ramachandran and Varoquaux 2011) which overlays the 4D track representations with a downscaled 4D representation of the raw data. The raw data is downscaled and a maximum intensity projection is created in the T dimension. This can be used to visually inspect whether tracks are correctly overlaying with the data (figure 5.11).

Github URL for validation script can be found in section 3.4.3.

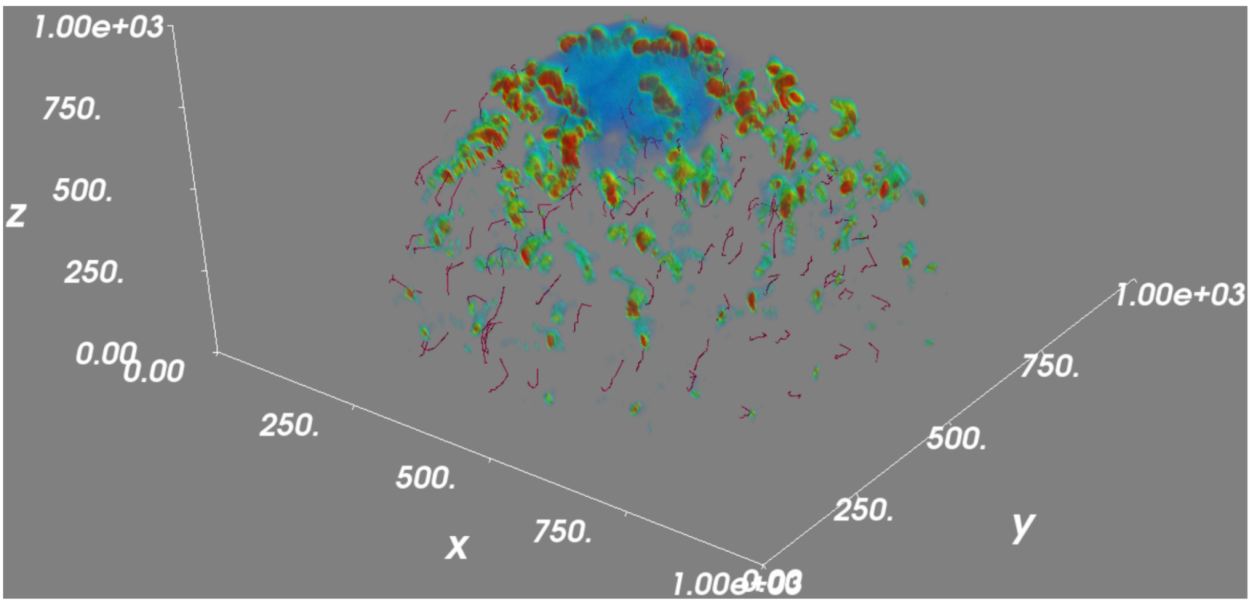


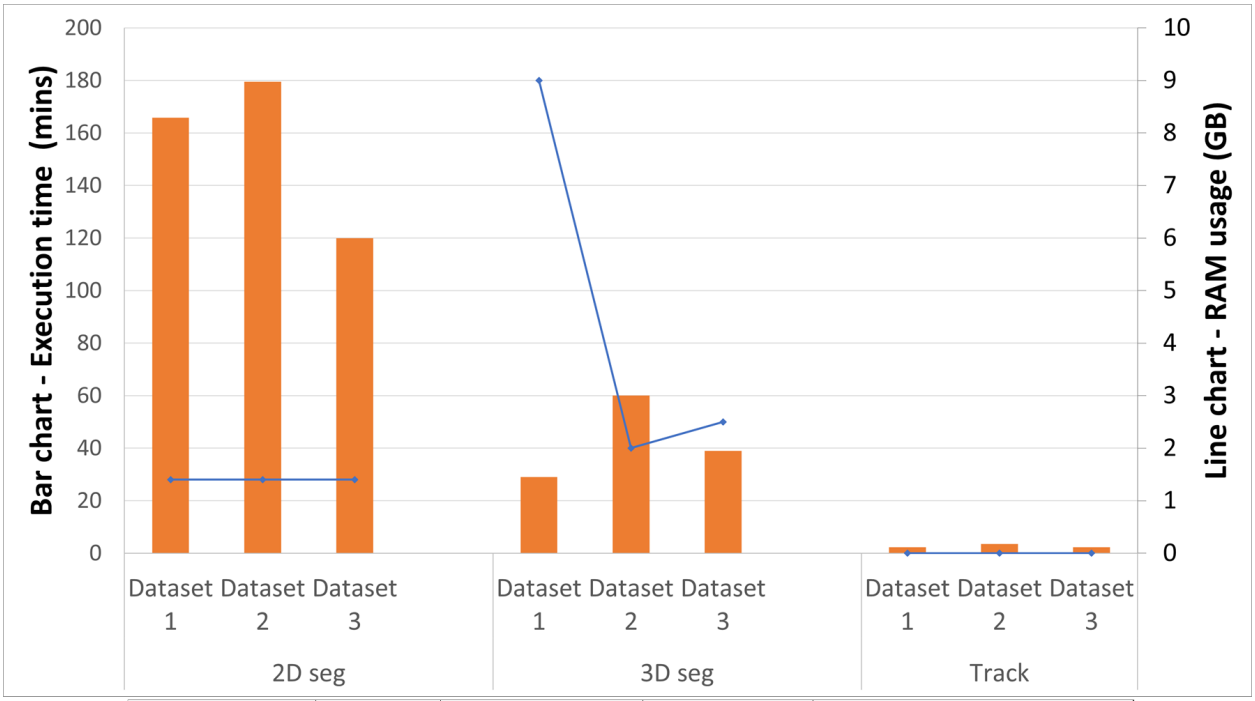
Figure 5.11: Output of validation, overlaying the tracks with a 4D maximum intensity projection. Green/red intensities represent maximum intensity of nuclei per time frame, resulting in a static image showing the position of nuclei in every time point. Blue cloud atop the animal cap is overthresholding from the visualisation. Only the first cell division is represented in the image rather than the full set of lineage tracks.

5.2.12 Resource utilisation

As part of the core design of both the 2D and 3D segmentation modules, I implemented multicore processing to parallelise the workload and increase speed. Data complexity was a notable factor which increased processing times in 2D segmentation.

To test the lightweight claim of our tool, I performed performance benchmarks to measure the system resource utilisation and the execution times for the various component modules which make up our tool (figure 5.12). The most time consuming portion is the segmentation, the timing of which is tied to data size, and the number of features which are detected. The most RAM intensive portion of the pipeline is the 3D segmentation. The 2D segmentation time is correlated to the number of z slices per frame with the longest 2D segmentation time belonging to dataset 2, which has the highest total z slices (figure 5.12 bottom). 3D segmentation time does not follow this same trend. Though not measured, this may be due to the complexity of the 3D shapes segmented, or even the number of false positives caused by low signal:noise.

These figures are samples of times and usage on a single system, it is likely these values will vary depending on different computer configurations, as well as ongoing updates to the source code. A like-for-like comparison with existing open source software packages such as Fiji ImageJ or Icy Bioimage was not possible as these software require loading data into the RAM on the processing computer, and such resource was unavailable. This highlights a key difference between this designed solution and existing solutions for segmentation and tracking.



Dataset	Size (GB)	t Length (frames)	Z slices	Total processed z slices
Dataset 1	153	65	320	20800
Dataset 2	180	200	122	24400
Dataset 3	72	64	144	9216

Figure 5.12: Performance benchmarks for 3 datasets. **Top.** 3 modules were benchmarked for their time taken to execute, and the peak amount of RAM (gigabytes) consumed during run time. The 3 modules benchmarked are 2D segmentation tool (2D seg), the 3D segmentation tool (3D seg) and the tracking tool (Track). **Bottom.** Table listing parameters for benchmarked datasets above. Total processed z slices is calculated as the product of t length x z slices.

5.3 Discussion

Our segmentation and tracking tool was inspired by the specific questions our laboratory faces, but also by general problems imaging scientists run into; handling bulky data, scattering of light within specimens and accessibility of tools for image processing. With regards to our laboratory specific problems, we needed to be able to identify rapidly moving nuclei in the developing zebrafish embryo, and track the lineages of said nuclei over multiple generations to gain insights into inherited characteristics of the cell cycle.

5.3.1 Overview of software design

I aimed to design a modular piece of software with each module being dedicated to optimally performing a specific task, in line with the UNIX software development philosophy of modular design where each program component is dedicated to do a single task effectively and then output to another program (McIlroy et al. 1978).

The first such module, the segmentation module, discerns nuclei through the scattered optical environment deep within embryo. Quantitation of number of nuclei detected per embryonic suggests this segmentation technique is able to capture the majority of nuclei present.

The module designed for tracking nuclei and identifying lineages of nuclei is able to track fast moving nuclei following mitosis, and keep tracks going over multiple generations, while recording metadata providing spatial information regarding tracked nuclei. The tracking module takes into account potential reductions in cell size with successive divisions, the parameters for which can be customised via the GUI (Reisser et al. 2018).

This spatial metadata carries information regarding the time and position of mitotic events, which makes it possible to probe the coordination of mitosis in time and space. This offers scope to investigate putative coordinated 'waves' of mitosis in the pre-MBT zebrafish embryo (Anderson et al. 2017, Deneke et al. 2016). This is a notable line of inquiry as it follows from the core questions this tool was designed to answer, and will be addressed in the following chapter, of how the timing of cell cycle lengths are coordinated and correlated within cycles. Varied cell cycles may influence the varied timings of mitosis observed in chapter 4, and any potential coordination of cell cycles may manifest as these theoretical mitosis waves.

The more general purpose of this tool pertains to investigations involving the isolation of lineages in noisy 3D datasets. The threshold free segmentation algorithm has utility as a binarization step in general image processing scenarios, beyond the uses investigated in this thesis. The segmentation module can be used in isolation, without the specific purpose of tracking.

5.3.2 The segmentation and tracking tool is a dedicated, lightweight solution to segmenting and tracking 3D datasets

Compared to existing software solutions which offer segmentation and/or tracking capabilities, our tool has the advantage of being optimised to work on demanding datasets on accessible office hardware, as that is the environment in which it was designed and tested. This contrasts with recommendations of using computing clusters for the use of available tools such as MaMut from Fiji ImageJ or XPIWIT (Wolff et al. 2018, Bartschat et al. 2016). The requirement for RAM during segmentation of very large datasets is highly

manageable on conventional office hardware, and can utilise a user defined number of processing cores in the CPU. This is made accessible through a GUI explicitly designed to show precisely what it requires of the user in order to produce segmentations or tracks. This makes the tool more niche in its purpose, but the intention is to fill its niche effectively.

3D segmentation is the most RAM heavy portion of this pipeline, by virtue of the use of binary connected components MATLAB function, yet this portion of the program is designed to clear the memory cache as it sequentially processes stacks of 2D segmented images to produce 3D segmented outputs. This portion of the program can be tailored to result in an on screen 3D visualisation of segmentations, which consumes more RAM, and at this time this is only recommended to be tested on small samples to make sure that segmentation is occurring normally. The tradeoffs of not loading large datasets into RAM, involve requiring additional storage space and the potentially longer processing times which may come with not keeping data in the high speed memory of RAM.

The tracking module uses the 3D segmentation data processed transformation step, which outputs a csv containing geometrical parameters of detected features. This is computationally lightweight representation of 3D data which is easy task for tracking module to utilise, resulting in negligible RAM usage. This aspect of the tool offers a tangible advantage over the semi-manual method of tracking nuclei in 2D described in 4.2.2. It offers time savings, removes the need to convert data to 2D as 3D data can be parsed by accessible hardware, and increases the number of nuclei we have access to for analysis. Analysis can be scaled up to tracking hundreds of lineages across multiple embryos.

5.3.3 Advantages and limitations

The principal bottleneck to effective segmentation and tracking remains to be the base quality of raw microscopy data. Despite a core strength of this tool lying in its ability to detect nuclei in highly scattered environments, the precision of such detections will always be determined by the image quality. Poorly defined nuclei will present with more irregularly segmented geometries, making volume measurements less reliable. Volume measurements are incorporated in the design of the metadata, though haven't been included in analyses for this thesis. Potential ways to improve this aspect may lie in advanced image pre-processing methods involving deconvolution.

By comparing with currently available image processing solutions, such as XPIWIT, our novel 2D segmentation method uses fewer image transforms on the raw data, and focuses more on user defined segmentation parameters to match the dynamic fluorescence environment of the light sheet image. This offers user friendly, fine control of a computer vision framework to identify areas of interest within a specimen. This function will be useful when parsing datasets of varying imaging parameters, allowing us to quickly change

segmentation and tracking settings for each dataset, as required.

A recommendation which emerged from software testing was that of testing the segmentation parameters on a sample dataset, typically taking 2 timeframes from the beginning and end of the timelapse, and observing the segmentation outputs. It is part of the development roadmap for this tool to integrate such an option within the GUI to streamline this, and potentially automate the testing process. A rough understanding of the biological specimen being segmented or tracked helps the user to discern which parameters to apply to the software. Estimates for cell cycle length and nucleus size are often easily determined from a quick look at imaging data, or perusing the literature.

This tool has limitations, the algorithms can be improved to be more efficient although this may require further expertise and collaboration. Another limitation is that the software is optimised in the early embryonic zebrafish, when the underlying algorithms may be adapted to work in a variety of models and developmental contexts.

In this chapter I have presented a tool that is capable of detecting discrete features in a noisy medium and following them until they divide, while generating family histories. This pipeline is not just limited to use in the zebrafish embryo but may be employed wherever nuclear stained timelapses require tracking. However, the recording of cell cycle durations, family histories and points of mitosis puts us in prime position to address the questions posed in section 4.3 can take this tool to query whether cell cycle lengths correlate with lineages, and look for potential organisation to how to how cell cycles vary across the embryo.

6 Investigating heritable cell cycle features within pre-MBT embryonic cell lineages

6.1 Introduction

In chapter 4, 2D image and data analysis yielded a relationship between the length of transcription time and cell cycle length. I also noted that cells expressing transcription had longer cell cycle, and cells showing an earlier transcription signal were more likely to produce early transcribing daughter cells before global ZGA at MBT. Additional experiments using transcription block and synthetic cell cycle elongation posited the length of cell cycle as permissive to increased transcription length (section 4.3.5). Injection of Chk1 RNA introduced checkpoints into S-phase, and elongated cell cycle, and with it transcription duration. On the other hand, transcription block did not inhibit cell cycle elongation with successive pre-MBT embryonic stages. Taking this outcome with the additional finding that transcribing parents were more likely to produce transcribing daughters, this suggests that an aspect of cell cycle length may be heritable in early embryonic cells.

6.1.1 Epigenetic memory

Transcription is largely dependent on the state of chromatin. Open chromatin conformations make DNA accessible for transcription machinery. Chromatin modifications have been associated with transcriptional state, in particular methylations at the sites of gene promoters and enhancers have been shown to increase the speed of transcription reactivation in the cell cycles of daughter cells (Muramoto et al. 2010, Ferraro et al. 2016) .

Heritable variations obtained over the course of the cell cycle may manifest in daughter cells in the form of cell cycle elongation. One of the proposed models for the influence of inherited features on cell cycle is a stochastic model which proposes stochastic variations in the production of mRNA, and protein production and degradation (Sandler et al. 2015, Dolbniak et al. 2015, Mura et al. 2019). Another model suggests inequality in the distribution of cytoplasmic constituents between dividing daughters, leading to variations in cell cycle (Xie et al. 2017). The distribution of factors which contribute to cell cycle elongation may be uneven in such a case (C. B. Kimmel et al. 1995).

Observations in yeast present compelling evidence of inherited cell cycle variability from a clonal cell population. These experiments correlated lineage fitness with these variations, and showed a link between catabolic gene activity and cell cycle length variability and inheritance (Mosheiff et al. 2018, Cerulus et al.

2016).

