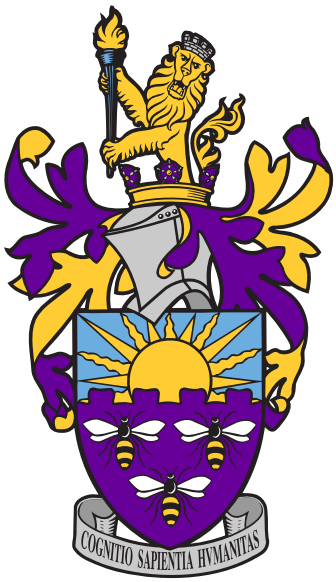
# Decoding the BMP signalling gradient at single cell resolution during *Drosophila* embryogenesis




A thesis submitted to the University of Manchester
for the degree of Doctor of Philosophy
in the Faculty of Biology, Medicine and Health

2019

By
Caroline Hoppe
School of Biological Sciences
Division of Cell Matrix Biology and Regenerative Medicine

# Contents

**Word Count: ∼ 60,000**

# List of Figures

# List of Tables

# List of Abbreviations

**Anterior-Posterior (AP)** Body axis.

**Bone morphogenetic protein (BMP)** Extracellular ligand of the TGF-β family.

**brinker (brk)** Dorsal target and DV patterning gene.

**Chromatin-Immunoprecipitation Sequencing (ChIP-Seq)** Method to detect genome-wide DNA binding sites of transcription factors.

**Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR)** Viral derived DNA sequences present throughout the genome.

**CRISPR associated protein 9 (Cas9)** Enzyme that uses CRISPR sequences as a guide for DNA cleavage.

**decapentapleigic (dpp)** BMP-type extracellular ligand and BMP2/4 orthologe.

**Dorsal (Dl)** Morphogen protein forming a nuclear concentration gradient to active threshold genes.

**Dorsal-Ventral (DV)** Body axis.

**DRB Sensitivity-Inducing Factor (DSIF)** Involved in PPP, cooperatively represses transcription elongation with NELF.

**Fluorescent In Situ Hybridization (FISH)** Fluorescently labelled probes visualise complementary DNA or RNA regions.

**Glass Bottom Boat (Gbb)** BMP-type ligand and BMP 5/6/7/8 ortholog.

**Global Run-On Sequencing (GRO-Seq)** Mapping of genome-wide RNA Pol II distribution.

**Hidden Markov Model (HMM)** Statistical model used here as a memory-adjusted Hidden Markov Model to investigate promoter parameters.

**hindsight (hnt)** DV patterning gene, also known as *pebbled*.

**In Situ Hybridization (ISH)** Labels complementary DNA or RNA regions.

**Initiator Sequence (Inr)** Promoter element overlapping with the TSS.

**maternal-to-zygotic transition (MZT)** Developmental phase of zygotic genome activation.

**Medea (Med)** *Drosophila* co-Smad protein and S-SMAD 2/4 ortholog.

**Mediator of RNA polymerase II transcription (Mediator)** Multiprotein complex that functions as a transcriptional co-activator.

**Mothers against Dpp (Mad)** *Drosophila* receptor Smad protein and R-SMAD 1/5/9 ortholog.

**MS2 Capsid Protein (MCP)** Bacteriophage derived coat protein used in MS2 live

imaging approaches.

**Negative Elongation Factor (NELF)** Involved in PPP, cooperatively represses transcription elongation with DSIF.

**Notch intracellular domain (NICD)** Intracellular domain of transmembrane Notch transduces signalling to the nucleus.

**nuclear cycle (nc)** Rapid mitotic division cycles in *Drosophila* embryogenesis.

**Pannier (Pnr)** BMP target and DV patterning gene.

**phosphorylated Mad (pMad)** Phosphorylation by type II receptor, activation for complex formation with Med.

**Promoter-Proximal Pausing (PPP)** Regulates transcriptional elongation.

**Related to angiotensin converting enzyme (Race)** BMP target and DV patterning gene, also known as *Ance*.

**RNA Polymerase II (Pol II)** Catalyses transcription of DNA.

**Saxophone (Sax)** Transmembrane receptor binging Scw and ACVR1 ortholog.

**Screw (Scw)** BMP-type extracellular ligand and BMP 5/6/7/8 ortholog.

**short gastrulation (sog)** Dorsal target and DV patterning gene.

**single molecule inexpensive FISH (smiFISH)** Quantitative imaging approach visualising single mRNAs using primary and secondary oligos.

**single molecule FISH (smFISH)** Quantitative imaging visualising single mRNAs.

**snail (sna)** Dorsal target gene and DV patterning gene.

**tail-up (tup)** BMP target and DV patterning gene.

**Thickveins (Tkv)** Transmembrane receptor binging Dpp and BMPR1 ortholog.

**Tolloid (Tld)** Protease cleaving Sog and releasing Dpp from shuttling complex.

**Transcription Start Site (TSS)** Location of 5' transcription start.

**Transcription factor (TF)** Sequence-specific DNA-binding factor.

**Transforming growth factor-$\beta$ (TGF-b)** Secreted signalling protein and cytokine.

**twisted gastrulation (tsg)** Dorsal target gene implicated in Dpp shuttling.

**u-shaped (ush)** BMP target and DV patterning gene.

**Zinc-finger early *Drosophila* activator (Zelda) (Zld)** Pioneer transcription factor.

# Abstract

Cells need to accurately decode and integrate information from signalling molecules to regulate cellular processes during development and adult homeostasis. The highly conserved Bone Morphogenetic Protein (BMP) signalling pathway is essential during development. In the *Drosophila* embryo, a gradient of BMP signalling patterns the dorsal ectoderm through the differential regulation of gene expression. It is well established that spatial patterns of gene expression underpin the specification of different cell types. However, the temporal dynamics with which cells respond to cell signals at the transcriptional level is poorly understood.

This project aims to investigate how the BMP signalling gradient is decoded at single-cell resolution to generate transcriptional responses. This is accomplished through the use of quantitative live and fixed imaging complemented with computational modelling of burst kinetics. Static images reveal a graded transcriptional response to the BMP gradient, depending on a cell's position within the expression domain. Highest mRNA numbers are found in cells residing in regions exposed to peak BMP signalling levels. In contrast, some cells at the expression domain border are unable to maintain active transcription from both alleles resulting in reduced transcript numbers per cell. Moreover, evidence is provided that expression of a BMP target gene from both alleles in response to peak signalling requires the full complement of early embryonic enhancers. Correspondingly, investigation of transcriptional bursting parameters, based on live imaging of endogenous BMP target gene loci, showed that cells receiving low BMP signalling levels have poor transcriptional burst kinetics that generate only short, low frequency bursts.

The burst profiles of two BMP target genes, *u-shaped* and *hindsight*, differ significantly in their profiles but both decode BMP signalling levels by modulating promoter occupancy and burst amplitude, suggesting a common mechanism for BMP gradient interpretation. Furthermore, BMP signalling influences promoter occupancy even in the presence of a heterologous promoter, suggesting that the signal is interpreted by the enhancer, which in turn regulates the rate at which the target promoter switches on. In terms of the promoter's contribution to transcriptional bursting, data is presented that the promoter sequence regulates the transcriptional response primarily by altering burst amplitude.

Based on these findings a mRNA threshold model is proposed in which a minimum number of mRNA molecules needs to be produced to ensure robustness of cell fate decisions. Moreover, the results presented here provide a platform for understanding how signals are decoded by individual cells in other contexts during both development and adult homeostasis.

# Declaration

No portion of the work referred to in the thesis has been submitted in support of an application for another degree or qualification of this or any other university or other institute of learning.

# Copyright

i. The author of this thesis (including any appendices and/or schedules to this thesis) owns certain copyright or related rights in it (the "Copyright") and she has given The University of Manchester certain rights to use such Copyright, including for administrative purposes.

ii. Copies of this thesis, either in full or in extracts and whether in hard or electronic copy, may be made only in accordance with the Copyright, Designs and Patents Act 1988 (as amended) and regulations issued under it or, where appropriate, in accordance with licensing agreements which the University has from time to time. This page must form part of any such copies made.

iii. The ownership of certain Copyright, patents, designs, trademarks and other intellectual property (the "Intellectual Property") and any reproductions of copyright works in the thesis, for example graphs and tables ("Reproductions"), which may be described in this thesis, may not be owned by the author and may be owned by third parties. Such Intellectual Property and Reproductions cannot and must not be made available for use without the prior written permission of the owner(s) of the relevant Intellectual Property and/or Reproductions.

iv. Further information on the conditions under which disclosure, publication and commercialisation of this thesis, the Copyright and any Intellectual Property and/or Reproductions described in it may take place is available in the University IP Policy (see `http://documents.manchester.ac.uk/Docu\Info.aspx?DocID=24420`), in any relevant Thesis restriction declarations deposited in the University Library, The University Library's regulations (see `http://www.library.manchester.ac.uk/about/regulations`) and in The University's policy on Presentation of Theses.

Für meine Eltern

Ihr seid mein Ursprung, mein Vertrauen, meine Insel und mein Schatz.
Mein Mund formt euer Lachen, mein Herz schlägt euren Takt.

# Acknowledgements

Undertaking this PhD would not have been possible without the support and guidance that I received from many people. I would like to thank the Wellcome Trust and the President's Doctoral Scholar award for funding this research project.

I would first like to thank my supervisor Hilary Ashe for her guidance and support throughout this PhD, especially during the stressful last months. Thank you for your positivity and encouragement. I would also like to thank Magnus Rattray and Mark Ashe for their valuable insight and discussion.

Thank you to the past and present members of the Ashe lab. You have made this time joyful and caffeinated. Thanks goes to Pryjanka for starting this project, to Lauren for joining me on the transcription endeavours, to Jonathan for his help with computational modelling, and to Sophie for her companionship during late hours. Special thanks goes to Catherine, the fly whisperer for all of her help.

Thank you to Tom for being a great friend and an invaluable collaborator. Thanks for sticking with me and here is to future joint papers! Thank you to the "wiglets" for your camaraderie. I could not have asked to be part of a better group of people on this adventure. Thanks to Stefano for all the runs we went on and being a soundboard for my ideas. Thank you to Beverley and Kathryn for your friendship, pep talks, theatre visits and for keeping me sane during the difficult times.

Thank you, Dan, for teaching me how to be a scientist. Thanks to my Jacobs family, especially Hannah, Franziska, Veronika, Anca, Victoria and Tobias, you keep inspiring me. I have to thank Anna, Irina, Debbie and Dario for always believing in me. I am deeply thankful to my American family and friends who taught me the importance of being adventurous. Thank you to Johanna for a lifelong friendship and together with Enrico for reminding me that there are more important things in life than PhDs and journal articles.

And finally, thank you to my whole family who have all played a part in helping me get to where I am today. Thanks to my sister Franziska, for teaching me how to be curious. To my parents, Astrid and Thomas, you have inspired me in more ways than you could know, and I am grateful for your never-ending love and support. I hope that I have made you proud.

# 1 | General Introduction

## 1.1 Early embryonic *Drosophila* development

How organisms develop tissues, form independent cell fates and organise their body plan has been at the centre of developmental biology, and more specifically the embryology field, for decades. Many lessons have been learned from studying embryonic development in invertebrate organisms, including the model organism *Drosophila melanogaster*. Numerous gene families and pathways are highly conserved between *Drosophila* and vertebrates, making *Drosophila* a valuable model to study developmental processes as well as dysfunction and disease (Reiter et al. 2001, Wangler et al. 2015).

Following fertilisation, *Drosophila* embryogenesis is completed within 24 hours and is partitioned into 17 distinct developmental stages, referred to as Bownes stages (Bownes 1975). Early embryonic development is driven by maternally deposited proteins and RNA before zygotic transcription is initiated (Lefebvre et al. 2018). During the first five stages of development, the embryo undergoes 13 rapid mitotic nuclear division and cleavage cycles, which are synchronised across the embryo (Figure 1.1 A) (Zalokar & Erk 1976). During this time the embryo exists as a syncytial blastoderm, where nuclei share a common cytoplasm enabling molecules to diffuse freely through the cytoplasm, therefore in these divisions karyokinesis occurs without cytokinesis. The first eight mitotic divisions occur rapidly and produce 256 nuclei, most of which then migrate to the periphery, form a monolayer and undergo another 5 mitotic divisions at a slower rate (Foe & Alberts 1983). At the periphery mitotic divisions occur metachronously, where the most anterior and posterior nuclei divide first and mitosis spreads wavelike towards the middle of the embryo (Foe & Alberts 1983, Foe 1989). The cell cycle during nuclear cycle (nc) 1-13 is synchronous and oscillates between M- and S-phase without G1 and G2 phases (Figure 1.1 B) (O'Farrell 2001). During these cell cycles, maternal String (Cdc25 in vertebrates) phosphatase is abundant and keeps the phosphorylated form of Cyclin dependent kinase 1 (Cdk1) to a minimum (Edgar et al. 1994, Stumpff et al. 2004). Cell cycle regulation is first apparent during nc14, where maternal String phosphatase protein is degraded, allowing phosphorylated Cdk1 to accumulate, halting mitosis and introducing a G2 arrest (Figure 1.1 B) (Edgar & O'Farrell 1989, 1990, Edgar et al. 1994).

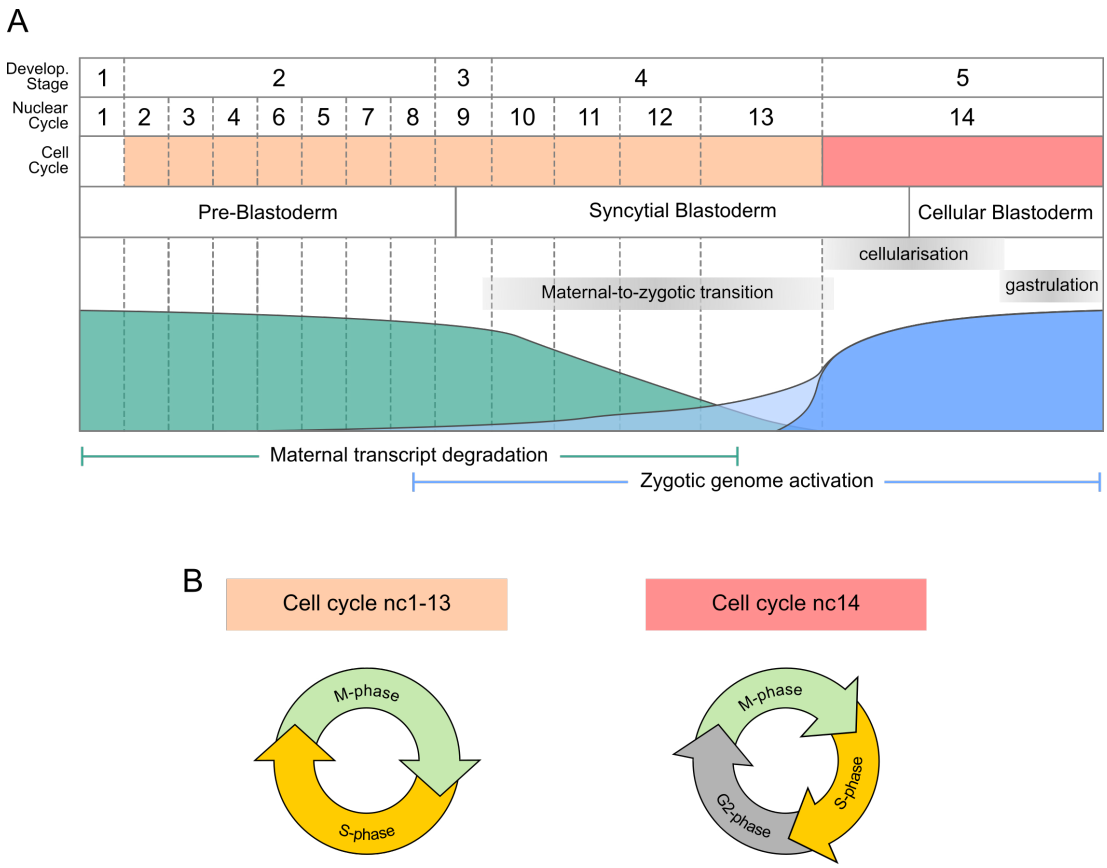



**Figure 1.1: Overview of early *Drosophila* development during nuclear cycles 1-14.** (A) Schematic highlighting mitosis and cleavage cycles, developmental stages (Bownes stages), and key events such as maternal-to-zygotic transition and cellularisation. (B) Cell cycles during early development. During nc1-13, nuclei oscillate between M-and S-phase. The first G2 phase occurs during nc14 and allows cellularisation to occur.

From nc10 onwards, each nuclear cycle increases in length and during nc14, the longest nuclear cycle, cellularisation occurs and cells are partitioned off in apical-to-basal direction, forming the cellular blastoderm (Loncar & Singer 1995). In addition, during nc14 mitotic synchronicity terminates and cells start to form patterns (Foe 1989). At this stage the embryo consists of approximately 6000 cells which were formed over three hours (Zalokar & Erk 1976). Formation of the cephalic furrow, which forms during nc14 and marks early gastrulation, can be used as a morphogenetic marker. The end of nc14 marks the beginning of gastrulation (stage 6), where large morphological changes start to produce multilayered tissues and distinguish the presumptive mesoderm, endoderm, and ectoderm (Figure 1.1 A, B).

### 1.1.1 *Drosophila* early embryonic patterning events

Segmentation and patterning are orchestrated by the spatially restricted activation of zygotic genes that are regulated by maternally supplied molecules. Two independent processes pattern the Anterior-Posterior (AP) and the Dorsal-Ventral (DV) axis. Most genes, that act as key players in axis patterning were identified by an extensive genetic screen in 1984 (Nüsslein-Volhard et al. 1984, Jürgens et al. 1984, Wieschaus et al. 1984). Within a tissue, a cell's fate is determined by its exact three dimensional position and exposure to different signalling molecules. The polarity of tissues and patterning of the body axes is often controlled by molecular signalling gradients, generated by cytokines or morphogens and regulated by concentration thresholds. Morphogens are secreted molecules that work in concentration-dependent manners by establishing gradients in target tissues. Based on the local morphogen concentration, cells gain positional information and modify gene expression accordingly, leading to distinct cell fates (Turing 1952, Wolpert 1969).

#### Dorsal-Ventral axis patterning

The DV axis is initially set up by a group of maternally deposited transcripts and proteins, termed the dorsal-group, which work in concert to form a nuclear morphogen gradient of Dorsal (Dl) protein. In summary, first, the maternally provided Spätzle precursor protein is cleaved by a proteolytic cascade to form an active ligand. This occurs in the perivitelline space at the ventral side of the egg (Stein & Nüsslein-Volhard 1992, Morisato & Anderson 1994, Schneider et al. 1994). The Spätzle ligand then binds to transmembrane Toll receptors, which are present uniformly in the plasma membrane (Hashimoto et al. 1988). This results in the activation of Toll receptors at the ventral side of the embryo, which in turn, leads to the release of Dl from a complex with its inhibitor Cactus. When released, Dl translocates from the cytoplasm to the nucleus (Roth et al. 1991, Geisler et al. 1992, Kidd 1992). The overall concentration of Dl remains unchanged in the embryo but its redistribution leads to the formation of a nuclear concentration gradient. Peak levels of Dl are found in ventral regions, low levels in lateral regions, and very little to none in dorsal regions (Figure 1.2 A) (Roth et al. 1989, Rushlow et al. 1989, Steward 1989).

Nuclear Dl functions as a morphogen and directly regulates target genes through both transcriptional activation and repression. Activation of zygotic gene transcription depends on the nuclear concentration as well as the number and affinity of Dl

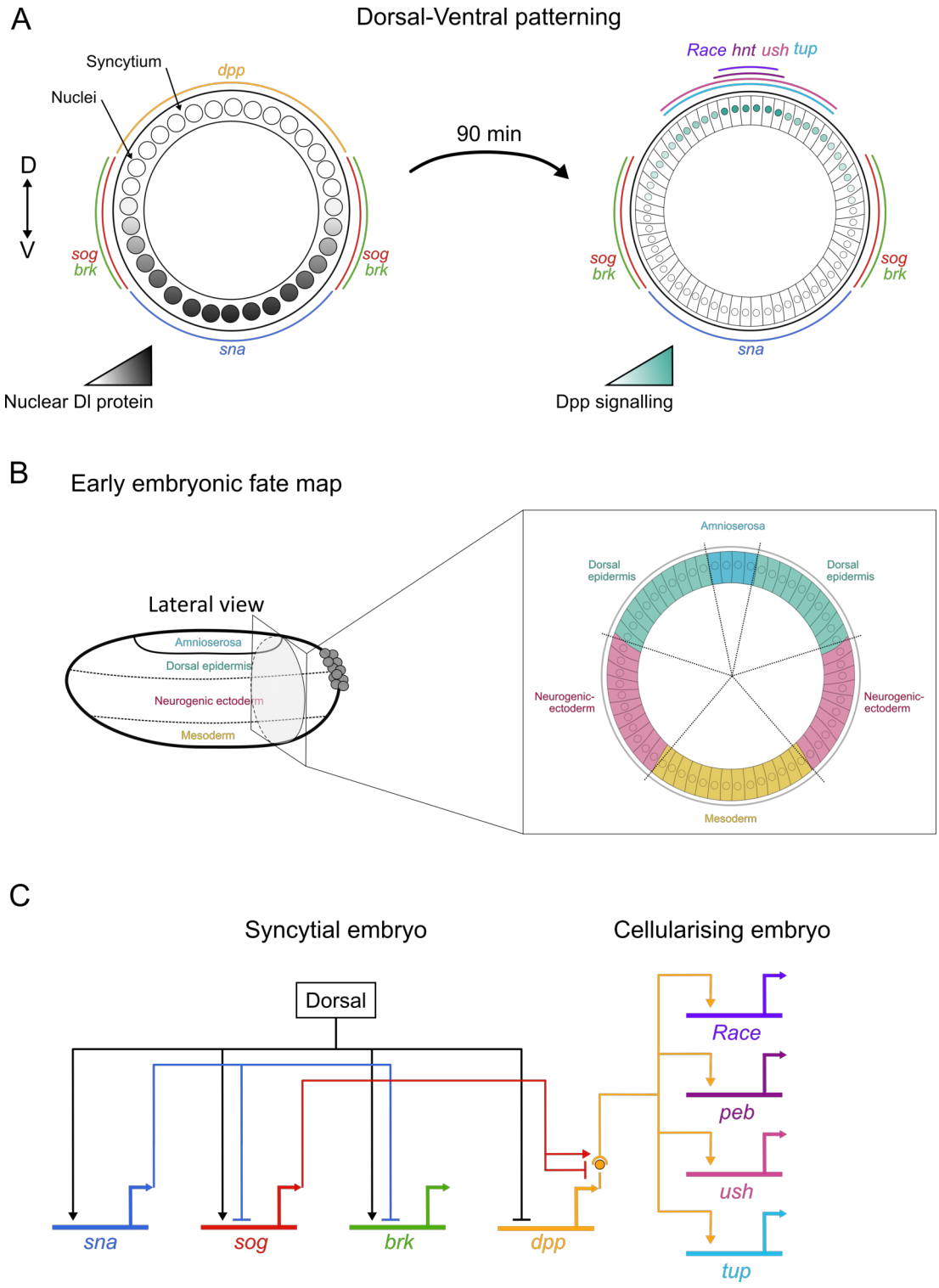



**Figure 1.2: Overview of the early embryonic fate map in *Drosophila*, Dorsal-Ventral patterning and the gene-regulatory network involved in this process.** (A) Cross-section of an embryo. A nuclear gradient of Dl protein partitions the embryo into regions of defined gene expression during the syncytial stage. Absence of nuclear Dl results in *dpp* expression and a gradient of Dpp signalling, in the dorsal region, activating Dpp target genes during later stages. (B) Lateral view and cross-section of the presumptive cell fates before gastrulation. (C) Gene regulatory network which is involved in dorsal-ventral patterning and is important for this study. Including genes expressed in the syncytial and cellularised embryo. The complete gene regulatory network can be found in Levine & Davidson (2005).

binding sites within promoters and/or enhancers of target genes (Rusch & Levine 1996, Stathopoulos et al. 2002). Threshold activated genes turn the information provided by Dl into a much more refined pattern of transcriptional responses, partitioning the embryo into the primary tissues: presumptive mesoderm, neuroectoderm and dorsal ectoderm (St Johnston & Nüsslein-Volhard 1992, Rusch & Levine 1996). These tissues will be refined later in embryonic development to form different cell types and mature tissues (Figure 1.2 B).

Peak levels of nuclear Dl activate expression of the Transcription factor (TF)s *snail (sna)* and *twist* which define the limits of the presumptive mesoderm. Sna acts as a repressor of genes involved in lateral tissue specification, such as *short gastrulation (sog)* and *brinker (brk)* (Figure 1.2 A,C). The dorsal and dorsal-lateral regions of the embryo are patterned by the activation of dpp-group genes. Genes included in this group are *decapentapleigic (dpp)* (St Johnston et al. 1990), *twisted gastrulation (tsg)* (Mason et al. 1994) and sog (Francois et al. 1994) (Figure 1.2 A). The importance of the dpp-group genes in patterning has been shown by their mutant phenotypes. Null alleles resulted in the loss of all dorsal patterning and the expansion of both lateral and ventral tissues (Arora & Nüsslein-Volhard 1992, Mason et al. 1994, Rushlow & Roth 1996). Expression of *sog* is found in a broad, lateral domain, as *sog* promoter elements respond to lower levels of nuclear Dl (Figure 1.2 A). Sog protein has been shown to both, attenuate and enhance Dpp signalling (Francois et al. 1994, Ashe & Levine 1999). Similar to *sog*, *brk* is expressed as a lateral stripe in the presumptive neuroectoderm, but acts here as a transcriptional repressor of Dpp target genes (Jaźwińska et al. 1999, Ashe et al. 2000, Zhang et al. 2001).

The dorsal region of the embryo is not solely patterned by the Dl gradient but is reliant on Dpp mediated target gene activation. In ventral and lateral domains, Dl acts as a repressor of *dpp* (with the help of corepressors) and restricts *dpp* expression to the dorsal-most 40% of the embryo, where its transcription is initiated around stage 5 in a broad pattern (Ray et al. 1991, Huang et al. 1993). Therefore, expression of *dpp* is characterised by the absence of nuclear Dl (Figure 1.2 A). Different threshold levels of Dpp activity partition the presumptive dorsal ectoderm into two tissues, the dorsal epidermis and the amnioserosa (Figure 1.2 B) (Ferguson & Anderson 1992, Wharton et al. 1993). The gene regulatory network important for this study has been summarised in Figure 1.2 C and a detailed description of the Dpp signalling pathway can be found in Section 1.2.

### 1.1.2 Amnioserosa tissue development

The dorsolateral region of the *Drosophila* embryo forms the dorsal epidermis and is initially about 16 cells in width. The dorsal epidermis is characterised by lower levels of Dpp signalling than in the dorsal most part of the embryo, and during later developmental phases forms the tracheal and the peripheral nervous system (Rushlow & Roth 1996). The dorsal-most part of the embryo, which experiences peak Dpp signalling levels, forms the amnioserosa tissue (Figure 1.2 B). The ability of different Dpp signalling levels to specify different cell fates, and the intricate regulation underlying it, was shown in embryos where the *dpp* gene dosage was increased. An experiment by Fergurson and Anderson showed that high levels of injected dpp mRNA into the embryo converted all dorsal ectodermal cells to become amnioserosa (Ferguson & Anderson 1992). Initially, the specification of dorsal tissues is achieved through the coordinated activity of many genes which are part of the dpp-group. Mutations in any of these genes lead to the adoption of lateral cell fates in those which were supposed to adopt the amnioserosa cell fate, leading to the complete loss of the amnioserosa tissue or a reduction of amnioserosa cells (Ray et al. 1991, Arora & Nüsslein-Volhard 1992). A group of genes, called the u-shaped group, is crucial for amnioserosa development and maintenance. Mutant embryos of any gene within this group will lead to premature amnioserosa cell apoptosis or reduced size (Frank & Rushlow 1996). The importance and function of these genes will be reviewed in Section 1.2.2.