Replication stress from conflicts between the DNA replication fork and transcription machinery can induce DNA damage which may be inherited following mitosis. Checkpoints triggered by DNA damage can contribute to variations in cell cycle length in embryonic cell populations, such as those in the very early embryo(Brauchle et al. 2003). In human cell culture, inherited lesions have been shown to prolong G1 phase post mitosis (Barr et al. 2017). Lesions incurred during S-phase are marked with caretaker protein 53BP1 and are passed down to offspring (Lezaja and Altmeyer 2018).

Whereas this is not an embryonic model, it shows the possibility of DNA damage influencing inherited states across cellular generations. This is of relevance to the gap-phase deficient embryonic cells in that the concurrence of DNA replication and transcription may lead to increased DNA damage events.

The corpus of work regarding epigenetic memory of cell cycle length in particular comes from single cell organisms or in vitro models. This relationship is yet to be explored in an *in vivo* vertebrate model, and the general applicability of the finding from these models to the early embryo is uncertain. In an embryonic model, the variations in cell cycle lengths will affect when the cells divide. In the highly coordinated environment of the early embryo the coordination of cell position and growth is of paramount importance to growth. The organisation of when and where mitosis occurs is another question we hope to be able to probe using our newly developed tools.

6.1.2 Coordination of mitosis with cell cycle

The spatiotemporal coordination of mitosis in the early embryo is a phenomenon that has seen extensive study in *Xenopus* models. This offers useful background for our own investigations as *Xenopus* is a developmentally similar model to zebrafish. Cell cycles across the *Xenopus* embryo are coordinated by Cdk1 waves which spread from the top of the animal pole and travel across the egg in the direction of gravity (Chang and Ferrell 2013). These early waves are involved in facilitating mitosis. Prior to this, a calcium wave emanates from the sperm entry point and spreads across the egg (Whitaker 2008). The role of this wave is to reactivate the cell cycle and to block fertilisation by other sperm cells. These findings indicate that waves of mitosis originate from a focal point and spread in a concentric manner away from the origin. The involvement of Cdk1 in mitosis and cell cycle advancement, in addition to the varying cell cycle lengths of the MZT zebrafish embryo hints at a link between the coordination of mitosis and cell cycle lengths.

6.1.3 Nucleocytoplasmic ratio

Another aspect of spatially analysing trends in cell cycles lengths is the potential influence of NCR, and how this may influence maternally deposited factors with involvement in cell cycle elongation. Cell cycle

elongation is a key feature of both MZT and MBT. The leading model for MZT onset is the NCR model, which suggests an increase in nucleoplasmic:cytoplasmic ratio is responsible for many of the notable changes during this period (Newport and Kirschner 1982). One of these changes is cell cycle length. As discussed in the introduction one such mechanism may lie in the activation of Chk1 which progresses the onset of M-phase. Our own experiments showed Chk1 does indeed elongate zebrafish cell cycle lengths, and in turn elongates transcriptional length also (Hadzhiev et al. 2019). Chk1 activation has been shown to be influenced by the titration of unbound histone 3, suggesting a mechanistic link between the NCR and cell cycle elongation (Shindo and Amodeo 2020).

The pre-MBT cell divisions of the zebrafish embryo preserve total animal cap volume, resulting in drastic reductions of the cell volume with each division. Nuclei, however lose less volume per division due to ongoing DNA synthesis. Kimmel's suggestion of uneven cell divisions may explain the variation in cell cycles, through the creation of uneven cell volumes, resulting in reduced cell cycle and transcription repression in a daughter cell, thus influencing elongation (Kane and Charles B Kimmel 1993). Cell size in the pre-MBT zebrafish is heterogeneous, although cell sizes are maintained within a range (T. Menon et al. 2020). This heterogeneity may also contribute to cell cycle variations and loss of mitotic synchrony during MZT.

6.1.4 Direction and aims

Our early data from chapter 4 (figures 4.18, 4.19) hinted at a potential correlation between ancestral and descendant cell cycle lengths. The literature suggests there may be a spatial component to how epigenetic memory is generated, with uneven distribution of cellular constituents, such as histones, possibly producing local lineages with similar cell cycle dynamics (Xie et al. 2017). The asycnhrony of mitoses could be coordinated, as in *Xenopus* where this results in a mitotic wave emanating from a focal point (Anderson et al. 2017). With a tool now established to deeply probe embryonic nucleus dynamics, we may address questions of cellular memory and inheritance. Is there underlying order to how cell cycle lengths vary and elongate? If yes, we then ask how this variation may be presented through the asynchrony of mitotic timing in the pre-MBT embryo. Is there a spatial component to how cells of varying lengths are organised in the embryo? To address this, I asked the following questions:

- Is cell cycle length correlated within cell lineages?
- Are mitosis events spatially and temporally coordinated?
- Are cells of varying cell cycle lengths spatially organised?

6.2 Results

All laboratory and microscopy work used for the following analyses was conducted by Kacper Wozniak and Yavor Hadzhiev. My role in this chapter was producing data handling scripts, and image processing and analysis.

Preparing data processing: The next logical step following the creation of a segmentation and tracking tool, was to develop methods to parse the outputs of the tool. This required constructing more scripts to organise data into dataframes so they could be analysed. A challenge from generating so much data through tracking was finding ways to visualise and inspect data for trends. This was addressed by creating scripts which could find and compare lineages based on their ID strings, and creating scripts which could visualise lineages in flexible multidimensional views.

6.2.1 Parameter selection for tracking algorithm

The Zeiss Z.1 light sheet microscope outputs data in the proprietary .czi format, which is difficult to handle using custom scripts. Fiji ImageJ was used to create .tiff series out of .czi files. A single tiff file was created for every z slice in every frame, and Fiji ImageJ outputs all of these into a single folder.

Here I designed a script to parse the names of each resultant tiff file and create folders for every time point to sort them into. A second script moves each file into its corresponding folder. These time point folders are read by the python 2D segmentation tool. Github URL for this script is in section 3.4.3.

Selection of parameters for the tracking algorithm required prior inspection of raw data to understand which cell stage the data begins at. This is because, in its current iteration, the tracking tool requires estimates for the cell diameter at the beginning and end of the dataset. These diameter measurements are provided in microns and are denoted by ‘scan radius’ headings. The decay factor takes into account the loss of cellular volume with each division, if there is any. The mitosis buffer indicates the period after a cell division during which no division is expected. This was built in to prevent newly divided nuclei from detecting each other and initiating another mitosis event (figure 6.1).

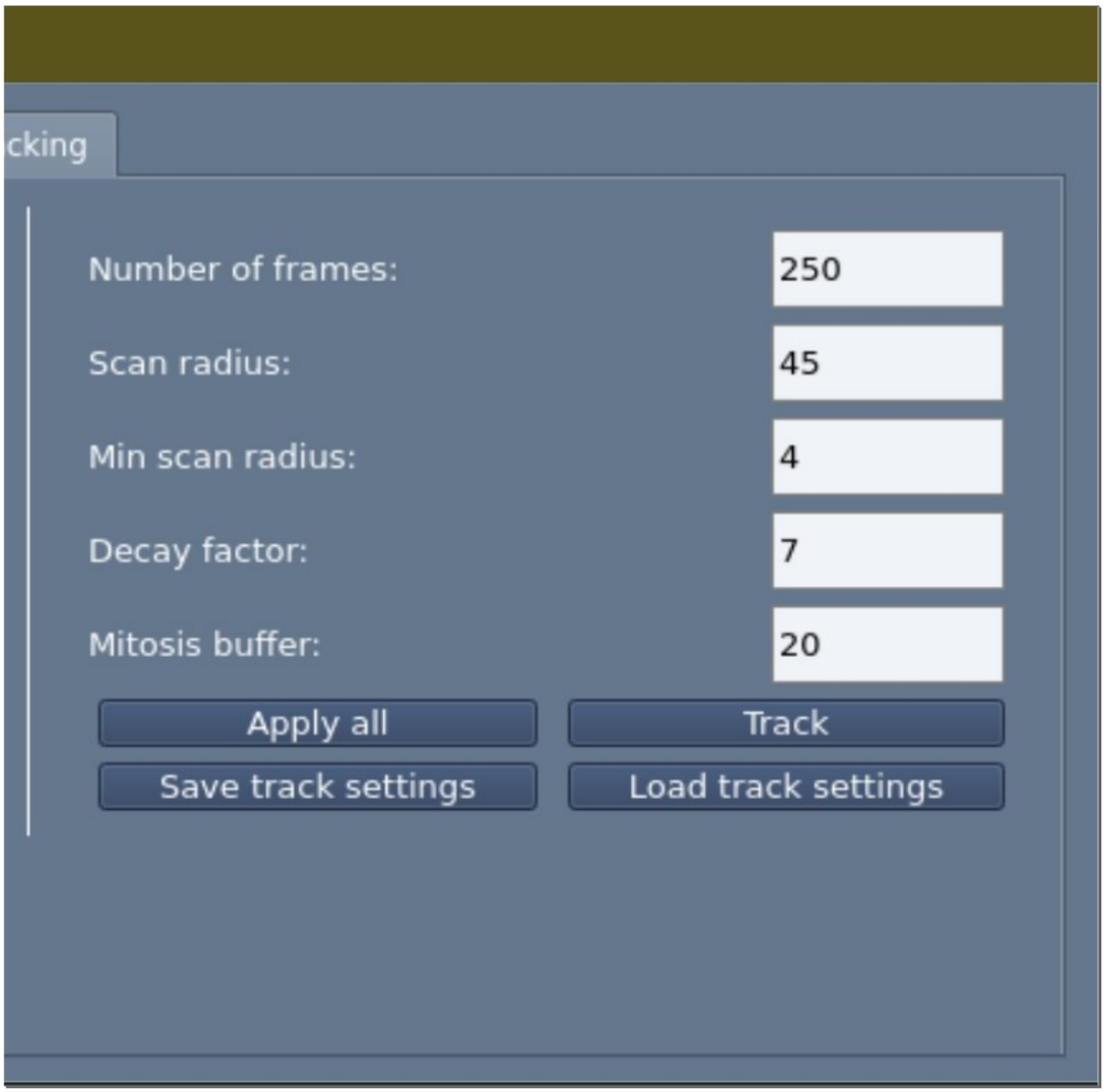


Figure 6.1: Screenshot of tracking module tracking parameters menu. The tracking parameters are indicated on the left, and text boxes on the right receive user inputs.

Are variations in cell cycle length correlated across generations?

6.2.2 Parent daughter cell cycle lengths show correlation in preMBT embryonic stages, within a single embryo

Upon implementing tracking on 5 embryos, I was now able to inspect cell cycle lengths within lineages (using cell cycle comparison script, Github URL in section 3.4.3). In section 4.2.16 we noted that the durations of both cell cycle and transcription were linked. Additional experiments within the lab (as discussed in

section 4.3), showed that cell cycle length determined duration of transcription, and not the other way round. We also noted that nuclei which showed transcriptionally active states earlier in MZT, were more likely to produce daughter cells showing pre-MBT transcription. This raises the question of whether cell cycle length inheritance drives this inheritance of transcriptional state. To begin probing this question, I asked whether the cell cycle lengths of parent cells are correlated with the cell cycle lengths of their daughter cells. This is to see whether cell cycle states may be inherited (figure 6.2).

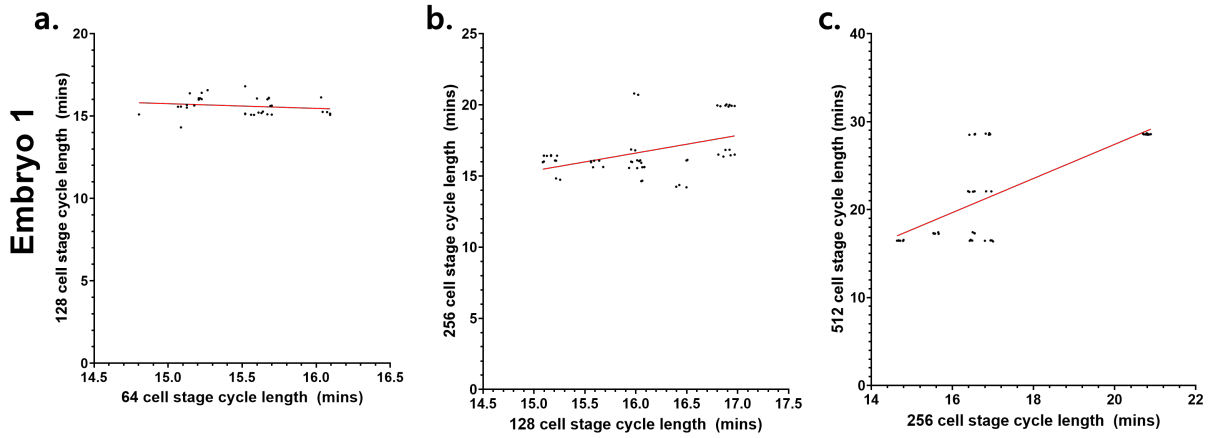


Figure 6.2: Comparison of parent (x axis) cell cycle length with daughter cell cycle lengths (y axis) in the next embryonic stage. Data has been jittered for visibility **a.** Comparison between 64 cell parent cell cycle length with 128 cell daughter cell cycle length. r^2=0.031 P>0.05. **b.**. Comparison between 128 cell parent cell cycle length with 256 cell daughter cell cycle length. r^2=0.2 P<0.001. **c.** Comparison between 256 cell parent cell cycle length with 512 cell daughter cell cycle length. r^2=0.54 P<0.0001.

6.2.3 Correlations lost between parent daughter cell cycle lengths in preMBT embryonic stages, across multiple embryos

Cross embryo analysis of parent-daughter cell cycle relationships between 128-256 cell stages shows a correlation between parent-daughter cell cycle lengths when raw cell cycles are compared. Scrambling parent daughter pairs to create a randomised control abolishes any kind of relationship in every dataset randomised, suggesting the effect observed is not stochastic (appendix, figure 8.3). A weak but not significant relationship is observed when comparing the cell cycle lengths of 64 cell stage nuclei with their 128 cell stage progeny. This may be due to cell cycle synchrony persisting across the embryo in the early stages of MZT. A similar relationship is also observed between 256 cell parents with their 512 cell stage offspring, potentially due to 512 cell stage being the final stage before MBT, and most cells are desynchronised and entering longer cell cycles.

Examination of these relationships within individual embryos yields mixed results (appendix, figures 8.1,

8.2). 80% of embryos show very clear parent-daughter cell cycle correlations when comparing 128 cell stage with 256 cell stage, with the remaining embryo showing only a weak, yet positive correlation in line with the other embryos. This relationship is more varied in comparisons between 64 cell parents with 128 cell daughters. Clusters of nuclei in the statistically non-correlating datasets do appear to show a relationship between parent and daughter cell cycle lengths, however. The relationship these clusters exhibit is lost in the global embryo view, suggesting there are other factors to consider when making comparisons of cell cycle lengths.

To address the effects of inter embryo variation, where one embryo may consistently show longer cell cycles, and thus skew over all cell cycle comparison trends in one direction, I compared distributions of cell cycle lengths per stage, per embryo (appendix, figures 8.1, 8.2). Distributions of cell cycle lengths fall within similar ranges across embryos, negating effects of embryo to embryo variation in cell cycle length where some embryos present longer cell cycles than others. Whereas median values for cell cycle length distribution do show variation in embryo-embryo comparisons, the overall distributions of cell cycle lengths over different stages do not show consistency in cell cycle duration levels (figure 6.3). This suggests that there is no consistent correlation between parent/offspring cell cycles between embryos. Randomisng the parent-daughter cell pairs and plotting their cell cycle relationship eliminates any semblance of correlation, suggesting any minor correlation we see may not be random. This would require further study, however.

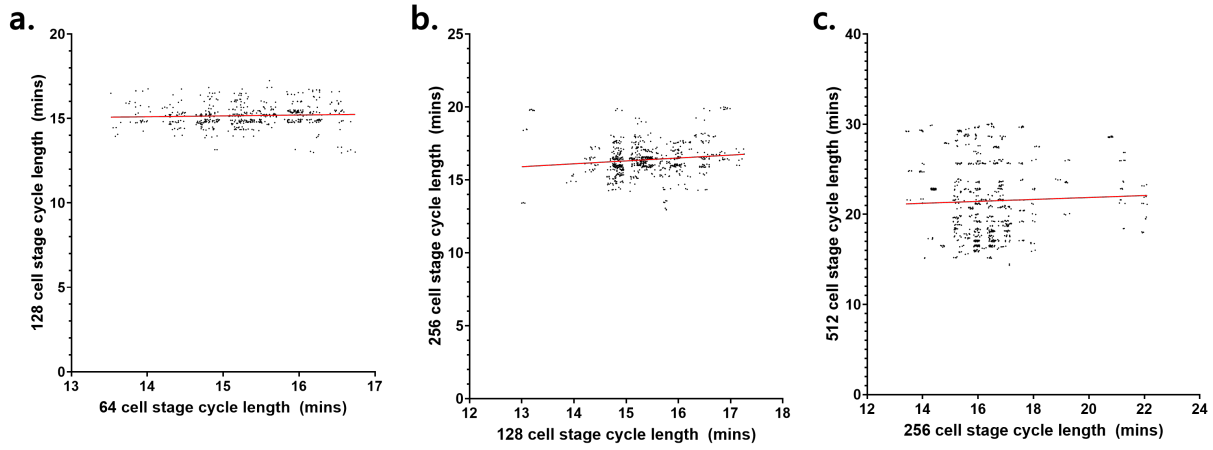


Figure 6.3: Aggregation of parent-daughter cell cycle relationships from 5 embryos. For each chart, the cell cycle length of the parent nucleus is measure against the cell cycle lengths of its daughter cells, with each representing a parent-daughter comparison of successive pre-MBT embryonic stages. Data has been jittered for visibility. **a.** Comparison of 64cell stage nuclei cell cycle lengths with their daughter cells (128 cell stage). N=552 nuclei. r^2=0.003 P>0.05. **b.** Comparison of 128 cell stage nuclei cell cycle lengths with their daughter cells (256 cell stage). N=977 nuclei. r^2=0.018 P<0.001. **c.** Comparison of 256cell stage nuclei cell cycle lengths with their daughter cells (512 cell stage). N=887 nuclei. r^2=0.002 P>0.05.

6.2.4 Timing of cell division, and cell cycle length are correlated with position in the embryo

Nuclei do not remain in one position in the rapidly dividing embryo. Regular divisions push nuclei around in their pre-migratory state. As a result, nuclei are represented in the tracking data by a set of coordinates representing their path through space. To pinpoint a location for an abstract concept such as the path taken by a nucleus in 3D space, I used the point of mitosis as the proxy for the location of a track. This is a discrete phenomenon which is easily and efficiently processed based on the parameters of the data obtained, and the tools available.

Using a custom script designed to read tracking metadata, the positions of mitosis from one embryonic stage to the next were extracted and plotted in 2D space on the X,Y, or X/Y,Z plane. This perspective of the embryo is either looking down from the above the animal cap, or viewing laterally (figure 6.4). To visually compare how different parameters vary in space, these 2D maps of mitosis points are colour coded based on a 3rd parameter. As established in literature (C. B. Kimmel et al. 1995, Kane and Charles B Kimmel 1993) and further verified by my own analysis shown in section 4.2.5, the timing of mitosis is variable and plotting this variation in space allows us to explore whether there is coordination of mitotic events across the embryo. How this analysis is constructed is further elaborated upon in figure 6.5.

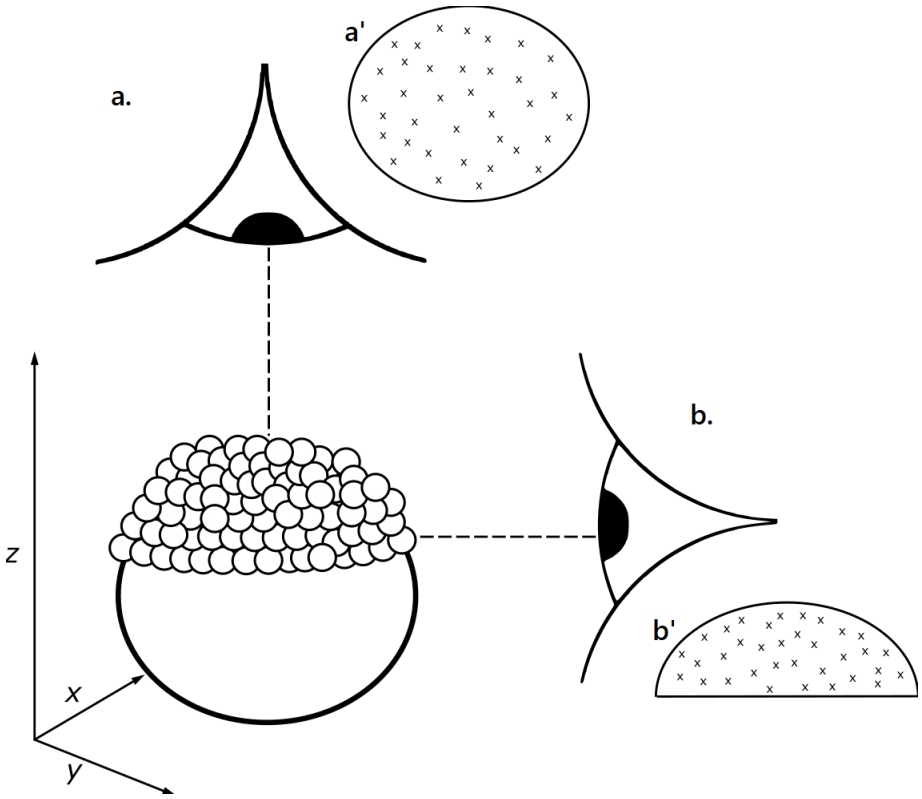


Figure 6.4: Multi-Dimensional views of mitosis in the embryo. To analyse spatiotemporal relationships between cell cycle length and position in the embryo, a custom script was created to map cell cycle cytokinesis points in planes (x,y), (x,z) or (y,z) (Github URL located in section 3.4.3). The different planes produce differently shaped plots, but convey similar information. Points are plotted on the same plane regardless of depth, similarly to how a maximum intensity projection may work. **a.** Top view. (x,y) plane offers a top down view of the embryo. **a'** The top view produces an oval/spherical map of mitosis locations (denoted by 'x's) in the transition between 2 embryonic stages. **b.** Planes (x/y,z) in a lateral view. **b'** This produces a hemispherical map of mitosis locations in the embryo.

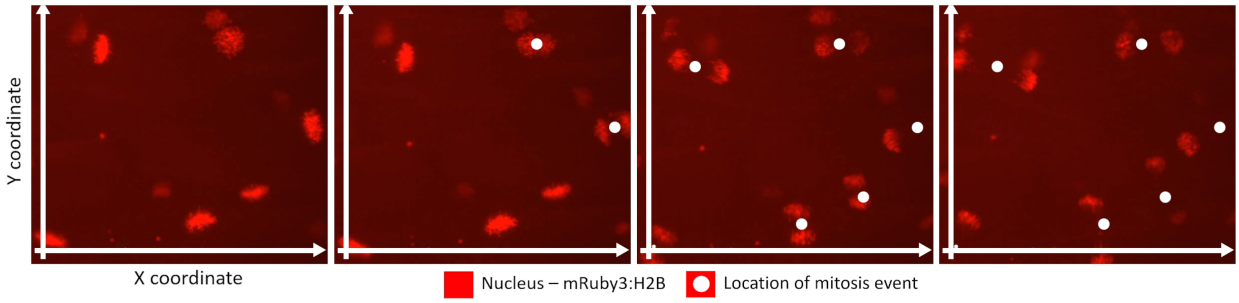


Figure 6.5: Example of how mitosis plotting from figure 6.4 is constructed. The timelapse shows cell divisions in 2D, although the software will do this in 3D, by viewing from the top or lateral planes and registering mitosis events on a 2D plot, where the depth of each cell division is not considered

6.2.5 Mitosis times are staggered across an axis in the embryo

To understand how variations in cell cycle length may manifest spatially, I used the method described in figure 6.4 to make a 2D plot of nuclei at 256 cell stage (figure 6.6). The points on the plot are located at the

x,y coordinates of mitosis for each nucleus, and are shaded based on when the division occurred. Shading can be customised to represent any field of metadata attached to a nucleus track. The shading on the plot gives a qualitative view of whether there may be a wave of mitosis to further study.

Following inspection of the 2D plot (figure 6.6a), the data can be recast into 1 spatial dimension, by extracting the coordinates of the axis of interest and plotting against a parameter of interest to see if there is directional correlation along the selected axis (figures 6.6b, c). In the case of figure 6.6b, I selected the parameter of time each nucleus underwent mitosis, and plotted it against its position in the embryo. This transformation revealed a polarity in where mitosis would begin and end in a pre-MBT embryo, developing from 256 cell stage to 512 cell stage. This wave of mitosis was consistently seen in the XY plane (using the top view illustrated in figures 6.4, 6.5) in all 5 embryos analysed (only one embryo is shown here, further data in appendix: figure 8.4).

The analysis shows that nuclei on polar ends of the embryo divide at different times, and this is driven by nuclei which divide later having longer cell cycles. Cell cycle length could be a heritable trait (Brauchle et al. 2003, Barr et al. 2017), although my own data has not robustly shown this (figure 6.3). This finding that cell cycles in one region of an embryo are shorter than in another region suggested another approach to take regarding inherited cell cycle length. Next I would investigate local cell clusters in the early embryo to see if they share cell cycle lengths and lineages.

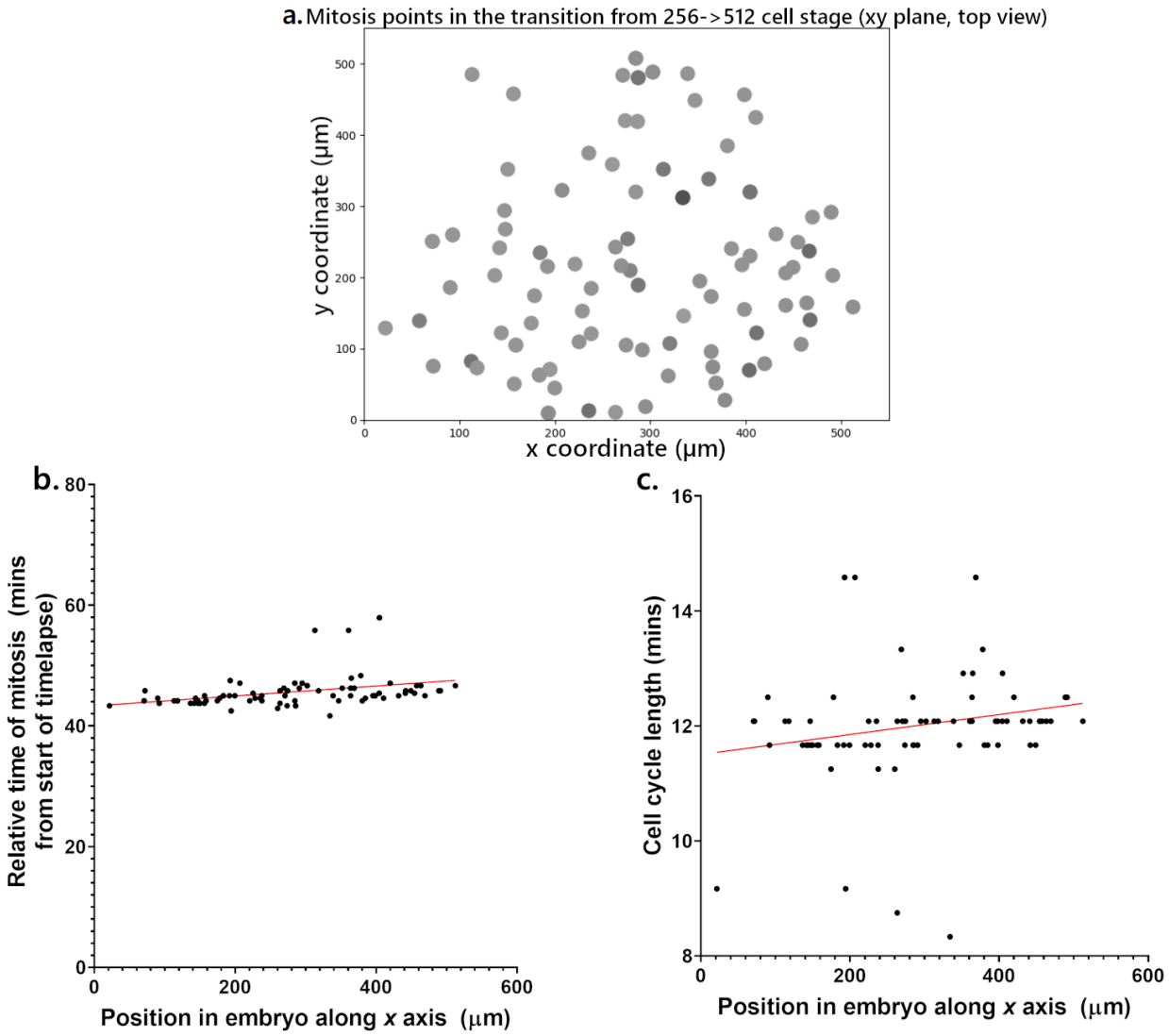


Figure 6.6: Finding the wave of mitosis using 2D scatter plots to represent 4D data. Single embryo, plotted mitosis points represent nuclei of 256 cell stage (mitosis point is therefore their division into 512 cell stage nuclei) **a.** The 2D output from implementing method from figure 6.4. showing the points of mitosis across a single embryo. **b.** Plotting data and parameters from mitosis occurrences plotted in panel **a** onto a single spatial axis to determine presence of correlation between time of mitosis and location of mitosis occurrences, along a single axis. The time at which mitosis occurs is plotted against where mitosis occurs in x. r^2=0.125 P<0.001. **c.** Same method as in panel **b**. The cell cycle length of a nucleus is plotted against its mitosis position in x. r^2=0.04 P>0.05.

Nucleus diameter declines with successive cell cycles

6.2.6 Can we discern a role for nucleocytoplasmic ratio in the spatial aspects of memory?

For further insights into the spatial organisation of inherited variations in cell cycle lengths, I directed my analysis to a contributing factor to cell cycle elongation, the NCR. The increase in NCR contributes to cell cycle elongation over successive cell cycles, thus I asked whether this phenomenon would be reflected in

the cell cycle variations within an embryonic stage (figure 6.7). Initially, I characterised the global decline in nucleus and cell diameter, to gauge a trend for the increase in NCR over the course of the time lapse. With this correlation established over several cell cycles, I would then ask how this manifests within a single cell cycle, not over time, but in space. To investigate this, I would examine where nuclei place across the embryo, and what the cell cycle length of these nuclei are (figure 6.7).

Cell diameter was calculated during tracking by calculating the average distance to the three nearest nuclei. This method is based off the technique published by Jevtić and Levy 2017 (Jevtić and Levy 2017) Cell diameters were taken at the midpoint of the track, and this diameter was used to represent the cell diameter for a single nucleus track. In the pre-MBT stages, cell diameter is not expected to grow or decline significantly until mitosis. Cell diameters were not sampled at every frame as it was not deemed useful to increase processing times for this purpose when cell diameter is expected to remain mostly consistent during a cell cycle.

Nucleus diameters were automatically sampled per frame as a part of segmentation metadata. It was assumed nuclei and cells take on approximate spherical shapes in the early stages of development (C. B. Kimmel et al. 1995). Based off this assumption, comparing nucleus diameter with cell diameter was seen as an approximation of NCR. Volume can be expressed as a function of diameter, so this is seen as a reasonable approximation to use in lieu of actual volume measurements.

In line with expected theory regarding nucleus and cell diameter measurements (Triveni Menon et al. 2019), the nuclear and cell diameter declines with time, with nuclei from later in development exhibiting smaller diameter. The decline in volume is much sharper in cell diameter, and this results in the ratio of nucleus:cell diameter rising with time. This trend appears to be continuous, even within timescales representing a single embryonic stage, showing a general trend of smaller nucleus diameter with nuclei which divide later.