The amnioserosa is derived from the dorsal most region of 200 cells in a 5-7 cell-wide domain (Hartenstein & Campos-Ortega 1985, St Johnston & Nüsslein-Volhard 1992). During gastrulation, the cells, destined to form the amniosersa, exist as a thin band and no longer divide post cellularisation but instead undergo cycles of endoreduplication, where replication of the nuclear genome occurs in the absence of mitosis leading to cells becoming polyploid (Figure 1.3). The cells extend and flatten out during germ band extension and spread to form two elongated "arms", which cover a large proportion of the lateral areas of the embryo during developmental stages 8 to 11 (Figure 1.3) (Pope & Harris 2008). At developmental stage 12, approximately 8 hours after egg laying, the aminoserosa cells retract and reshape into an oval sheet which is located dorsally (Schöck & Perrimon 2002). At this point the amnioserosa assists in dorsal closure, where the lateral epidermal tissues are pulled over the invaginating amnioserosa cells (Kiehart et al. 2000). Dorsal closure is achieved by "zipping" together the lateral epidermal cells. Subsequenctly,

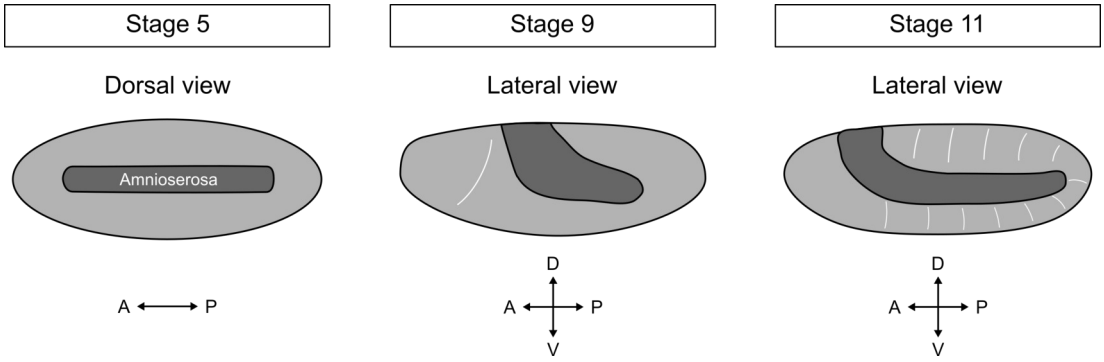



**Figure 1.3: Schematic overview of amnioserosa development.** The amnioserosa tissue is derived from the most dorsal cells during stage 5. Next, cells form two elongated "arms" due to germband extension, which cover lateral areas at stage 9. At stage 11, the amnioserosa is fully extended and cell counts are used to study cell fate decisions.

the apical amnioserosa degenerates through apoptosis at developmental stage 15. Therefore, the aminoserosa tissue is classified as an extraembryonic epithelial tissue, as it is not present in the mature embryo (Hartenstein & Jan 1992, Abrams et al. 1993, Lacy & Hutson 2016, Frank & Rushlow 1996).

Aminoserosa tissue has been shown to support morphogenetic processes through cell signalling events, including RhoA and Jun kinase signalling (Reed et al. 2001, Lamka & Lipshitz 1999), and direct force generation aiding in the cell shape changes (Schöck & Perrimon 2002, 2003, Kiehart et al. 2000). It is highly flexible and morphogenetically active, which allows it to accommodate great changes in cell morphology and cell arrangement during germ band movement. This feature is crucial in allowing it to fulfill its role in germ band movement and dorsal closure as embryos lacking amnioserosa tissue present defects in both processes (Pope & Harris 2008, Kiehart et al. 2000, Frank & Rushlow 1996). The amnioserosa's function in dorsal closure has remarkable similarities to the vertebrate processes of wound healing and neural tube closure and is, therefore, used as a model to study these processes (Gorfinkiel et al. 2011, Heisenberg 2009, Wood et al. 2002, Kiehart et al. 2000).

### Amnioserosa as a model for cell fate

The fully extended aminoserosa tissue at stage 11 is used as a model for cell fate decisions. Since the amnioserosa and dorsal epidermis are specified by signalling strength, the number of amnioserosa cells present in an embryo can be used as a direct readout of the cell fate decisions made earlier during development. At

embryo stage 11, wildtype embryos contain around 130 amnioserosa cells (Figure 1.3) (Wharton et al. 1993, Miles et al. 2008, Newton et al. 2015, Deignan et al. 2016).

Altered levels of Dpp signalling can shift the balance between both cell types resulting in more or fewer cells committing to either fate. The amnioserosa development is very robust with little variation observed in cell number under normal conditions, most likely due to cell specification and patterning events being buffered against changes in environmental conditions or genetic variation (Gavin-Smyth et al. 2013). Embryos can tolerate an increase of up to 250% in amnioserosa cell number but only a 20% decrease without compromising viability (Gavin-Smyth et al. 2013). An increase in variability of amnioserosa cell number is suggested to reflect decanalisation of Dpp signalling and amnioserosa specification (Gavin-Smyth et al. 2013). Through the number of amnioserosa cells in mutant embryos, inferences can be made about how changing levels of Dpp signalling alter cell fate decisions, the importance of specific genes in the process and the effects on dorsal-ventral patterning.

### 1.1.3 Zygotic genome activation

The fact that maternal mRNAs and proteins control early development in multicellular organisms is a universal feature that is shared between animals and plants. Maternal factors (mRNAs and proteins) are known to be loaded into the transcriptionally silenced oocyte during oogenesis (Tadros & Lipshitz 2009). In *Drosophila*, maternally provided mRNAs represent 50 to 75% of the protein-coding transcriptome (7000-10,000 genes) (Tadros et al. 2007, Thomsen et al. 2010). The cell cycle length during the syncytial phases of early *Drosophila* development are too short to allow for successful zygotic transcription. Therefore, the foundation laid by the mother allows embryos to implement fundamental molecular and cellular processes, before zygotic transcription is initiated. The control over the transcriptional processes is handed over to the zygotic genome in a process called maternal-to-zygotic transition (MZT) (Vastenhouw et al. 2019).

The decay of maternal transcripts takes place by targeting the 3'UTRs and occurs in waves with later later waves of decay requiring zygotic decay pathways (Figure 1.1 A) (Bashirullah et al. 1999, 2001, Vastenhouw et al. 2019). Similarly, the zygotic genome is activated in waves starting as early as nc8 (Figure 1.1 A). Experimentally, zygotic transcripts were first identified using in situ hybridisation

(Pritchard & Schubiger 1996), and subsequently they have been studied further using advances in technology such as single embryo mRNA sequencing (Lott et al. 2011) and nano-string analysis (Sandler & Stathopoulos 2016). Lott and colleagues identified approximately 200 genes which are transcriptionally active during nc8, of which 70% are short genes which lack introns. A high proportion of these genes encode TFs, early patterning genes and other developmental regulators. After this transcription gradually increases over the following nuclear cycles. Increase in cell cycle duration impacts on the length and number of mRNAs that can be transcribed in the time window before the next M-phase (Edgar & Schubiger 1986, Rothe et al. 1992). A second wave of zygotic transcription onset is observed during nc14, which coincides with the mid-blastula transition (Figure 1.1 A) (Tadros & Lipshitz 2009). The extent to which cell cycle duration influences genome activation is still unclear (discussed in Vastenhouw et al. 2019).

Only one third of activated genes are strictly zygotic and were not previously provided maternally (De Renzis et al. 2007). Furthermore, it has been observed, that the onset of transcription occurs stochastically in many genes. Studies in *D. rerio* (Stapel et al. 2017) and in *Drosophila* (Boettiger & Levine 2009, Little et al. 2013, Dufourt et al. 2018) showed stochasticity in gene expression and resulting in large cell-to-cell variability. It is suggested that spatial averaging, by diffusion through the syncytium, and temporal averaging, through increased cell cycle duration, establish uniform gene expression (Boettiger & Levine 2009, Little et al. 2013, Stapel et al. 2017).

**Chromatin accessibility**

The early zygotic genome requires the presence of gene specific, regulatory proteins that bind and activate zygotic transcription. In *Drosophila* the pioneer factor regulating transcription during early embryogenesis is maternally deposited Zinc-finger early *Drosophila* activator (Zelda) (Zld), also known as Vielvältig (Staudt et al. 2006, Liang et al. 2008). Zld is a TF and binds DNA regions at consensus motifs, the main one being CAGGTAG. Its wide ranging activity and essential role during MZT was shown in embryos lacking *zld*. *zld* mutant embryos fail to form an intact cellular blastoderm and cannot activate genes important in cellularisation and patterning beyond MZT (Liang et al. 2008). As expected, Zld occupancy is present on the regulatory regions of early zygotic genes, such as Dl targets, and can be detected as early as nc8. Regions bound by Zld are associated with open and accessible chromatin and Zld acts as a predictor of subsequent TF binding by priming enhancers

(Li et al. 2008, MacArthur et al. 2009, Harrison et al. 2011). In its role as pioneer factor, Zld binding is associated with remodelling of the chromatin landscape and increased accessibility. Chromatin opening is achieved through the depletion of nucleosomes allowing spatially controlled enhancer activation and TF recruitment (Harrison et al. 2011, Sun et al. 2015). Using live imaging techniques, the influence and dynamics of Zld binding has recently been investigated *in vivo*. Dufourt et al showed that an increase in Zld binding sites resulted in increased transcriptional synchronicity and hence, accelerated temporal coordination. Furthermore, they showed that Zld accumulates in dynamic local nuclear hubs exhibiting very low residence time and therefore is highly dynamic (Dufourt et al. 2018).

## 1.2 The BMP signalling pathway

The Bone morphogenetic protein (BMP) signalling pathway orchestrates tissue patterning in organism and is facilitated by extracellular BMPs. BMPs are evolutionarily conserved and exist in both vertebrate and invertebrate organisms (Hogan 1996). Vertebrate BMPs were initially identified as factors in demineralised bone extracts that could induce ectopic bone and cartilage formation in mice (Urist 1965, Wozney et al. 1988, Ozkaynak et al. 1992, Chang et al. 1994). Today, it is apparent that BMPs are pivotal signalling molecules and regulators coordinating developmental processes and adult tissue homeostasis (Wozney et al. 1988, Wang et al. 2014, Gaarenstroom & Hill 2014). BMPs belong to the Transforming growth factor-$\beta$ (TGF-b) superfamily. Based on structural and sequence similarities, TGF-$\beta$ superfamily members have been classified into three subgroups- TGF-$\beta$s, BMPs, and activins, with the exception of BMP1, which encodes a metalloprotease (Parker et al. 2004, Ge & Greenspan 2006).

BMPs act through Type I and Type II threonine/serine kinase receptors, which activate intracellular signalling cascades. They signal through canonical SMAD protein pathways and non-canonical SMAD-independent pathways (Shi & Massagué 2003, Wang et al. 2014). During canonical signalling, signal transduction from receptors to the nucleus occurs through SMAD TFs. Individual SMAD proteins continuously shuttle between cytoplasm and nucleus in both the absence and presence of BMP signalling. They are stabilised in the nucleus upon complex formation in response to BMP signal, leading to increased nuclear residence time and regulation of gene expression. This helps to buffer noise and allows dynamic changes and flexibility to respond to signals (Inman et al. 2002, Hill 2009).

As BMPs are involved in many different processes in organism development, mutations have been reported in a number of diseases. Altered levels of BMP signalling and mutations in BMP pathway components have been associated with chronic kidney disease (Simon et al. 1999, Wang et al. 2001), various tumors (Kleeff et al. 1999, Jin et al. 2001, Langenfeld et al. 2003), vascular disorders (Waite & Eng 2003) and skeletal disorders (Salazar et al. 2016).

As TGF-β ligands, BMPs share a common peptide structure. They are synthesised as precursor polypeptides that contain a signal peptide, a C-terminal mature peptide, and a N-terminal pro-domain that is responsible for folding. Dimerisation of ligands occurs through a disulphide bond and maturation involves protease cleavage of the pro-domain by pro-protein convertases, yielding a C-terminal mature form (Gray & Mason 1990, Bragdon et al. 2011).

### 1.2.1 The canonical BMP/Dpp signalling pathway in *Drosophila*

In *Drosophila*, BMP signalling is responsible for patterning the embryonic DV axis (Hogan 1996), the imaginal wing disc (Lecuit et al. 1996, Singer et al. 1997), the formation of pupal wing veins (Yu et al. 1996, De Celis 1997) and for germline stem cell maintenance (Xie & Spradling 1998, Wilcockson & Ashe 2019). In *Drosophila*, three BMP-type ligands exist. Dpp is the vertebrate BMP2/4 orthologue, Glass Bottom Boat (Gbb) and Screw (Scw) are vertebrate 5/6/7/8 relatives. Gbb is found in adult tissues, whereas Scw is found in the early embryo (Padgett et al. 1987, Doctor et al. 1992, Arora et al. 1994). The Dpp protein contains two cleavage sites, generating a full length and a shorter alternative ligand peptide with similar receptor binding affinities (Künnapuu et al. 2009). The different cleavage forms of Dpp could regulate signalling in tissue-specific contexts (Sopory et al. 2010).

#### Dpp morphogen gradient formation

The concept that morphogens regulate tissue patterning is based on the observation that they induce different cell fates in a concentration-dependent manner. Therefore, the existence of a BMP signalling gradient is derived from a BMP ligand concentration gradient. *dpp* and *scw* mRNA however, are uniformly distributed throughout the dorsal 40% of the embryo, suggesting that gradients are established post-transcriptionally and that protein gradients are formed across areas of homogeneous ligand expression (Ferguson & Anderson 1992, Wharton et al. 1993, Arora et al. 1994). BMP homo and heterodimers have different signalling strengths. *In*

*vitro* studies showed that Scw and Dpp homodimers exert only 10% of the signalling activity of the Dpp-Scw heterodimer (Shimmi et al. 2005). Interestingly, Dpp-Scw heterodimers are more stable, have a higher signalling activity, show a greater affinity towards Tsg and Sog and induce Tolloid (Tld)- mediated Sog processing more efficiently (see below) (Shimmi et al. 2005). Together this evidence suggests that Dpp-Scw heterodimers are the primary ligand involved in active transport and strong synergistic signalling and that homodimers are more likely to act locally and initiate low and moderate level signalling and regulating low threshold genes (Figure 1.4) Shimmi et al. (2005).

The formation of the extracellular BMP ligand gradients depends on active transport of BMP ligands to the dorsal midline (Huang et al. 1993). Active transport of ligands is most likely achieved through shuttling of Dpp-Scw to the dorsal midline and is facilitated with the help of BMP inhibitors Sog, Tsg and the metalloprotease Tld (Holley et al. 1996). In lateral domains, Sog and Tsg bind Dpp through their cysteine-rich motifs and prevent binding of sequestered Dpp to its receptors (Shimell et al. 1991, Francois et al. 1994, Ross et al. 2001, Wang & Ferguson 2005). This extracellular complex is then shuttled to dorsal parts of the embryo where ligands are released from the tripartite complex by Tld cleavage of Sog. As a result, the released ligands can signal through receptor complexes in dorsal regions (Shimell et al. 1991, Ashe & Levine 1999, Marqués et al. 2002, Srinivasan et al. 2002). Collagen IV is suggested to act as a scaffold for the Sog, Tsg, Tld and Dpp complex formation (Sawala et al. 2015, Wang et al. 2008, Winstanley et al. 2015).

The resulting extracellular protein gradient of Dpp has been visualised using epitope and fluorescent tags. A genomic epitope-tagged Dpp-HA construct was used by Shimmi et al to visualise Dpp enrichment in a narrow dorsal stripe of the same width as phosphorylated Mad (pMad) (*Drosophila* Smad, see Section 1.2.1) just before gastrulation (Shimmi et al. 2005). In another approach, anti-GFP antibody was injected into the privitelline space to bind to secreted, extracellular Dpp-GFP expressed under the control of *eve* stripe 2 enhancer. Using this approach the extracellular shuttling of Dpp was investigated and a positive feedback mechanism identified where previous BMP signalling promotes future signalling activity. As a result of these studies a spatial bistable receptor model was hypothesised where peak BMP signalling levels result from ligand shuttling and signalling feedback (Wang & Ferguson 2005).

pMad concentrations are are often used as a direct indicator of Dpp signalling strength. pMad levels can be visualised using a phospho-specific antibody (Tanimoto et al. 2000). Initially, pMad is detected in a broad domain, centred around the dorsal midline during the early and mid part of embryo stage 5. In stage 6 the shallow gradient refines to form a steep gradient. This steep gradient is still centred on the dorsal midline, and the area with the highest pMad levels spans a 6 cell width with lower levels of pMad being found in the adjacent 3-4 cells. Therefore, pMad concentrations mirror the BMP signalling step-gradient (Figure 1.5) (Ross et al. 2001, Dorfman & Shilo 2001, Raftery & Sutherland 2003, Xu et al. 2005, Sawala et al. 2015).

**Molecular mechanism of BMP/Dpp signalling transduction**

Initial studies identified the key BMP signalling components through genetic approaches. Based on mutant phenotypes that showed embryonic dorsal-ventral patterning defects, BMP signalling components were characterised and the signalling cascade was dissected. Mature BMP-type ligands bind to heterotetrameric receptor complexes which consist of two heterodimers made up of type I and type II transmembrane receptors. Each contains a serine or threonine intracellular kinase domain. In *Drosophila* Saxophone (Sax) and Thickveins (Tkv) were identified as Type I receptors, which share functional similarities with mammalian TGF-$\beta$ receptors (Nellen et al. 1994). Two type II receptors exist and transduce signalling in *Drosophila*. Punt is most active during embryonic patterning and *Wishful thinking* in the nervous system (Ruberte et al. 1995, Letsou et al. 1995, Marqués et al. 2002). Ligands display preferences in the type I receptor they bind to. Dpp was shown to preferably signal through Tkv, whereas Gbb/Scw bind to Sax receptors. Despite their different receptor affinities, dimers function synergistically (Haerry et al. 1998).

In the *Drosophila* embryo, mature extracellular Dpp/Scw heterodimers assemble a heteromeric complex of Tkv and Sax type I receptors at the plasma membrane. Whereas, mature Dpp or Scw homodimers assemble homomeric dimers of either Tkv or Sax. Ligand-bound type 1 receptors recruit dimers of Punt Type II receptors, forming a heterotetrameric complex (Figure 1.4). Upon ligand binding to the extracellular domain of the receptor complex, type I receptors become activated through trans-phosphorylation by constitutively active Punt (Figure 1.4) (Shi & Massagué 2003). Next, the activated type I receptors recruit and phosphorylate the TF Mothers against Dpp (Mad) (R-SMAD 1/5/9 ortholog). The pMad protein forms a com-

plex with the TF Medea (Med) (vertebrate co-SMAD 4 ortholog). Vertebrate SMADs were shown to form heterotrimers (2 SMADs: 1 co-SMAD) (Inman & Hill 2002). The high conservation of BMP signalling suggests a similar stoichiometry in *Drosophila*. Next, the transcriptionally active pMad and Med complexes accumulate in the nucleus, where they bind to the GC-rich DNA sequences of BMP target genes to regulate their gene expression (Figure 1.4) (Gao et al. 2005).

### 1.2.2 BMP downstream target genes

By regulating the expression of target genes, BMPs influence developmental decisions. These target genes are activated in specific regions of the embryo and their products drive the developmental programs which will define particular structures and organs. Target genes are activated in response to different BMP signalling thresholds, which are defined by their distance to the signalling peak at the dorsal midline. Through the active transport of ligands to the dorsal midline and receptor bistability, the BMP signalling gradient is not continuous but a step gradient (Ashe 2005). The abrupt, step-like changes in concentration are thought to activate target genes in a threshold response manner and sharpen their expression domain boundaries (Raftery & Sutherland 2003, Ashe & Briscoe 2006).

The amnioserosa marker *Related to angiotensin converting enzyme (Race),* also known as *Ance*, is expressed in the dorsal most 5-6 cell wide regions of the embryo, centred around the dorsal midline, and responds to peak levels of BMP signalling (Figure 1.5) (Rusch & Levine 1996, Xu et al. 2005). Race is functionally associated with germ band elongation and shortening. Alongside its role in amnioserosa development, Race is involved in heart morphogenesis and is expressed in the gut, in a similar way to its mammalian homologue *angiotensin converting enzyme* (Tatei et al. 1995, Rusch & Levine 1997).

The zinc finger TF *hindsight (hnt)* is expressed in the presumptive amnioserosa and in cells that border the epidermis, where it is important for the communication between the two tissues (Figure 1.5) (Yip et al. 1997, Lamka & Lipshitz 1999). Hnt is also required for dorsal closure (Reed et al. 2001), and its deletion is lethal highlighting its importance in development (Perrimon et al. 1989). In later stages of development, Hnt is involved in retinal (Pickup et al. 2002) and tracheal (Wilk et al. 2000) development.

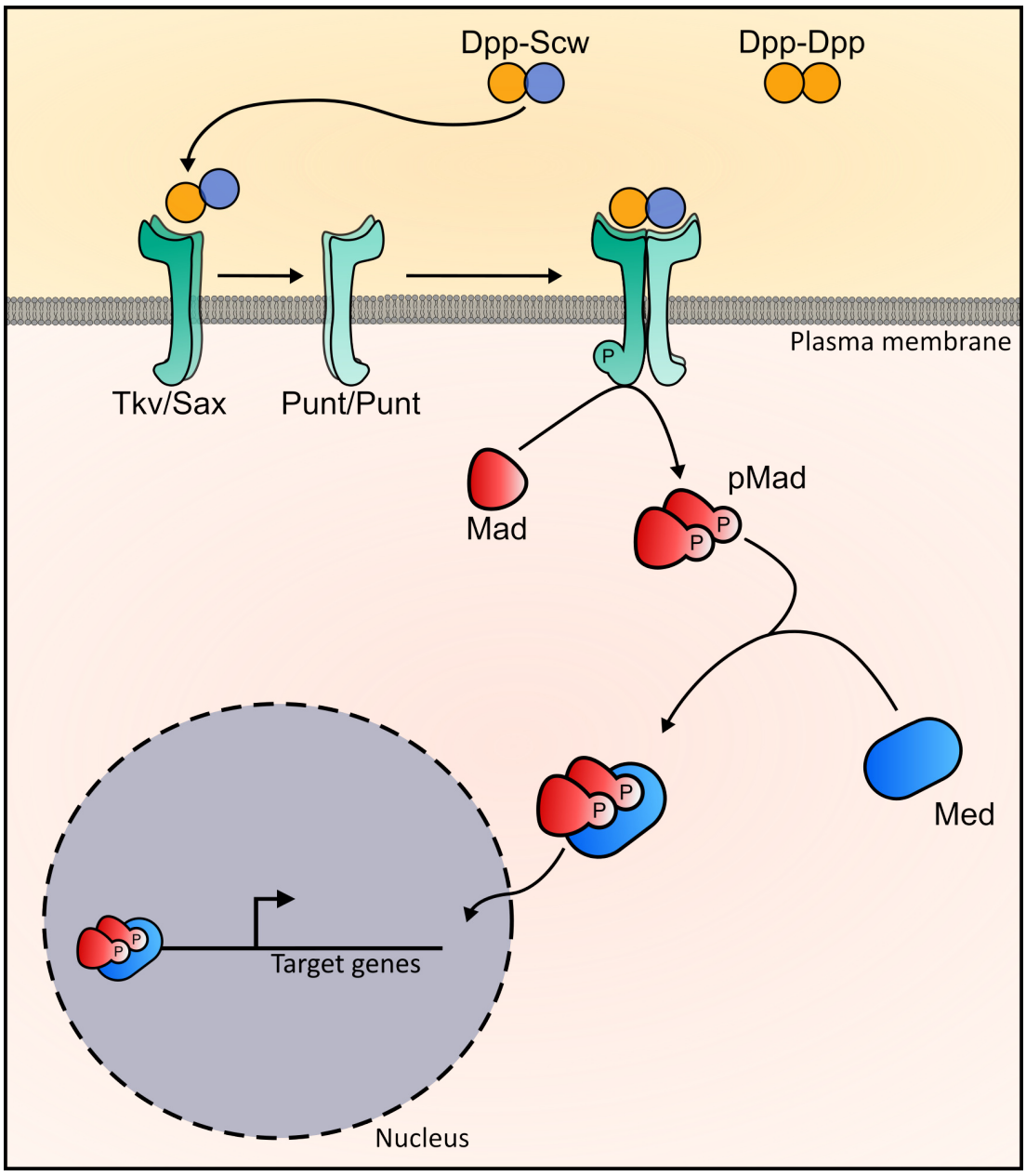



**Figure 1.4: Schematic representation of the canonical Dpp signalling pathway in the *Drosophila* embryo.** Secreted homo- or heterodimers of BMP-type ligands Dpp and Scw bind to Type I Tkv/Sax receptor dimers. Ligand bound Type I receptors recruit two Punt Type II receptors (Wit in the nervous system) to form a heterotetrameric receptor complex. Punt activates Tkv/Sax by trans- phosphorylation, promoting Mad (R-SMAD) phosphorylation. Two pMad molecules form a heterotrimeric complex with the TF Medea, which shuttles to the nucleus where it can regulate target gene expression. At the midline Dpp-Scw heterodimers produce a synergistic signal, regulating high-threshold targets. In lateral regions Dpp homodimers produce moderate, and Scw homodimers low BMP signalling, regulating low threshold genes.

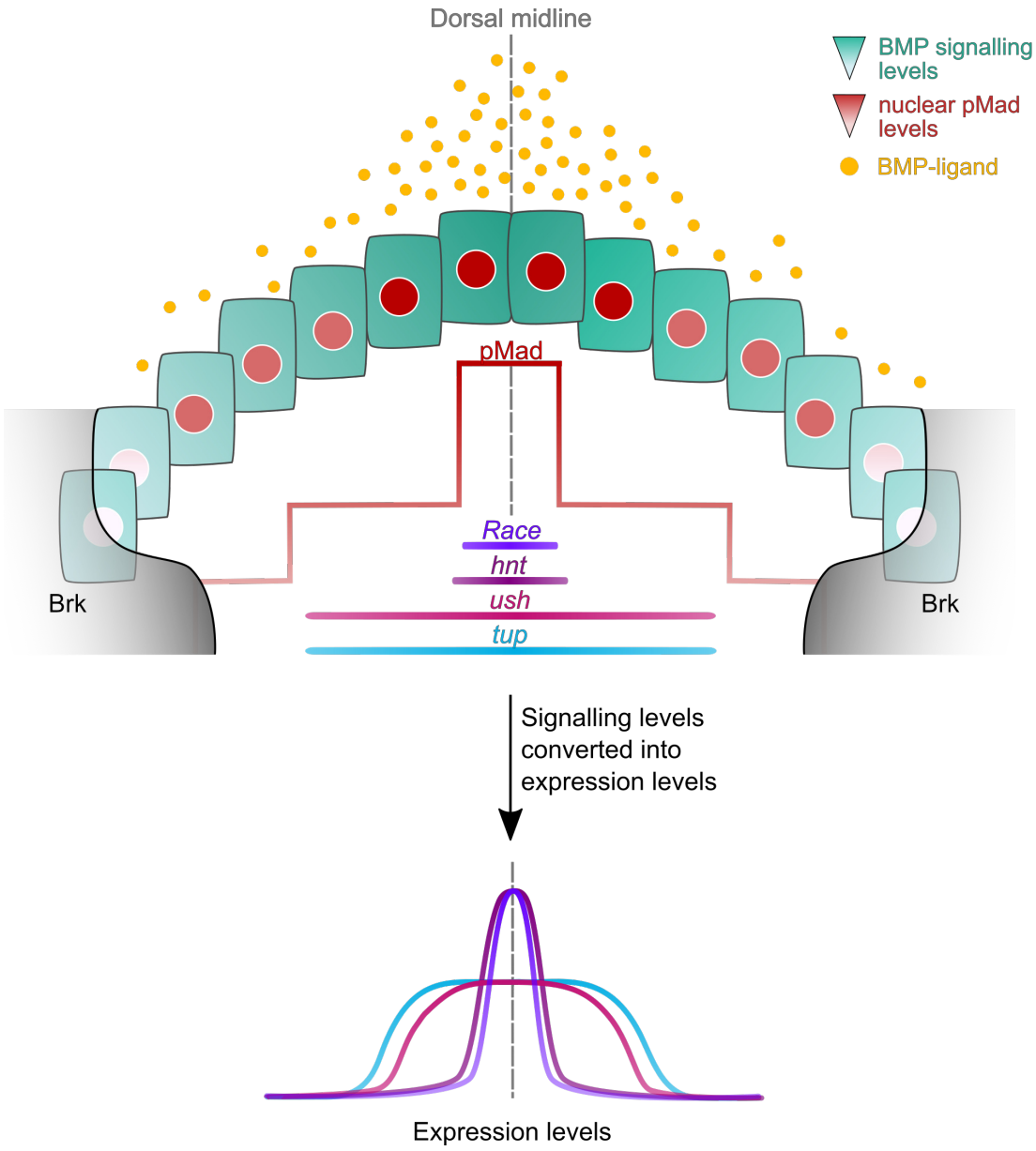



**Figure 1.5: Schematic of an embryo cross-section showing the BMP signalling gradient and downstream target genes.** Active shuttling transports BMP-type ligand dimers to the dorsal midline where they are released from an extracellular complex (not shown), forming a concentration gradient. The differences in extracellular protein concentration are transduced by cells through intracellular pMad levels. pMad assembles a complex with the TF Med (not shown) and forms a step-like gradient, which initiates target gene transcription through binding to regulatory elements. Target genes respond to different signalling thresholds, leading to varying expression domain widths. Direct BMP targets, important for this study, are shown. Target gene expression is limited laterally by the transcriptional repressor Brk.

Transcription of other BMP target genes is initiated in broader expression patterns that span the amnioserosa and expand, to different extents, into dorsolateral regions, indicating activation in response to lower BMP signalling levels. Genes of the u-shaped group belong to this set of genes and include the TF encoding genes *u-shaped (ush)* and *tail-up (tup)* (Figure 1.5) (Nüsslein-Volhard et al. 1984, Frank & Rushlow 1996). The expression domains of *ush* and *tup* are 12-14 cells wide and encompass regions with lower pMad levels (Xu et al. 2005, Sawala et al. 2015). All u-shaped group genes, in addition to other functions, are required for amnioserosa maintenance, with germ band retraction defects observed in null mutant backgrounds (Frank & Rushlow 1996, Goldman-Levi et al. 1996).