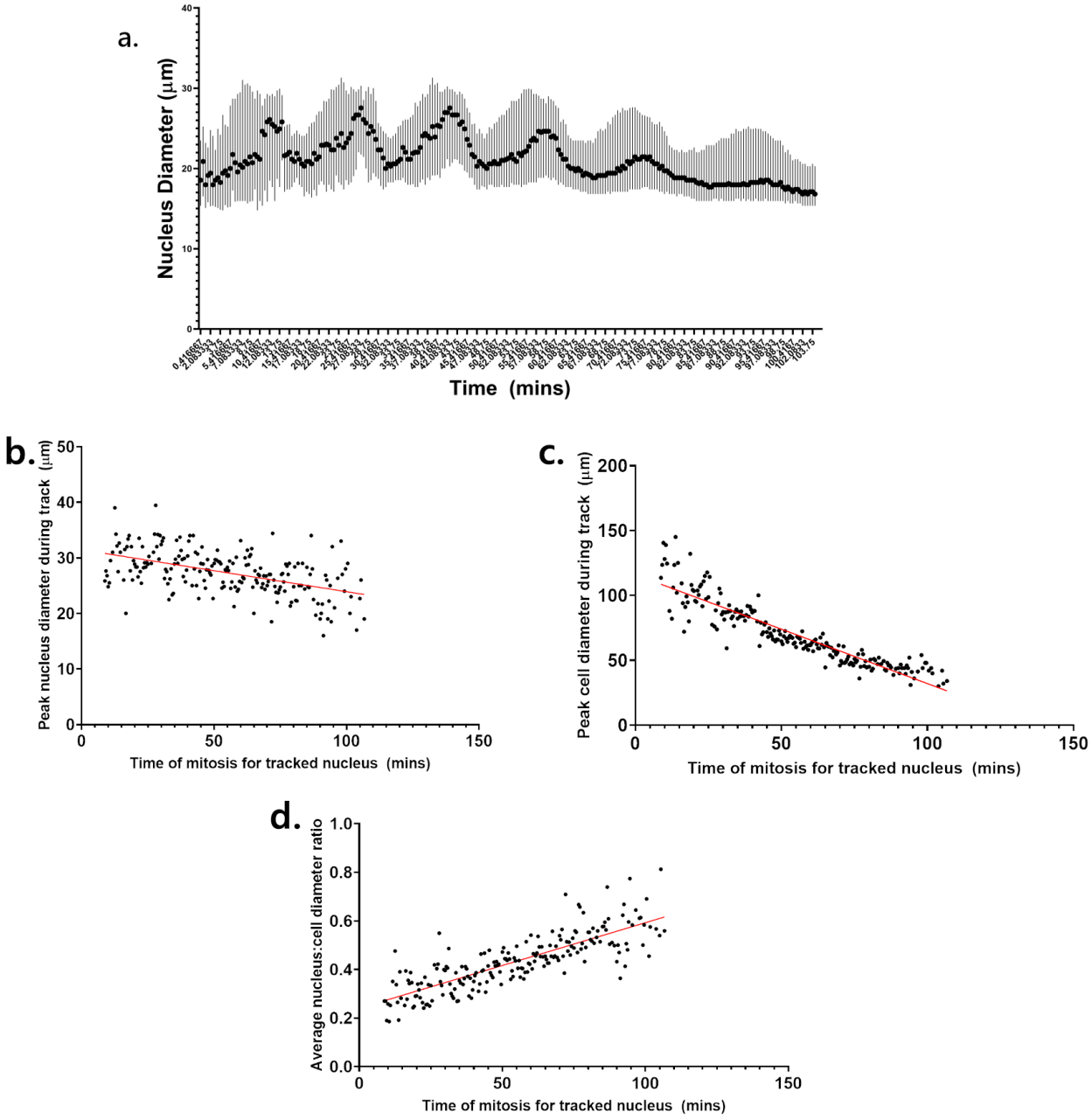


Figure 6.7: **All panels:** longitudinal trends in nucleus and cell diameters over the full time course of an imaging timelapse, representing individual detected nuclei from embryonic stages from 32 cell stage through 1000 cell stage. Data from a single embryo. Data starts part way through 32 cell stage. Frame rate = 25s. **a.** Median nucleus diameter (black points) at every frame in timelapse. Whiskers denote interquartile range.
b-d: Spatial characteristics of tracked nuclei, each nucleus represented in time via its time of mitosis, relative to the start of the time lapse (x axis). Data points on charts represent every nucleus tracked between 32 and 1000 cell stage. **b.** Peak nucleus diameter per tracked nucleus over entire timeline. r^2=0.253 P<0.0001. **c.** Maximum cell diameter over the course of its cell cycle, per tracked nucleus. r^2=0.831 P<0.0001. **d.** Nucleus:cell diameter ratio for each tracked nucleus, plotted against its relative division time. r^2=0.661 P<0.0001.

6.2.7 Nucleus diameter correlates with position in embryo

Having established the trends in nuclear and cell size change over a period encompassing multiple cell cycles, the next step was to query this phenomenon in individual embryonic stages. I sought to investigate how nucleus diameter varies across the embryo, and whether this relates to cell cycle length variation. In line with NCR model, I would expect relatively larger nuclei within an embryonic stage to yield longer cell cycles as these are more likely to present a greater NCR, based on the NCR theory of MBT (Newport and Kirschner 1982).

Another aspect of spatial positioning of nuclei in relation to cell cycle variations is seeing which regions cell progeny occupy in the growing embryo. This expands on investigations into the mitosis wave by seeing if lineages occupy a local region of the embryo, and may place the cell cycle memory of longer lineages in the context of staggering mitoses in a spatial wave. Measurements of nucleus diameter and cell cycle length follow the same trends as each other; nuclei measured on one extreme of the axis of the measurement exhibit earlier mitosis, shorter cell cycle length and smaller nucleus size (figure 6.8a-c). This is weakly reflected in nucleus diameter ratio – a trend consistent across all embryos measured in this way (5 embryos examined in total, only 1 shown in figure) (figure 6.8d). These data show a staggering of mitotic events across an axis within the embryo, which is consistent with observations of a wave of mitosis across the embryo.

Nucleus families, or whole lineages, remain localised without spreading across the embryo in the pre-MBT stages (figure 6.9). The analysis of where nucleus lineages reside in the embryo shows that different members of a lineage remain localised in one region of the embryo and do not exhibit long distance separation. When taken with the observation that cell cycle lengths vary across the embryo, these data may suggest a potential role for localised clusters of nuclei which divide together and may have similar cell cycle lengths. This could be a possible manifestation of local cell cycle memory. Prior to MBT and epiboly, cells do not exhibit migratory behaviour so it is expected that cells may not actively move large distances across the embryo (C. B. Kimmel et al. 1995). However, cells are constantly moving between each other during mitosis, so a question remained as to just how much mixing between cell lineages occurs pre-MBT. This data shows that this effect may be very little, and opens up the possibility for shared familial factors to influence the coordination of mitosis events across the embryo.

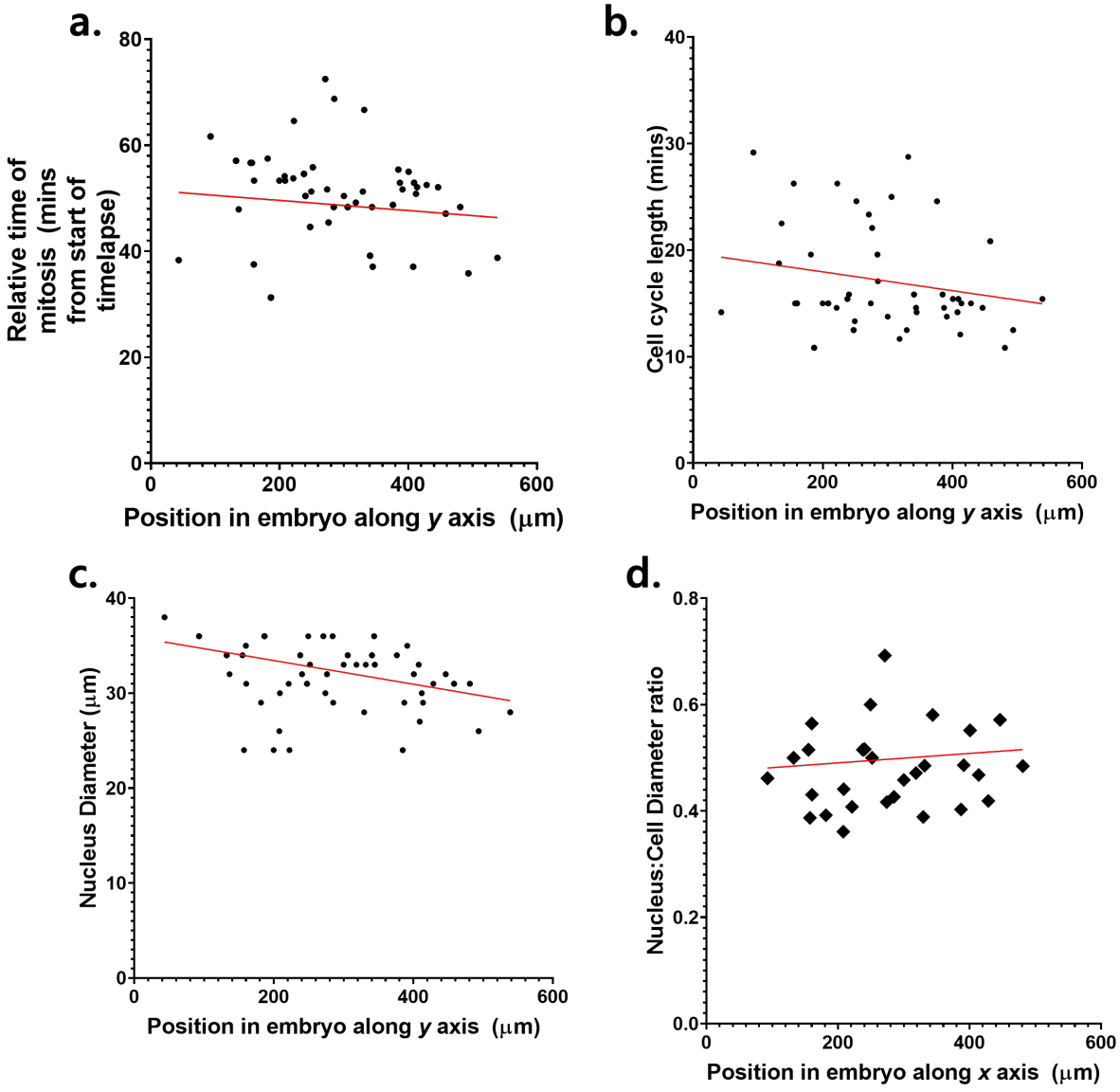


Figure 6.8: Spatial trends of nucleus size show a relationship with cell cycle length and position in embryo within a single embryonic stage. Presented using methodology from figure 6.4. Data from 256 cell stage from same embryo, each nucleus represented in space by the coordinate where it underwent mitosis. **a.** Position of cell division along defined x-axis in top view, where mitosis occurred, and the time at which mitosis occurred. r^2=0.014, P>0.05. **b.** Cell cycle length of 256 cell stage nuclei and their location in x axis. r^2=0.041, P<0.05. **c.** Maximal diameter reached by nuclei during the 256 cell stage cell cycle, and the position along the defined x axis at which they underwent mitosis. r^2=0.191, P<0.0001. **d.** Nucleus:cell diameter ratio (approximation of NCR) of 256 cell stage nuclei, and their position in x. r^2=0.019, P>0.05.

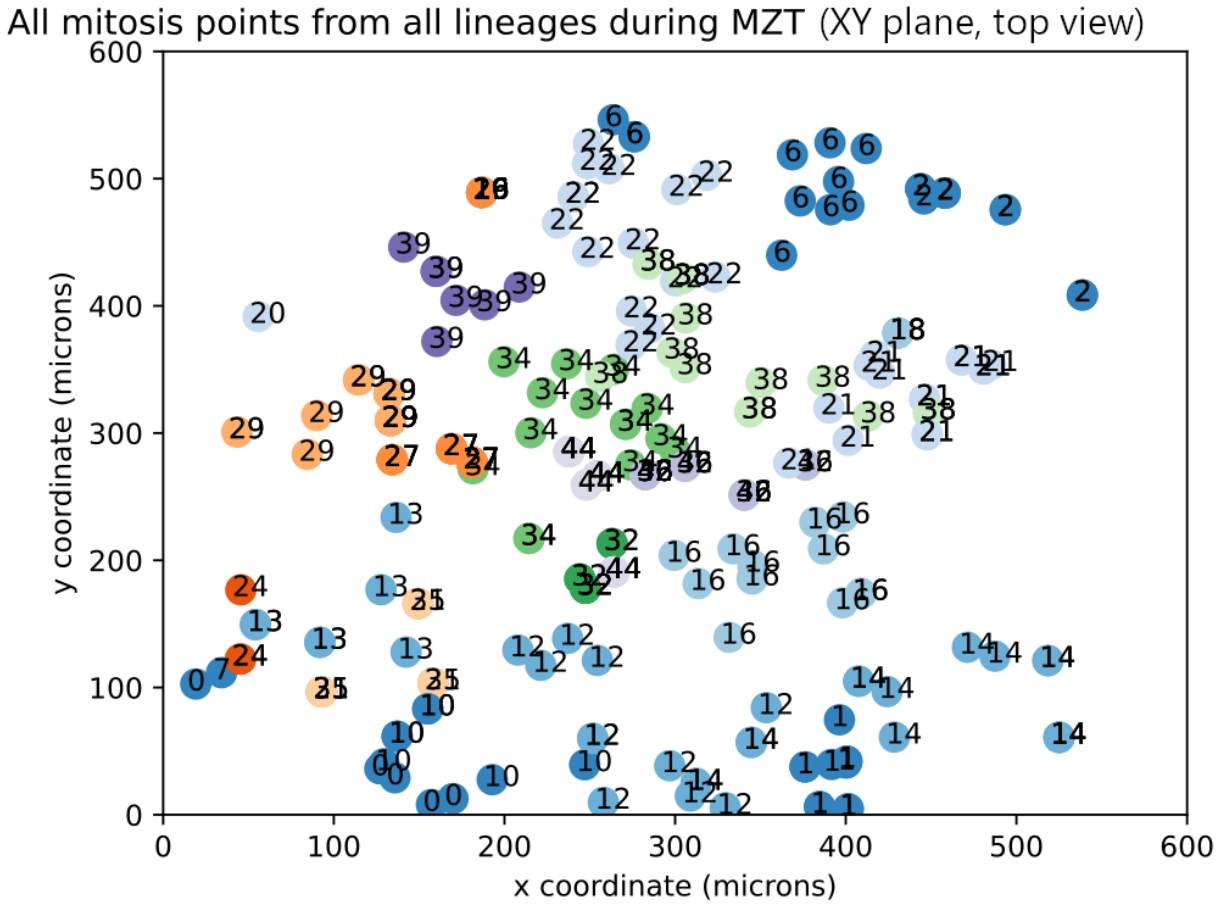


Figure 6.9: Map of 2D mitosis location of all tracked nuclei across all tracked stages in single embryo (depicted using method described in figure 6.4, same embryo as used in figure 6.8). Families are colour coded for visual differentiation and labelled with the family ID assigned to them by the tracking tool. Nuclei from embryonic stages from 64-512 cell stages represented.

6.3 Discussion

6.3.1 Cell cycle memory

The primary question we sought to address in this chapter was confirming whether cell cycle length was a heritable characteristic in the pre-MBT zebrafish. The data and analytical approach implemented here was far more granular, detailed and comprehensive than our initial pilot investigation into this in chapter 4. The conclusions from this follow up are as yet unclear. Any kind of global embryonic relationship is difficult to discern when examining relationships across embryos, as well as within individual embryos, although clusters of nuclei in the datasets do appear to show a relationship between parent and daughter cell cycle lengths (figures 6.2, 6.3). The relationship these clusters exhibit is lost in the global embryo view, suggesting

there are other factors to consider when making comparisons of cell cycle lengths.

The observation of clusters of nuclei with parent-daughter correlations suggests sub populations may exhibit the effects of epigenetic cell cycle inheritance more than others. The relationship between transcription and cell cycle explored in section 4.2.9 may provide a starting point when unpicking which other factors to examine to investigate this relationship; in particular, the relationship between S-phase and transcription, as S-phase elongation defined longer pre-MBT cell cycles (section 4.2.10). Whereas the duration of early zygotic transcription may be determined by cell cycle length, the capacity for transcription itself may have involvement in S-phase duration. A potential candidate for this is the RNA binding protein smaug, which in addition to enabling zygotic transcription, is also involved in cell cycle elongation (Benoit et al. 2009).

Another option to consider when explaining the loss of parent-daughter relationship between 64 and 512 cell stages is a potential diminished returns effect of heritable cell cycle characteristics reported by Lezaja and Altmeyer 2018. In their studies in human cell culture, they saw the heritable effects of cell cycle variation become less apparent after multiple divisions (ibid.).

The variability in cell cycle correlation output reflected high levels of variability between embryos. A portion of this effect may be due to fundamental biological variability in how cell cycle affecting factors may be inherited, but one must also consider technical limitations. Embryos were not all of equal imaging quality so the quality of segmentation and tracking was not consistent. Indeed, it is apparent from some plots (8.1) that far fewer lineages were availabl to track. Another side to this limitation is the quality of the tracks themselves, which will also be limited by quality.

6.3.2 Mitosis wave

There is a consistent observation of a gradient of cell cycle lengths across the embryo. Varied cell cycle lengths appear to be coordinated spatially, occurring as a wave. A wave of mitosis spreads across the embryo with nuclei with different cell cycle lengths contributing to the progress of the wave. This is visible with one end of the embryo showing earlier divisions which spreads along an axis to an opposing pole.

To measure whether variations in cell cycle length established in chapter 4 are spatially coordinated to produce a wave of mitoses across the embryo, I examined the position of nuclei within the embryo and measured their cell cycle length (alongside, maximal nucleus diameter). I investigated whether there was an axis along which cell divisions would get progressively later in comparison to nuclei measured spatially before them. A mitosis wave like effect can be found, although it does not manifest as a very pronounced difference between nuclei at polar ends of the embryo (poles as defined by the axis along which I was measuring mitosis events).

One potential reason for this is that pre-MBT cell cycle lengths do not vary greatly within an embryonic

stage (figure 4.12) particularly before cell cycle elongation and the onset of the major wave of ZGA. This may give the effect of only a modest difference between earliest and latest dividing nuclei.

Another possible reason for the weak correlations between cell cycle length and location in the embryo is that there may be a masking effect from the nuclei which will go on to form the yolk syncytial layer (YSL). The YSL forms at 512 cell stage when nuclei from the blastoderm drop into the yolk cell to form a multinucleate cell (Chu et al. 2012, Trinkaus 1992, Trinkaus 1993). Once the YSL has formed, the nuclei within it are known to divide in a wave-like manner (ibid.). These nuclei have abnormal dynamics compared to nuclei of the blastoderm, and this may interfere with the cell cycle length measurements taken for the mitosis wave analysis. The segmentation and tracking software could not differentiate potential YSL nuclei as precisely which nuclei will become part of the YSL is not known.

An additional consideration is that the pilot study from figure 4.19 was not just the caveat that it was done in 2D in a single embryo, but that the nuclei analysed were taken from a specific region in the embryo, as is liable to occur with the usage of the 'bright-zone' methodology I outlined in figure 4.4. On the one hand, this leads to a bias in which nuclei are measured, but it may also have an advantage in that it may exclude potential YSL nuclei, which by their adjacency to the yolk, and thus falling on the periphery or deep within the embryo are likely to be excluded from the bright zone. This could provide a clearer view of a blastoderm mitotic wave, but the caveat remains that it is still a very limited dataset by its numbers and the shortcomings of analysing MIP data.

Although statistically significant measurements of correlation between mitosis events and location in the embryo were made, significance on its own is not enough to strongly evidence this phenomenon. These data provide an interesting pilot to produce more imaging to more comprehensively explore this phenomenon, and produce a study with reasonable statistical power.

6.3.3 Nucleocytoplasmic ratio

Cell cycle length and nucleus size appear correlated (figure 6.7). I see this manifest in 2 divergent ways. Firstly, a gross overview of nucleus diameter and cell cycle length across multiple cell cycles shows decreasing nucleus size and increasing cell cycle lengths. Within a single cell cycle stage in individual embryos, decreased nucleus size consistently shows a correlation with decreased cell cycle length, and thus earlier mitosis time in the mitosis wave. The key difference between these divergent observations is the difference in cell diameter. My longitudinal analysis of nuclear diameter and cell diameter shows both decreasing over embryonic stages, with cell diameter decreasing at a much sharper rate than the nucleus, resulting in a progressively increasing NCR with time. Within a single cell cycle, approximate cell diameters show a very weak correlation with cell cycle lengths or direction of mitosis wave. This means that smaller nuclear sizes

will have slightly increased NCR, even within an embryonic stage. This data supports NCR model and is consistent with the overall trends seen in longitudinal analysis of cell and nucleus diameters.

The observation that lineages remain locally clustered in regions of the embryo provides further insight into how a mitotic wave may coordinate. Based on the findings that cell cycle length may correlate across lineages (figures 6.2, 6.3), the spatial positioning of lineages may be partially responsible for how mitosis events are staggered across cells. Longer lineages may cluster together locally, forming clusters of later dividing cells, contributing to how the mitosis wave presents across the embryo. The next step on from this would be to see if transcription activation follows this wave, and if it respects the cell cycle variations. This may inform about the role of transcription in cell cycle length variation. The geographical location of transcription initiation would be a useful piece of complementary information for further characterising waves of mitosis across the pre-MBT embryo.

7 Summary and conclusions

The first question we asked was how transcription is activated across the embryo during zygotic genome activation (ZGA). I sought to explore this by constructing an image analysis pipeline for 2D microscopy data, and using it to examine the temporal dynamics of transcription detection across the embryo, and the geometrical properties of transcription hotspots. This spatiotemporal analysis provided an overview of pre MBT transcription activity, and presented compelling evidence suggesting a relationship between parent-daughter cell cycle lengths. Transcription duration and cell cycle length appeared linked, with additional experimental data suggesting cell cycle length may be permissible to transcriptional activity. Furthermore, cell cycle length appeared to correlate within lineages, with descendants of cells with longer cell cycle length also showing longer cell cycle length in comparison to descendants from parents of shorter cell cycle length. This, in conjunction with the findings of cell cycle length preceding transcriptional activity, suggests heritable components which may impact both cell cycle length and transcriptional activity. However, another key conclusion from this study was the limitations of the method implemented on the large light sheet microscopy datasets.

The results from the first set of analyses suggested a form of cell cycle memory, hinting at an epigenetic memory.To discern whether parent-daughter cell cycle relationships were present in an *in vivo* developing early zebrafish embryo, we needed the relevant computational infrastructure to be able to parse our data in 3D. To address this, I developed a lightweight nucleus segmentation and tracking tool capable of working in large, noisy datasets to produce lineage trees of dividing nuclei. This allowed me to overcome the bottlenecks presented in section 4.3, and perform a more comprehensive analysis. This tool provides another open source avenue for researchers to segment and track nuclei without the need for expensive or hardware draining software packages. This software has its limitations but it will continue to undergo revisions and upgrades as newer segmentation and tracking libraries become available.

The rise of advanced microscopy has created great opportunities for researchers to probe biological processes *in vivo*, but has presented a logistical problem in data handling. A significant portion of my work in this thesis was developing ways around such problems. When the data outputs are large, one must be creative in how they are able to feasibly extract information from them. Through the development of both 2D and 3D pipelines, I hope to have contributed to the scientists' toolkit when dealing with such issues.

With a usable tool in hand, I could begin investigating the existence of heritable cell cycle traits. Using the nucleus tracking tool, I was able to probe multiple *in vivo* datasets in search of potential relationships between parent and daughter nuclei in the pre-MBT embryo. The precise relationships between parent and daughter cells are unclear, although the early data obtained here does hint at a relationship between

ancestors and descendants. Through the use of the tool, I was also able to make measurements of NCR, and contribute to the existing corpus of MBT literature which shows how NCR is tied to cell cycle length, and embryonic stage. Through this, I was able to begin to shed light on relationships between the location of a nucleus in the embryo, and its cell cycle dynamics. This may hint at a potential mitosis wave across the pre-MBT blastoderm, and this may be linked to NCR effects, although any data linking NCR to the mitosis wave is extremely preliminary. The study of early embryonic mitosis waves has largely been confined to *Xenopus*, *Drosophila* and the *Fundulus* YSL (Anderson et al. 2017, Deneke et al. 2016, Trinkaus 1993). If indeed mitosis waves do exist in the early zebrafish embryo, these data will contribute another model system in which the coordination of early cell divisions may be examined.

The understanding of coordination of rapid cell cycles (as in the pre-MBT zebrafish embryo) has applications beyond the niche of ZGA. The organisation of key cellular processes such as transcription, in the context of a rapid cell cycle may provide additional understanding to how aggressive cancers may proliferate (Chen et al. 2019). Additionally, gaining insight into how cells balance DNA synthesis with transcription during rapid cell cycles may offer further insight into how DNA damage may occur.

In this thesis, using the data generated by my laboratory colleagues, I contributed understanding to how cells developing in a fast paced developmental environment juggle the needs of DNA synthesis and transcription, through the generation of novel nuclear compartments to provide escape from genomic repression of transcription. I advanced the ways in which we may interrogate complex data outputs from modern microscopy, and I probed the possibility that early embryonic cells may pass on an, as yet undefined, heritable trait which may influence cell cycle.

8 Appendix

Title page. Appendix figures take the whole page so will continue on the next page.

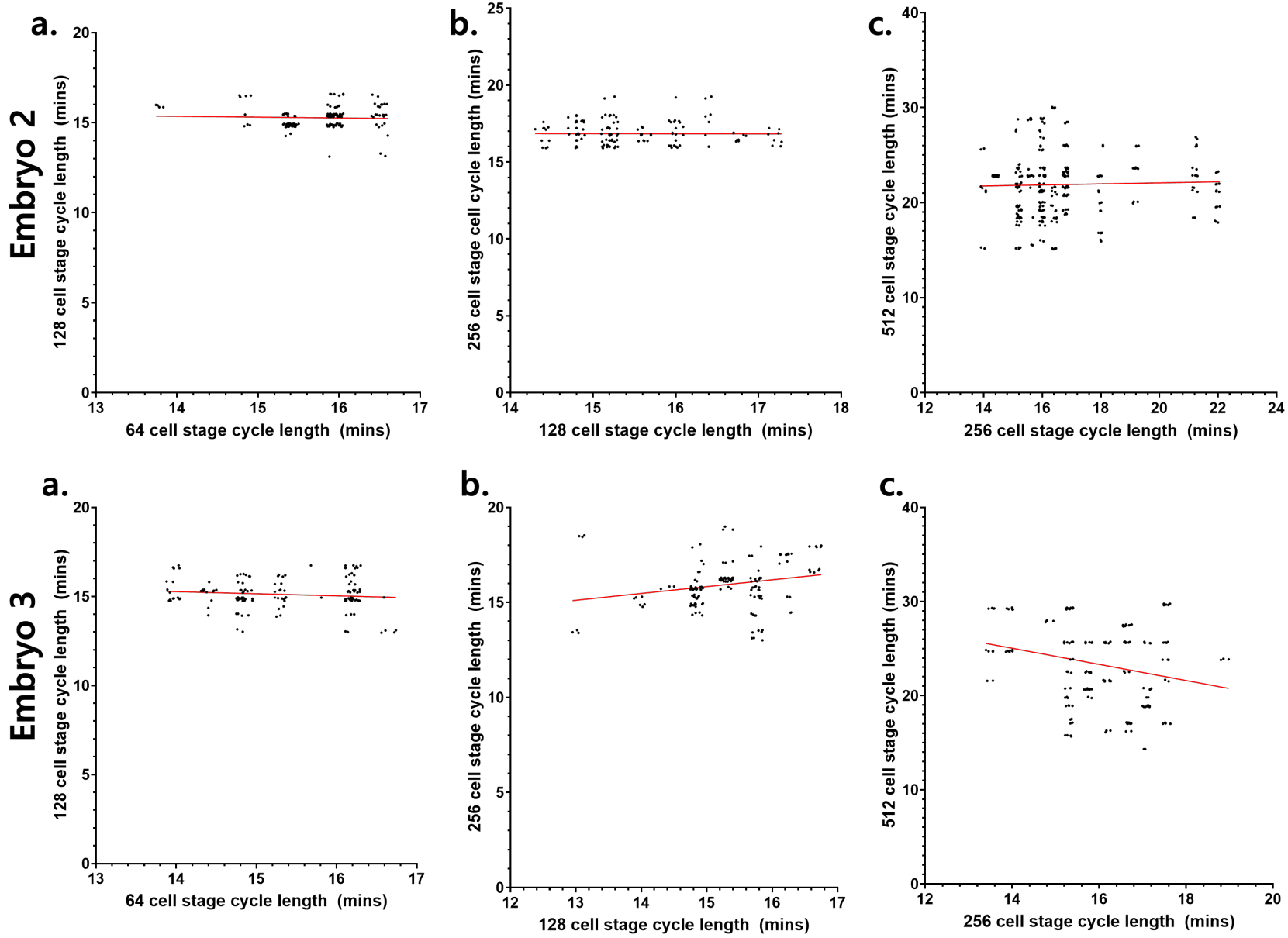


Figure 8.1: Additional data for figure 6.3. Parent-daughter cell cycle correlations per embryo and embryonic stage transition. **Embryo 2a.** r^2=0.002 P>0.05. **b.** r^2=0 P>0.05. **c.** r^2=0.001 P>0.05. **Embryo 3a.** r^2=0.013 P>0.05. **b.** r^2=0.051 P<0.01. **c.** r^2=0.056 P=0.001.