The *ush* gene encodes a multitype zinc-finger protein and is a member of the Friend of GATA (FOG) protein family (vertebrate homolog, FOG2) (Cubadda et al. 1997). As such, it can bind GATA factors including other BMP targets Serpent and Pannier (Pnr). *ush* null mutants are embryonic lethal (Nüsslein-Volhard et al. 1984), and multiple *ush* mutant alleles have been described in regards to their effect on bristle patterning on the head and thorax (Cubadda et al. 1997). Ush is also important during cardiac cell specification (Fossett et al. 2000). Additionally, Ush is targeted by the micro RNA mir-8 and together they are implicated in the regulation of *Drosophila* body size, as mir-8 null mutants are smaller in size (Hyun et al. 2009). The *ush* expression domain width corresponds to the shoulder of weak pMad concentration (Figure 1.5) (Raftery & Sutherland 2003).

The gene *tup* encodes a TF that is involved in early embryonic patterning as well as in the development of the imaginal wing disc (De Navascués & Modolell 2007). Additionally, as the homologue of vertebrate Islet1, Tup has important roles in axon pathfinding and motor neuron differentiation (Thor & Thomas 1997). Similar to *ush*, its expression domain borders the second step in BMP signalling levels (Figure 1.5).

The different width of *hnt* and *tup/ush* expression domains represent distinct thresholds of BMP signalling readout. These threshold response genes are limited laterally by Sog inhibition of Dpp, and in the case of *tup* possibly by Brk repression, as shown by ectopic expression using an *eve* stripe 2 enhancer (Ashe et al. 2000).

The large number of BMP target genes involved in *Drosophila* embryogenesis contributes to a highly complex gene regulatory network (part of the network is shown in Figure 1.2 C). Within these networks, many interactions exist between BMP tar-

gets. For example, an interaction between the GATA protein Pnr and Ush was shown during cardiogenesis (Fossett et al. 2000) and thoracic bristle patterning (Haenlin et al. 1997, García-García et al. 1999). Ush modulates the function of Pnr by direct interaction and the formation of a heterodimer through the amino-terminal zinc finger of Pnr, thereby preventing its activity (Haenlin et al. 1997). Additionally, Pnr was shown to positively regulate *ush* and Ush was shown to negatively regulate its own expression (Fromental-Ramain et al. 2008). The network of downstream BMP effectors is very complex, ensuring that gene expression is spatially and temporally regulated.

## 1.3 Transcription, a regulated and dynamic process

In order to adapt to changing environments, cells must have a dynamic system to regulate and coordinate gene expression. This is especially important in, but not limited to, developmental processes. The gene expression dynamics of single cells define their cell-specific signatures. In response to stimuli, cells must adapt gene expression both temporally and spatially. Many studies have investigated how gene expression dynamics are regulated. For example the addition of Tumour Necrosis Factor to cells was shown to induce inflammatory gene expression. Three groups of genes were shown to be simultaneously transcribed but their mature RNA was detected sequentially. The temporal order was found to be coordinated through splicing delays (Hao & Baltimore 2013).

### 1.3.1 Transcription initiation and regulation

Transcription is a highly complex and regulated process. It is controlled through the activity of cis-regulatory elements that bind activating or repressing TFs to modulate gene expression. In order to initiate RNA synthesis, a series of events need to be coordinated, and the transcriptional machinery and pathway for this is very similar among eukaryotes (Thomas & Chiang 2006).

The core promoter is usually defined as the region ~50bp upstream and downstream of the TSS. The core promoter sequence serves as a a platform for the basal transcription machinery and general TFs to assemble and form the pre-initiation complex, and to initiate transcription. As such, it defines the directionality of transcription as well as the Transcription Start Site (TSS). Core promoters themselves have low basal activity but can be activated by enhancers (Smale & Kadonaga 2003,

Haberle & Stark 2018). The pre-initiation complex consists of RNA Polymerase II (Pol II) and a number of general TFs to allow the identification of the exact location for transcription initiation (Hahn 2004). By itself the pre-initiation complex is capable of eliciting low level basal transcription *in vitro* (Roeder 1996). However, in order to initiate transcription in response to a stimulus, the information of enhancer bound activators has to be relayed to the promoter and the pre-initiation complex. This function is fulfilled by co-activator multiprotein-complexes. First identified, and widely conserved between eukaryotes, was the Mediator of RNA polymerase II transcription (Mediator) complex (Flanagan et al. 1991). It became clear that activating transcription factors recruit Pol II to promoters in the form of a holoenzyme containing the pre-initiation complex and the Mediator complex (Malik & Roeder 2000).

After assembly, the transcription complex escapes the promoter and enters elongation. During elongation, the RNA remains stably associated with the transcription complex (Dvir 2002). Traditionally, recruitment of the pre-initiation complex has been considered as the rate-limiting step, and therefore, as the main point of temporal control. However, recent discoveries have challenged this view and evidence suggests that the rate-limiting step varies between genes and can either occur at transcription initiation (Wade & Struhl 2008) or during early elongation through a process called Promoter-Proximal Pausing (PPP) (Krumm et al. 1995, Lis 1998). Occurring shortly after polymerase release from the promoter PPP can be used to synchronise transcription over a field of cells (Lis 1998).

#### 1.3.1.1 Promoter-proximal pausing

**Pause induction, release and the importance of the promoter sequence**

Pausing can occur at multiple sites from +20 to +40 bp downstream of the TSS (Rasmussen & Lis 1993, Giardina et al. 1992) and is facilitated by the binding of pausing factors such as DRB Sensitivity-Inducing Factor (DSIF) and Negative Elongation Factor (NELF). *In vitro* studies showed that these two factors are sufficient to inhibit early elongation (Yamaguchi et al. 2002, Narita et al. 2003). Additional proteins could also be involved and contribute to Pol II stalling (Cheng et al. 2012). Pol II pausing is important to keep promoters in an active and open conformation while the paused polymerase competes with the +1 nucleosome (Gilchrist et al. 2010). The degree of PPP varies among genes, is dictated by the gene's promoter sequence (Narita et al. 2003, Lagha et al. 2013) and PPP is measured as the ratio

of Pol II density in the promoter region to the the Pol II density in the gene body, this is referred to as the pausing index (Core et al. 2008). PPP is a highly dynamic processes where the polymerase can escape rapidly from the pause site to resume elongation, enabling rapid changes in transcriptional output (Lis 1998).

**Gene regulation through promoter-proximal pausing**

Gene expression is often stochastic due to variable recruitment and availability of Pol II and the other protein complexes which are necessary for transcription (Raj et al. 2006). PPP is an effective way to transform stochastic gene activation into a synchronous process by pre-loading stalled polymerases onto a gene. Pol II release from pausing, in response to a stimulus, reduces cell-to-cell variation in gene expression (Boettiger & Levine 2009). Synchronicity, elongation speed and the robustness to environmental fluctuations are important properties of transcription, cell fate and tissue formation (Boettiger et al. 2011). Multiple studies have shown that pausing directly controls elongation speed and gene expression synchrony (Yao et al. 2007, Boettiger & Levine 2009, Boettiger et al. 2011).

High levels of pausing, measured by pausing indices, were shown to lead to more synchronous gene activation (Boettiger & Levine 2009). Conversely, gene expression of *sna* was transformed from synchronous to stochastic by reducing the degree of PPP, which was shown to be detrimental to development (Lagha et al. 2013). Based on these findings, it was hypothesised that different degrees of pausing are important for the correct temporal expression of genes, on the time-scale of minutes, required for tissue patterning. Pausing influences many BMP pathway regulated genes including the previously discussed TFs involved in AS maintenance (Table 1.1) (Saunders et al. 2013). Genome wide analysis has reported that genes exhibiting pausing are enriched in signalling pathways and play important roles in stress response, cell proliferation, and cell differentiation (Zeitlinger et al. 2007, Guenther et al. 2007, Muse et al. 2007, Saunders et al. 2013). In the *Drosophila* embryo, it is estimated that at least half of all developmental control genes are paused, suggesting that PPP may act as a checkpoint that controls the fine tuning of gene expression during developmental processes (Levine 2011). A high number of these genes continue to be paused throughout embryogenesis independent of their expression status (Gaertner et al. 2012).

**Table 1.1: Pausing indices and their respective percentile placement for Dpp target genes involved in the DV axis patterning.** Data was retrieved from a Global Run-On-Sequencing study by Saunders et al. 2013.

| Gene Name | Pausing index | Pausing index percentile |
|---|---|---|
| *Race* | 72 | Top 25% |
| *hnt* | 102 | Top 25% |
| *ush* | 5 | Bottom 50% |
| *tup* | 288 | Top 25% |

#### 1.3.1.2 Promoter elements

Core promoter sequences have been shown to display structural and functional diversity (Maston et al. 2006, Butler & Kadonaga 2002). As such, the core promoter sequence was found to contribute and dictate the degree of PPP a gene is subjected to (Tang et al. 2000, Hendrix et al. 2008, Lagha et al. 2013).

The core promoter can contain a number of motifs. The TATA box was the first motif to be described, is usually found 24-31 nt upstream of the TSS but it is only present in a minority of promoters (Breathnach & Chambon 1981, Haberle & Stark 2018). In *Drosophila* it is found in 5% of all core promoters (Ohler et al. 2002, FitzGerald et al. 2006). The TATA box is identified by its consensus sequence TATAWAWR and is bound by the TATA-box binding protein (TBP), which is a general transcription factor (Haberle & Stark 2018). The TATA box has been implicated in PPP. *Drosophila* genes with a low pausing index are enriched for the TATA box sequence (Gaertner et al. 2012, Shao & Zeitlinger 2017) and studies in mammalian cells have suggested that the TATA box assists in Pol II release from pausing (Amir-Zilberstein et al. 2007, Montanuy et al. 2008). However, the *Drosophila hsp70* promoter contains a TATA box and is known to be highly paused (Gilchrist et al. 2010, Buckley et al. 2014). Hence, the connection between TATA box presence and PPP is complex. Studies have suggested that besides the presence, the optimal spacing of promoter elements is important to predict their influence on PPP. Genes with a TATA box near the -30 position showed higher degrees of PPP than genes with a TATA at -40 or more (Kwak et al. 2013). Additionally, TATA box presence was shown to negatively influence the Pol II half life in some cases but the overall promoter element context remained important (Shao et al. 2019).

Another core promoter motif is the Initiator Sequence (Inr), which encompasses the TSS. This motif is more abundant than the TATA box and can function independently of it. The Inr consensus sequence is TCAGTY (Smale & Baltimore 1989, Haberle & Stark 2018). The Inr sequence was shown to be correlated with long Pol II half-life and a sub-sequence was identified to play a dominant role in stabilising paused Pol II during PPP (Shao et al. 2019).

A number of other downstream motifs exist, besides the TATA-box and Inr. The pause button, motif ten element and downstream core element were found to correlate with strong Pol II pausing and are enriched in highly paused genes in *Drosophila* (Hendrix et al. 2008, Gaertner et al. 2012, Lenhard et al. 2012, Shao & Zeitlinger 2017). Pol II pausing can be stabilised by changing the downstream promoter regions containing these motifs (Shao et al. 2019).

### 1.3.2 Enhancers and dynamic activity

Enhancers are sequences that regulate transcription and therefore control spatio-temporal patterns of gene expression. They contain TF binding motifs which can either have activating or repressive functions. Enhancer sequences can have long-range activities and can be found upstream, downstream or even within introns of the neighbouring gene (Maston et al. 2006, Levine 2010). In *Drosophila*, some enhancer-promoter interactions occur over a length of 20kb or more (Kvon et al. 2014).

Genes are often regulated by multiple enhancers. Novel enhancers were identified in Chromatin-Immunoprecipitation Sequencing (ChIP-Seq) assays and where there are multiple these are termed "shadow enhancers". Previously identified enhancers were termed primary enhancers (Hong et al. 2008). This nomenclature does not define the importance or strength of the enhancer sequence. Shadow enhancers initially seemed redundant in function but were shown to be required for the robustness of gene expression in development (Bothma et al. 2015, Frankel et al. 2010, Cannavò et al. 2016).

The functional significance of shadow enhancers was reported in studies where additivity and buffering capacities were investigated. Shadow enhancers buffer transcription against environmental factors and confer robustness in forming gene expression patterns (Frankel et al. 2010, Bothma et al. 2015, Cannavò et al. 2016). When embryos were subjected to high temperature or genetically reduced levels

of Dl, the full complement of enhancers was needed for normal embryogenesis to occur. The removal of either enhancer led to gastrulation defects under stress conditions (Dunipace et al. 2011).

According to the conventional model of enhancer-promoter interactions, transcription is initiated through DNA looping which brings the enhancer and promoter into close proximity with physical contact (Bulger & Groudine 1999, Carter et al. 2002, Levine 2010). DNA looping is supported using chromosome conformation capture assays (Rao et al. 2014, Ramírez et al. 2018, Schoenfelder & Fraser 2019). With recent evidence, however, the model of direct enhancer-promoter contact is being challenged. Live imaging studies have shown that a shared enhancer can produce coordinated transcriptional bursts from two genes concurrently. Alternating activation of transcription would be expected if looping was occurring (Fukaya et al. 2016). In agreement, another study provided evidence that promoters of different genes do not compete for direct contact with a shared enhancer (Lim et al. 2018). Moreover, others have hypothesised that no direct physical contact can occur due to molecular crowding around the gene's promoter (Heist et al. 2019). Instead it has been postulated that a "transcription hub" is formed, where enhancers, promoters, TFs and polymerases form a microenvironment, which is sufficient for activation (Hnisz et al. 2017, Chen et al. 2018, Lim et al. 2018, Heist et al. 2019).

Such transcriptional hubs could be formed within the nucleus by phase separation (Hnisz et al. 2017). Liquid-liquid phase separation has been postulated to form membraneless micro compartments in cells. This physicochemical process helps to separate fluids into dense and less dense phases and therefore separate molecules into condensates, examples include the nucleolus and stress granules (Sabari et al. 2018). The existence of phase separation compartments could explain the function of another type of enhancer, termed super-enhancers. They are defined as large clusters of enhancer elements that regulate the function of important genes through the assembly of transcription machinery at a very high density (Whyte et al. 2013, Hnisz et al. 2013). Super-enhancer hubs are proposed to assemble through phase separation as the high density of transcription machinery, including Mediator, was shown to rapidly form and dissolve (Lovén et al. 2013, Brown et al. 2014, Mansour et al. 2014), which are properties associated with phase separated compartments (Sabari et al. 2018). Phase separated compartments could explain recent observations such as a single enhancer activating two genes simultaneously (Fukaya et al. 2016), but more quantitative and functional studies are needed (Hnisz et al. 2017). In *Drosophila*, maternal proteins Biocid and Zld have

been shown to bind DNA very transiently. They have also been shown to form highly dynamic, high-concentration hubs within nuclei, suggesting that the formation of dynamic multi-factor hubs could help cells to regulate dynamic transcription spatially and temporally (Mir et al. 2018).

## 1.4 Studying transcriptional dynamics

### 1.4.1 Quantitative imaging systems

#### 1.4.1.1 Single molecule In Situ Hybridisation

New quantitative imaging techniques have allowed the investigation of single cell transcription kinetics, revealing how gene expression in neighbouring cells can be highly stochastic (Kaufmann & van Oudenaarden 2007). Historically, gene expression domains were visualised using conventional In Situ Hybridization (ISH) or Fluorescent In Situ Hybridization (FISH), which often rely on enzymatic reactions or other signal amplification steps. These techniques generate a qualitative image of mRNA localization within cells or expression domains and have contributed to a large number of discoveries in relation to tissue specific gene regulation (Driever et al. 1989, Ashe et al. 2000, Gregor et al. 2014). However, these techniques are limited as the products of enzymatic reactions are small molecules or precipitates that can diffuse away from their production site. If transcripts are labelled directly with fluorophores they maintain their spatial information but their sensitivity is often poor (Gregor et al. 2014).

To combat these shortcomings, Raj and colleagues (Raj et al. 2008) improved a method that uses short nucleic acid probes, pioneered by the Singer laboratory (Femino et al. 1998), to visualise individual transcripts. This methods is called single molecule FISH (smFISH). Up to 48 different short oligonucleotide probes are synthesised, complementary to the target transcript and labelled with a single fluorophore moiety at the probe's 3' terminus. Using these probes Raj and colleagues were successful in gaining information about the spatial localization of the detected mRNA molecules as well as quantitative information about the number of transcripts present within a cell (Figure 1.6 A). Besides testing their system in cell culture; they presented proof of concept experiments in *C. elegans* and the *Drosophila* wing disc. The signal- to-noise ratio of smFISH probes is much improved compared to conventional FISH techniques. By using different fluorophores, mul-

tiple genes can be targeted within the same experiment, allowing simultaneous assessment of a number of gene transcripts (Raj et al. 2008). Similarly, two different fluorophores can be used to synthesise probes against intronic and exonic gene regions, allowing discrimination of pre-mRNA and mature mRNA molecules respectively (Figure 1.6 A) (Bahar Halpern & Itzkovitz 2016).

The development of smFISH has allowed researchers to draw quantitative conclusions about gene expression. Studies have used this system to address a wide range of research questions in *Drosophila*, highlighting changes in gene expression activity, copy number variation, stochastic and synchronous gene expression and mRNA localization (Boettiger & Levine 2009, Lagha et al. 2013, Garcia et al. 2013, Little et al. 2013, Boettiger & Levine 2013). Since the fluorescent intensity of a single transcript can be determined, inferences can be made about the number of nascent transcripts present at active transcription sites in nuclei at the time of fixation (Little et al. 2013). In 2016 the single molecule inexpensive FISH (smiFISH) system was described, which combines unlabelled primary probes with fluorescently labelled secondary detector probes. Using unlabelled oligonucleotides for probe detection offers a cost advantage, as probe synthesis without attached fluorophores is less expensive and fluorescently labelled detector probes can be used for multiple experiments (Tsanov et al. 2016).

Other methods capable of quantitative single molecule detection in fixed tissue are In Situ Hybridisation Chain Reaction (HCR) (Dirks & Pierce 2004) and RNAscope. The latter uses multiple tandem probes that hybridise to the target transcript. The tree like structure pre-amplifies the signal by an adapter and is often used for whole mount tissue preparation (Wang et al. 2012).

#### 1.4.1.2 Stem loop based live imaging systems

Although smFISH offers insights into gene expression levels and temporal gene control, it only captures a snap shot image of the cell's state at time of fixation. Therefore, new live imaging techniques have been developed to follow transcription events live and capture the dynamic nature of gene transcription.

**Development and initial studies**

A new live imaging system was pioneered by the Singer laboratory to study ASH1 mRNA localisation in *Saccharomyces cerevisiae*. They used a mRNA tagging system,

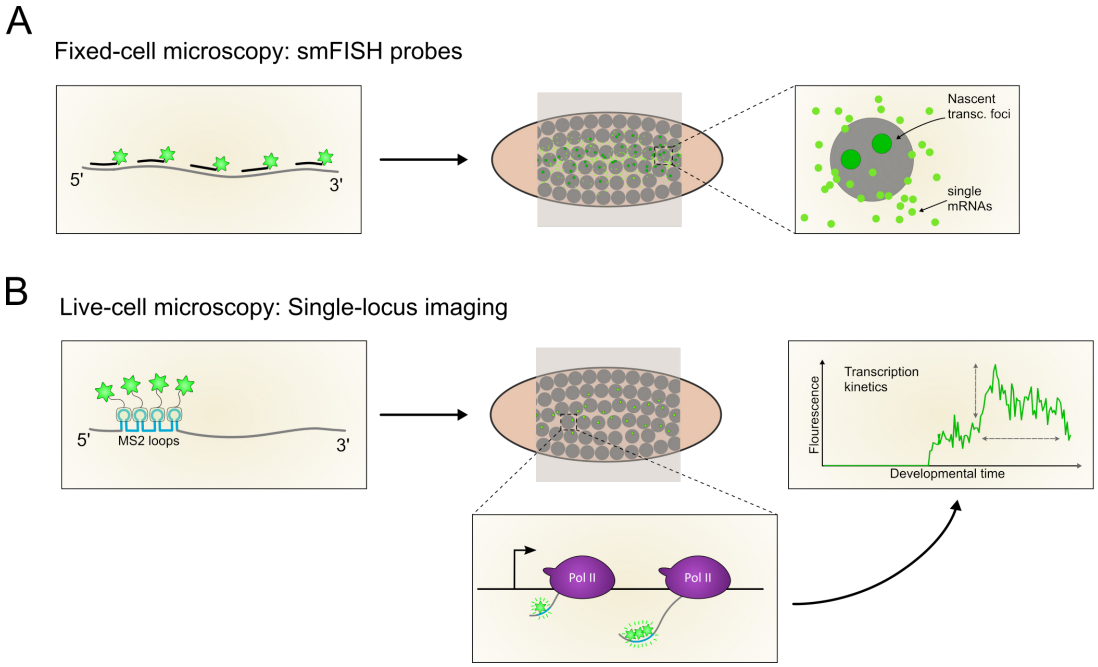



**Figure 1.6: Quantitative imaging techniques.** (A) Single molecule FISH is used to visualise and quantify nascent nuclear transcription sites and cytoplasmic mature mRNAs in fixed embryos. (B) Real-time transcription dynamics can be quantified using stem-loop based live imaging systems. MS2 stem loops are bound by a coat protein-eGFP fusion and transcription can be monitored through fluorescent traces.

based on components of the *MS2 bacteriophage* (Bertrand et al. 1998). In this system, the bacteriophage derived MS2 Capsid Protein (MCP) is expressed and fused to a fluorophore for detection. The MCP specifically recognises a 19 nucleotide RNA MS2 stem loop sequence and binds each MS2-loop cooperatively as a dimer (LeCuyer et al. 1996). The gene in question can be tagged with multiple MS2 stem loop sequence repeats. Once transcribed, the stem loops are recognised and bound by the MCP and detected by the fluorescence of the attached fluorophore (Figure 1.6 B). The high affinity of MCP to MS2 stem loops means that the binding is almost irreversible (Carey & Uhlenbeck 1983, Bertrand et al. 1998, Urbinati & Long 2011). Homologous systems exist that use similar mechanisms involving RNA stem loops. Examples are the PP7 loops and their respective coat protein, derived from the *Pseudomonas* phage (Lim et al. 2001, Lim & Peabody 2002), U1A RNA and its binding protein (Oubridge et al. 1994, Moras & Poterszman 1995) and other alternative labelling systems, reviewed in Weil et al. (2010).

### Applications

After initial studies in yeast cells, this MS2 system has been slightly modified to accommodate experiments in mammalian cells, hippocampal neuron cultures, plant cells, *D. rerio* embryos, and *Drosophila* oocytes and whole embryos (Rook et al. 2000, Fusco et al. 2003, Forrest & Gavis 2003, Hamada et al. 2003, Garcia et al. 2013, Campbell et al. 2015). Fusing a target gene to 24 MS2 stem-loop sequence cassettes was sufficient to detect single molecules in cell culture (Fusco et al. 2003, Shav-Tal et al. 2004). Advances were made by cloning constructs with increased numbers of stem loops leading to an increase in fluorescence. A reporter with 120 functional stem loop cassettes, inserted into an intron, was used to visualise single HIV-1 RNAs and allowed the detection of Pol II convoys in HeLa cells (Tantale et al. 2016). Detecting single transcripts in multicellular embryos, however, is still challenging and may require a few hundred fluorescent molecules per transcript (Gregor et al. 2007, Garcia et al. 2013).

The MS2 imaging system was established in *Drosophila* embryos in 2013. Studies presented evidence that spatial and temporal averaging of gene expression modulates developmental patterns using transgenic reporters. Specifically, transcription, driven by the proximal hunchback enhancer, throughout early nuclear division cycles and the increase of *hunchback* transcription periods in response to Bicoid were visualised (Garcia et al. 2013, Lucas et al. 2013). Transcription profiles were shown to be mostly constant between transcription initiation and termination approximately an hour later (Garcia et al. 2013). In comparison, a similar experimental set-up showed that the expression pattern generated by the *even skipped* stripe two enhancer is highly dynamic and features episodic transcription profiles of 4-10 minutes, termed transcriptional bursts (Bothma et al. 2014).

Transcriptional regulation was further characterised by a study investigating the contribution of individual enhancers to a gene's transcription profile. *knirps* enhancers were shown to act in an additive or even super-additive fashion, where the combined activity of two enhancers is greater or equal to the predicted sum of two individual enhancers. Interestingly, *hunchback* enhancers were shown to work in a sub-additive manner, where the presence of both enhancers resulted in reduced transcriptional output compared to either enhancer alone in the anterior region of the embryo. At the posterior *hunchback* expression domain border, however, enhancers transition from a sub-additive to an additive behaviour in response to sharply decreasing Bicoid activator concentrations. This behaviour

was concluded to represent different enhancer strengths, where weaker enhancers (producing fewer transcripts) function additively and stronger enhancers function sub-additively as they are more likely to interfere with each other (Bothma et al. 2015).

Moreover, dynamic enhancer-promoter interactions were investigated by visualising the process of transvection using the MS2 system. Here, a gene without an enhancer in *cis*, was activated by an enhancer in *trans*. Interestingly, the two alleles (one in *trans* and one in *cis*) could be activated at the same time by one enhancer, suggesting that enhancers do not directly contact promoter regions (Lim et al. 2018). These and other live imaging studies contributed significantly to the understanding of gene expression kinetics and highlighted the importance of dynamic transcription during early embryonic development (Ferraro et al. 2016, Dufourt et al. 2018, Lucas et al. 2018).

### 1.4.2 Studying transcriptional bursting

Since stochasticity and noise is inherent to biological systems, heterogeneity is expected when it comes to transcription events. However, for a long time the process of transcription was treated as a continuous and smooth process, with a constant probability of transcription at each time point (Raj & van Oudenaarden 2008, Wang et al. 2019). This model would imply a constant rate of transcription, proportional RNA degradation, and small variance, therefore transcription would follow a Poisson distribution (Nicolas et al. 2017). Traditionally, RNA expression was measured and analysed using bulk methods such as RT-PCR, microarrays and bulk RNA-Seq, where thousands of cells were assayed simultaneously. In these approaches all dynamic information is lost due to averaging over a large number of cells, and this often resulted in the interpretation of this data as a continuous transcription rate (Chubb et al. 2006, Wang et al. 2019). Recent technological advances, such as single cell RNA-Seq, smFISH and the MS2 live imaging system have enabled measurement of single cell transcription rates.

Single cell analysis revealed an intrinsic stochasticity in many genes, where transcription occurs in non-continuous bursts of different time intervals, separated by periods of transcriptional inactivity. This mode of transcription can be found in bacteria and multicellular organisms (Elowitz et al. 2002, Golding et al. 2005, Chubb et al. 2006, Fukaya et al. 2016). Bursting has also been shown to arise from cell cycle effects (Padovan-Merhar et al. 2015), short term oscillations (Bonev et al. 2012,

Corrigan & Chubb 2014, Phillips et al. 2017) and long term oscillations, such as the circadian rhythm (Hurley et al. 2016).

**Burst kinetics**

Transcriptional bursting can be described by burst parameters. Burst frequency describes the number of transcription events and burst size describes the number of transcripts that are produced during a single burst. The burst size is calculated based on the burst amplitude and duration (Figure 1.7) (Chubb et al. 2006, Dar et al. 2012). Despite studies investigating the underlying kinetics of transcriptional bursting, no unified model exists of the molecular mechanisms which drive this process (Fujita et al. 2016).

Theoretical studies have proposed that transcriptional bursting occurs in a two-step mechanism, which switches between ON-states, where mRNA is produced, and OFF-states, without mRNA production (Peccoud & Ycart 1995, Lionnet & Singer 2012, Munsky et al. 2012). Computational modelling approaches have implicated two state models to analyse transcription kinetics (Larson et al. 2009, Lammers et al. 2019, Larsson et al. 2019). Recently, a three-state model has been proposed to account for the activity of sister-chromatids after DNA replication (Lammers et al. 2019).

Changes in transcriptional output are modulated by underlying changes in transcriptional burst kinetics. A study by Dar et al. showed that transcriptional bursting is the dominating mode of transcription across the human genome and that both burst frequency and/or burst size are modulated in order to adjust transcriptional output (Figure 1.7) (Dar et al. 2012). Over 20 years ago, a study postulated that enhancers control transcription probability, i.e. burst frequency, but not the expression level, i.e. burst size (Walters et al. 1995). This is supported by data showing that burst frequency is increased when the β-*globin* enhancer was forced to interact with its promoter through chromatin looping. The deletion of the β-*globin* enhancer resulted in reduced burst frequency (Bartman et al. 2016). Another study has shown that the addition of estrogen to MCF7 cells increased burst frequency of responding enhancers (Fritzsch et al. 2018).

Further evidence that enhancers regulate burst frequency, comes from studies in *Drosophila* embryos. The burst frequency but not the burst size was shown to differ between enhancers controlling early patterning (Fukaya et al. 2016). A recent

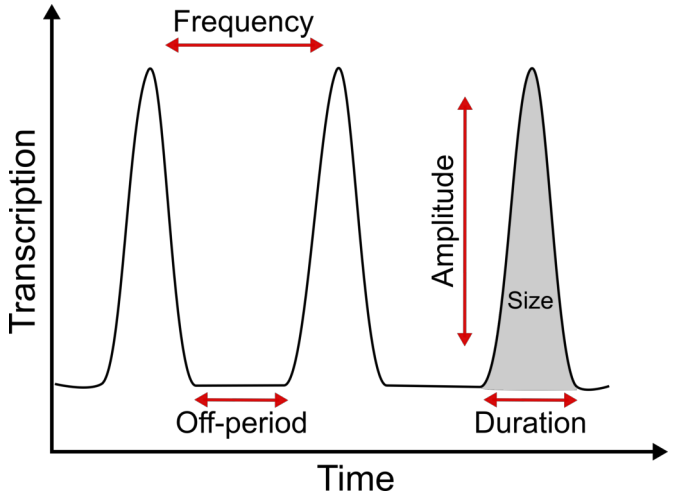



**Figure 1.7: Transcriptional bursting.** Transcription can occur in non-continuous stochastic bursts. Changes in transcriptional output can be explained by transitions in burst kinetics. Any change in burst duration, frequency or amplitude will alter the number of mRNAs produced.

study evaluated genome wide burst parameters by single cell RNA-Seq in primary mouse fibroblasts. Again, this study showed that gene expression is primarily defined and shaped by enhancers through burst frequency control. Additionally, the authors suggest that burst size is encoded by promoters (Larsson et al. 2019). Some evidence implicates promoter structure modulation in affecting burst kinetics (Raj et al. 2006, Jones et al. 2014), however, the role of promoters in burst dynamics has not yet been addressed through live imaging.