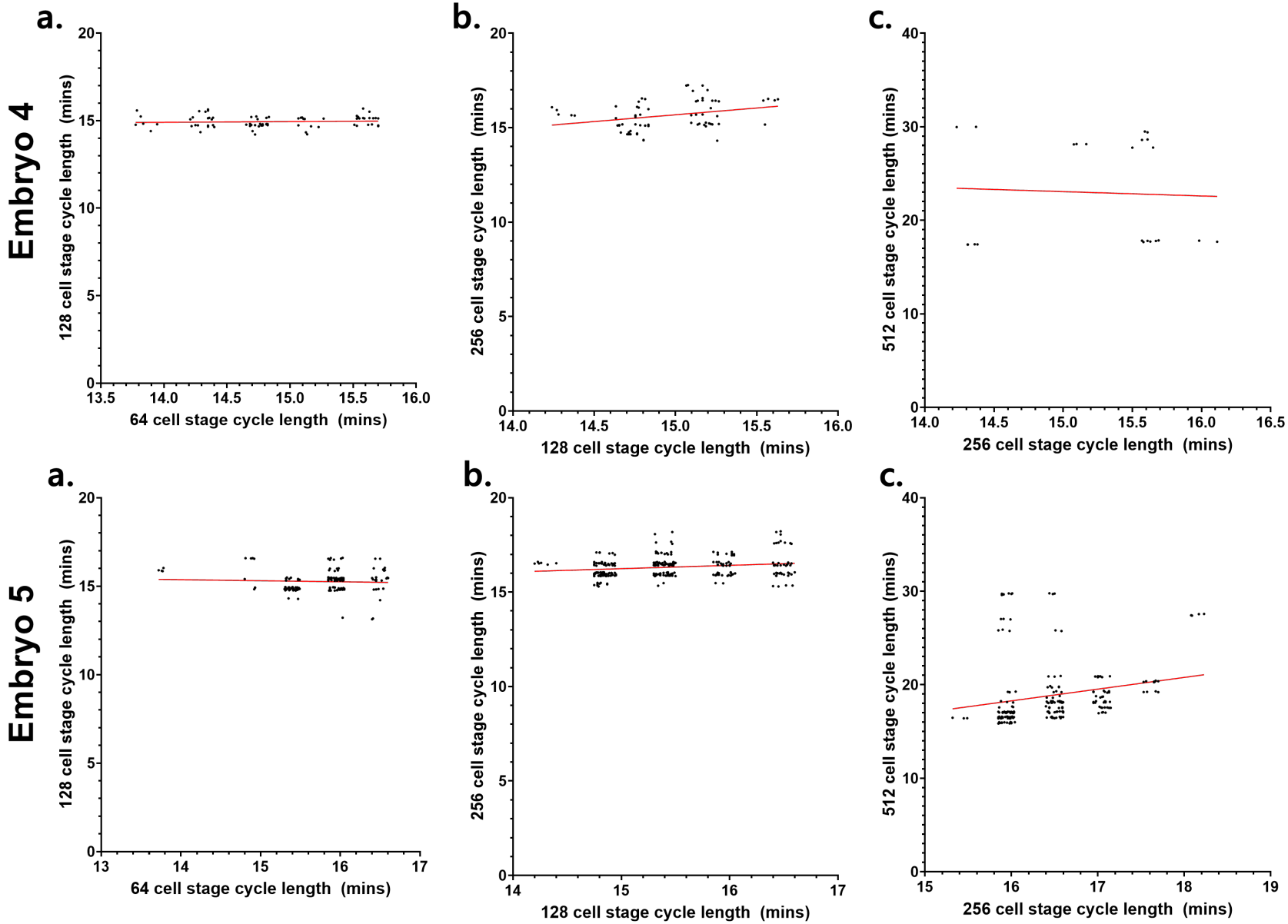


Figure 8.2: Additional data for figure 6.3 Parent-daughter cell cycle correlations per embryo and embryonic stage transition. **Embryo 4a.** r^2=0.005 P>0.05. **b.** r^2=0.094 P<0.05. **c.** r^2=0.002 P>0.05. **Embryo 5a.** r^2=0.003 P>0.05. **b.** r^2=0.038 P<0.01. **c.** r^2=0.041 P<0.001.

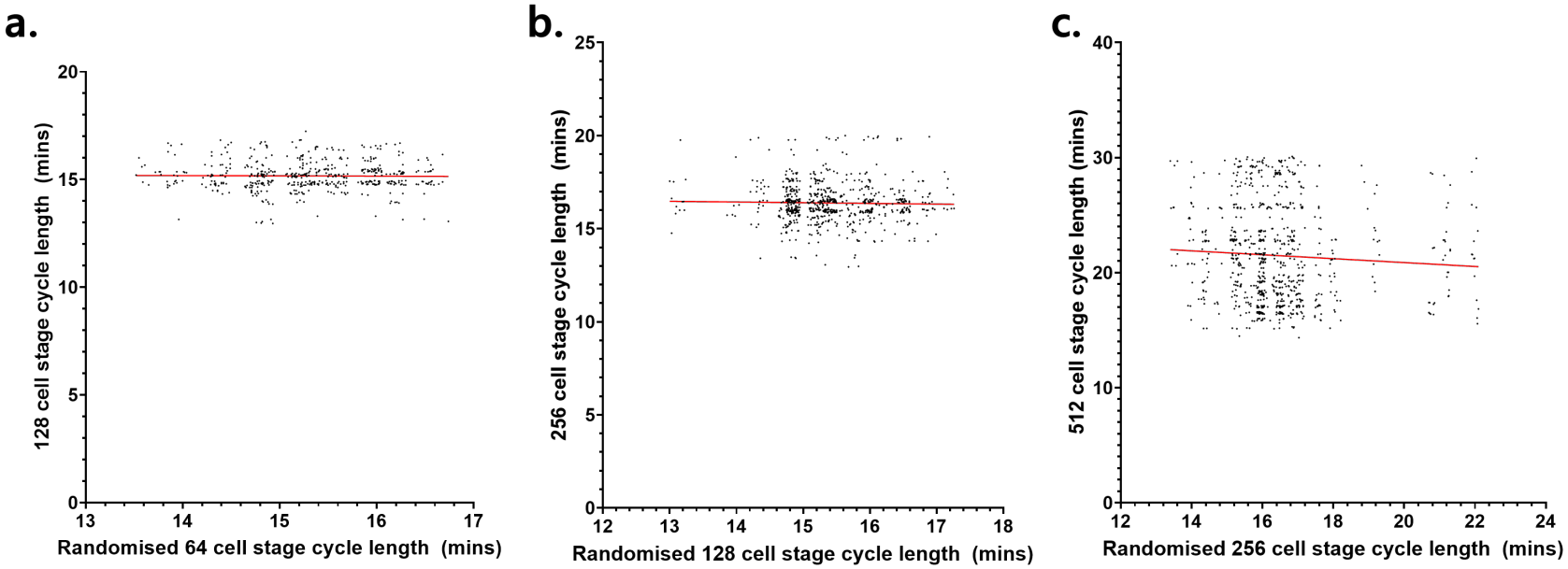


Figure 8.3: Additional data for figure 6.3. Randomised parent-daughter cell cycle lengths (nuclei from 5 embryos). Parents paired with randomly chosen (scrambled using Microsoft Excel) daughters, to use as a control for correlation between parent-daughter cell cycles. Data has been jittered for visibility. **a.** Comparison of 64 cell stage nuclei cell cycle lengths with their daughter cells (128 cell stage). N=552 nuclei. r^2=0.0002 P>0.05. **b.** Comparison of 128 cell stage nuclei cell cycle lengths with their daughter cells (256 cell stage). N=977 nuclei. r^2=0.0006 P>0.05. **c.** Comparison of 256cell stage nuclei cell cycle lengths with their daughter cells (512 cell stage). N=887 nuclei. r^2=0.004 P>0.05.

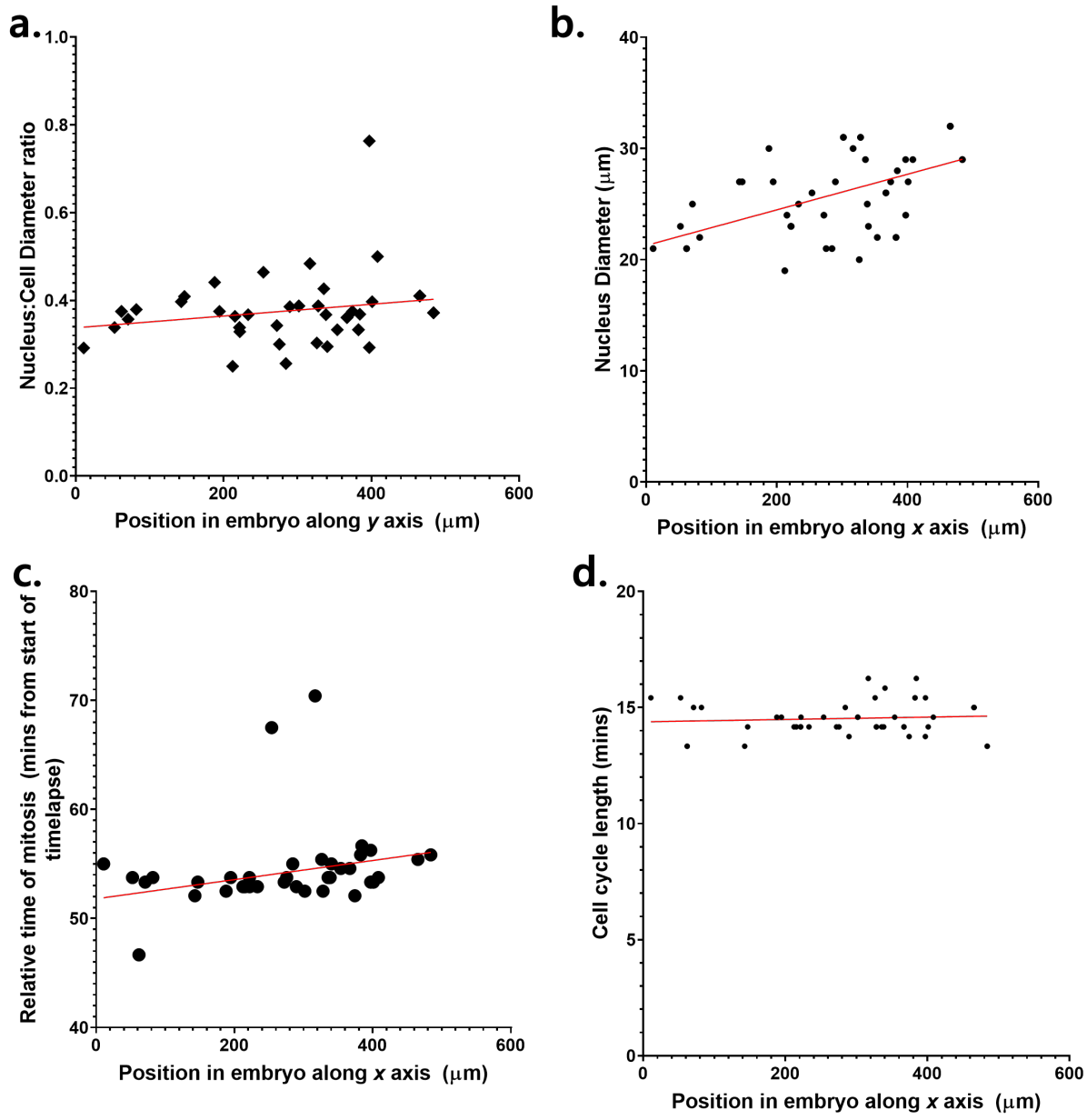


Figure 8.4: Additional data for figure 6.6. Single embryo, plotted mitosis points represent nuclei of 256 cell stage (mitosis point is therefore their division into 512 cell stage nuclei) Showing the points of mitosis across a single embryo, from a top view, as in figure 6.6a. **a.** Maximal NCR of 256 cell stage nucleus across the embryo, and its relationship with spatial position (position of mitosis into 512 cell stage nucleus) along defined axis. r^2=0.048 P>0.05. **b.** Maximal nucleus diameter attained by 256 nucleus in embryo, plotted against the position along the axis in which it underwent mitosis. r^2=0.281 P=0.0001. **c.** Timing (relative to start of time lapse) and positioning (along axis) of 256 stage nuclei. r^2=0.091 P<0.05. **d.** Cell cycle length of dividing 256 cell stage nuclei, and the position along x axis they underwent mitosis. r^2=0.007 P>0.05.

9 References

Abe, K. et al. (2015). “The first murine zygotic transcription is promiscuous and uncoupled from splicing and 3’ processing”. In: *EMBO J* 34.11, pp. 1523–37. ISSN: 1460-2075 (Electronic) 0261-4189 (Linking). DOI: 10.15252/embj.201490648.

Alberts, Bruce (2008). *Molecular biology of the cell*.

Alzu, Amaya et al. (Nov. 2012). “Senataxin Associates with Replication Forks to Protect Fork Integrity across RNA-Polymerase-II-Transcribed Genes”. In: *Cell* 151.4, pp. 835–846. DOI: 10.1016/j.cell.2012.09.041.

Anderson, G. A., L. Gelens, J. C. Baker, and Jr. Ferrell J. E. (2017). “Desynchronizing Embryonic Cell Division Waves Reveals the Robustness of Xenopus laevis Development”. In: *Cell Rep* 21.1, pp. 37–46. ISSN: 2211-1247 (Electronic). DOI: 10.1016/j.celrep.2017.09.017.

Barr, A. R., S. Cooper, F. S. Heldt, F. Butera, H. Stoy, J. Mansfeld, B. Novak, and C. Bakal (2017). “DNA damage during S-phase mediates the proliferation-quiescence decision in the subsequent G1 via p21 expression”. In: *Nat Commun* 8, p. 14728. ISSN: 2041-1723 (Electronic) 2041-1723 (Linking). DOI: 10.1038/ncomms14728.

Bartman, C. R., N. Hamagami, C. A. Keller, B. Giardine, R. C. Hardison, G. A. Blobel, and A. Raj (2019). “Transcriptional Burst Initiation and Polymerase Pause Release Are Key Control Points of Transcriptional Regulation”. In: *Mol Cell* 73.3, 519–532 e4. ISSN: 1097-4164 (Electronic) 1097-2765 (Linking). DOI: 10.1016/j.molcel.2018.11.004.

Bartschat, Andreas, Eduard Hübner, Markus Reischl, Ralf Mikut, and Johannes Stegmaier (2016). “XPIWIT—an XML pipeline wrapper for the Insight Toolkit”. In: *Bioinformatics* 32.2, pp. 315–317. ISSN: 1367-4803. DOI: 10.1093/bioinformatics/btv559.

Benoit, B., C. H. He, F. Zhang, S. M. Votruba, W. Tadros, J. T. Westwood, C. A. Smibert, H. D. Lipshitz, and W. E. Theurkauf (2009). “An essential role for the RNA-binding protein Smaug during the Drosophila maternal-to-zygotic transition”. In: *Development* 136.6, pp. 923–32. ISSN: 0950-1991 (Print) 0950-1991 (Linking). DOI: 10.1242/dev.031815.

Boettiger, Alistair N. and Michael Levine (July 2009). “Synchronous and Stochastic Patterns of Gene Activation in theiDrosophila/iEmbryo”. In: *Science* 325.5939, pp. 471–473. DOI: 10.1126/science.1173976.

Brauchle, Michael, Karine Baumer, and Pierre Gönczy (2003). “Differential Activation of the DNA Replication Checkpoint Contributes to Asynchrony of Cell Division in C. elegans Embryos”. In: *Current Biology* 13.10, pp. 819–827. ISSN: 09609822. DOI: 10.1016/s0960-9822(03)00295-1.

Brewer, Bonita J. and Walton L. Fangman (Nov. 1988). “A replication fork barrier at the 3′ end of yeast ribosomal RNA genes”. In: *Cell* 55.4, pp. 637–643. DOI: 10.1016/0092-8674(88)90222-x.

Canals-Hamann, Ana Z, Ricardo Pires das Neves, Joyce E Reittie, Carlos Iñiguez, Shamit Soneji, Tariq Enver, Veronica J Buckle, and Francisco J Iborra (Feb. 2013). “A biophysical model for transcription factories”. In: *BMC Biophysics* 6.1. DOI: 10.1186/2046-1682-6-2.

Celis, J E and A Celis (1985). “Cell cycle-dependent variations in the distribution of the nuclear protein cyclin proliferating cell nuclear antigen in cultured cells: subdivision of S phase.” In: *Proceedings of the National Academy of Sciences* 82.10, pp. 3262–3266. DOI: 10.1073/pnas.82.10.3262. eprint: https://www.pnas.org/doi/pdf/10.1073/pnas.82.10.3262.

Cerulus, Bram, Aaron M New, Ksenia Pougach, and Kevin J Verstrepen (2016). “Noise and epigenetic inheritance of single-cell division times influence population fitness”. In: *Current Biology* 26.9, pp. 1138–1147.

Chang, J. B. and Jr. Ferrell J. E. (2013). “Mitotic trigger waves and the spatial coordination of the Xenopus cell cycle”. In: *Nature* 500.7464, pp. 603–7. ISSN: 1476-4687 (Electronic) 0028-0836 (Linking). DOI: 10.1038/nature12321.

Cheedipudi, Sirisha, Oriana Genolet, and Gergana Dobreva (2014). “Epigenetic inheritance of cell fates during embryonic development”. In: *Frontiers in Genetics* 5. DOI: 10.3389/fgene.2014.00019.

Chen, Xinyue et al. (Dec. 2019). “MLL-AF9 initiates transformation from fast-proliferating myeloid progenitors”. In: *Nature Communications* 10.1. DOI: 10.1038/s41467-019-13666-5.

Chu, Lee-Thean, Steven H. Fong, Igor Kondrychyn, Siau Lin Loh, Zhanrui Ye, and Vladimir Korzh (June 2012). “Yolk syncytial layer formation is a failure of cytokinesis mediated by Rock1 function in the early zebrafish embryo”. In: *Biology Open* 1.8, pp. 747–753. DOI: 10.1242/bio.20121636.

Cole, Richard W, Tushare Jinadasa, and Claire M Brown (2011). “Measuring and interpreting point spread functions to determine confocal microscope resolution and ensure quality control”. In: *Nature protocols* 6.12, pp. 1929–1941.

Coleman, Mathew L, Christopher J Marshall, and Michael F Olson (2004). “RAS and RHO GTPases in G1-phase cell-cycle regulation”. In: *Nature reviews Molecular cell biology* 5.5, pp. 355–366.

Collart, C., G. E. Allen, C. R. Bradshaw, J. C. Smith, and P. Zegerman (2013). “Titration of four replication factors is essential for the Xenopus laevis midblastula transition”. In: *Science* 341.6148, pp. 893–6. ISSN: 1095-9203 (Electronic) 0036-8075 (Linking). DOI: 10.1126/science.1241530.

Collart, Clara et al. (May 2014). “High-resolution analysis of gene activity during the iXenopus/i mid-blastula transition”. In: *Development* 141.9, pp. 1927–1939. DOI: 10.1242/dev.102012.

Corey, David R and John M Abrams (2001). “Morpholino antisense oligonucleotides: tools for investigating vertebrate development”. In: *Genome biology* 2.5, pp. 1–3.

Dalle Nogare, D. E., P. T. Pauerstein, and M. E. Lane (2009). “G2 acquisition by transcription-independent mechanism at the zebrafish midblastula transition”. In: *Dev Biol* 326.1, pp. 131–42. ISSN: 1095-564X (Electronic) 0012-1606 (Linking). DOI: 10.1016/j.ydbio.2008.11.002.

Deneke, V. E., A. Melbinger, M. Vergassola, and S. Di Talia (2016). “Waves of Cdk1 Activity in S Phase Synchronize the Cell Cycle in Drosophila Embryos”. In: *Dev Cell* 38.4, pp. 399–412. ISSN: 1878-1551 (Electronic) 1534-5807 (Linking). DOI: 10.1016/j.devcel.2016.07.023.

Dima, A. A. et al. (2011). “Comparison of segmentation algorithms for fluorescence microscopy images of cells”. In: *Cytometry A* 79.7, pp. 545–59. ISSN: 1552-4930 (Electronic) 1552-4922 (Linking). DOI: 10.1002/cyto.a.21079.

Ding, Xiaoyan, Peter Hausen, and Herbert Steinbeisser (Jan. 1998). “Pre-MBT patterning of early gene regulation in Xenopus: the role of the cortical rotation and mesoderm induction”. In: *Mechanisms of Development* 70.1-2, pp. 15–24. DOI: 10.1016/s0925-4773(97)00163-9.

Dolbniak, Marzena, Marek Kimmel, and Jaroslaw Smieja (2015). “Modeling epigenetic regulation of PRC1 protein accumulation in the cell cycle”. In: *Biology direct* 10.1, pp. 1–15.

Duret, Laurent, Franck Dorkeld, and Christian Gautier (1993). “Strong conservation of non-coding sequences during vertebrates evolution: potential involvement in post-transcriptional regulation of gene expression”. In: *Nucleic acids research* 21.10, pp. 2315–2322.

Eckersley-Maslin, Melanie A, Celia Alda-Catalinas, and Wolf Reik (2018). “Dynamics of the epigenetic landscape during the maternal-to-zygotic transition”. In: *Nature Reviews Molecular Cell Biology* 19.7, pp. 436–450.

Edgar, Bruce A, Caroline P Kiehle, and Gerold Schubiger (1986). “Cell cycle control by the nucleo-cytoplasmic ratio in early Drosophila development”. In: *Cell* 44.2, pp. 365–372.

Ferg, M. et al. (2007). “The TATA-binding protein regulates maternal mRNA degradation and differential zygotic transcription in zebrafish”. In: *EMBO J* 26.17, pp. 3945–56. ISSN: 0261-4189 (Print) 0261-4189 (Linking). DOI: 10.1038/sj.emboj.7601821.

Ferraro, Teresa et al. (2016). “Transcriptional memory in the Drosophila embryo”. In: *Current Biology* 26.2, pp. 212–218.

Fowler, Boyd, Chiao Liu, Steve Mims, Janusz Balicki, Wang Li, Hung Do, Jeff Appelbaum, and Paul Vu (Feb. 2010). “A 5.5Mpixel 100 frames/sec wide dynamic range low noise CMOS image sensor for scientific applications”. In: *SPIE Proceedings*. Ed. by Erik Bodegom and Valérie Nguyen. SPIE. DOI: 10.1117/12.846975.

Giraldez, A. J. (2010). “microRNAs, the cell’s Nepenthe: clearing the past during the maternal-to-zygotic transition and cellular reprogramming”. In: *Curr Opin Genet Dev* 20.4, pp. 369–75. ISSN: 1879-0380 (Electronic) 0959-437X (Linking). DOI: 10.1016/j.gde.2010.04.003.

Giraldez, A. J., R. M. Cinalli, M. E. Glasner, A. J. Enright, J. M. Thomson, S. Baskerville, S. M. Hammond, D. P. Bartel, and A. F. Schier (2005). “MicroRNAs regulate brain morphogenesis in zebrafish”. In: *Science* 308.5723, pp. 833–8. ISSN: 1095-9203 (Electronic) 0036-8075 (Linking). DOI: 10.1126/science.1109020.

Hadzhiev, Y. et al. (2019). “A cell cycle-coordinated Polymerase II transcription compartment encompasses gene expression before global genome activation”. In: *Nat Commun* 10.1, p. 691. ISSN: 2041-1723 (Electronic) 2041-1723 (Linking). DOI: 10.1038/s41467-019-08487-5.

Hamperl, Stephan and Karlene A. Cimprich (Dec. 2016). “Conflict Resolution in the Genome: How Transcription and Replication Make It Work”. In: *Cell* 167.6, pp. 1455–1467. DOI: 10.1016/j.cell.2016.09.053.

Hari-Gupta, Yukti et al. (Apr. 2020). “Nuclear myosin VI regulates the spatial organization of mammalian transcription initiation”. In: DOI: 10.1101/2020.04.21.053124.

Harlen, Kevin M and L Stirling Churchman (2017). “The code and beyond: transcription regulation by the RNA polymerase II carboxy-terminal domain”. In: *Nature reviews Molecular cell biology* 18.4, pp. 263–273.

Harrison, M. M., X. Y. Li, T. Kaplan, M. R. Botchan, and M. B. Eisen (2011). “Zelda binding in the early Drosophila melanogaster embryo marks regions subsequently activated at the maternal-to-zygotic transition”. In: *PLoS Genet* 7.10, e1002266. ISSN: 1553-7404 (Electronic) 1553-7390 (Linking). DOI: 10.1371/journal.pgen.1002266.

Harvey, Steven A., Ian Sealy, Ross Kettleborough, Fruzsina Fenyes, Richard White, Derek Stemple, and James C. Smith (July 2013). “Identification of the zebrafish maternal and paternal transcriptomes”. In: *Development* 140.13, pp. 2703–2710. DOI: 10.1242/dev.095091.

Heyn, P., M. Kircher, A. Dahl, J. Kelso, P. Tomancak, A. T. Kalinka, and K. M. Neugebauer (2014). “The earliest transcribed zygotic genes are short, newly evolved, and different across species”. In: *Cell Rep* 6.2, pp. 285–92. ISSN: 2211-1247 (Electronic). DOI: 10.1016/j.celrep.2013.12.030.

Hilbert, Lennart, Yuko Sato, Hiroshi Kimura, Frank Jülicher, Alf Honigmann, Vasily Zaburdaev, and Nadine L. Vastenhouw (2018). In: DOI: 10.1101/234112.

Hnisz, Denes, Krishna Shrinivas, Richard A Young, Arup K Chakraborty, and Phillip A Sharp (2017). “A phase separation model for transcriptional control”. In: *Cell* 169.1, pp. 13–23.

Howe, Kerstin et al. (2013). “The zebrafish reference genome sequence and its relationship to the human genome”. In: *Nature* 496.7446, pp. 498–503.

Huff, Joseph (2015). *The Airyscan detector from ZEISS: confocal imaging with improved signal-to-noise ratio and super-resolution*.

Hwang, Young Sun, Minseok Seo, Sang Kyung Kim, Sohyun Bang, Heebal Kim, and Jae Yong Han (Oct. 2018). “Zygotic gene activation in the chicken occurs in two waves, the first involving only maternally derived genes”. In: *eLife* 7. DOI: 10.7554/elife.39381.

Iborra, F.J., A. Pombo, D.A. Jackson, and P.R. Cook (June 1996). “Active RNA polymerases are localized within discrete transcription “factories’ in human nuclei”. In: *Journal of Cell Science* 109.6, pp. 1427–1436. DOI: 10.1242/jcs.109.6.1427.

Jallow, Zainab, Ulrike G. Jacobi, Daniel L. Weeks, Igor B. Dawid, and Gert Jan C. Veenstra (Sept. 2004). “Specialized and redundant roles of TBP and a vertebrate-specific TBP paralog in embryonic gene regulation in iXenopus/i”. In: *Proceedings of the National Academy of Sciences* 101.37, pp. 13525–13530. DOI: 10.1073/pnas.0405536101.

Jemielita, M., M. J. Taormina, A. Delaurier, C. B. Kimmel, and R. Parthasarathy (2013). “Comparing phototoxicity during the development of a zebrafish craniofacial bone using confocal and light sheet fluorescence microscopy techniques”. In: *J Biophotonics* 6.11-12, pp. 920–8. ISSN: 1864-0648 (Electronic) 1864-063X (Linking). DOI: 10.1002/jbio.201200144.

Jevtić, Predrag and Daniel L Levy (2017). “Both nuclear size and DNA amount contribute to midblastula transition timing in Xenopus laevis”. In: *Scientific reports* 7.1, pp. 1–9.

Jukam, D., S. A. M. Shariati, and J. M. Skotheim (2017). “Zygotic Genome Activation in Vertebrates”. In: *Dev Cell* 42.4, pp. 316–332. ISSN: 1878-1551 (Electronic) 1534-5807 (Linking). DOI: 10.1016/j.devcel.2017.07.026.

Kane, Donald A and Charles B Kimmel (1993). “The zebrafish midblastula transition”. In: *Development* 119.2, pp. 447–456.

Karman, GP, MW Beijersbergen, A Van Duijl, D Bouwmeester, and JP Woerdman (1998). “Airy pattern reorganization and subwavelength structure in a focus”. In: *JOSA A* 15.4, pp. 884–899.

Kastan, Michael B and Jiri Bartek (2004). “Cell-cycle checkpoints and cancer”. In: *Nature* 432.7015, pp. 316–323.

Keller, Philipp J., Annette D. Schmidt, Joachim Wittbrodt, and Ernst H. K. Stelzer (2008). “Reconstruction of Zebrafish Early Embryonic Development by Scanned Light Sheet Microscopy”. In: *Science* 322.5904, p. 1065. DOI: 10.1126/science.1162493.

Kermi, C., A. Aze, and D. Maiorano (2019). “Preserving Genome Integrity During the Early Embryonic DNA Replication Cycles”. In: *Genes (Basel)* 10.5. ISSN: 2073-4425 (Print) 2073-4425 (Linking). DOI: 10.3390/genes10050398.

Kim, AeRi and Ann Dean (May 2012). “Chromatin loop formation in the -globin locus and its role in globin gene transcription”. In: *Molecules and Cells* 34.1, pp. 1–5. DOI: 10.1007/s10059-012-0048-8.

Kimmel, C. B., W. W. Ballard, S. R. Kimmel, B. Ullmann, and T. F. Schilling (1995). “Stages of embryonic development of the zebrafish”. In: *Dev Dyn* 203.3, pp. 253–310. ISSN: 1058-8388 (Print) 1058-8388. DOI: 10.1002/aja.1002030302.

Klages, Natacha and Michel Strubin (1995). “Stimulation of RNA polymerase II transcription initiation by recruitment of TBP in vivo”. In: *Nature* 374, pp. 822–823.

Korobchevskaya, Kseniya, B Christoffer Lagerholm, Huw Colin-York, and Marco Fritzsche (2017). “Exploring the potential of airyscan microscopy for live cell imaging”. In: *Photonics*. Vol. 4. 3. Multidisciplinary Digital Publishing Institute, p. 41.

Kwasnieski, J. C., T. L. Orr-Weaver, and D. P. Bartel (2019). “Early genome activation in Drosophila is extensive with an initial tendency for aborted transcripts and retained introns”. In: *Genome Res* 29.7, pp. 1188–1197. ISSN: 1549-5469 (Electronic) 1088-9051 (Linking). DOI: 10.1101/gr.242164.118.

Langley, A. R., J. C. Smith, D. L. Stemple, and S. A. Harvey (2014). “New insights into the maternal to zygotic transition”. In: *Development* 141.20, pp. 3834–41. ISSN: 1477-9129 (Electronic) 0950-1991 (Linking). DOI: 10.1242/dev.102368.

Larson, Daniel R., Daniel Zenklusen, Bin Wu, Jeffrey A. Chao, and Robert H. Singer (Apr. 2011). “Real-Time Observation of Transcription Initiation and Elongation on an Endogenous Yeast Gene”. In: *Science* 332.6028, pp. 475–478. DOI: 10.1126/science.1202142.

Lau, Y., H. P. Oamen, and F. Caudron (2020). “Protein Phase Separation during Stress Adaptation and Cellular Memory”. In: *Cells* 9.5. ISSN: 2073-4409 (Electronic) 2073-4409 (Linking). DOI: 10.3390/cells9051302.

Laue, K., S. Rajshekar, A. J. Courtney, Z. A. Lewis, and M. G. Goll (2019). “The maternal to zygotic transition regulates genome-wide heterochromatin establishment in the zebrafish embryo”. In: *Nat Commun* 10.1, p. 1551. ISSN: 2041-1723 (Electronic) 2041-1723 (Linking). DOI: 10.1038/s41467-019-09582-3.

Lee, M. T., A. R. Bonneau, and A. J. Giraldez (2014). “Zygotic genome activation during the maternal-to-zygotic transition”. In: *Annu Rev Cell Dev Biol* 30, pp. 581–613. ISSN: 1530-8995 (Electronic) 1081-0706 (Linking). DOI: 10.1146/annurev-cellbio-100913-013027.

Lee, M. T., A. R. Bonneau, C. M. Takacs, A. A. Bazzini, K. R. DiVito, E. S. Fleming, and A. J. Giraldez (2013). “Nanog, Pou5f1 and SoxB1 activate zygotic gene expression during the maternal-to-zygotic transition”. In: *Nature* 503.7476, pp. 360–4. ISSN: 1476-4687 (Electronic) 0028-0836 (Linking). DOI: 10.1038/nature12632.

Lezaja, Aleksandra and Matthias Altmeyer (2018). “Inherited DNA lesions determine G1 duration in the next cell cycle”. In: *Cell Cycle* 17.1, pp. 24–32.

Margottin, Florence, Geneviève Dujardin, Matthieu Gérard, Jean-Marc Egly, Janine Huet, and André Sentenac (Jan. 1991). “Participation of the TATA Factor in Transcription of the Yeast U6 Gene by RNA Polymerase C”. In: *Science* 251.4992, pp. 424–426. DOI: 10.1126/science.1989075.

McIlroy, M. D., E. N. Pinson, and B. A. Tague (July 1978). “UNIX Time-Sharing System: Foreword”. In: *Bell System Technical Journal* 57.6, pp. 1899–1904. DOI: 10.1002/j.1538-7305.1978.tb02135.x.

Menon, T., A. S. Borbora, R. Kumar, and S. Nair (2020). “Dynamic optima in cell sizes during early development enable normal gastrulation in zebrafish embryos”. In: *Dev Biol* 468.1-2, pp. 26–40. ISSN: 1095-564X (Electronic) 0012-1606 (Linking). DOI: 10.1016/j.ydbio.2020.09.002.