## 1.5 Aims and Objectives

In summary, BMP signalling plays a crucial role in embryogenesis and developmental patterning. The BMP activity gradient is transduced through an intracellular signalling cascade that regulates downstream target genes. Their tightly regulated gene expression patterns contribute to cell fate decisions and tissue specification.

The precise understanding of how morphogen gradients are decoded at the single cell level and converted into transcriptional programmes, remains an unanswered question. Investigating transcription dynamics requires quantitative experimental approaches, as many of the molecular processes occur on the time scale of seconds. This project aims to define how target genes interpret the BMP signalling gradient at single cell resolution. To achieve this I will address the following objectives:

1. **Quantitate the transcriptional response of target genes in single cells using smFISH.**

2. **Define the temporal dynamics of gradient interpretation using live imaging.**

3. **Investigate how promoter and enhancer elements modulate the transcription output in individual cells.**

# 2 | Materials and Methods

## 2.1 Fly stocks, embryo collection and fixation

### 2.1.1 Fly husbandry and stocks

*Drosophila* of the genotype $y^{67c23}w^{118}$ were used as wildtype. Flies were maintained at 18°C or 25°C in vials containing standard yeast medium or for embryo collections in collection bottles with apple juice agar plates containing yeast paste. A full list of *Drosophila* stocks used in this study can be found in Table 2.1.

### 2.1.2 Embryo collection

Embryos were collected from overnight plates. For other experiments embryos were aged to 2-4h on apple juice agar plates before fixation. Embryos were dechorionated in 2% bleach for 2 minutes. The bleach was washed off with water and embryos were transferred into fixation vials.

**Fixation for RNA in situ hybridisation:**

Embryos were fixed in a mixture of 3ml fixing buffer (1.3 x phosphate buffered saline (PBS), 67mM EGTA pH 8), 1ml of 37% formaldehyde and 4ml heptane on a shaker for 20 min. After the formaldehyde and heptane mix was removed, embryos were devitellinised in 8ml methanol through shaking and stored at 4°C in methanol.

**Fixation for immunostain:**

Embryos were fixed in 3.6ml Buffer B (10mM $KPO_4$ pH 6.8, 15mM NaCl, 45mM KCl, 2mM $MgCl_2$, 10mM $NaH_2PO_4$, 10mM $Na_2HPO_4$) 0.4ml 37% formaldehyde and 4ml heptane on a shaker for 12 min. The vitelline membrane was removed in 8ml of methanol through shaking and the embryos were stored in methanol at 4°C.

**Table 2.1: List of *Drosophila melanogaster* lines used in this study.**

| Genotype | Origin | Identifier |
|---|---|---|
| $y^{67c23}w^{118}$ | Stathalopolous Lab, used as wildtype strain | N/A |
| **Live imaging lines** | | |
| $y^{67c23}w^{118}$; $\Delta$*Push/cyO* | Priyanka Upadhyai, Ashe Lab | N/A |
| $y^{1}M${*vas-Cas9*}*ZH-2A*$w^{118}$, $\Delta$P*hnt*/FM7; | Priyanka Upadhyai, Ashe Lab | N/A |
| $y^{67c23}w^{118}$; *ushP-24xMS2-ush* | Priyanka Upadhyai, Ashe Lab | N/A |
| $y^{67c23}w^{118}$; *hntP-24xMS2-ush* | Priyanka Upadhyai, Ashe Lab | N/A |
| $y^{1}M${*vas-Cas9*}*ZH-2Aw* $^{118}$, *hntP-24xMS2-hnt;* | Priyanka Upadhyai, Ashe Lab | N/A |
| $y^{1}M${*vas-Cas9*}*ZH-2A*$w^{118}$, *ushP-24xMS2-hnt;* | Priyanka Upadhyai, Ashe Lab | N/A |
| $y^{67c23}w^{118}$;*ushP-*$\Delta$*24xMS2-ush* | Caroline Hoppe | N/A |
| $y^{67c23}w^{118}$; *hntP-*$\Delta$*24xMS2-ush* | Caroline Hoppe | N/A |
| $y^{1}w^{*}$;*P{His2Av-mRFP1}II.2; P{nos-MCP.EGFP}2* | Bloomington Stock Center | FBst00060340 |
| $y^{1}w^{67c23}$; *MKRS, P{ry[+t7.2]=hsFLP}86E/TM6B, P{w[+mC]=Crew}DH2, Tb*$^{1}$ | Bloomington Stock Center | FBst00001501 |
| $y^{1}$ $w^{67c23}$; sna[Sco]/CyO, Pw[+mC]=CrewDH1 | Bloomington Stock Center | FBst0001092 |
| $y^{1}$ *M{vas-Cas9}ZH-2A*$w^{118}$ | Bloomington Stock Center | FBst00051323 |
| **Mutant lines** | | |
| $y^{1}$ *sog*$^{S6}$*/FM7c, sn*$^{+}$;; | Bloomington Stock Center | FBst0002497 |
| *st2-dpp1 Flpd* | Ashe Lab | FBal0118266 Ashe et al. (2000) |
| ;;*ru*$^{1}$ *h*$^{1}$ *Pry*$^{+t7.2}$*=neoFRT82B sr*$^{1}$ *e*$^{s}$ *Med*$^{13}$*/TM3, Sb*$^{1}$ | Bloomington Stock Center | FBst00007340 |
| $w^{118}$; *Df(2L)BSC332/CyO;* | Bloomington Stock Center | FBrf0204503 |
| *yw*; *dpp*$^{Hin37}$ */ Dp(2;2) DTD48* ; *dpp*$^{ho2}$ *Gla* | Bloomington Stock Center | FBal0003057 Irish & Gelbart (1987) |
| **Transgenic lines** | | |
| $y^{67c23}w^{118}$ ;; *tup prox enh evePyellow* | Catherine Sutcliffe, Ashe Lab | N/A |
| $y^{67c23}w^{118}$ ;; *tup sh-enh 1 evePyellow* | Catherine Sutcliffe, Ashe Lab | N/A |

| | | |
|---|---|---|
| $y^{67c23}w^{118}$ ;; *tup sh-enh 2 evePyellow* | Catherine Sutcliffe, Ashe Lab | N/A |
| $y^{67c23}w^{118}$ ;; *tup triple enh evePyellow* | Catherine Sutcliffe, Ashe Lab | N/A |
| $y^{67c23}w^{118}$ ;; *ush prox enh evePyellow* | Catherine Sutcliffe, Ashe Lab | N/A |
| $y^{67c23}w^{118}$ ;; *ush sh-enh 1 evePyellow* | Catherine Sutcliffe, Ashe Lab | N/A |
| $y^{67c23}w^{118}$ ;; *ush sh-enh 2 evePyellow* | Catherine Sutcliffe, Ashe Lab | N/A |

## 2.2 Transgenics and genome modifications

### 2.2.1 Transgenic reporter fly lines

#### Reporter constructs of *tup* shadow enhancers

The following lines were designed and generated by Catherine Sutcliffe. The plasmid pLacZattB (kind gift from the Stathalopolous lab, described in Deignan et al. (2016)) was used as the plasmid backbone for cloning. This plasmid contains a lacZ gene under the minimal *even-skipped* promoter with a SV40 late polyadenylation signal. All primers used in this cloning scheme can be found in Table A.1. The backbone was amplified excluding the lacZ gene and a SphI restriction site was introduced on either side. The *yellow* gene, containing both exon and intron sequences, was amplified from the plasmid pmCD8-GFPy+ (Addgene, 24350) and SphI restriction sites were introduced. After digestion with SphI (NEB, R3182), the backbone and *yellow* gene insert were ligated together using infusion ligase (Takara, 639650). Next, *tup* enhancer sequences were generated. A schematic of the genomic *tup* region can be found in Figure 5.25. The proximal *tup* shadow enhancer, located within the first intron, was identified by the Levine lab (Zeitlinger et al. 2007), amplified from wildtype genomic fly DNA and inserted using a XbaI restriction site. The next enhancer, here called shadow enhancer 1, was identified by Deignan et al., amplified and inserted into the plasmid backbone using a BglII restriction site. The third enhancer, here called shadow enhancer 2, was also identified by Deignan et al, amplified from wildtype genomic DNA and inserted into the plasmid using an AgeI restriction site. This cloning scheme resulted in a triple enhancer cassette upstream of the even-skipped minimal promoter driving yellow gene expression. This plasmid was injected by the Cambridge Injection Service into

the 86Fb attP landing site on the 3rd chromosome (BL 24749). Additionally, single enhancer constructs were generated using the same cloning scheme and primers.

**Reporter constructs of *ush* shadow enhancers**

Reporter constructs for presumptive *ush* enhancer sequences were created by Catherine Sutcliffe in an identical cloning scheme to the one described above for *tup* reporter constructs. *ush* enhancer sequences were identified from pMad ChIP-Seq peaks (Deignan et al. 2016). All primer sequences that were used in this cloning scheme can be found in Table A.2.

## 2.2.2 Genome modifications by CRISPR editing

Life imaging fly lines of the genotype ushP-24xMS2-ush, hntP-24xMS2-hnt, hntP-24xMS2-ush and ushbP-24xMS2-hnt were generated by Priyanka Upadhyai in the Ashe Lab using Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR)-CRISPR associated protein 9 (Cas9) genome editing. Their design will be described, as they were used for experiments in this study. Additionally, precursors of these fly lines were used to generate transgenic lines during this study.

### 2.2.2.1 MS2 cassette insertion into the endogenous 5'UTR

**Strategy overview**

A two step approach was used to insert 24x MS2-loops into the endogenous 5'UTR of the Dpp target genes *ush* and *hnt*. First, a knockout of the endogenous promoter and 5'UTR region was created and replaced with an attP site. In the second step MS2-loop cassettes together with the promoter sequence were inserted through φC31 P-element reintegration. The MS2 stem loop cassettes used in this study was obtained from the plasmid pCR-24xMS2SL (Addgene, PL 31865). A single MS2 cassette contains two non-identical loop sequences, tacggtacttattgccaagaaaGCACGAGCATCAGCCGTGCctccaggtcgaatcttcaaaCGACGACGATCACGCGTCGctccag tattccagggttcatc. Here, the non-identical stem loops are shown in uppercase letters. These stem loops were initially synthesized to reduce sequence redundancy and the possibility of recombination in bacteria (Bertrand et al. 1998).

### Gene targeting

In short, homologous recombination was used to allow for the deletion of endogenous genomic sequences. Two PAM sites were used to generate double strand breaks and a homology containing donor plasmid was supplied for homology- directed repair. The dsDNA donor plasmid was injected together with plasmids encoding two guide RNAs (gRNA), targeting the endogenous locus (strategy based on Baena-Lopez et al. (2013)). The PAM sites and guide RNA sequences were determined with the flyCRISPR Optimal Target Finder tool:
(http://flycrispr.molbio.wisc.edu/tools). To study gene transcription during early embryogenesis, the *ush* isoform RC and the *hnt* RA and RB isoforms were targeted in this approach.

The following genomic target sites were used to create double strand breaks. Cas9 nuclease cut sites, 3nt upstream from PAM, highlighted in red and PAM sequence in upper case cyan:
*ush* genomic target sequence 1 (plus strand): 5' gtctcgtctcgtccccgxctcCGG 3'
*ush* genomic target sequence 2 (minus strand): 5' CCCacgxggaagtgacaacataat 3'

*hnt* genomic target sequence 1 (minus strand): 5' CCAatgxtgtaatcctatttgcgc 3'
*hnt* genomic target sequence 2 (minus strand): 5' CCGtcgxcaactgttgaacacaa 3'

The plasmid pTV-cherry (kind gift from the Vincent lab; DGRC, 1338) was used as the donor plasmid. The strategy is summarised in Figure 2.1. The plasmid contains membrane-cherry and a mini-white gene for screening as well as an attP reintegration site and is flanked on either side by homology arms (HA) (Figure 2.1A). HAs for *ush* and *hnt* were amplified from genomic DNA, obtained from the *Drosophila* line BL 51323 (Table 2.1) and their specifications can be found in Table 2.2. HA1 and HA2 were inserted into the pTV-cherry plasmid using KpnI and SpeI restriction sites respectively Figure 2.1A.

The pU6-BbsI-chiRNA plasmid (Addgene, 45946) was used to deliver gRNAs. 5' phosphorylated oligos were annealed and ligated into BbsI digested pU6-BbsI-chiRNA plasmid (Table 2.2).

Together, gRNA plasmids and the donor plasmid were injected into Cas9 expressing flies (BL 51323, Table 2.1) by the Cambridge University injection service (Figure 2.1B). All surviving flies were crossed to wildtype flies and the offspring screened

for red eyes originating from the mini-white marker. Successful *ush* CRISPR flies were then balanced by crossing them to a 2nd chromosome balancer and successful *hnt* flies to a 1st chromosome balancer. Afterwards, the marker region was removed by crossing the flies to a fly line expressing cre-recombinase (BL1501, Table 2.1, Figure 2.1C). The generated ΔP*ush* fly line contains a 465bp genome deletion (Chr 2L: 523446-523902) and the ΔP*hnt* fly line contains a 705bp deletion (Chr X: 4617319-4618023). All genome annotations were obtained from the *Drosophila* dm6 genome.

**Table 2.2: Primer sequences for genome editing of endogenous 5'UTR regions for Dpp target genes *hnt* and *ush*.**

| Gene | | Sense primer 5'-3' | Antisene primer 5'-3' | Product size (bps) |
|---|---|---|---|---|
| *ush* | HA1 | gctagcacatatgcaggtaccgt gcatagccacgacgttagg | cagttggggcactacggtacccgg ggacgagacgagacctctta | 1092 |
| *ush* | HA2 | acgaagttatcactagtgg aagtgacaacataattgcc | ggagatctttactagttccaagccttca ctccactc | 1152 |
| *ush* | gRNA1 | cttcgtctcgtctcgtccccgctc | aaacgagcggggacgagacgagac | 20 |
| *ush* | gRNA2 | cttcgattatgttgtcacttcccgt | aaacacgggaagtgacaacataatc | 20 |
| *hnt* | HA1 | cgctaccgcgggctagcgaag ggttgctggtcacc | cctgcatatgtgctagccattgggtgcgtgt gtgtg | 1028 |
| *hnt* | HA2 | acgaagttatcactagtcaactg ttgaacacaatttcac | tggagatctttactagtcacacatgcataca tccagtc | 1096 |
| *hnt* | gRNA1 | cttcgcgcaaataggattacacat | aaacatgtgtaatcctatttgcgc | 20 |
| *hnt* | gRNA2 | cttcgattgtgttcaacagttgcga | aaactcgcaactgttgaacacaatc | 20 |

### φC31 integrase-mediated site-specific transgenesis

Next, the attB-attP system was used for site-specific reintegration, after integration of the P-element attP into the genome. In this case wildtype promoter sequences were reintegrated together with the MS2-loop cassettes. The reintegration fragments were inserted into the RIV$^{cherry}$ plasmid (gift from the Vincent lab; DGRC, 1331) and summarised in Table 2.2. Wildtype sequences of promoter and 5'UTR regions, previously removed in the CRISPR process, were inserted into the multicloning site of RIV$^{cherry}$ using NotI to reconstitute the wildtype loci. Additionally, the MS2-loop cassettes were extracted from the pCR-24xMS2L (Addgene, PL 31865) plasmid by BamHI and BglII digestion and inserted into the BglII digested 5'UTR sequences (Figure 2.1D). The RIV$^{cherry}$ plasmid was co-injected with a φC31 integrase plasmid (Cambridge injection service) into the balanced CRISPR fly lines by

the Cambridge University injection service (Figure 2.1E). All surviving flies were outcrossed to wildtype flies and the offspring was screened for red eyes, generated by the *eyeless* promoter driven membrane cherry marker. Successful transformants were balanced and the marker region was removed Figure2.1E by crossing to a cre-recombinase expressing fly line (BL1501, Table 2.1).

**Promoter swaps**

Using the previously generated ΔP*ush* and ΔP*hnt* fly lines, promoter swap lines were generated by Priyanka Upadhyai. Accordingly, core promoter sequences from *ush* and *hnt* were amplified and inserted into the RIV$^{cherry}$ plasmid for site-directed recombination into ΔPromoter flies (see Figure 2.2A). Core promoter sequences and annotations were determined based on peaks from a Global Run-On Sequencing (GRO-Seq) study (Saunders et al. 2013). The core promoter region was specified as 200bps around the peak region. The *ush* isoform C promoter region was defined as Chr 2L: 523,636 - 523,835 and the hnt promoter region as Chr X: 4,617,464 - 4,617,663 (dm6 *Drosophila* genome). Primers for region amplification can be found in Table A.3. Primers to amplify upstream and downstream sequences contained NotI restriction sites which were used for insertion into RIV$^{cherry}$. After successful co-injection with the φC31 plasmid by the Cambridge University injection service, flies were outcrossed to wildtype flies and screened for correct insertion by red eye colour. Successful transformants were homozygosed and crossed to a cre-recombinase expressing fly line (BL1501, Table 2.1). Finally, the promoter swap lines carry the core promoter sequence of a different gene together with all endogenous up- and downstream 5'UTR regions (see Figure 2.2B).

#### 2.2.2.2 Control lines without MS2 for 5'UTR insertions

To control for the effect of MS2-loops on gene expression, fly lines lacking the MS2-loop cassettes were generated during this study. Again, the previously described ΔPromoter fly lines were used as a starting point. Primers (forward: 5' aaggAGATCTgctagcgtttaaacac 3', reverse: 5' ggaaAGATCTgcggccgccg 3') were used to PCR the backbone of existing RIV$^{cherry}$ constructs (*ushPush, ushPhnt*). Primers contained BglII restriction sites (capital letters in primer sequences), which were then used to digest the PCR products and re-ligate them using T4 DNA ligase (NEB, M0202S). The resulting constructs were injected into ΔP*ush* fly lines by the Cambridge injection service together with a ψC31 integrase plasmid. All surviving flies were outcrossed to wildtype flies, the offspring was screened for the membrane

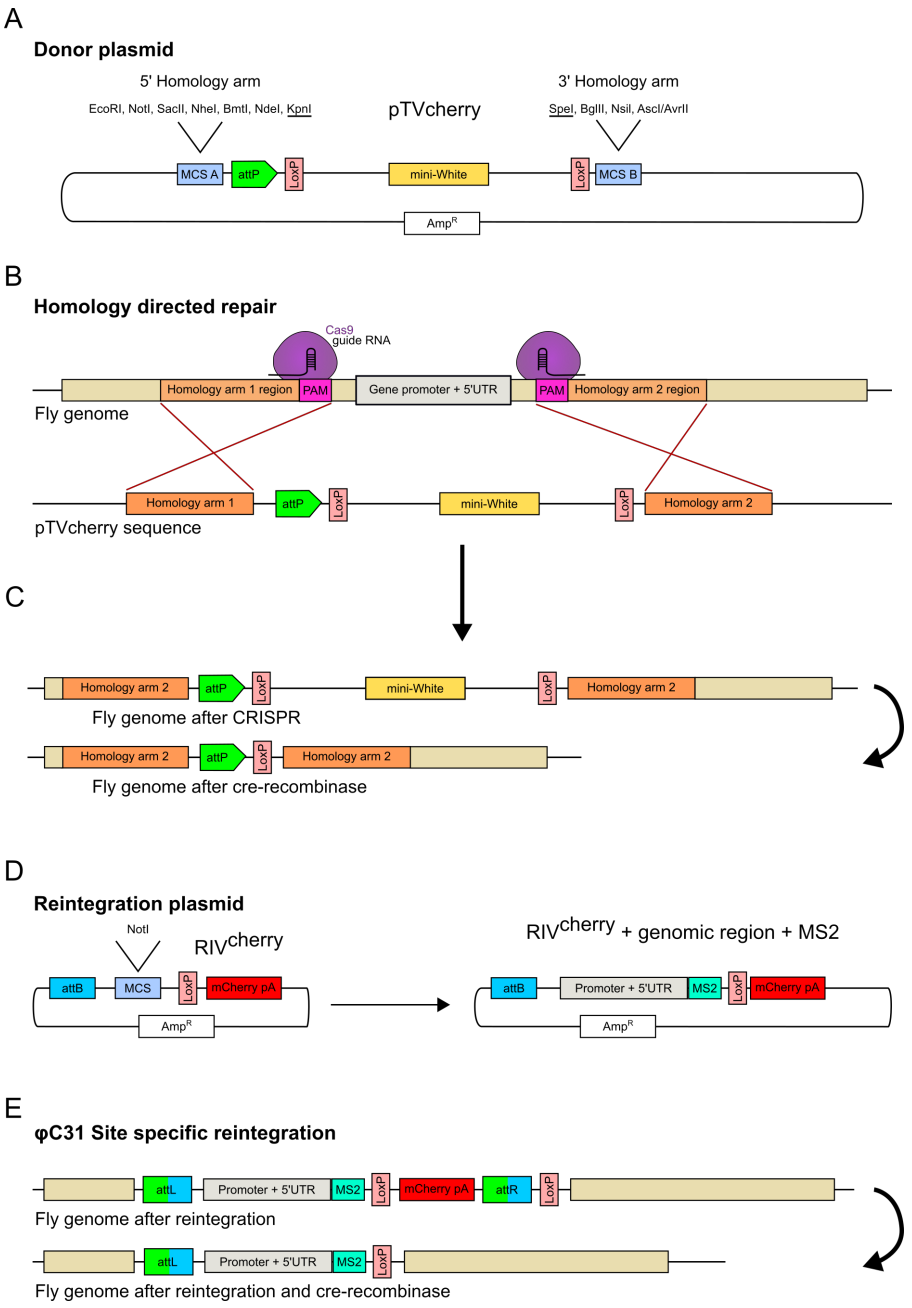



**Figure 2.1: Schematic overview of the genome modification strategy used to generate live imaging fly lines.**(A) The plasmid pTV$^{cherry}$ was used as a donor plasmid for CRISPR Cas-9 genome editing. Homology arm sequences were inserted into multi cloning sites (MCS). (B) After two double strand breaks, the homology containing donor plasmid was used for homology-directed repair, creating a genome deletion. (C) Successful CRISPR flies were identified by the mini-white marker, which was removed based on LoxP sites by cre-recombinase. (D) Genomic sequences, previously removed by CRISPR, were inserted into RIV$^{cherry}$ using a NotI restriction site. Additionally, 24x MS2-loops were inserted into the 5'UTR sequence. (E) φC31 integrase-mediated site-specific transgenesis allowed for the recombination of attB and attP sites and the insertion of target DNA. The marker region was removed by cre-recombinase. A scar of the attL and one loxP site remains in the genome.

cherry marker and successful transformants were homozygosed. The marker region was removed by crossing the flies to a cre-recombinase expressing fly line (BL1092, Table 2.1).

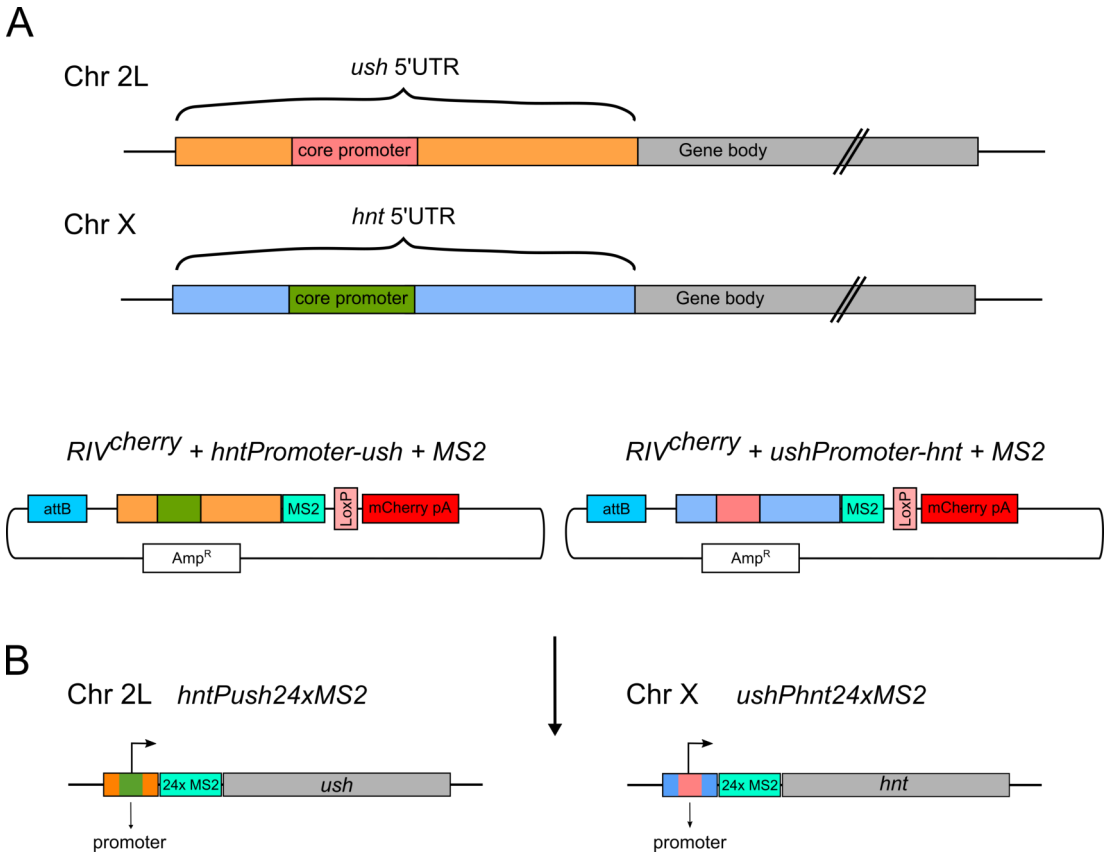



**Figure 2.2: Schematic of promoter swap line creation using previously produced ΔPromoter flies.** (A) Genomic regions of the Dpp target genes *hnt* and *ush* were amplified to be reinserted into CRISPR deletion lines. The promoter upstream and downstream regions were kept whereas the core promoter sequences (200bps) were swapped between the two genes. All regions were merged in the RIV$^{cherry}$ reintegration plasmid and then injected into the ΔPromoter fly lines. (B) The resulting fly lines now transcribe *ush* under the control of the *hnt* core promoter and *hnt* under the control of the core *ush* promoter.

## 2.3 Immunostain and (Fluorescent) In Situ Hybridisation

### 2.3.1 Immunostain

Embryos were washed for 10 min in 50% methanol/50% PBS + 0.1% Tween-80 (PBT), followed by five 6 min washes in PBT. Embryos were then blocked 30 min in 10% Bovine Serum Albumin (BSA; Sigma, A2058) in PBT four times, rinsed four times in PBT, followed by incubation in primary mouse anti Hnt antibody (1:40; DSHB, 1G9) in 1% BSA overnight at 4°C. Samples were then washed four times 30 min in 1% BSA, incubated for 2h in alkaline phosphatase conjugated anti-mouse secondary antibody (1:500; Promega, S372B) in 0.1% BSA. Following incubation, embryos were washed three times 15 min in PBT, followed by two 5 min washes in staining buffer (100mM NaCl, 50mM MgCl2, 100mM Tris pH 9.5, 20% Tween-80 in $dH_2O$). Embryos were transferred into Pyrex spot plates in 400µl staining solution (0.675mg/mL nitro blue tetrazolium (NBT; in 70% dimethylformamide; Sigma, 11585029001), 0.35mg/mL 5-bromo-4-chloro-3-indolyl-phosphate (BCIP; in 100% dimethylformamide; Sigma, B6149) in staining buffer). After sufficient staining was achieved, embryos were rinsed in PBT and washed in 100% ethanol until excess staining was removed. Then, embryos were rinsed in xylenes and mounted in Permount (Thermo Fisher Scientific, SP15100).

### 2.3.2 Whole mount In Situ Hybridisation

Fixed embryos were rehydrated for 1h in PBT, followed by a 5 min wash in 50% PBT/ 50% hybridisation buffer (50% formamide, 0.1 mg/ml ssDNA, 0.05 mg/ml Heparin, 0.1% Tween-80 in 5x Saline-Sodium Citrate buffer (SSC)) and a 2 min wash in 100% hybridisation buffer. Embryos were then pre-hybridised in hybridisation buffer for 1h at 55°C and then incubated with digoxygenin-UTP labelled antisense probes targeting *ush*, *hnt*, *tup* or *yellow*, generated from cDNA (1:500, Ashe Lab), in hybridisation buffer at 55°C for 18h. After hybridisation, embryos were rinsed once, followed by three 30 min washes in pre-heated hybridisation buffer, two 15 min washes in 50% hybridisation buffer/50% PBT and five 10 min washes in PBT. Afterwards, embryos were incubated in pre-blocked alkaline phosphatase-conjugated anti-digoxigenin antibody (1:250; Sigma, 11093274910) in PBT overnight at 4°C. Embryos were subsequently washed four times for 15 min in PBT and

two times for 5 min in staining buffer before they were stained with NBT and BCIP and mounted in Permount as described in Section 2.3.1.