Menon, Triveni, Asfa Sabrin Borbora, Rahul Kumar, and Sreelaja Nair (2019). “Dynamically Evolving Cell Sizes During Early Development Enable Normal Gastrulation Movements In Zebrafish Embryos”. In: *bioRxiv*, p. 481325.

Mitchell, Jennifer A. and Peter Fraser (Jan. 2008). “Transcription factories are nuclear subcompartments that remain in the absence of transcription”. In: *Genes &amp Development* 22.1, pp. 20–25. DOI: 10.1101/gad.454008.

Mosheiff, Noga et al. (2018). “Inheritance of cell-cycle duration in the presence of periodic forcing”. In: *Physical Review X* 8.2, p. 021035.

Mura, Marzena, Céline Feillet, Roberto Bertolusso, Franck Delaunay, and Marek Kimmel (2019). “Mathematical modelling reveals unexpected inheritance and variability patterns of cell cycle parameters in mammalian cells”. In: *PLoS computational biology* 15.6, e1007054.

Muramoto, Tetsuya, Iris Müller, Giles Thomas, Andrew Melvin, and Jonathan R Chubb (2010). “Methylation of H3K4 Is required for inheritance of active transcriptional states”. In: *Current Biology* 20.5, pp. 397–406.

Newport, John and Marc Kirschner (1982). “A major developmental transition in early Xenopus embryos: I. characterization and timing of cellular changes at the midblastula stage”. In: *Cell* 30.3, pp. 675–686.

O’Farrell, Patrick H. (Aug. 2015). “Growing an Embryo from a Single Cell: A Hurdle in Animal Life: Figure 1.” In: *Cold Spring Harbor Perspectives in Biology* 7.11, a019042. DOI: 10.1101/cshperspect.a019042.

O’Farrell, Patrick H., Jason Stumpff, and Tin Tin Su (Jan. 2004). “Embryonic Cleavage Cycles: How Is a Mouse Like a Fly?” In: *Current Biology* 14.1, R35–R45. DOI: 10.1016/j.cub.2003.12.022.

Olivier, N. et al. (2010). “Cell lineage reconstruction of early zebrafish embryos using label-free nonlinear microscopy”. In: *Science* 329.5994, pp. 967–71. ISSN: 1095-9203 (Electronic) 0036-8075 (Linking). DOI: 10.1126/science.1189428.

Orsztynowicz, M., D. Lechniak, P. Pawlak, B. Kociucka, S. Kubickova, H. Cernohorska, and Z. E. Madeja (2017). “Changes in chromosome territory position within the nucleus reflect alternations in gene expression

related to embryonic lineage specification". In: *PLoS One* 12.8, e0182398. ISSN: 1932-6203 (Electronic) 1932-6203 (Linking). DOI: 10.1371/journal.pone.0182398.

Padgett, Joe and Silvia D.M. Santos (July 2020). "From clocks to dominoes: lessons on cell cycle remodelling from embryonic stem cells". In: *FEBS Letters* 594.13, pp. 2031–2045. DOI: 10.1002/1873-3468.13862.

Palozola, Katherine C, Greg Donahue, Hong Liu, Gregory R Grant, Justin S Becker, Allison Cote, Hongtao Yu, Arjun Raj, and Kenneth S Zaret (2017). "Mitotic transcription and waves of gene reactivation during mitotic exit". In: *Science* 358.6359, pp. 119–122.

Pampaloni, Francesco, Nariman Ansari, and Ernst HK Stelzer (2013). "High-resolution deep imaging of live cellular spheroids with light-sheet-based fluorescence microscopy". In: *Cell and tissue research* 352.1, pp. 161–177.

Parsons, Glenn G. and Charlotte A. Spencer (Oct. 1997). "Mitotic Repression of RNA Polymerase II Transcription Is Accompanied by Release of Transcription Elongation Complexes". In: *Molecular and Cellular Biology* 17.10, pp. 5791–5802. DOI: 10.1128/mcb.17.10.5791.

Pauli, Andrea, John L Rinn, and Alexander F Schier (2011). "Non-coding RNAs as regulators of embryogenesis". In: *Nature Reviews Genetics* 12.2, pp. 136–149.

Pedregosa, Fabian et al. (2011). "Scikit-learn: Machine Learning in Python". In: *J. Mach. Learn. Res.* 12.null, pp. 2825–2830. ISSN: 1532-4435.

Pei, Jimin and Nick V Grishin (2013). "A new family of predicted Krüppel-like factor genes and pseudogenes in placental mammals". In: *PloS one* 8.11, e81109.

Petrenko, Natalia, Yi Jin, Liguo Dong, Koon Ho Wong, and Kevin Struhl (2019). "Requirements for RNA polymerase II preinitiation complex formation in vivo". In: *Elife* 8, e43654.

Preim, Bernhard and Charl Botha (2014). "Chapter 4 - Image Analysis for Medical Visualization". In: *Visual Computing for Medicine (Second Edition)*. Ed. by Bernhard Preim and Charl Botha. Boston: Morgan Kaufmann, pp. 111–175. ISBN: 978-0-12-415873-3. DOI: https://doi.org/10.1016/B978-0-12-415873-3.00004-3.

Prescott, D. M. and R. F. Kimball (May 1961). "RELATION BETWEEN RNA, DNA, AND PROTEIN SYNTHESES IN THE REPLICATING NUCLEUS OF EUPLOTES". In: *Proceedings of the National Academy of Sciences* 47.5, pp. 686–693. DOI: 10.1073/pnas.47.5.686.

Prescott, D.M. and M.A. Bender (Mar. 1962). "Synthesis of RNA and protein during mitosis in mammalian tissue culture cells". In: *Experimental Cell Research* 26.2, pp. 260–268. DOI: 10.1016/0014-4827(62)90176-3.

Pritchard, David K and Gerold Schubiger (1996). “Activation of transcription in Drosophila embryos is a gradual process mediated by the nucleocytoplasmic ratio.” In: *Genes & development* 10.9, pp. 1131–1142.

Probst, Aline V., Elaine Dunleavy, and Geneviève Almouzni (Mar. 2009). “Epigenetic inheritance during the cell cycle”. In: *Nature Reviews Molecular Cell Biology* 10.3, pp. 192–206. DOI: 10.1038/nrm2640.

R Core Team (2018). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. Vienna, Austria.

Ramachandran, P. and G. Varoquaux (2011). “Mayavi: 3D Visualization of Scientific Data”. In: *Computing in Science Engineering* 13.2, pp. 40–51. ISSN: 1558-366X. DOI: 10.1109/MCSE.2011.35.

Reinberg, Danny and Lynne D Vales (2018). “Chromatin domains rich in inheritance”. In: *Science* 361.6397, pp. 33–34.

Reisser, M., A. Palmer, A. P. Popp, C. Jahn, G. Weidinger, and J. C. M. Gebhardt (2018). “Single-molecule imaging correlates decreasing nuclear volume with increasing TF-chromatin associations during zebrafish development”. In: *Nat Commun* 9.1, p. 5218. ISSN: 2041-1723 (Electronic) 2041-1723 (Linking). DOI: 10.1038/s41467-018-07731-8.

Rhind, N. and D. M. Gilbert (July 2013). “DNA Replication Timing”. In: *Cold Spring Harbor Perspectives in Biology* 5.8, a010132–a010132. DOI: 10.1101/cshperspect.a010132.

Rieder, Dietmar, Zlatko Trajanoski, and James McNally (2012). “Transcription factories”. In: *Frontiers in genetics* 3, p. 221.

Sandler, Oded, Sivan Pearl Mizrahi, Noga Weiss, Oded Agam, Itamar Simon, and Nathalie Q Balaban (2015). “Lineage correlations of single cell division time as a probe of cell-cycle dynamics”. In: *Nature* 519.7544, pp. 468–471.

Sankar, T Sabari, Brigitta D Wastuwidyaningtyas, Yuexin Dong, Sarah A Lewis, and Jue D Wang (2016). “The nature of mutations induced by replication–transcription collisions”. In: *Nature* 535.7610, pp. 178–181.

Schindelin, Johannes et al. (2012). “Fiji: an open-source platform for biological-image analysis”. In: *Nature Methods* 9.7, pp. 676–682. ISSN: 1548-7105. DOI: 10.1038/nmeth.2019.

Schönenberger, Felix, Anja Deutzmann, Elisa Ferrando-May, and Dorit Merhof (May 2015). “Discrimination of cell cycle phases in PCNA-immunolabeled cells”. In: *BMC Bioinformatics* 16.1. DOI: 10.1186/s12859-015-0618-9.

Schulz, K. N. and M. M. Harrison (2019). “Mechanisms regulating zygotic genome activation”. In: *Nat Rev Genet* 20.4, pp. 221–234. ISSN: 1471-0064 (Electronic) 1471-0056 (Linking). DOI: 10.1038/s41576-018-0087-x.

Schulz, Katharine N, Eliana R Bondra, Arbel Moshe, Jacqueline E Villalta, Jason D Lieb, Tommy Kaplan, Daniel J McKay, and Melissa M Harrison (2015). “Zelda is differentially required for chromatin accessibility, transcription factor binding, and gene expression in the early Drosophila embryo”. In: *Genome research* 25.11, pp. 1715–1726.

Shindo, Yuki and Amanda A. Amodeo (June 2020). “Excess histone H3 is a Chk1 inhibitor that controls embryonic cell cycle progression”. In: DOI: 10.1101/2020.06.09.142414.

Stiffler, Lisa A, Jun-Yuan Ji, Susanne Trautmann, Cory Trusty, and Gerold Schubiger (1999). “Cyclin A and B functions in the early Drosophila embryo”. In: *Development* 126.23, pp. 5505–5513.

Sutherland, Heidi and Wendy A. Bickmore (July 2009). “Transcription factories: gene expression in unions?” In: *Nature Reviews Genetics* 10.7, pp. 457–466. DOI: 10.1038/nrg2592.

Tadros, W. and H. D. Lipshitz (2009). “The maternal-to-zygotic transition: a play in two acts”. In: *Development* 136.18, pp. 3033–42. ISSN: 1477-9129 (Electronic) 0950-1991 (Linking). DOI: 10.1242/dev.033183.

Tang, Jiabin, Raymond L Erikson, and Xiaoqi Liu (2006). “Checkpoint kinase 1 (Chk1) is required for mitotic progression through negative regulation of polo-like kinase 1 (Plk1)”. In: *Proceedings of the National Academy of Sciences* 103.32, pp. 11964–11969.

Thatcher, Elizabeth J, Jordan Bond, Ima Paydar, and James G Patton (2008). “Genomic organization of zebrafish microRNAs”. In: *BMC genomics* 9.1, pp. 1–9.

Trinkaus, J. P. (Apr. 1992). “The midblastula transition, the YSL transition and the onset of gastrulation in iFundulus/i”. In: *Development* 116.Supplement, pp. 75–80. DOI: 10.1242/dev.116.supplement.75.

— (Mar. 1993). “The yolk syncytial layer ofFundulus: Its origin and history and its significance for early embryogenesis”. In: *Journal of Experimental Zoology* 265.3, pp. 258–284. DOI: 10.1002/jez.1402650308.

Tsirkas, Ioannis, Daniel Dovrat, Manikandan Thangaraj, Ineke Brouwer, Amit Cohen, Zohar Paleiov, Michael M Meijler, Tineke Lenstra, and Amir Aharoni (Feb. 2022). “Transcription-replication coordination revealed in single live cells”. In: *Nucleic Acids Research* 50.4, pp. 2143–2156. DOI: 10.1093/nar/gkac069.

Tutucci, Evelina, Maria Vera, Jeetayu Biswas, Jennifer Garcia, Roy Parker, and Robert H Singer (Nov. 2017). “An improved MS2 system for accurate reporting of the mRNA life cycle”. In: *Nature Methods* 15.1, pp. 81–89. DOI: 10.1038/nmeth.4502.

Vastenhouw, N. L., W. X. Cao, and H. D. Lipshitz (2019). “The maternal-to-zygotic transition revisited”. In: *Development* 146.11. ISSN: 1477-9129 (Electronic) 0950-1991 (Linking). DOI: 10.1242/dev.161471.

Veenstra, Gert Jan C, Olivier HJ Destrée, and Alan P Wolffe (1999). “Translation of maternal TATA-binding protein mRNA potentiates basal but not activated transcription in Xenopus embryos at the midblastula transition”. In: *Molecular and cellular biology* 19.12, pp. 7972–7982.

Virtanen, P. et al. (2020). “SciPy 1.0: fundamental algorithms for scientific computing in Python”. In: *Nat Methods* 17.3, pp. 261–272. ISSN: 1548-7105 (Electronic) 1548-7091 (Linking). DOI: 10.1038/s41592-019-0686-2.

Wallace, Callen T, Claudette M St. Croix, and Simon C Watkins (2015). “Data management and archiving in a large microscopy-and-imaging, multi-user facility: Problems and solutions”. In: *Molecular reproduction and development* 82.9, pp. 630–634.

Wang, Jianming, Patricia Rojas, Jingwen Mao, Martina Mustè Sadurnì, Olivia Garnier, Songshu Xiao, Martin R. Higgs, Paloma Garcia, and Marco Saponaro (Feb. 2021). “Persistence of RNA transcription during DNA replication delays duplication of transcription start sites until G2/M”. In: *Cell Reports* 34.7, p. 108759. DOI: 10.1016/j.celrep.2021.108759.

Weng, Xiaoli and Jie Xiao (2014). “Spatial organization of transcription in bacterial cells”. In: *Trends in genetics* 30.7, pp. 287–297.

Whitaker, M. (2008). “Calcium signalling in early embryos”. In: *Philos Trans R Soc Lond B Biol Sci* 363.1495, pp. 1401–18. ISSN: 0962-8436 (Print) 0962-8436 (Linking). DOI: 10.1098/rstb.2008.2259.

Wolff, Carsten et al. (2018). “Multi-view light-sheet imaging and tracking with the MaMuT software reveals the cell lineage of a direct developing arthropod limb”. In: *Elife* 7, e34410.

Wu, Yicong et al. (2016). “Simultaneous multiview capture and fusion improves spatial resolution in wide-field and light-sheet microscopy”. In: *Optica* 3.8, pp. 897–910.

Xiaowei, Chen, Zhou Xiaobo, and S. T. C. Wong (2006). “Automated segmentation, classification, and tracking of cancer cell nuclei in time-lapse microscopy”. In: *IEEE Transactions on Biomedical Engineering* 53.4, pp. 762–766. ISSN: 1558-2531. DOI: 10.1109/TBME.2006.870201.

Xie, Jing, Matthew Wooten, Vuong Tran, and Xin Chen (July 2017). “Breaking Symmetry - Asymmetric Histone Inheritance in Stem Cells”. In: *Trends in Cell Biology* 27.7, pp. 527–540. DOI: 10.1016/j.tcb.2017.02.001.

Xu, Mo, Weixiang Wang, She Chen, and Bing Zhu (Nov. 2011). “A model for mitotic inheritance of histone lysine methylation”. In: *EMBO reports* 13.1, pp. 60–67. DOI: 10.1038/embor.2011.206.

Yao, Jie, M. Behfar Ardehali, Christopher J. Fecko, Watt W. Webb, and John T. Lis (Dec. 2007). “Intranuclear Distribution and Local Dynamics of RNA Polymerase II during Transcription Activation”. In: *Molecular Cell* 28.6, pp. 978–990. DOI: 10.1016/j.molcel.2007.10.017.

Yasuda, G (Apr. 1992). “Temporal regulation in the early embryo: is MBT too good to be true?” In: *Trends in Genetics* 8.4, pp. 124–127. DOI: 10.1016/0168-9525(92)90369-f.

Yuan, Kai, Charles A. Seller, Antony W. Shermoen, and Patrick H. O'Farrell (Aug. 2016). "Timing the Drosophila Mid-Blastula Transition: A Cell Cycle-Centered View". In: *Trends in Genetics* 32.8, pp. 496–507. DOI: 10.1016/j.tig.2016.05.006.