### 2.3.3 Nascent RNA Fluorescent In Situ Hybridisation

#### Single fly genomic DNA preparation

Genomic DNA from single flies was prepared by grinding a single fly in 200µl Buffer A (100mM TrisHCL, 100mM EDTA, 100mM NaCl, 0.5% (w/v) SDS) with a pipette tip followed by a 30 min incubation at 65°C. After incubation, 800µl LiCl/KAc solution (1:2.5 5M KAc: 6M LiCl) were added and the samples were incubated on ice for 10 min before they were centrifuged for 15 min at 13000g. Next, 600µl 2-propanol were added to 1ml of supernatant in order to precipitate the DNA. The solution was mixed and centrifuged at 13000g for 15 min. The supernatant was removed and the DNA was resuspended in 150µl Tris/EDTA buffer and stored at 4°C.

#### RNA probe preparation

RNA probes for FISH were generated by *in vitro* transcription using primers containing either T7 or T3 promoter sequences. Table 2.3 lists primer sequences for all RNA probes. Genomic DNA from single fly preparations was used as a DNA template. *In vitro* transcription reactions were set up in 10µl volumes. Reactions contained 1µl 10x transcription buffer (Sigma; 10999644001), 1µl 10x digU-NTP or biotin-NTP mix (10mM ATP, 10mM GTP, 10mM CTP, 6mM UTP, 4mM digoxigenin-11-UTP (Sigma, 11277073910) or biotin-16-UTP (Sigma, 11685597910)), 1µl T7 (Sigma, RPOLT7-RO) or T3 RNA polymerase (Sigma, RPOLT3-RO), 1µl RNase inhibitor (Sigma, RNAINH-RO) and 2-5µl purified DNA template. Reactions were incubated at 37°C for 2h. In order to verify successful transcription, 1µl of each reaction was run on a 1.5% agarose gel. Partial hydrolysis of probes was achieved by the addition of 15µl $H_2O$ and 25µl 2x carbonate buffer (120mM Na2CO3, 80mM $NaHCO_3$, pH 10.2) to the remaining reaction mixtures, followed by an incubation at 65°C for 10 min. Hydrolysis was stopped by the addition of 50µl 2x stop buffer (200mM sodium acetate, pH 6.0) and probes were precipitated using 10µl 4M LiCl, 5µl 10mg/ml tRNA and 300µl 100% ethanol for 15 min at -20°C. RNA was pelleted by centrifugation for 20 min at 4°C and 13000g. RNA probes were washed with 1ml of 70% ethanol for 5 min, followed by a 5 min centrifugation at 13000g. Pellets were resuspended in 150µl hybridisation buffer and stored at -20°C. Before use, the

RNA probes were shortly heated at 85°C to denature any secondary structures and probe complexes.

**Table 2.3: Nascent RNA in situ hybridisation probe sequences.** Gene specific complementary DNA sequences in capitals, T3 promoter sequence in purple and T7 promoter sequence in blue.

| Probe name | Probe sequence 5'-3' |
|---|---|
| *Race* sense | attaaccctcactaaagggaAGTAGAAACATTATTGCAAT |
| *Race* antisense | gaattaatacgactcactatagggaAAGCAAAAAATTACGTTTTT |
| *hnt* sense | attaaccctcactaaagggaATTCCCAAAACCCCTCCCTT |
| *hnt* antisense | gaattaatacgactcactatagggaCCAGTCTTCGATTGTCGCG |
| *ush* sense | attaaccctcactaaagggaGTGAGAATTATTCATAC |
| *ush* antisense | gaattaatacgactcactatagggaATTAAACTACAGT ACCCA |
| *tup* sense | attaaccctcactaaagggaTAATTACAAACAAATTAA |
| *tup* antisense | gaattaatacgactcactatagggaATTAAATATTTACC AGC |
| *sog* sense | attaaccctcactaaagggaAATTTTATTTTCAATCTATT |
| *sog* antisense | gaattaatacgactcactatagggaAAAAAACGAGAAAATA |
| *brk* sense | attaaccctcactaaagggaGAACAGTTGAACGGATCGGGAGCTT |
| *brk* antisense | gaattaatacgactcactatagggaCGATTCTCCAAATAGCCATGCAG |
| *sna* sense | attaaccctcactaaagggaACACCGGAAAGGAACTCCAG |
| *sna* antisense | gaattaatacgactcactatagggaTCTGTTTGTTTGGTCTTCGCC |
| *yellow* sense | attaaccctcactaaagggaCCAATCACAAACAATAACTTATGGC |
| *yellow* antisense | gaattaatacgactcactatagggaTATCGGAAAATGTTGGCAATTCTGT |

### Sample preparation

Fixed embryos that were stored in methanol were rinsed once in 50% methanol/50% ethanol, followed by one rinse in 100% ethanol and a 5 min wash in 100% ethanol. All washes unless stated otherwise were performed with agitation. The ethanol was removed except for 100µl and 900µl of xylenes were added. Embryos were incubated for 1h, followed by two rinses in 100% ethanol and a 5 min wash in 100% ethanol. Afterwards, embryos were rinsed twice and washed once in 100% methanol, followed by one wash in 50% methanol/ 50% PBT with 5% formaldehyde for 5 min. Embryos were fixed in PBT with 5% formaldehyde for 25 min. After fixation, embryos were washed four times for 5 min in PBT and once in 50% PBT/ 50% hybridisation buffer (50% formamide, 0.1 mg/ml ssDNA, 0.05 mg/ml Heparin, 0.1% Tween-80 in 5x Saline-Sodium Citrate buffer (SSC)) for 10 min. Next, embryos were pre-hybridised in hybridisation buffer at 55°C for 1h. During this hour the hybridisation buffer was exchanged twice. RNA probes, targeting intronic sequences (Table 2.3), were diluted in hybridisation buffer (1:100), and

embryos were incubated with the probe solution at 55°C overnight. After hybridisation, embryos were washed 1x 5 min and 2x 30 min in pre-heated hybridisation buffer at 55°C, followed by a 10 min wash in 50% hybridisation buffer/50% PBT at room temperature. Embryos were washed four times for 5min in PBT, blocked in PBT with 1X blocking reagent (Sigma, 11921673001) for 30 min and incubated with primary antibody (sheep anti-digoxigenin (Sigma, 11333089001) or mouse anti-biotin (Sigma, 1297597)) in PBT with 1X blocking reagent at 4°C overnight. Then, embryos were washed 4x 5min in PBT, blocked in PBT with 1X blocking reagent for 30 min and incubated with fluorescently conjugated secondary antibody (donkey anti-sheep Alexa 555 (1:500; Thermos Fisher Scientific A-21436), donkey anti-mouse Alexa 647 (1:500; Sigma A-31571)) in PBT with 1X blocking reagent for 90 min at room temperature in the dark. After incubation, the embryos were rinsed once and washed three times for 15 min in PBT (one PBT wash included DAPI (1:500; NEB, 4083)), mounted in ProLong Diamond Antifade Mountant (Thermo Fisher, P36965) or ProLong Gold Antifade Mountant (Thermo Fisher, P36930) and dried. Sense probes were used as controls and antisense probes were used to generate images for result figures.

### 2.3.4 Single molecule (inexpensive) FISH

#### smiFISH primary probe preparation

Custom gene specific oligonucleotides were designed using the online smFISH design tool (https://www.biosearchtech.com/support/tools/design-software/stellaris-probe-designer) and ordered from Sigma at 100μM concentration according to Tsanov et al. 2016. Their sequences can be found in Tables A.9 A.10 A.11 A.12. Each oligonucleotide carries a respective 5' FLAP primer sequence which is complementary to one FLAP sequence (Figure 3.1 A). For probe hybridisation, an equimolar mixture of all individual oligonucleotides was prepared at 100μM.

#### FLAP preparation

A fluorescently labelled FLAP acts as a secondary probe that is pre-hybridised to the primary probe before use and consists of an oligonucleotide with two fluorophores attached per molecule, one to its 5' and one to its 3' end (Tsanov et al. 2016). FLAPs with two different sequences were used in this study (kind gift from the Ronshaugen lab, Eurofins). FLAP oligonucleotides are shipped lyophilised and reconstituted to a final concentration of 100μM. All FLAP specifications can be found in Table 2.4

and an overview of the method in Figure 3.1 A.

First, FLAPs were hybridised to the gene specific probes in a PCR reaction (Tsanov et al. 2016). The PCR reaction consisted of the probe set at 200pM, FLAP at 50 pM, 1x Buffer 3 (NEB, B7003S) in $H_2O$. This reaction mix creates a 4μM probe/FLAP duplex. The thermocycler programme used was as follows: 85°C 3 min, 65 °C 3 min, 25 °C 5 min, Lid: 99°C

**Table 2.4: FLAP sequences, complementary FLAP primer sequences (attached to the 5' end of gene specific primary probes) and the fluorescent label for smiFISH experiments.**

| FLAP name | FLAP sequence (labelled) 5'-3' | Complementary primer seq 5'-3' | Quazar Label |
|---|---|---|---|
| X | CACTGAGTCCAGCTCGAAACTT AGGAGG | CCTCCTAAGTTTCGAGCTGGA CTCAGTG | 546 |
| X | CACTGAGTCCAGCTCGAAACTT AGGAGG | CCTCCTAAGTTTCGAGCTGGA CTCAGTG | 647 |
| Z | CTTATAGGGCATGGATGCTAGAA GCTGG | CCAGCTTCTAGCATCCATGCCC TATAAG | 647 |

### smFISH probe design

Custom, gene specific Stellaris® RNA FISH probe sets were designed using the online stellaris design tool (https://www.biosearchtech.com/support/tools/design-software/stellaris-probe-designer) and purchased through 2BScientific. All probes were designed with the highest stringency to eliminate possible cross-reaction. Probe sequences can be found in Tables A.4, A.5, A.6, A.7, A.8 and an overview of the method in Figure 3.1 B. Primary smFISH probes are directly conjugated to a single fluorophore.

### Sample preparation

Fixed embryos, staged to be 2-4h old, were washed 5 min in 50% methanol/50% PBT, followed by four 10 min washes in PBT, a 10 min wash in 50% PBT/5% wash buffer (10% formamide in 2X SSC; 300mM NaCl and 30mM trisodium citrate adjusted to pH 7) and two 5 min washes in 100% wash buffer. Next, embryos were rinsed once and incubated 2h at 37°C in smFISH hybridisation buffer (2.5mM dextran sulphate, 10% formamide in 2X SSC). During that time the hybridisation buffer

was exchanged twice. Probes were diluted in hybridisation buffer to a final concentratiin of 1.25μM for smFISH Stellaris probes, and 4μM probe/FLAP duplex for smiFISH probes. Embryos were incubated in probe solution for 14h at 37°C, washed 30 min in pre-warmed hybridisation buffer at 37°C, followed by three 15 min washes in pre-warmed wash buffer at 37°C. At room temperature, embryos were washed 15 min in wash buffer and three times 15 min in PBT in the dark. One of the PBT washes included DAPI (1:500) and in some cases Wheat Germ Agglutinin (WGA, 1:1000; Thermo Fisher Scientific, W11261). Embryos were then mounted in ProLong Diamond or Gold Antifade Mountant.

When immunofluorescence was performed, embryos were washed four times for 10 min in PBT after overnight hybridisation and blocked for 1h in PBT-blocking solution (20% (v/v) blocking reagent in PBT; PBT-B). Preabsorbed primary mouse anti spectrin antibody (1:50; DSHB, 3A9) was diluted in PBT-B and incubated with the embryos at 4°C overnight. This was followed by four 10 min washes in PBT and 1h blocking in PBT-B. Embryos were then incubated with secondary anti mouse antibody conjugated to Alexa 488 (1:500; Thermo Fisher Scientific, A-21422) for 1h and washed four times for 10 min in PBT. One PBT wash included DAPI at 1:500. Embryos were mounted in ProLong Diamond or Gold Antifade Mountant.

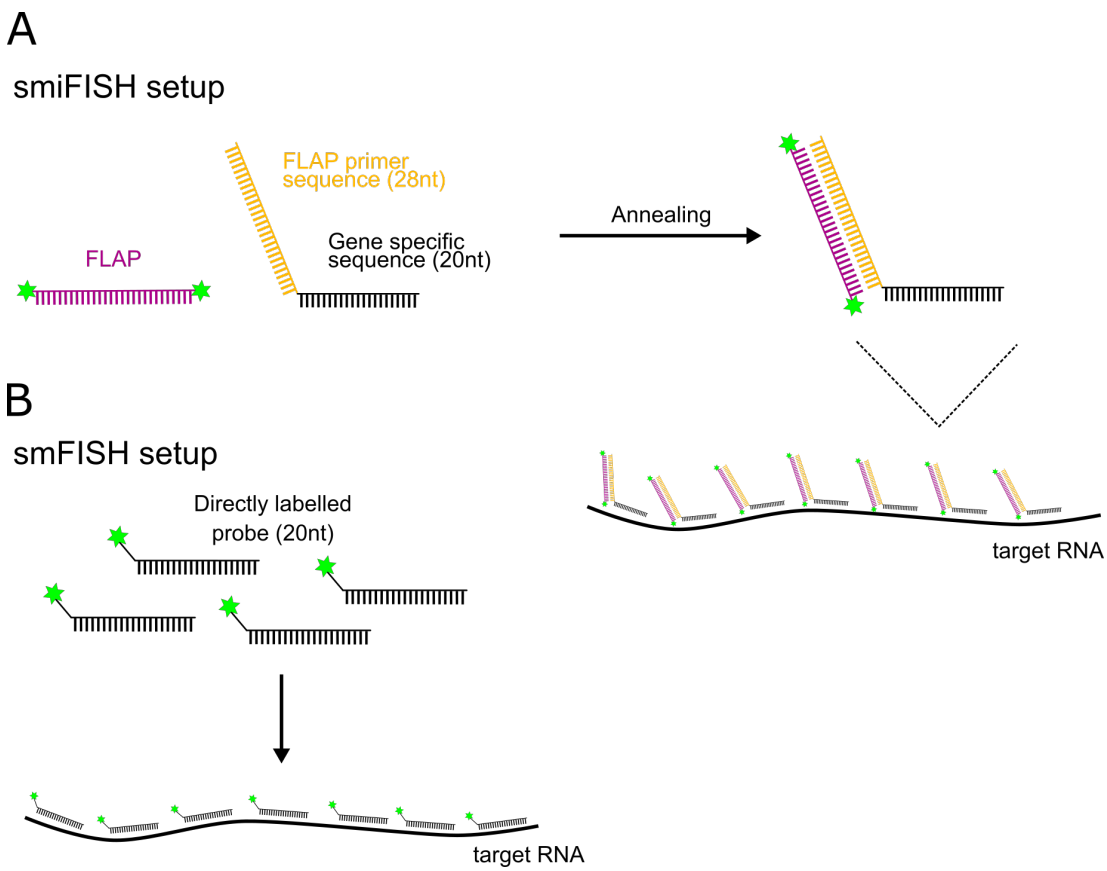



**Figure 2.3: Overview of smFISH and smiFISH labelling techniques.** (A) smiFISH primary gene probes are fused to a FLAP primary sequence of 28nt. In a hybridisation step the FLAP primer sequence is annealed to FLAP oligonucleotides, which carry single fluorophores on their 5' and 3' end. Afterwards the probes are used in in situ hybridisation experiments to anneal to sequence specific regions of target RNA molecules. (B) smFISH probes are directly labelled with single fluorophores and hybridise to complimentary target RNA sequences.

## 2.4 Live imaging

### Sample preparation

Crosses were set up at least two days prior to imaging in collection cages and were not kept for longer than one week. Imaging crosses were set up as follows:

1. First generation crosses to test fluorescence background levels in Chapter 4: Females homozygous for His-RFP;MCP-GFP were crossed to males homozygous for MS2 tagged imaging locus.

2. Second generation crosses used for *ush* and *hnt* imaging: Females homozygous for His-RFP;MCP-GFP were crossed to wt males. Female offspring were crossed to males homozygous for MS2 tagged imaging locus.

3. Homozygous *ush* imaging: Females homozygous for His-RFP;MCP-GFP were crossed to males homozygous for MS2 tagged imaging locus. Female offspring were crossed to to males homozygous for MS2 tagged imaging locus.

4. *ush* transcription in *st2-dpp* embryos: His-RFP;MCP-GFP expressing females were crossed to *st2-dpp* expressing males and resulting female offspring was crossed to homozygous *ush-MS2* males.

For embryo collection, the apple juice agar plate was changed and the cage was returned to the 25˚C incubator for egg lay. After 1h of egg lay, embryos were dechorinated for 2 minutes in 2% bleach, washed with water and then lined up on a fresh apple juice agar plate. Next, embryos were oriented with their dorsal side up and a coverslip (Nr. 1, 18x 18 mm; Deltalab, D101818), coated with a thin layer of heptane glue, was positioned on top.

### Mounting

A drop of halocarbon oil mix (4:1, halocarbon oil 700: halocarbon oil 27; Sigma H8898 and H8773) was placed in the middle of a Lumox imaging dish (Sarstedt, 94.6077.305) and two coverslips (Nr. 0, 18x 18mm; Scientific Laboratory Supplies, PK200) were placed on either side of the oil drop, creating a bridge. The coverslip with the embryos glued to it was then inverted into the oil, sandwiching the embryos between the imaging dish membrane and the coverslip. The coverslips were secured in place with a drop of nail polish.

## 2.5 General microscopy

### 2.5.1 Live imaging microscopy

**Imaging of embryos heterozygous for MS2 imaging loci**

Embryos were imaged on a Leica TCS SP8 AOBS inverted confocal microscope with a resonant scan head, using a 40x/ 1.3 HC PL apochromatic oil objective. Images were obtained with the following confocal settings, pinhole 1.3 airy units, scan speed 8000Hz bidirectional, format 1024 x 700 pixels at 8 bit and at 1.25 x optical zoom. Images were collected using the white laser with 488nm (8%) and 574nm (2%) at 8x line averaging and detected with hybrid detectors. Three-dimensional optical sections were acquired at 1μm distance, a final depth of 55μm and a final temporal resolution of 20 seconds per time frame. Images were processed with the Leica lightning deconvolution software. The mounting medium refractive index was estimated to be 1.41. Maximum intensity projections of 3D stacks are shown in the result sections. Embryos were imaged for 70-90 min and included the mitotic division of nc14 and the onset of gastrulation.

**Imaging of embryos homozygous for the *ush*-MS2 imaging locus**

Embryos were imaged on the microscope setup described above with the following alterations. Images were obtained at a 2.15 x confocal zoom and optical sections were obtained at 1μm with a final depth of 40μm and a final temporal resolution of 15.2 sec.

### 2.5.2 Nascent Fluorescent In Situ Hybridization images

Images were acquired on two different systems. Images were taken on an Olympus IX83 inverted microscope using Lumencor LED excitation, a 60x/ 1.42 Plan Apo or 40x/ 1.0 UplanApo objective and the Sedat filter set Chroma 89000. The images were collected using a R6 Qimaging CCD camera with a z optical spacing of 300nm. Alternatively, images were acquired using a Leica TCS SP8 AOBS inverted microscope using a 40x/ 1.3 HC Pl Apo CS2 or 63x/ 1.4 Plan APO objective with 2x line averaging. The confocal settings were as follows, pinhole 1 airy unit, scan speed 400Hz and format 2048 x 2048 pixels. Images were collected with either Photon Multiplying Tube Detectors or Hybrid Detectors and illuminated using a white laser.

The following detection mirror settings were used: Photon Multiplying Tube Detector at 405nm (4.66%); Hybrid Detectors: 490nm (10%, 0.3 to 6us gating), 548nm (26.1%, 0.3 to 6us gating) and 647nm (17%, 0.3 to 6us gating). All images were collected sequentially and optical stacks were acquired at 300nm spacing.
Raw images were then deconvolved using Huygens Pro software (SVI) and maximum intensity projections are shown in the result sections.

### 2.5.3 Single molecule Fluorescent in Situ Hybridisation

For quantitative purposes all images were acquired using the Leica SP8 microscope described in Section 2.5.2

### 2.5.4 Light microscopy

Images of embryos were acquired for immunostainings using a Leica upright microscope with a 10x objective and a Coolsnap EC camera.

## 2.6 Image processing, modelling and statistics

### 2.6.1 Image analysis of static images

Nuclei and RNA puncta were initially detected using the Imaris (Bitplane, Oxford Instruments, Concord MA) software 9.2. RNA puncta were then assigned to nuclei in a proximity based method using python scripts developed by Thomas Minchington.

#### Nuclei and spot identification in IMARIS

In detail, nuclei were identified and segmented using the Imaris "surface" function. Nascent transcription foci were identified using Imaris "spots" function and estimated to be 0.6μm in diameter with a z-axis point spread function of 1μm. Single mRNA puncta were identified with spot volumes of 0.3μm across and 0.6μm in the z direction.

**Spot assignment**

Customised Python scripts were used to analyse the data extracted from Imaris. In short, transcription foci were assigned to their closest nucleus in a three dimensional space along the ellipsoid axis of the nucleus. The middle of the expression domain was determined as follows: The embryo was divided into 10 equal regions in two dimensions. Centroids were calculated for each region within the expression domain and used to determine the long axis of the embryo. The six middle centroids were then used to calculate the line of best fit using the numpy least squares analysis (numpy.linalg.lstsq). All points for the equation of the line along the embryo axis were then plotted and the minimum distance to these points calculated for each nucleus within the expression domain. Next, the nuclear distance to the middle of the expression domain was computed in micrometer.

**Analysis of nuclear distance to dorsal midline**

The relative distance of a nucleus to the expression domain midline was calculated for analyses performed in Chapter 3. The relative distance was calculated by dividing the distance in μm by the median nuclear distance of the expression domain of the respective replicate. This scaling is done to control for differences that could be introduced by differences in age between replicates or differences in flattening of the embryo against the coverslip during mounting.

**Analysis of nascent mRNA number**

When the number of mRNAs contained in a nascent transcription site was calculated, fluorescent nascent transcription sites and single mRNAs were segmented using the Imaris "surface" function. The fluorescence determined for nascent transcription sites was divided by the mean fluorescence intensity of single cytoplasmic RNA molecules.

### 2.6.2 Live imaging analysis

Imaging movies were analysed using the Imaris software coupled with custom Python scripts.

### Nuclear tracking and spot identification for embryos with one MS2 locus

Nuclei were first smoothed and blurred using a wavelet filter (implemented by Egor Zindy) and then segmented using the Imaris "surface" function based on the His-RFP fluorescent channel. Nuclei were tracked through time using the inbuilt autoregressive motion with a maximum frame gap size of 5 and a maximum travel distance of 5μm. Active transcription sites were detected using the Imaris "spots" function. Transcription foci were estimated to be 1.8μm across with a z-axis point spread function estimation of 7.8μm. To determine the background fluorescence of the data set, a set of "spots" were generated for background correction. Here, four spots were inserted every third time frame, avoiding nascent transcription sites. The background correction spots had the identical volume as the transcription site spots.

### Nuclear tracking and spot identification for embryos with two MS2 loci

Transcription site tracking was performed manually for the time-lapse movie imaging an embryo homozygous for the *MS2-ush* locus. This was done to ensure that both alleles were tracked individually and no crossover occurred. Only nuclei with clearly distinct homologous alleles were chosen for this analysis to prevent the loss of alleles during tracking when they come into too close proximity.

### Spot assignment to nuclei and background correction

After tracking, intensity measurements and position files were exported and analysed using a series of custom Python scripts, written by Thomas Minchington. In short, nascent transcription spots were assigned to their closest nucleus in a three dimensional space using the nuclear ellipsoid axis. Next, background fluorescence was removed by defining, then subtracting and dividing by the background fluorescence for any given time frame. For that, a linear regression line was fitted based on fluorescent values from background spots. The fluorescent value in the transcription site channel was set to "0" when no spot was assigned to a nucleus and when the spot fluorescence was negative after background subtraction. A subset of expression domain nuclei was computed, based on the definition that nuclei had to be present in every time frame during gene transcription and were not allowed to exit and re-enter the field of view. Finally, a midline through the expression domain was computed as described for static images.

**Developmental time adjustments**

All datasets were adjusted in time to account for temperature differences during imaging as those can alter the speed of development. It has been shown that the early *Drosophila* development scales uniformly across temperature (Kuntz & Eisen 2014). Therefore nc14 was defined as the time between telophase of cleavage cycle 14 and the beginning of cephalic furrow formation. This time frame was defined to be 50 minutes long.

**Datasets used for figures in results chapters**

For each genotype of *ushPush, hntPush, hntPhnt* and *ushPhnt*, one full embryo (full field of view acquired on microscope) was analysed. Additionally, two more biological replicates of *ushPush* and *hntPush* were analysed that contained only the anterior subset of the imaging region, which was determined to be 77 μm wide along the AP axis. For flies with the *hntPhnt* and *ushPhnt* genotype, one more biological replicate was analysed for each containing the centre domain, which was determines to be 125 μm along the AP axis. Nuclei from one homozygous *ushPush* embryo were analysed and tracked manually. The *ushPush* allele was imaged in one embryo heterozygous for ectopic *st2-dpp*. A 41 μm wide (AP axis) strip of cells was analysed.

### 2.6.3 K-Means clustering

Clustering of fluorescence traces obtained from live imaging experiments was performed using the K-Means clustering algorithm and computed in Python. The number of clusters was chosen based on elbow graphs (not shown) and computational analysis was performed by Jonathan Bowles.

### 2.6.4 Memory-adjusted Hidden Markov Model

All computational modelling was performed by Jonathan Bowles and results are shown in result chapter figures. The adapted Hidden Markov Model (HMM) was developed by Lammers et al. (2019) and adjusted to work on fluorescent traces from long genes by Jonathan Bowles.

### 2.6.5 Statistics

Statistical analysis and comparisons were performed using two-tailed Student's t-tests, paired t-tests, Gaussian distribution curves, Lowess smoothing, Spearman or Pearson correlation coefficients, Mann-Whitney U tests, Kruskal-Wallis tests and one way ANOVA tests were performed (where indicated additional multiple comparisons tests were added) in GraphPad Prism or R.

# 3 | Morphogen gradient interpretation by single cells

## 3.1 Introduction

Cells in the dorsal ectoderm of the *Drosophila* embryo are subject to different levels of BMP signalling, which helps them to determine their position within the tissue and define their cell-specific transcriptome. However, it is poorly understood how cells interpret different morphogen concentrations over time and decode them at a transcriptional level. Do cells respond in a digital fashion, where transcription is initiated through a binary threshold switch, and the cell's output will be identical to its neighbours, which experience the same threshold? Or do cells interpret signalling levels in an analogue or graded fashion, where transcription is modulated and mirrors the concentration gradient the cell is experiencing? These questions can now be addressed, and new insights into transcriptional dynamics can be gained by studying single cells using quantitative fluorescent in situ techniques.

This chapter will explore the transcription dynamics that target genes experience in response to morphogen signalling and how their position within the gradient shapes it. The analysis will focus on high threshold response genes *Race* and *hnt* as well as intermediate targets *ush* and *tup*. Their spatial expression patterns have been characterised using traditional ISH techniques (Ashe et al. 2000), but very little is known about the heterogeneity in transcription within the tissue. Furthermore, their mRNA copy number and temporal expression dynamics have not been described. Transcription dynamics will be visualised using quantitative, single-molecule resolution techniques in fixed embryos.

## 3.2 Single cell dynamics visualised by RNA-FISH

### 3.2.1 BMP signalling gradient readout by target genes

Digoxigenin- or Biotin labelled FISH probes targeting intronic RNA sequences were used to visualise nascent transcription of BMP target genes and investigate their overall expression patterns. After staining fixed *Drosophila* embryos at stage 5, nascent transcription sites can be detected as bright fluorescent foci (Figure 3.1 A). All embryos show classic BMP target gene expression patterns as they were previously described using colourimetric ISH (Ashe et al. 2000). High threshold response genes *Race* and *hnt* show narrow transcription patterns. *Race* transcription is restricted to the anterior and middle of the embryo but absent from the posterior, whereas *hnt* transcription is absent from the anterior. The area in which the two genes overlap in their expression forms the presumptive amnioserosa. The expression patterns of *ush* and *tup* are wider, consistent with their activation by lower BMP signalling levels and encompass the presumptive amnioserosa as well as parts of the dorsal epidermis that will ultimately specify the leading edge during dorsal closure (Figure 3.1 A).

Typically, two transcription sites are visible, representing each allele of the gene (Figure 3.1 Bi). The two gene loci are often found separated by a considerable distance (Figure 3.1 A, Bi). Homologue pairing of alleles in *Drosophila* embryos is initiated during late nc14 (Hiraoka et al. 1993). During nc14, S-phase and therefore DNA replication starts shortly after cleavage cycle 14 and continues for the first 50 min (Shermoen et al. 2010). Following replication, sister chromatids can be visualised, resulting in three (Figure 3.1 Bii) or sometimes four fluorescent dots per nucleus (Figure 3.1 Biii). In most cases, sister chromatids cannot be resolved within one bright fluorescent spot, as they are kept in very close proximity to each other by cohesin complexes (Peters et al. 2008). If sister chromatids are observed as two fluorescence spots, they most likely differ in their lateral position as confocal imaging restricts the resolution in the axial-direction (Little et al. 2013). For all further image analysis, sister chromatids are scored as one active transcription site as illustrated in Figure 3.1 B, lower panels.

Analysis of RNA-FISH images in more detail identifies a proportion of nuclei where transcription of target genes occurred from a single allele. Single nascent transcription sites are observed in a subset of nuclei in the four target genes anal-

ysed, suggesting a general behaviour in response to BMP signalling (Figure 3.2 Ai). This study will refer to this mode of transcription as monoallelic. This term is used here to describe the activity of a single transcriptionally active allele but does not imply that the other homologous allele is stably silenced as is the case during imprinting for example (Reik & Walter 2001).

In order to visualise nuclei that show monoallelic transcription, images were generated with nuclei false coloured depending on their transcriptional states (Figure 3.2 Aii). When quantified, at least one-quarter of nuclei, that are part of the expression domain, display monoallelic transcription for all BMP target genes (Figure 3.2 B). *Race* shows the highest and *ush/tup* the lowest proportion of monoallelic nuclei. *Race* was also identified as the only gene that shows a higher proportion of nuclei with monoallelic than biallelic transcription. The overall percentage of nuclei with monoallelic expression is observed to decrease inversely with the width of the expression domain (Figure 3.2 B). Target genes *Race* and *hnt*, which are activated by peak levels of BMP signalling, show a higher percentage of monoallelic transcription but the absolute number of monoallelic nuclei is similar to *ush*. *hnt* is shown to have a slightly lower absolute number of monoallelic nuclei (Figure 3.2 C).

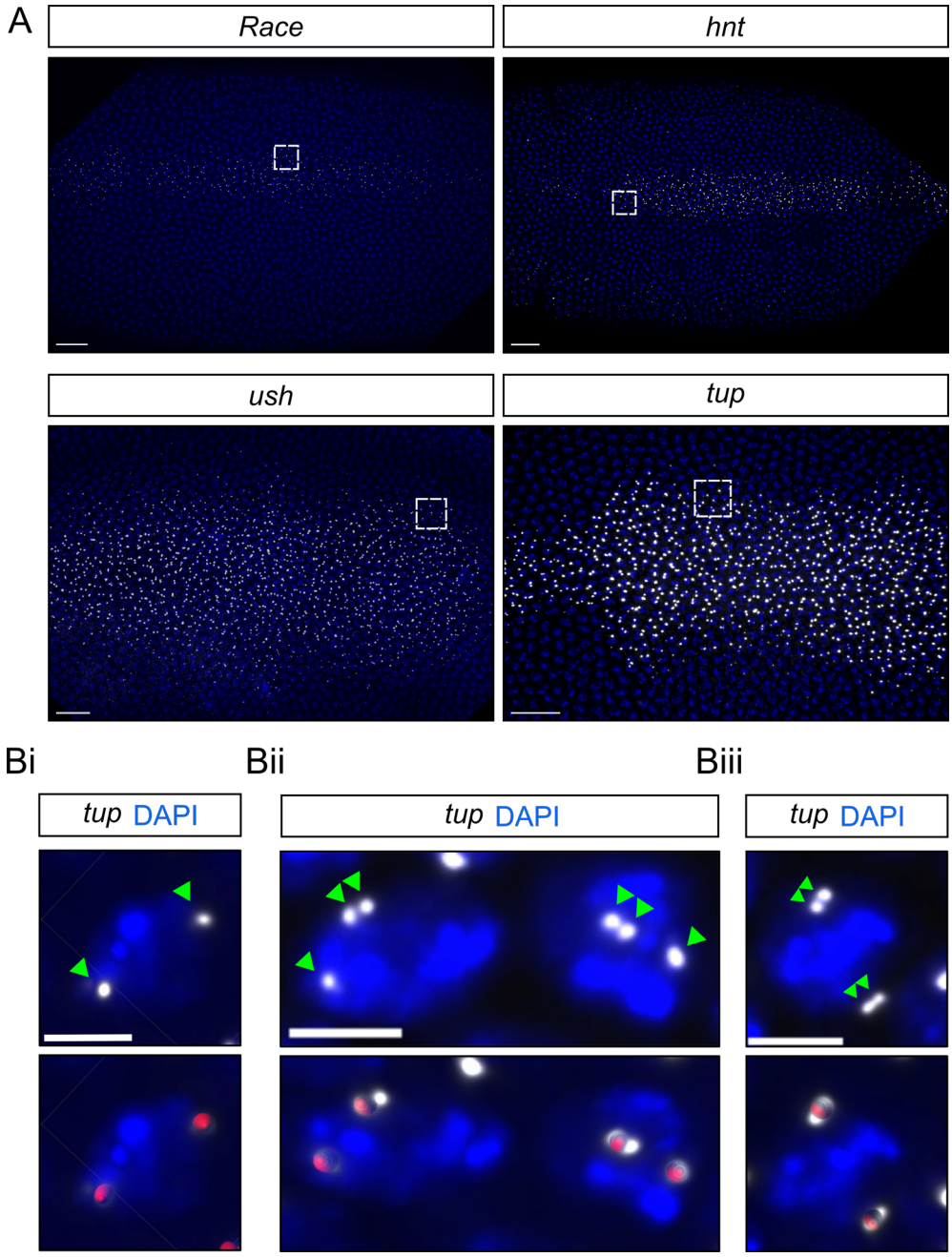



**Figure 3.1: Visualisation of BMP target gene expression domains by RNA-FISH.** Probes targeting intronic RNA sequences were used to visualise nascent transcription foci. (A) Images show maximum projections of embryos in the dorsal orientation with Dapi (blue) and depict classic expression patterns of *Race*, *hnt*, *ush* and *tup* genes (grey). Inset images can be found in Figure 3.2 Ai. Enlarged nuclei from the *tup* expression domain in (A), showing two transcription foci (Bi), and replicated sister chromatids to form three (Bii) or four (Biii) sites of active transcription. *tup* transcription in white and nuclei labelled with DAPI in blue. Lower panel shows transcription site analysis performed in the Imaris software and scoring of sister chromatids as one site of active transcription (red spots) by increasing the spot diameter and preventing sister chromatid detection.
All embryos oriented anterior to the left. Scale bar = 20 μm (A) and 3 μm in (Bi-iii).

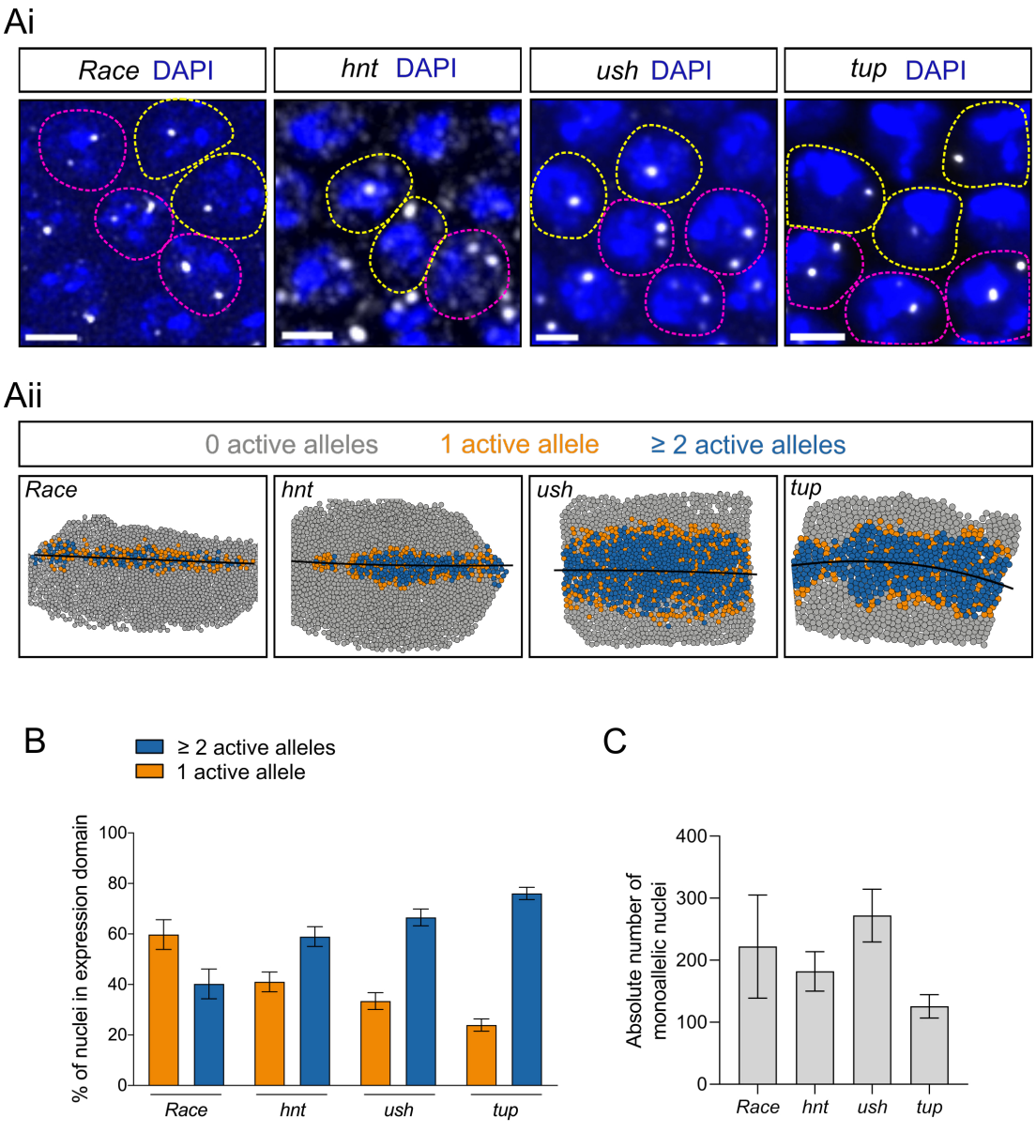



**Figure 3.2: Monoallelic gene expression in BMP target genes *Race*, *hnt*, *ush*, and *tup*.** (Ai) Enlarged regions from RNA-FISH images in Figure 3.1. Probe signal in white and nuclei labelled with DAPI in blue. Images identify nuclei that contain one active transcription site (yellow outline), while others display two transcriptionally active homologous gene alleles (pink outline). Scale bar = 3 μm. (Aii) Images show nuclei false coloured based on the number of active transcription foci, in a coordinate system replicating microscopy images (Figure 1.1 A). Monoallelic transcription in orange, biallelic in blue. A line of best fit was applied to determine the middle of the expression domain. (B) Proportion of nuclei displaying mono- vs biallelic transcription within the expression domain. (C) Absolute number of nuclei displaying monoallelic transcription using same data as in (B). Genes were ordered based on expression domain width.
All images in (Aii) oriented with anterior to the left. Mean and SD are shown in (B, C) and n = 3 biological replicates.

### 3.2.2 Monoallelic transcription occurs at the expression domain border

Following the initial identification of monoallelic nuclei, false colouring reveals that those nuclei predominantly localise towards the edge of their respective expression domain (Figure 3.2 Aii). In order to quantify the distance of individual nuclei to the expression domain centre, the expression domain midline was calculated computationally. This allowed the measurement of nuclear distances to the middle of the expression domain (Figure 3.3 A). Quantification of the number of active transcription sites and the nuclear distance to the midline shows minimal variation between biological replicates (Figure 3.3 B). Replicates were pooled and quantification shows that nuclei with one active transcription site are located significantly further away from the expression domain midline compared to nuclei with biallelic transcription (Figure 3.3 C).

The above data display transcription states as "snapshots" during the time of embryo fixation and suggest a highly dynamic behaviour at the expression domain border. The reduced transcriptional activity of alleles could be explained by limiting nuclear activator levels in response to lower BMP signalling in these domains.

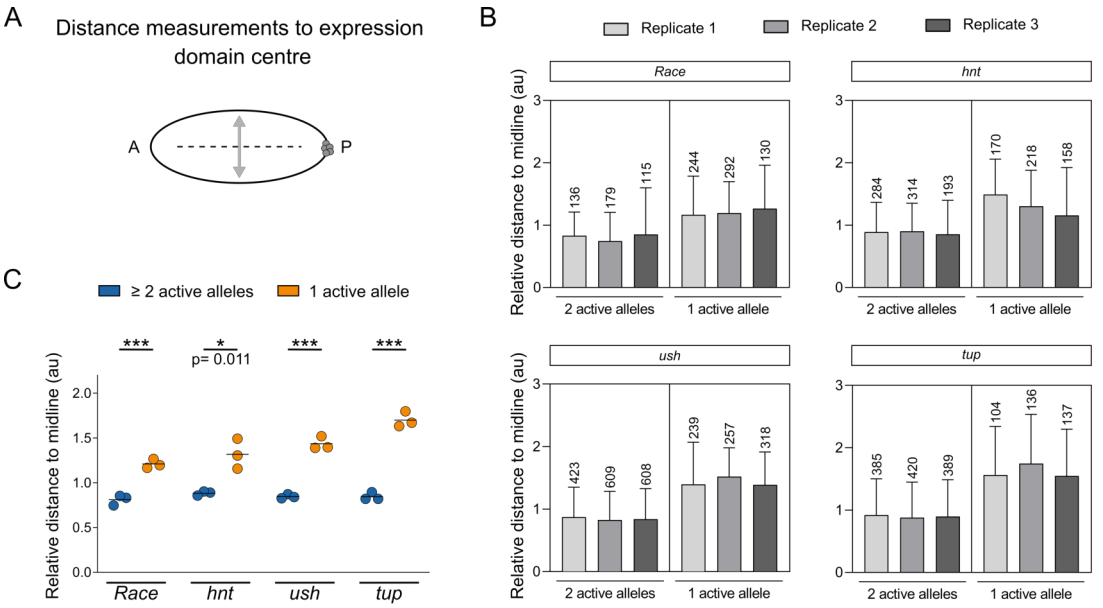



**Figure 3.3: Monoallelic transcription at the expression domain border in BMP target genes.** (A) The distance of each nucleus to the expression domain centre was measured and adjusted to the overall expression domain width, to account for potential embryo size/age differences between replicates. (B) Nuclear distances to the midline scored depending on the number of active transcription sites. All replicates are shown with the number of nuclei analysed per replicate. (C) Median values from (B) show that nuclei displaying monoallelic transcription are significantly further away from the midline and therefore closer to the expression domain border, than nuclei with biallelic transcription.
Error bars depict median and 95% confidence intervals in (B) and line in (C) depicts mean, n = 3 biological replicates. Significance tested with a Mann-Whitney test, *p < 0.05, ***p < 0.001.

### 3.2.3 Monoallelic transcription is a global phenomenon during morphogen gradient interpretation

To test whether monoallelic gene expression is a general feature during gradient interpretation, target genes of another morphogen gradient were analysed. Genes were chosen based on their activation in response to nuclear Dl concentrations, another important morphogen gradient during early *Drosophila* embryogenesis. *sna* transcription is initiated in response to peak levels of nuclear Dl, whereas *brk* and *sog* transcription is activated by lower Dl levels (Figure 1.2 A). As for BMP response genes, the expression patterns of Dl target genes were visualised by RNA-FISH. Probes targeting intronic sequences were generated for *sog* and exonic probes for the intronless genes *brk* and *sna*. Hence, *brk* and *sna* probes visualised both nascent transcription sites as well as cytoplasmic mRNA (Figure 3.4 Ai). *sog* and *brk* are expressed in a lateral stripe along the embryo (Figure 3.4 Ai), corresponding to the previously described expression patterns (Francois et al. 1994, Jaźwińska et al. 1999). On the dorsal side the *sog* and *brk* expression domains border the *dpp* domain, and on the ventral side they border *sna*. Expression of *sna* is detected on the ventral side of the embryo (Figure 3.4 Ai) (Kosman et al. 1991). Again, transcription is found to occur monoallelically in a sub-population of nuclei (Figure 3.4 Aii). Nuclei were falsely coloured in representative images according to the number of active transcription sites (Figure 3.4 Aiii). Consistent with most BMP targets, the proportion of monoallelic transcription in Dl target genes is approximately 25-30% (Figure 3.4 B). Additionally, the absolute number of monoallelic nuclei present in the microscopy images is found to be similar between the three Dl target genes with approximately 180 nuclei (Figure 3.4 C).

Following the same analysis pipeline as for BMP target genes, the relative distance of nuclei to the expression domain midline was calculated. In agreement with previous results, nuclei that show monoallelic transcription in Dl target genes, are located significantly further away from the expression domain centre, compared to nuclei that show biallelic transcription (Figure 3.5 A,B). Less obvious, while still being significant, is the distribution of *sna* transcribing nuclei (Figure 3.5 B). The slightly reduced number of monoallelic nuclei at the *sna* expression domain border suggests a sharper transcription boundary compared to *sog* and *brk*. In agreement, a previous study showed that *sna* expression output is restricted to a tight domain with the help of auto-repression (Boettiger & Levine 2013).

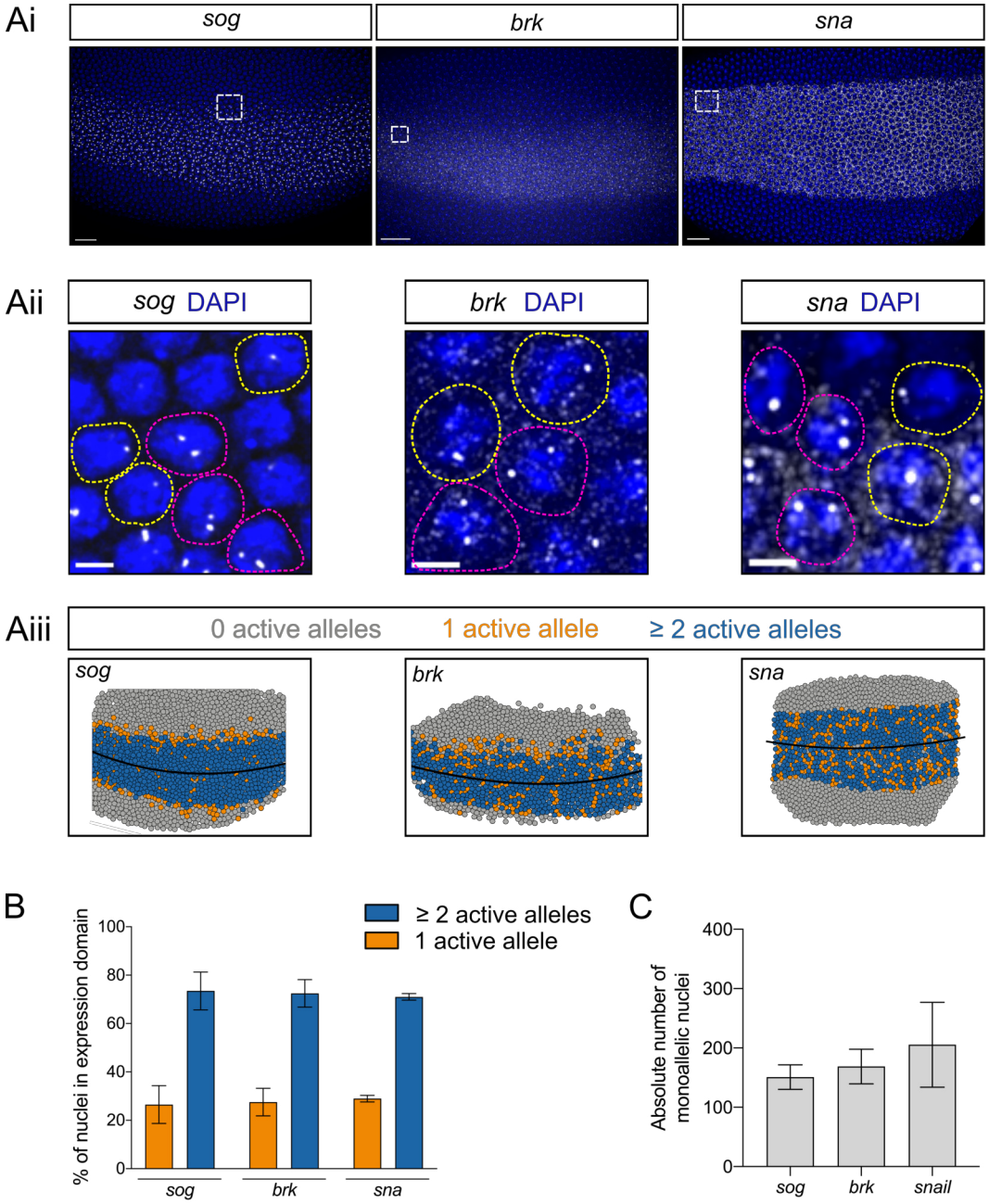



**Figure 3.4: Monoallelic gene expression is a global phenomenon.** (Ai) Maximum projected RNA-FISH images of Dl target genes *sog*, *brk* and *sna* (grey) and DAPI (blue). (Aii) Enlarged regions from images in (Ai). Probe signal shown in white and nuclei labelled with DAPI in blue. Sub-population of nuclei shows monoallelic transcription (yellow outline), while others show biallelic transcription (pink outline). (Aiii) Images recapitulate microscopy images by false colouring nuclei depending on the number of active transcription foci. A line of best fit determines the middle of the expression domain. (B) Analysis of the proportion of nuclei that display mono- vs biallelic transcription within the expression domain. (C) Graph shows the absolute number of monoallelic nuclei using data from (B).
All embryos in (Ai) and (Aiii) are oriented anterior to the left and positioned laterally (*sog*, *brk*) or ventrally (*sna*). Mean and SD shown in (B, C). Biological replicates: *sog* n = 5, *brk* n= 4 and *sna* n = 3. Scale bar = 20 μm in (Ai) and 3 μm in (Aii)

These findings suggest a general existence of monoallelic gene expression in response to morphogen gradients in both BMP- and Dl- responsive genes. Overall, expression domains display between 25% and 60% of monoallelic transcription, mostly at the expression domain borders.

**Reduced monoallelic transcription at repressor boundary**

Sna produces a sharp border between the presumptive mesoderm and neuroectoderm by repressing *sog* and *brk* transcription. Dorsally, *sog* and *brk* expression is limited by decreasing levels of nuclear Dl (Stathopoulos et al. 2002, Reeves & Stathopoulos 2009). If increased monoallelic transcription at the expression domain border is due to insufficient activator concentration, a higher proportion of monoallelic nuclei would be predicted to occur at the dorsal border of *sog* and *brk* domains where their transcription is not actively repressed. Indeed, a sharp expression domain ventral border and the more diffuse dorsal border of *sog* and *brk* transcription is visible in Figure 3.4 Aiii. The number of nuclei with monoallelic transcription at either expression domain border was quantified separately. A higher proportion of nuclei with monoallelic transcription is found at the dorsal expression domain border of *sog* and *brk* (Figure 3.6 Ai). In comparison, quantification of the symmetric BMP target gene and *sna* expression domains reveal an equal proportion of monoallelic nuclei on either side (Figure 3.6 Aii).

Overall, these data further suggest that single allele transcription arises from reduced activator concentrations that are insufficient to initiate transcription from both homologous alleles with high frequency. It can be hypothesised that the lower initiation frequency at the domain borders will result in the production of fewer transcripts and therefore, a heterogeneous expression pattern.

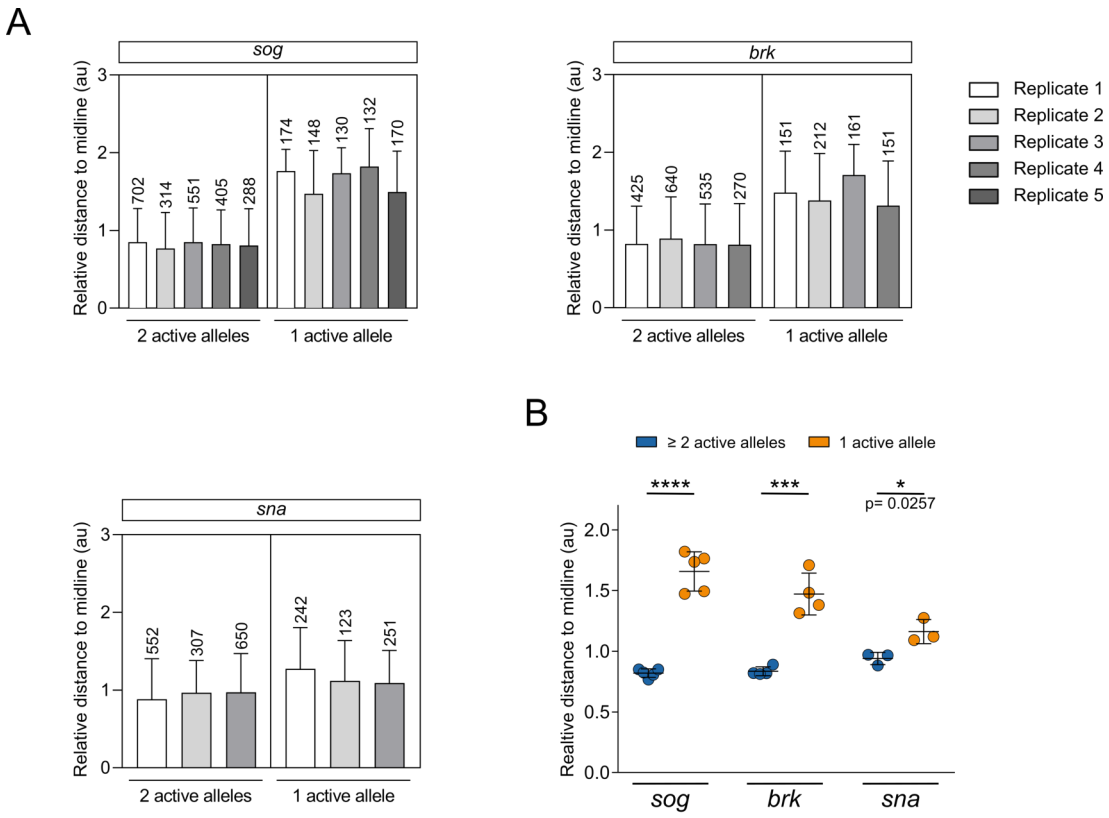



**Figure 3.5: Monoallelic transcription at the expression domain border in Dl target genes.** (A) Relative nuclear distances to the expression domain centre scored based on the number of active transcription sites. All biological replicates are shown with the number of nuclear measurements contained in each bar. (B) Median values from (A) show that nuclei with one active allele (orange) are positioned significantly further away from the midline than nuclei with biallelic transcription (blue).
Error bars in (A) depict median and 95% confidence intervals and lines in (C) show the mean. Biological replicates: *sog* n = 5, *brk* n= 4 and *sna* n = 3. Significance tested with a Mann-Whitney test, *p < 0.05, ***, p < 0.001, ****p < 0.0001.

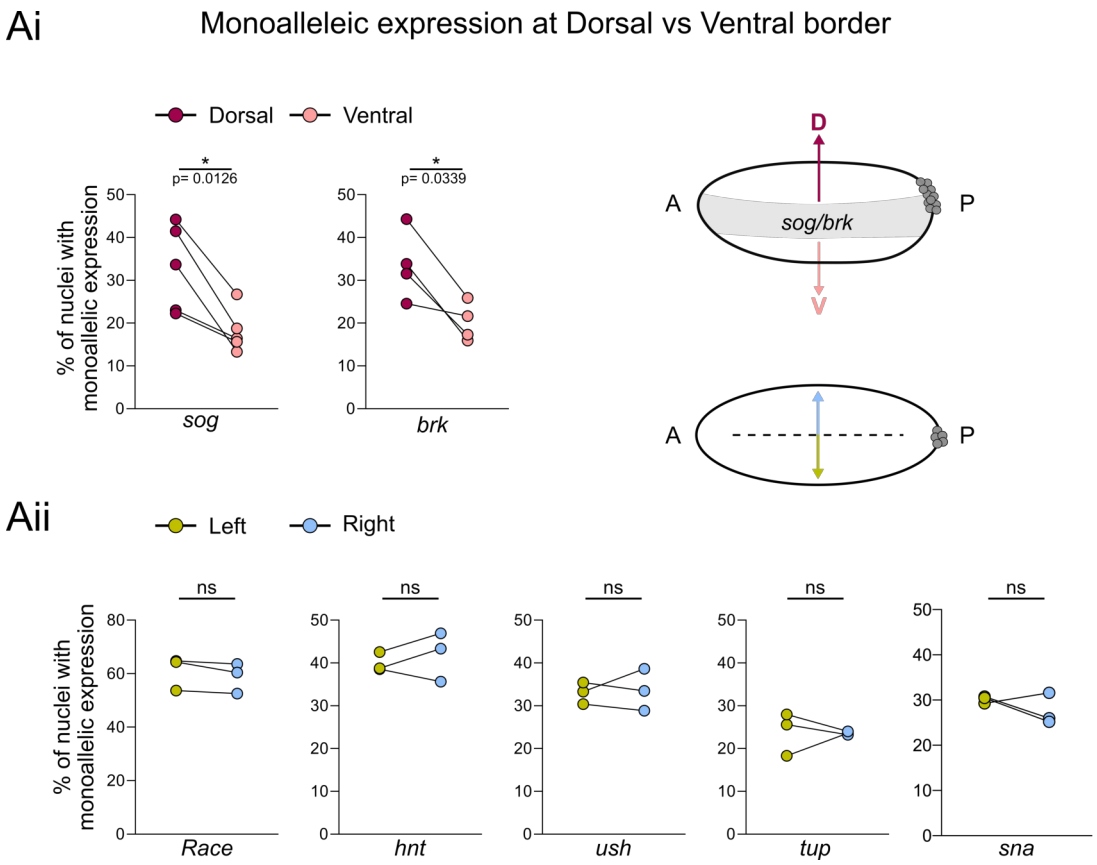



**Figure 3.6: Asymmetric distribution of monoallelic transcription due to border repression in Dl target genes.** (Ai) Proportion of monoallelic transcription in the dorsal (purple) versus the ventral (pink) expression domain half for each embryo. (Aii) Analysis of monoallelic transcription of BMP target genes and *sna* between the left (green) and the right expression domain half (blue). Biological replicates: *sog* n = 5, *brk* n= 4 and *sna* n = 3, *Race* n = 3, *hnt* n = 3, *ush* n = 3, *tup* n = 3. Significance tested with a paired Student's t-test. *p < 0.05; ns = not significant.

## 3.3 Cytoplasmic mRNA molecule distribution mirrors BMP gradient

### 3.3.1 Detection and counting of single transcripts

Following the identification of monoallelic nuclei and their primary location at the expression domain border, next their mRNA output was analysed. The number of mRNA transcripts produced by BMP target genes *hnt*, *ush* and *tup* was investigated as well as differences in mRNA number between monoallelic and biallelic nuclei. The number of cytoplasmic mRNA molecules can be detected using smFISH.

For smFISH detection, a mix of RNA probes, all complementary to the transcript in question and labelled with a single fluorophore, bind to a single transcript. The sum of probe fluorophores can be detected as a bright spot over the background fluorescence (Figure 3.7 A). Furthermore, a false-positive signal should be avoided, as a single probe binding to an off-target location will result in the fluorescence of one fluorophore, whereas the target RNA molecule will be bound by 48 fluorophores and therefore will be 48x brighter (Raj et al. 2008). Two types of signals can be observed in a fixed *Drosophila* embryo stained with smFISH probes targeting exonic RNA sequences (Figure 3.7 Bi). First, intense bright foci overlapping with the nuclear DAPI stain, which correspond to mRNA at active sites of transcription. Secondly, many small foci with weaker fluorescence are observed throughout the cytoplasm and the nucleus. These foci mark single mRNA transcripts (Figure 3.7 Bi).

Initial observation of this laterally positioned *Drosophila* embryo uncovers a *ush* transcript gradient, where the number of single mRNA molecules per cell decreases with distance further away from the midline (black arrow, midline on top of the image, Figure 3.7 Bi). Additionally, nuclei at the expression domain border are observed to show monoallelic (orange arrowheads) and biallelic transcription (white arrowheads), consistent with earlier observations using FISH probes (Figure 3.7 Bii). Together, these two observations suggest a spatial regulation of *ush* transcription and an overall heterogeneity in mRNA number, which will be further investigated in this chapter.

To be able to quantify the number of mRNA molecules, a custom image analysis pipeline was created. Individual molecules were detected in the microscopy images

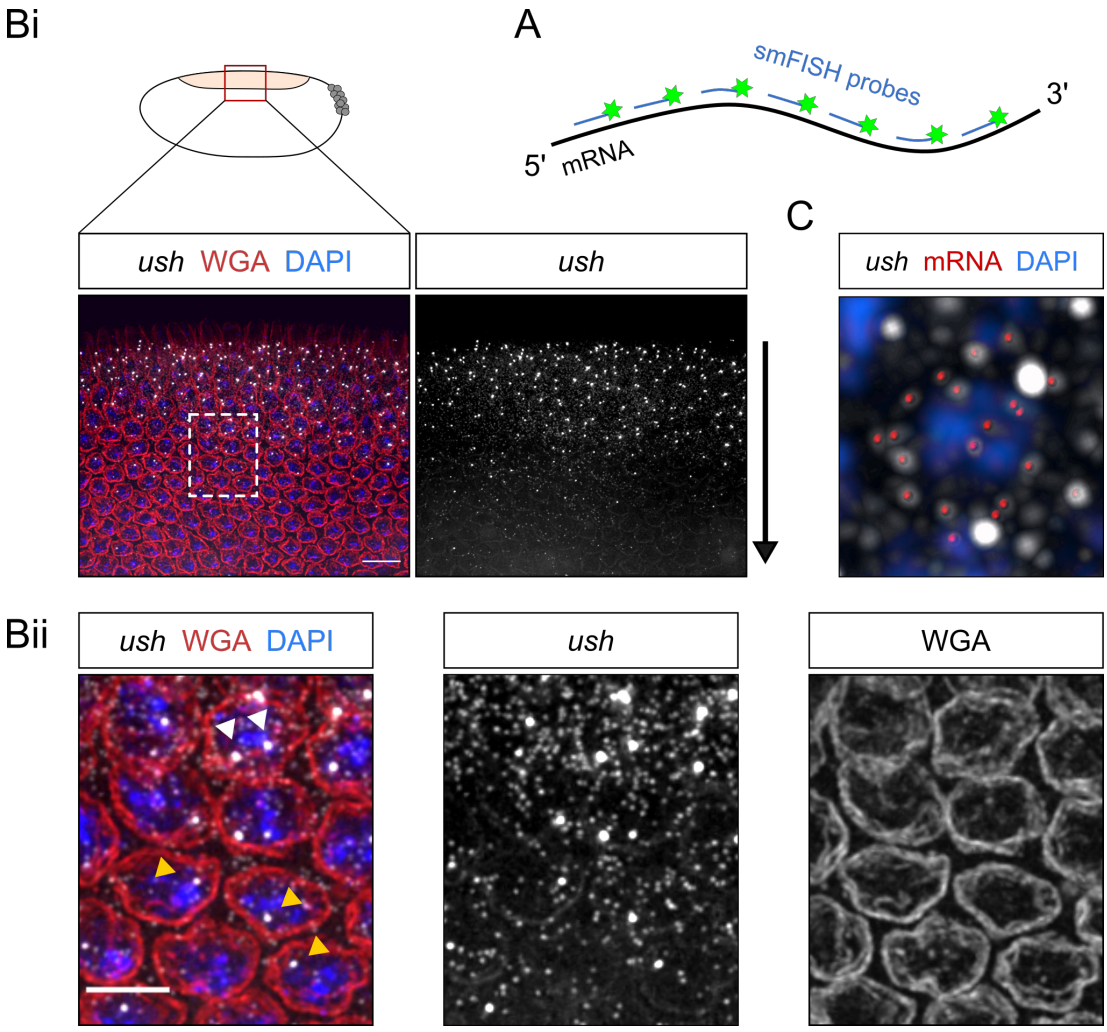



**Figure 3.7: SmFISH, a quantitative method to investigate transcription output.** (A) Schematic of smFISH method. 48 probes conjugated to individual fluorophores bind to different exonic regions within the targeted transcript. (Bi) Laterally positioned embryo, stained with smFISH probes targeting exonic *ush* sequences. Nuclei labelled with DAPI (blue) and the nuclear envelope with Wheat Germ Agglutinin (WGA). Transcript gradient from dorsal to more ventral region (black arrow). (Bii) Enlarged region from (Bi) showing transcription sites as bright foci and single mRNAs as small puncta. Monoallelic (orange arrowhead) and biallelic (white arrowhead) transcription is observed. (C) Enlarged image of a single nucleus (blue) and mRNA molecules (red) that are assigned during image analysis.
Embryo in (Bi) oriented anterior to the left and dorsal at the top. Scale bar = 15 μm in (Bi) and 10 μm in (Bii)

and then assigned to their closest nucleus in three dimensions similar to Boettiger & Levine (2013) (Figure 3.7 C). This way, all nuclei within the field of view could be analysed, and their number of mature RNA molecules and active transcription sites were recorded for analysis.

### 3.3.2 Time course of BMP target gene transcription

#### Transcription output increases with time

Given the differences in *ush* mRNA number observed, time course data were generated in order to link the temporal and spatial aspects of *hnt*, *ush* and *tup* transcription output. Sites of nascent transcription were identified using smiFISH probes targeting intronic RNA sequences. Exonic probes were used to quantify the absolute number of single mRNA molecules (Figure 3.8 A). Multiple embryos stained with probes against *hnt, ush* and *tup* were analysed to generate a time course data set. A representative microscopy image is shown for each gene (Figure 3.8 B). The age of embryos was staged relative to each other based on nuclear morphology. During the course of nc14, nuclei become more condensed as their shape changes from spherical to ellipsoid during cellularisation. Towards the end of nc14 and with the onset of gastrulation, nuclei start to migrate out of their cortical monolayer (Schejter & Wieschaus 1993).

When quantified, the maximum amount of mature RNAs detected in cells increases steadily throughout nc14 for all three genes analysed. The highest number of *hnt* mRNAs detected in a single cell is 222 molecules (Figure 3.9 Ai). Slightly higher numbers are detected for *ush* transcription with a maximum of 262 molecules per cell (Figure 3.9 Bi). The highest *tup* transcription output per cell was found to be 189 molecules (Figure 3.9 Ci). The amount of nuclei with monoallelic transcription is shown to decrease over time for all genes analysed and ranges from 36% to 18.5%. *hnt* transcription displays the highest proportion of monoallelic transcription (Figure 3.9 Aii-Cii). These numbers correlate with previous measurements using FISH probes (compare, Figure 3.2 B). Analysis of nuclear transcription state indicates that the initial transcription onset of *hnt, ush* and *tup* transcription is more stochastic than later stages, with increased occurrence of monoallelic transcription shortly after transcription initiation.

Next, mRNA numbers from the oldest embryos in the time course were analysed

in relation to the number of active nuclear transcription foci. A subset of nuclei is observed to contain low numbers of mature transcripts in the cytoplasm but no active transcription sites are detected (grey bars). These cells most likely transcribed the respective BMP target gene from at least one allele at an earlier time (Figure 3.9 Aiii-Ciii). The analysis shows that nuclei with two active transcription sites contain significantly higher numbers of mRNAs in their cytoplasm compared to monoallelic nuclei (Figure 3.9 Aiii-Ciii). This observation is consistent in all genes and suggests that the transcription state at the time of fixation is indicative of the cell's transcriptional output. Overall, the number of transcripts per cell correlates with the nuclear transcription state (correlation coefficient, Figure 3.9 Aiii-Ciii).

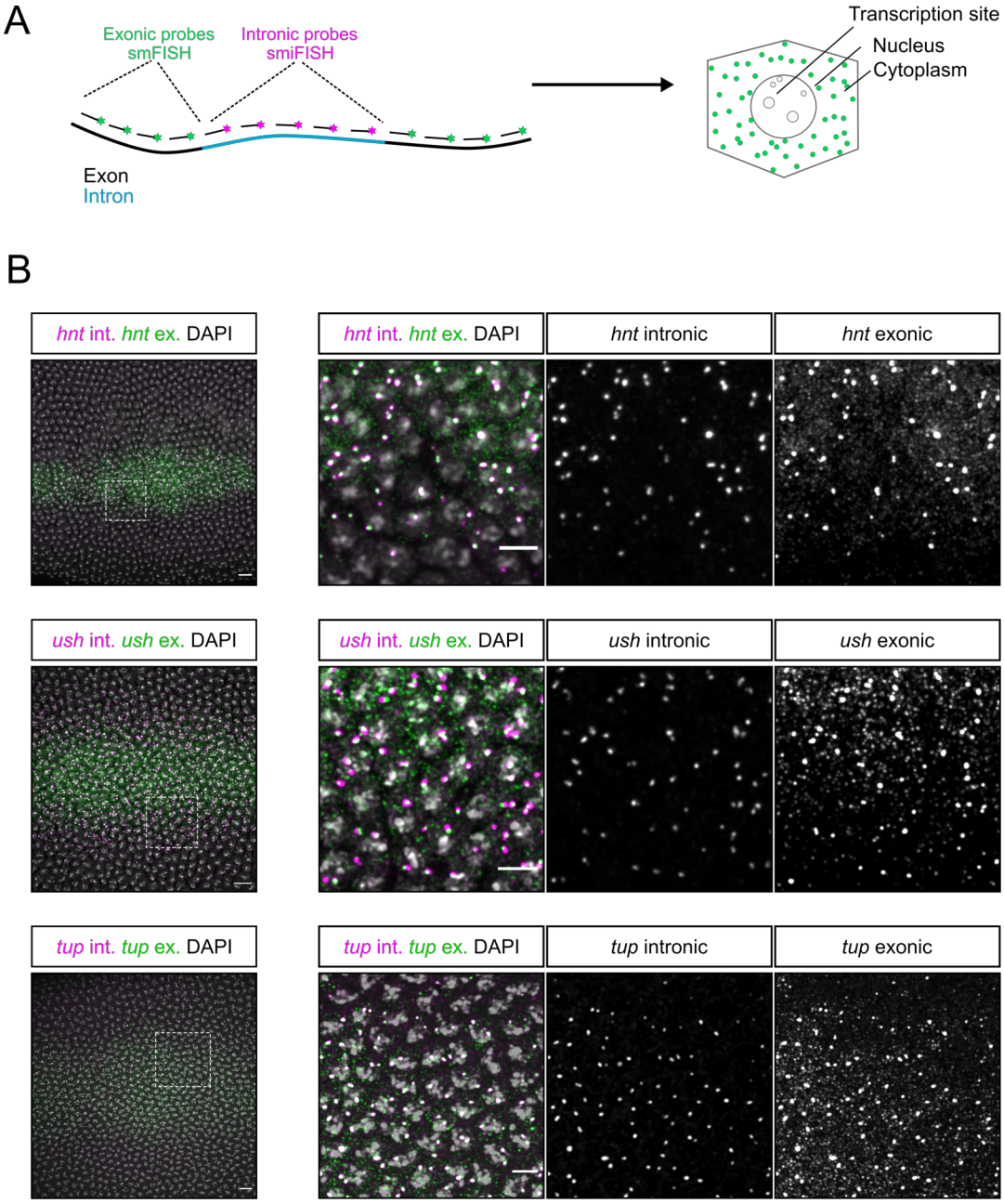



**Figure 3.8: Microscopy images of BMP target genes *hnt, ush* and *tup* labelled with smFISH probes.** (A) Schematic of smFISH approach where intronic probes target pre-mRNA in the nucleus and exonic probes label both, pre- and mature RNAs. (B) Maximum projected microscopy images of BMP target genes with DAPI labelling nuclei (white), intronic smFISH probes (magenta) and exonic probes (green) labelling transcription sites and single RNA molecules respectively.
All embryos in (B) are oriented dorsally with anterior to the left. Scale bar in (B) represents 10 μm and in enlarged image 5 μm.

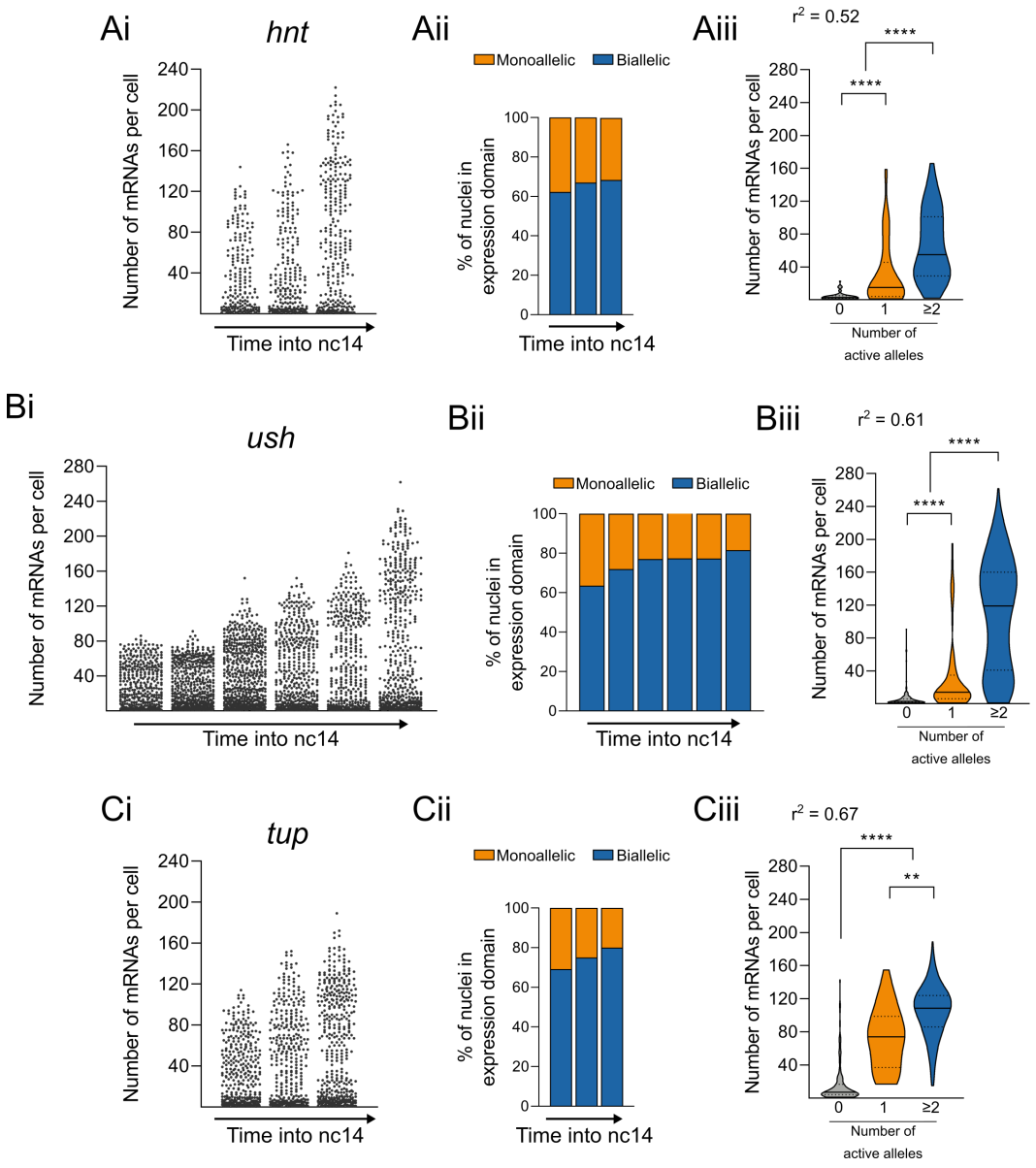



**Figure 3.9: Timecourse of BMP target gene transcription.** Time course of *hnt* (Ai), *ush* (Bi) and *tup* (Ci) transcription output where every column represents an embryo and every dot a single cell. (Aii-Cii) Corresponding percentages of monoallelic (orange) and biallelic nuclei (blue) in each time course embryo. (Aiii-Ciii) Data from oldest embryo in (Ai-Ci) plotted with the number of transcripts per cell based on nuclear transcription state. Spearman correlation coefficient calculated based on number of transcripts versus number of transcription foci.
Embryos were ordered according to increasing age in nc14. Lines in violin plots (Aiii-Ciii) depict median and 95% confidence intervals. Significance tested using a Kruskal-Wallis test with a Dunn's multiple comparisons test (Aiii-Ciii). **p< 0.01, ****p < 0.0001.

**Transcription output is correlated to position within the BMP gradient**

After the initial analysis of time course data, which showed differences in the mRNA number produced by monoallelic and biallelic nuclei, the datasets were further investigated towards potential differences based on the cell's position. The number of mRNA molecules per cell was plotted in relation to the cell's position across the expression domain within the BMP gradient. The visualisation of mRNA number based on position reveals an mRNA gradient. Highest transcript levels are found at the dorsal midline, which diminish in more dorsolateral regions. This behaviour is observed for all three BMP target genes analysed in this section. With increasing time into nc14, the absolute numbers of mature RNAs increase, sharpening the transcript gradient (Figure 3.10 Ai-Ci).

When embryos of different ages, are directly compared, it can be observed that their transcription profiles steepen as time progresses into nc14. Also, the expression domain widens, which is observed by the increase in curve width at the base (Figure 3.10 Aii-Cii). A two and three-dimensional representation of the *ush* transcription domain highlights the steep gradient of mRNA number and the slight variations around the dorsal midline (Figure 3.11 Ai, Aii). The transcription output differs slightly in a direct comparison between BMP target genes. The *hnt* transcription domain is narrower, compared to *ush* and *tup*, which is consistent with its activation by peak levels of BMP signalling (Figure 3.10 D). The transcription gradients in BMP target genes mirror that of BMP and pMad levels as shown in the literature (Dorfman & Shilo 2001). The difference in absolute numbers of mRNAs, between cells close to the dorsal midline and at the edges of the expression domains, are substantial (Figure 3.10 D). Previously these differences in expression could not be analysed as classical ISH is not a quantitative method (Ashe & Levine 1999). When estimating the total number of mRNAs produced by the respective expression domains (area under the curve), the *hnt* transcription domain produces the lowest, and *ush* the highest number of transcripts (Figure 3.10 D).This order is consistent with a nano-string study that evaluated transcript numbers during a time course of early Drosophila embryogenesis (Sandler & Stathopoulos 2016). Even though the relative order is consistent, the absolute number of transcripts produced by the respective expression domains is estimated by Nano-String analysis to be very different (see Discussion).

In addition to the number of mature cytoplasmic transcripts, the number of mRNAs present in an active transcription site can be estimated. At full probe occu-

pancy, every single mRNA molecule should emit the same fluorescence. Dividing the fluorescence of a nascent transcription site by the mean fluorescence of a single mRNA generates an estimate mRNA number (Figure 3.10 Ei). A large number of transcription sites was analysed for each BMP target gene, and the mean number of mRNAs present in a nascent spot is very similar. On average *hnt, ush* and *tup* transcription sites contain 20 mRNAs (Figure 3.10 Eii). Likely, these mRNAs were still being transcribed and therefore still associated with Pol II molecules at the time of fixation. Hence, it is possible that not all regions recognised by smFISH probes were fully transcribed and therefore, the count of fluorophores has an unknown error rate. The range of Pol II associated mRNA number is large and ranges from 4 to 110 in case of *tup* (Figure 3.10 Eii). This observation suggests dynamic transcription where the number of actively transcribing Pol II molecules varies between cells.

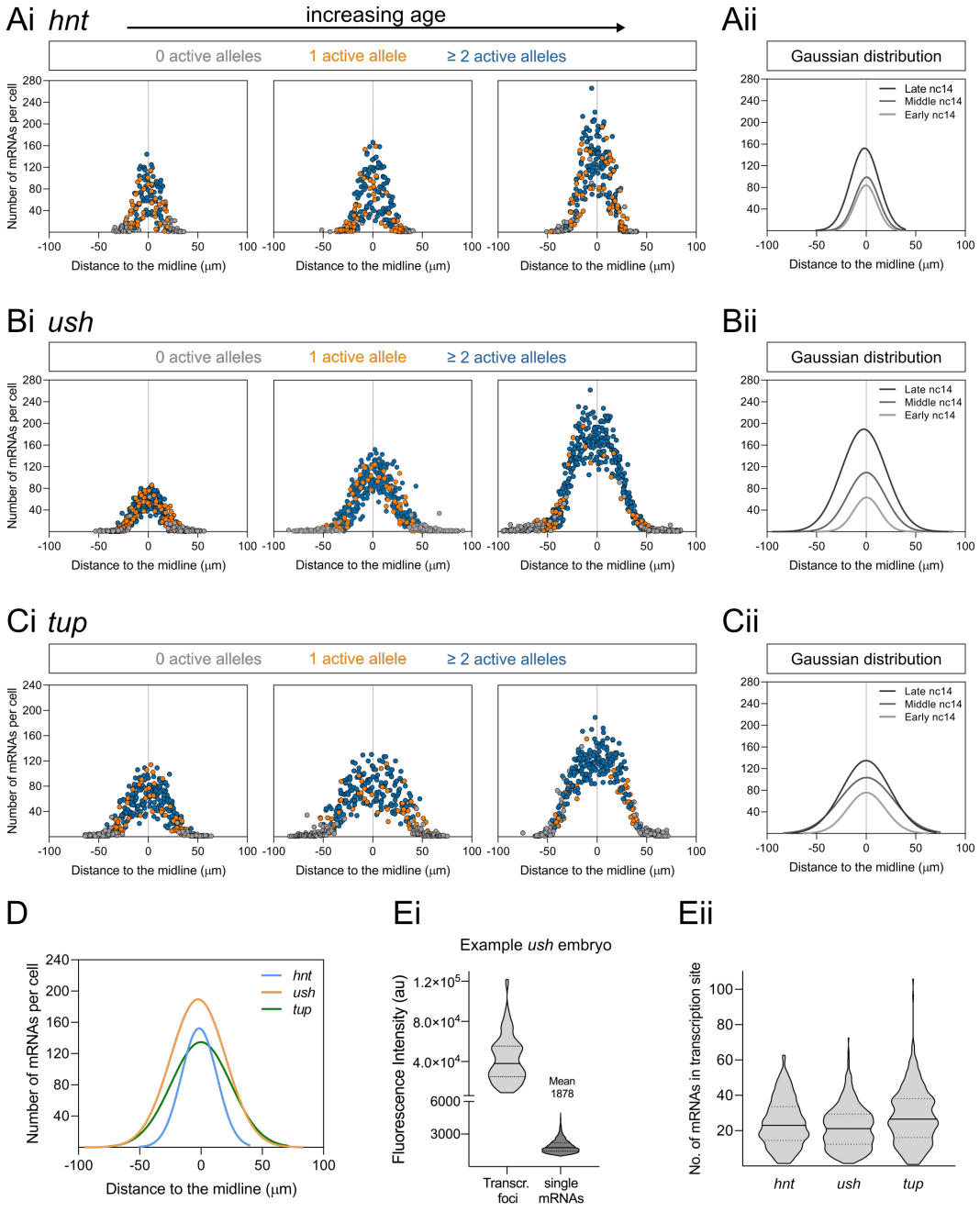



**Figure 3.10: Transcriptional output of BMP target genes mirrors BMP gradient.** Transcriptional output of time course data for *hnt, ush* and *tup* at three points during nc14. Number of mRNA transcripts per cell plotted according to their position across the dorsal midline for *hnt* (Ai), *ush* (Bi) and *tup* (Ci). Each plot shows data from one embryo and each point represents one cell. Embryos were ordered with increasing age and points are coloured according to the nuclear transcription state. mRNA output from different time points in nc14 is compared by plotting the Gaussian distributions from (Ai-Ci) for *hnt* (Aii), *ush* (Bii) and *tup* (Cii). (D) Gaussian curves from the oldest embryos in (Aii-Cii). (Ei) Example measurements to determine number of mRNAs present in transcription site. Fluorescence of transcription foci and single mRNA transcripts. (Eii) Number of single mRNAs present in nascent transcription foci. Number of foci analysed from n=3 biological replicates: *hnt* = 284, *ush* = 473, *tup* = 413.
Lines in violin plots (Ei, Eii) depict median and 95% confidence intervals.

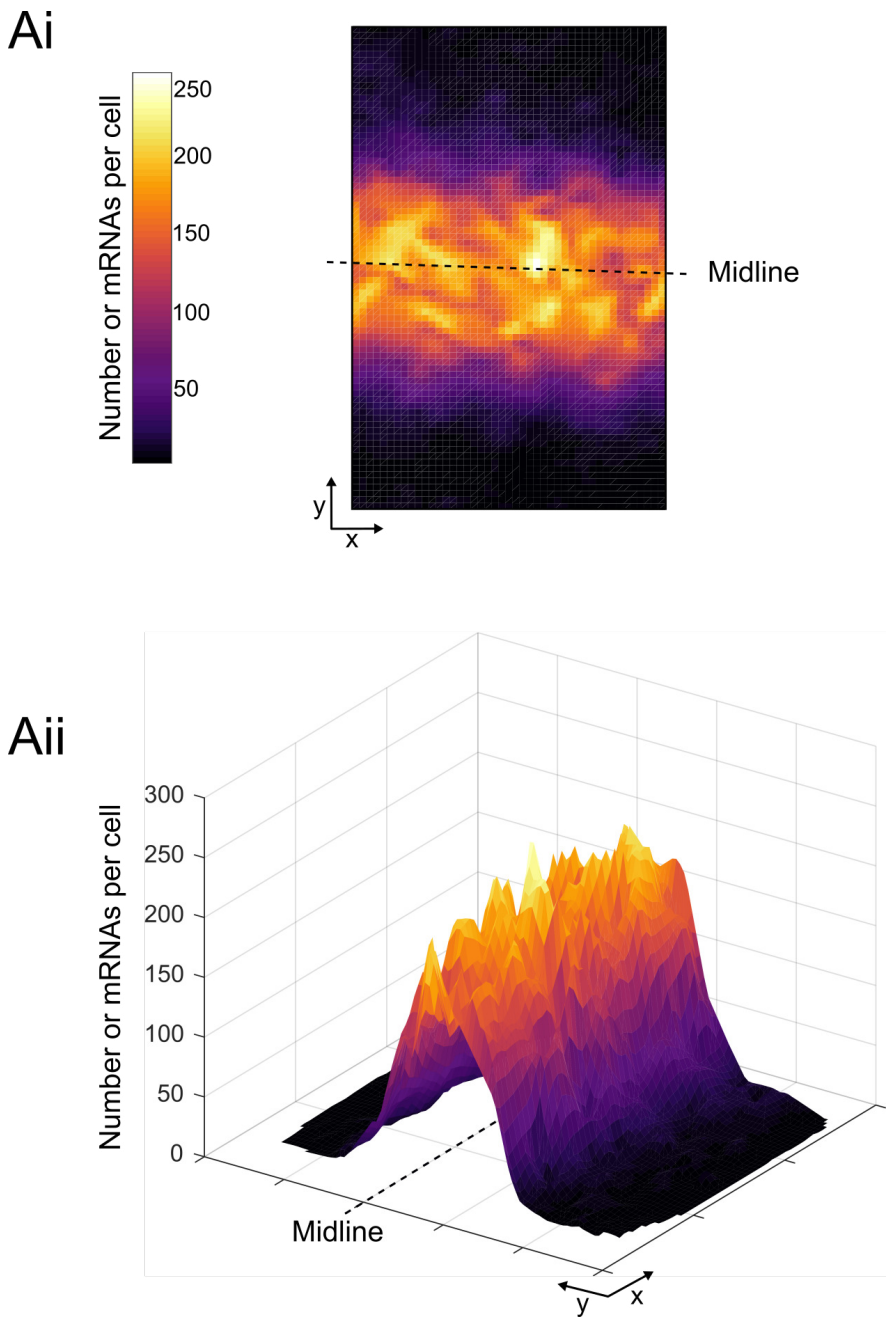



**Figure 3.11: Three dimensional visualisation of *ush* transcription gradient.** (Ai) Two dimensional projection of *ush* transcription. Data from oldest embryo in the *ush* time course. Nuclei coloured according to the number of mature mRNAs detected in their cytoplasm. (Aii) Three dimensional representation of *ush* transcription output.

**Disproportional contribution to transcription is based on cell position**

Next the data were considered in relation to BMP signalling. Peak BMP signalling levels are found in a narrow strip around the dorsal midline. This area is approximately 8-10 cells wide during stage 5, refines to 6 cells in stage 6 and represents the first "step" in the pMad/ BMP concentration gradient (Figure 3.12 Ai) (Dorfman & Shilo 2001, Mizutani et al. 2005). Since cells likely respond directly to the amount of BMP signalling they experience, this section will investigate the transcriptional output produced by each cell row. For this, data obtained from the *hnt*, *ush* and *tup* time courses were mirrored at the dorsal midline and separated into bins of one cell width. Cells were estimated to have a width of 5 μm (based on IMARIS image analysis). Data plotted for the oldest time course embryos reveal a considerable amount of transcripts produced in the 4 cell radius around the expression domain midline (Figure 3.12 B). In detail, the highest amount of *hnt* transcripts is produced directly adjacent to the dorsal midline and then drops rapidly until almost no transcripts are produced in cells positioned 5 cell rows away from the midline (Figure 3.12 B). Next, the proportion of transcripts produced in the 8 cell wide central domain is quantified in all three biological replicates. This central strip accounts for over 90% of all transcripts produced in the *hnt* expression domain (Figure 3.12 C). These data are in agreement with the narrow width of the overall *hnt* expression domain and its response to maximum levels of BMP signalling (Ashe et al. 2000) which should be encompassed within this 8 cell wide domain.

In comparison, the median number of *ush* transcripts produced within the 8 cell wide domain almost plateaus before it starts to decrease rapidly at the 8 cells wide perimeter and drops to a median of below 10 transcripts at 11 cells wide (Figure 3.12 B). The 8 cell wide peak BMP signalling domain still contributes approximately 70% of all produced *ush* transcripts (Figure 3.12 C). Hence, even though the expression domain of *ush* is much wider than *hnt*, most transcripts are produced in response to peak levels of BMP signalling and cells further away from the midline barely contribute to the overall transcription output.

Similarly, the median number of *tup* transcripts per cell is almost constant within the 1-4 cell domain adjacent to wither side of the midline. Starting in cell row 5 the number of transcripts steeply decreases with distance from the midline (Figure 3.12 B). Even though maximum *tup* transcript numbers per cell are slightly below that of *ush* the transcript distribution is very similar. The centre domain of *tup* contributes approximately 60% of the overall transcriptional output generated by the *tup* expression domain.

The spatial distribution of single mRNA molecules in the *Drosophila* blastoderm was shown to be variable between genes. *sna* transcripts (Boettiger & Levine 2013) and other developmental patterning genes (Davis & Ish-Horowicz 1991) have been reported to be asymmetrically distributed. Single *sna* transcripts are predominantly present in the apical compared to the basal cytoplasm (Boettiger & Levine 2013). This behaviour is suggested to limit the diffusion of proteins (Davis & Ish-Horowicz 1991) and at the same time probably mRNA within the *Drosophila* embryo before cellularisation. When *ush* transcripts are visualised in a zy projection, they are seen apically as well as basally in relation to the nuclear position (Figure 3.12 Aii). Quantification of the mRNA distribution in all time course embryos reveals that mRNAs are evenly concentrated at the apical and basal side for both *ush* and *tup*. tup transcripts show a trend towards basal localisation (Figure 3.12 D). In comparison, *hnt* shows a predominantly basal localisation (Figure 3.12 D). Hence, it is possible that a number of basal transcripts diffuse through the syncytium before and during cellularisation.

So far, measurements of transcript number and correlation to the transcription state and the cell's position within the BMP gradient suggest that BMP target gene expression occurs dynamically, where alleles in the expression domain centre initiate transcription more frequently or have longer burst duration. Moreover, a large range of mRNA number present in active transcription sites was detected, further suggesting a heterogeneous and possibly stochastic transcription behaviour. Often transcription dynamics are defined by stochastic transcription initiation resulting in transcriptional bursts (Raj et al. 2006). A key parameter indicating bursting is the fano factor which measures the variability in the number of produced mRNAs (Munsky et al. 2012). The fano factor is defined as the ratio between the variance and the mean (Thattai & van Oudenaarden 2001). If transcription occurs continuously without bursts, the fano factor equals one. If the distribution/variance of mRNA number is greater, as detected for bursting, the fano factor is greater than one (Munsky et al. 2012). The *hnt*, *ush* and *tup* time course replicates were analysed to determine if these BMP target genes are likely to display transcriptional bursting. All three genes display a much greater mean fano factor than one ($hnt = 45.55$, $ush = 42.39$, $tup = 36.80$) and therefore can be classed as super-Poisson (see Discussion) (Figure 3.12 E). These data suggest that BMP target gene transcription exhibits transcriptional bursting, and it can be hypothesised that changes in burst kinetics generate the changes in transcriptional output along the BMP gradient.

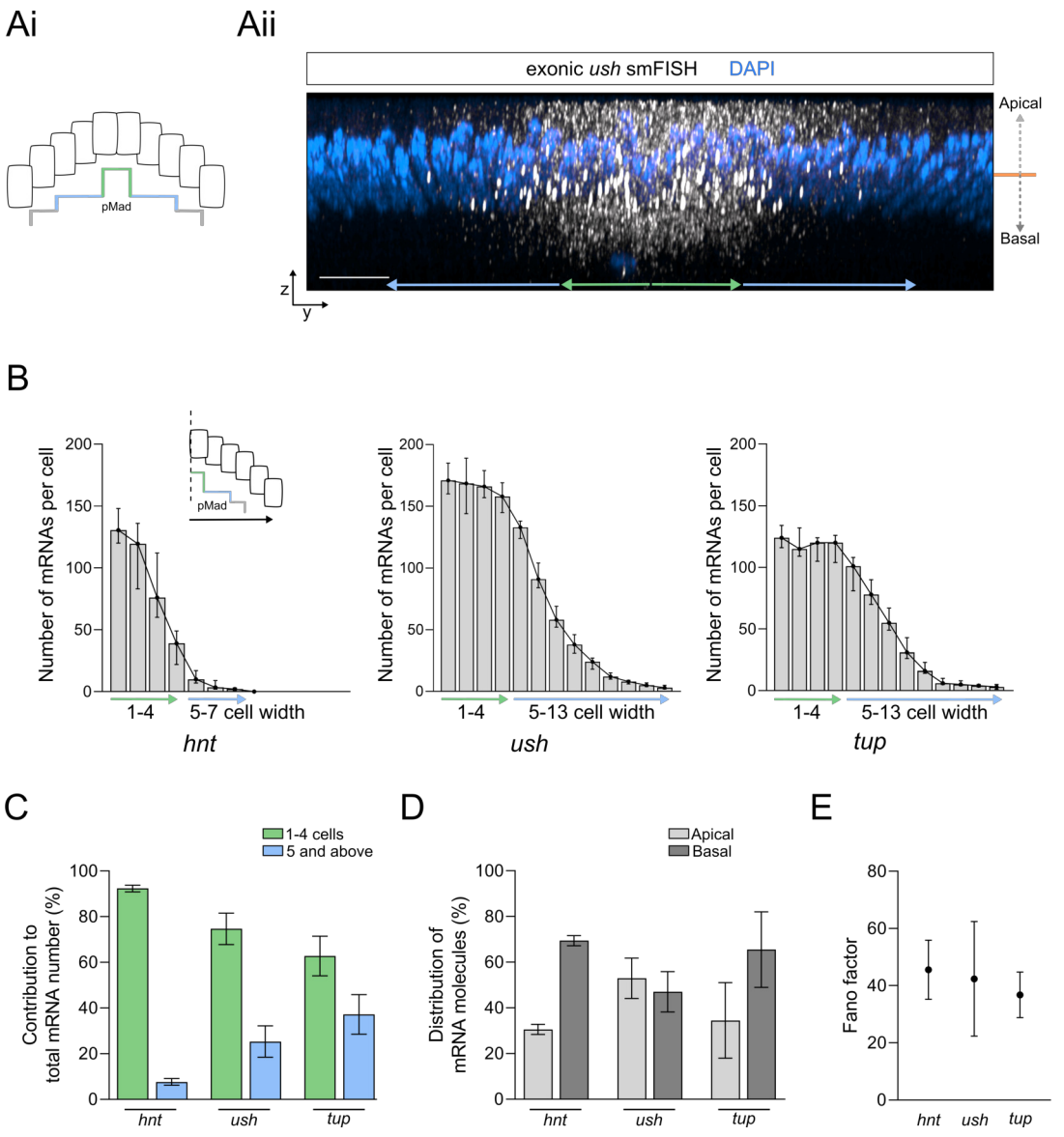



**Figure 3.12: Contribution of cells in different locations to overall transcription output.** (Ai) Schematic of embryo cross-section with pMad step-gradient coloured based on cell width from expression domain centre. (Aii) zy projection of a 10 μm thick slice from a *ush* time course embryo. Exonic smFISH probes labelling *ush* mRNA and DAPI labelling nuclei (blue). Green arrows indicate 0-4 cell width and blue arrows 5-13 cell width from expression domain centre. Apical and basal sides are indicated by grey arrows. (B) Data obtained from oldest embryos in *hnt*, *ush* and *tup* time courses. Median number of transcripts produced by cells within 1 cell wide bins. (C) Contribution of the central 8 cell wide domain and cell rows 5-7 or 13 of the respective expression domain. (D) Measurement of transcript localisation in relation to nuclear position. (E) Fano factors calculated as the ratio of variance to mean transcript number.
Error bars (B) depict median and 95% confidence intervals, mean and SD (D,E), n = 3 (*hnt, tup*) and 6 (*ush*) biological replicates (C-E). Scale bar in (Ai) = 10 μm.

## 3.4 Gene expression changes in response to genetic alteration

After characterising the wildtype transcription profiles of BMP target genes both spatially and temporally, their response to genetic changes was investigated. The transcription output of *tup* and *ush* was quantified in genetic backgrounds with altered BMP signalling levels. In this section, the *tup* transcription profile will be quantified in *sog* mutant embryos and in embryos that carry a *tup* deficiency. In a *sog*$^{-}$ background, there is no lateral inhibition of BMP signalling, and the steep BMP signalling gradient is lost due to loss of BMP ligand shuttling. As a result uniform low levels of BMP signalling, sufficient to activate intermediate and low level BMP target genes, expand towards lateral domains (Francois et al. 1994, Biehs et al. 1996).

*ush* transcription will be quantified in genetic backgrounds with increased or decreased BMP signalling by changing the dose of *dpp* in embryos. Furthermore, embryos that receive lower levels of maternal Med protein from mothers heterozygous for the *Med*$^{13}$ null allele experience reduced levels of BMP signalling (Hudson et al. 1998). Conversely, embryos that carry the *even skipped* stripe 2 enhancer (st2) dpp transgene show increased BMP signalling in areas where the *even skipped st2 enhancer* induces ectopic *dpp* expression (Ashe et al. 2000).

### 3.4.1 *tup* transcription profile

Different genetic backgrounds were analysed to characterise *tup* transcription behaviour. The *tup* wildtype expression profile was generated using three biological replicates that were determined to be of similar age based on nuclear morphology. One replicate has been used previously as part of the *tup* time course (medium age). The additional two replicates in this section have not been described in previous figures. The median number of transcripts per cell, produced by wt embryos, is very consistent at approximately 45 transcripts (Figure 3.13 Ai). The maximum number, however, varies, which represents biological variation (Figure 3.13 Ai). Next, the individual replicates were plotted according to the cell position across the midline to determine the *tup* wt transcription profile before comparing it to mutant backgrounds (Figure 3.13 Aii). Overlaying the individual Gaussian distribution curves shows the consistent width in expression domain and overall shape with small variation in the absolute curve peak (Figure 3.13 Aii).

The median number of transcripts produced per cells in embryos heterozygous for the *tup* deficiency, which removes the entire *tup* gene locus, is approximately 20 mRNAs and therefore reduced compared to wt levels (Figure 3.13 Bi). The reduced number of mRNAs in *tup* deficient heterozygous embryos is also visible when cells are plotted in relation to the midline based on their cross-section position (Figure 3.13 Bii).

Disruption of the extracellular BMP ligand gradient in *sog*$^{-}$ mutatants results in almost uniform extracellular BMP ligand concentration (Biehs et al. 1996, Mizutani et al. 2005). This study finds that the overall number of *tup* transcripts is reduced. It would be predicted that the *tup* transcription pattern is widened as Sog can no longer inhibit BMP signalling in dorsolateral domains. When plotted based on nuclear position, *tup* transcription shows a very shallow curve with a shallow peak and an overall broader domain than wt *tup* (Figure 3.13 Cii). Even though BMP signalling in *sog*$^{-}$ embryos is uniform, reduced mRNA levels at either side of the wide peak are visible. The tails at either side are possibly explained by Brk repression of Dpp (see Discussion).

A middle section of the embryo was analysed in terms of mono- and biallelic transcription for all embryos that are part of this data set (Figure 3.14 Ai). In direct comparison, the *tup* expression domain in embryos from different genetic backgrounds appears changed (Figure 3.14 Aii). In wt embryos, monoallelic nuclei are primarily found at the expression domain edge. These nuclei constitute approximately 20% of the overall expression domain (Figure 3.14 B). This proportion is consistent with results obtained earlier using FISH probe sets.

In $\Delta tup/+$ embryos, the expression domain width is slightly narrower compared to wt, but the overall expression domain shape is similar (Figure 3.14 Aii). The change in expression domain width can be predicted to arise from the loss of nuclei from the expression domain that would have previously shown monoallelic transcription at the domain edge in a wt embryo. All nuclei that are part of the *tup* deficiency embryo expression domain show monoallelic expression as it would be expected from embryos containing only one *tup* gene locus (Figure 3.14 B). Furthermore, the maximum transcriptional output is highest in wt embryos with 243 transcripts per cell. $\Delta tup/+$ embryos in this study produced a maximum of 195 mRNAs, a little more than two-thirds of the wt level. Even though the maximum number of transcripts produced per cell is higher than the expected 50% wt value, the median number produced by $\Delta tup/+$ embryos is approximately 50% of the

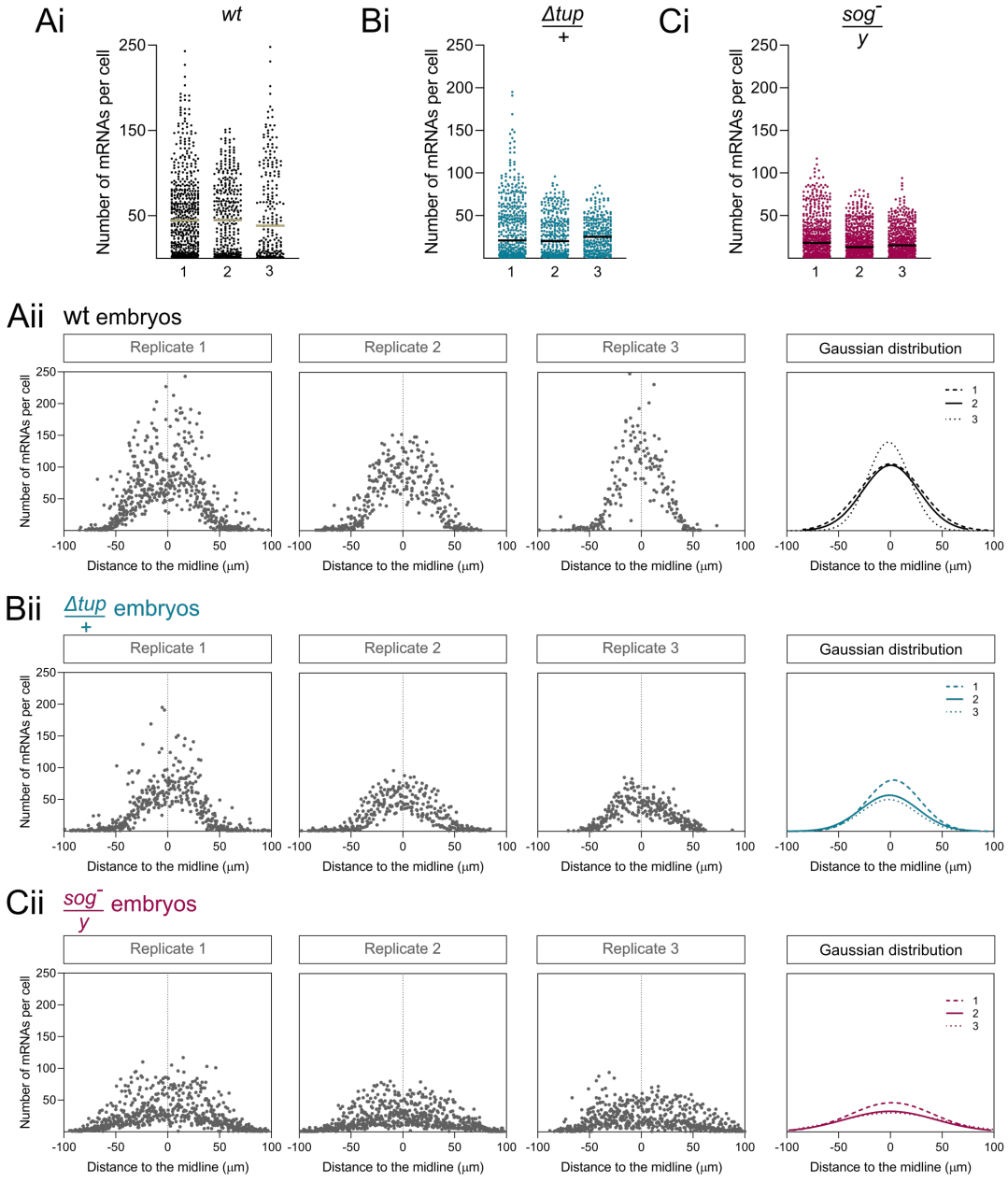



**Figure 3.13: Altered *tup* transcription behaviour in response to genetic changes.** *tup* transcription output per cell in wt embryos (Ai), *tup* deficient embryos (Bi) and with *sog*⁻ mutation (Ci). (Aii-Cii) Number of mature RNAs per cell plotted according to the cell position across the midline. Gaussian distribution curves overlaid for all replicates.
Bars (Ai-Ci) represent median.

profile of mRNA number in wt embryos (wt = 44.5, $\Delta tup/+ = 21$, Figure 3.14 D). The *tup* expression domain profile in heterozygous embryos is almost the same width as the calculated 1/2 wt profile. The peak height is slightly higher, possibly suggesting a small degree of compensation around the dorsal midline (Figure 3.14 E). Overall these results show that the *tup* deficiency results in approximately half the expression level and a narrower expression domain width.

The *tup* expression domain in *sog*$^{-}$ embryos is widened and nuclei, previously not part of the expression domain, now show active *tup* transcription, therefore increasing the number of nuclei that are part of the expression domain (Figure 3.14 Aii). When quantified, the proportion of monoallelic nuclei is slightly increased compared to wt embryos (Figure 3.14 B). This trend is even more obvious when considering the absolute number of monoallelic nuclei (Figure 3.14 Ci). The embryo regions (total number of nuclei in field of view) analysed in *wt* and *sog*$^{-}$ embryos are similar in size and the proportion of monoallelic nuclei is increased in *sog*$^{-}$ embryos when quantified in the whole embryo region (Figure 3.14 Cii). The increase in monoallelic transcription would be consistent with low BMP signalling in the whole expression domain simulating low signalling levels at the wt expression domain edge. *sog*$^{-}$ embryos only produced a maximum of 117 transcripts per cell and therefore half as much as the *wt* level (243, Figure 3.14 D). The *sog*$^{-}$ embryos show the widest expression domain at 141 $\mu$m with a wide peak representing more uniform low levels of BMP signalling (Figure 3.14 D).

### 3.4.2 ush transcription profile

The *ush* transcription profile was next analysed in embryos with increased or decreased levels of BMP signalling. First, the sensitivity of *ush* transcription to changes in *dpp* dosage was investigated. Embryos carrying the $dpp^{Hin37}$ allele contain a molecular lesion in the haploinsufficiency (hin) region of the *dpp* gene (Irish & Gelbart 1987). This mutation leads to haploinsufficiency and embryos of the genotype $dpp^{Hin37}/+$ experience embryonic lethality (Irish & Gelbart 1987). Therefore, the $dpp^{Hin37}$ *Drosophila* line contains an extra transgene copy of *dpp* on the wildtype allele to ensure stock viability (Irish & Gelbart 1987). Females of this stock were mated with wildtype males, resulting in embryos with either one (*1x dpp*) or three *dpp* copies (*3x dpp*).

The *1x dpp* embryos are haploinsufficient, but lethality does not occur until late embryonic stages. Hence, this cross was used to study *ush* transcription in *1x dpp*

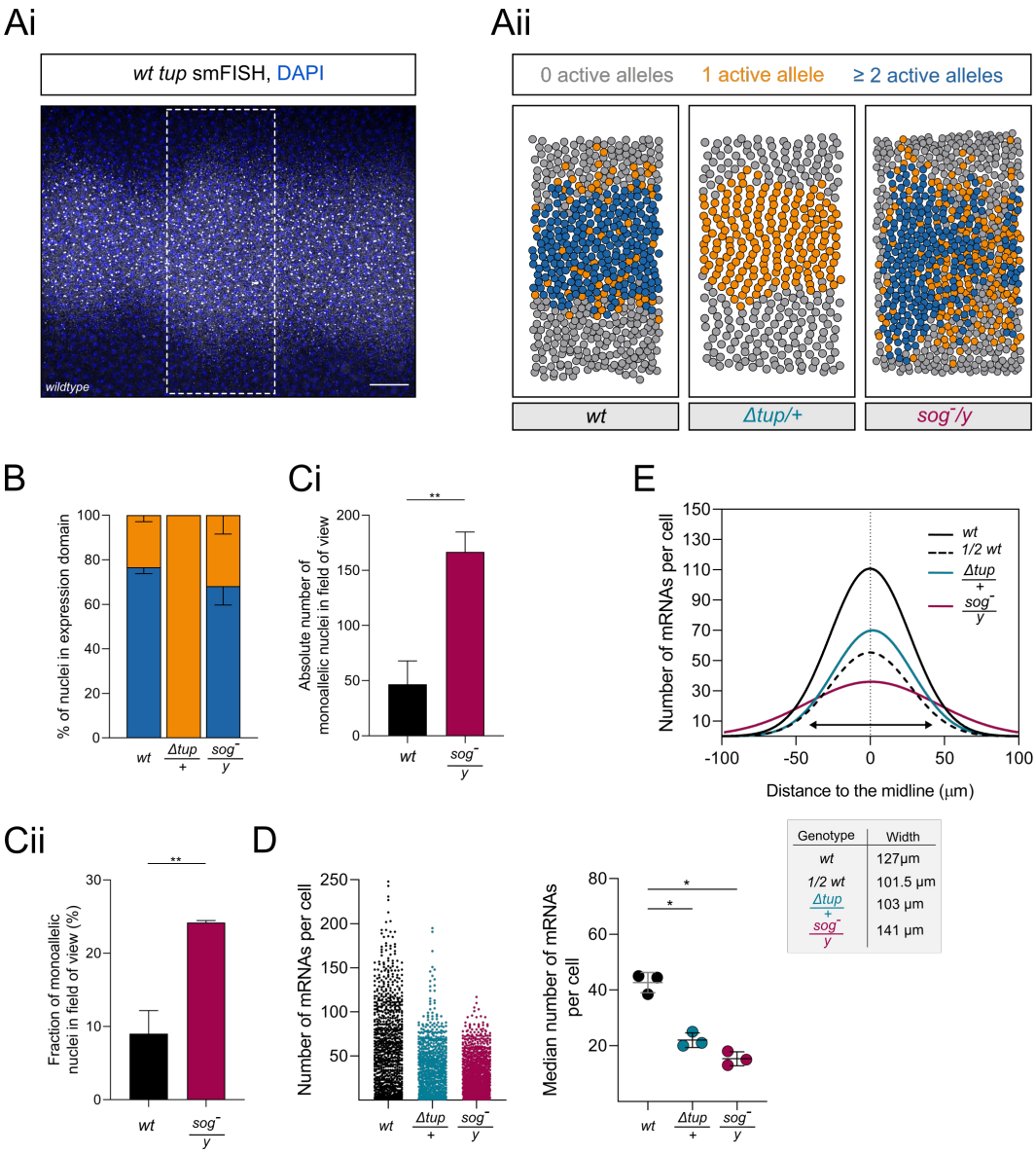



**Figure 3.14: *tup* transcription profile changes in different genotypes.** (Ai) Example microscopy image of *tup* transcription showing smFISH probes (white), DAPI staining (blue) and highlights the analysis region. (Aii) *tup* expression patterns of representative embryos. Nuclei were coloured depending on their transcription state. (B) Proportion of monoallelic (orange) and biallelic (blue) nuclei in expression domains analysed. (Ci) Absolute number of monoallelic nuclei in wt and *sog*⁻ embryos. (Cii) Proportion of monoallelic nuclei in wt and *sog*⁻ embryos in relation to the full analysis domain. (D) Number of mature *tup* transcripts per cell. Replicates of each genotype are pooled (left). Median number of transcripts per cell from individual replicates (right). (E) Comparison of expression domain shape between genotypes. Individual replicates are pooled to generate Gaussian distribution curves for different genotypes. 1/2 wt curve was calculated based on wt curve. Curve width was measured where curves intersected with a y-value of 10 transcripts per cell and recorded. Scale bar in (Ai) = 20 μm. Mean and SD (B, Ci, Cii, D) shown, n = 3 biological replicates. Significance tested with an unpaired Student's t-test (Ci, Cii) and with a one way ANOVA with Dunnett's multiple comparisons test (D). *p< 0.05, **p< 0.01, ns = not significant. Dorsal view of the embryo is shown in (Ai, Aii) with anterior to the left.

and *3x dpp* backgrounds, which should lead to decreased and increased BMP signalling levels.

In wt embryos, *ush* transcription is observed in a domain around the dorsal midline (Figure 3.15 Ai, Aii). In embryos with *1x dpp* and therefore reduced BMP signalling, very few mRNA molecules are observed in the cytoplasm and the expression domain appears flat (Figure 3.15 Aii). In comparison, the *ush* expression domain in *3x dpp* embryos is broader, and the overall number of cytoplasmic mRNAs appears to be increased, consistent with increased BMP signalling levels and previous RNA ISH studies (Figure 3.15 Aii) (Miles et al. 2008, Ashe et al. 2000). When quantified, the maximum number of mRNAs per cell detected in *1x dpp* embryos is very low at 29, 41 and 64 for three biological replicates compared to the wt control embryo (Figure 3.15 B). Conversely, in *3x dpp* embryos, the number of mRNAs per cell is greatly increased with maximum transcript counts per cell of 526 and 550 (Figure 3.15 B).

The substantial differences in mRNA number between genetic backgrounds are also illustrated by showing the distribution of mRNA number per cell according to the nuclear position across the midline. *1x dpp* embryos show a consistently low number of mRNA molecules per cell across the whole width of the imaging domain (Figure 3.15 C). Biological replicates 1 and 2 show no peak but uniformly low numbers of mRNAs, whereas biological replicate 3 shows a small peak at the dorsal midline smilar to previous observations by conventional ISH (Ashe et al. 2000) (Figure 3.15 C). The slight increase in mRNA number in the third *1x dpp* biological replicate, relative to the other two, possibly represents the variation that can be found in mutant backgrounds (Janssens et al. 2013). In contrast, the *3x dpp* embryo replicates show a broad peak of high transcript numbers per cell at the dorsal midline (Figure 3.15 D), relative to wt (Figure 3.15 E).

The direct comparison of the wt control curve with pooled data from *1x dpp* and *3x dpp* embryos highlights the severe differences in *ush* expression (Figure 3.15 F). The fold change in *ush* transcription output does not scale linearly with *dpp* dose, highlighting the sensitivity of the embryo to changes in *dpp* levels. Little or no signalling gradient is observed in *1x dpp* embryos which produce far fewer *ush* transcripts than half of the wt level (Figure 3.15 F). The transcription profile that is observed for *3x dpp* embryos is much wider than the that observed for wt embryos (Figure 3.15 F). Furthermore, the peak of mRNA number per cell at the dorsal midline is more than twice the height of the wt embryo and therefore, suggests

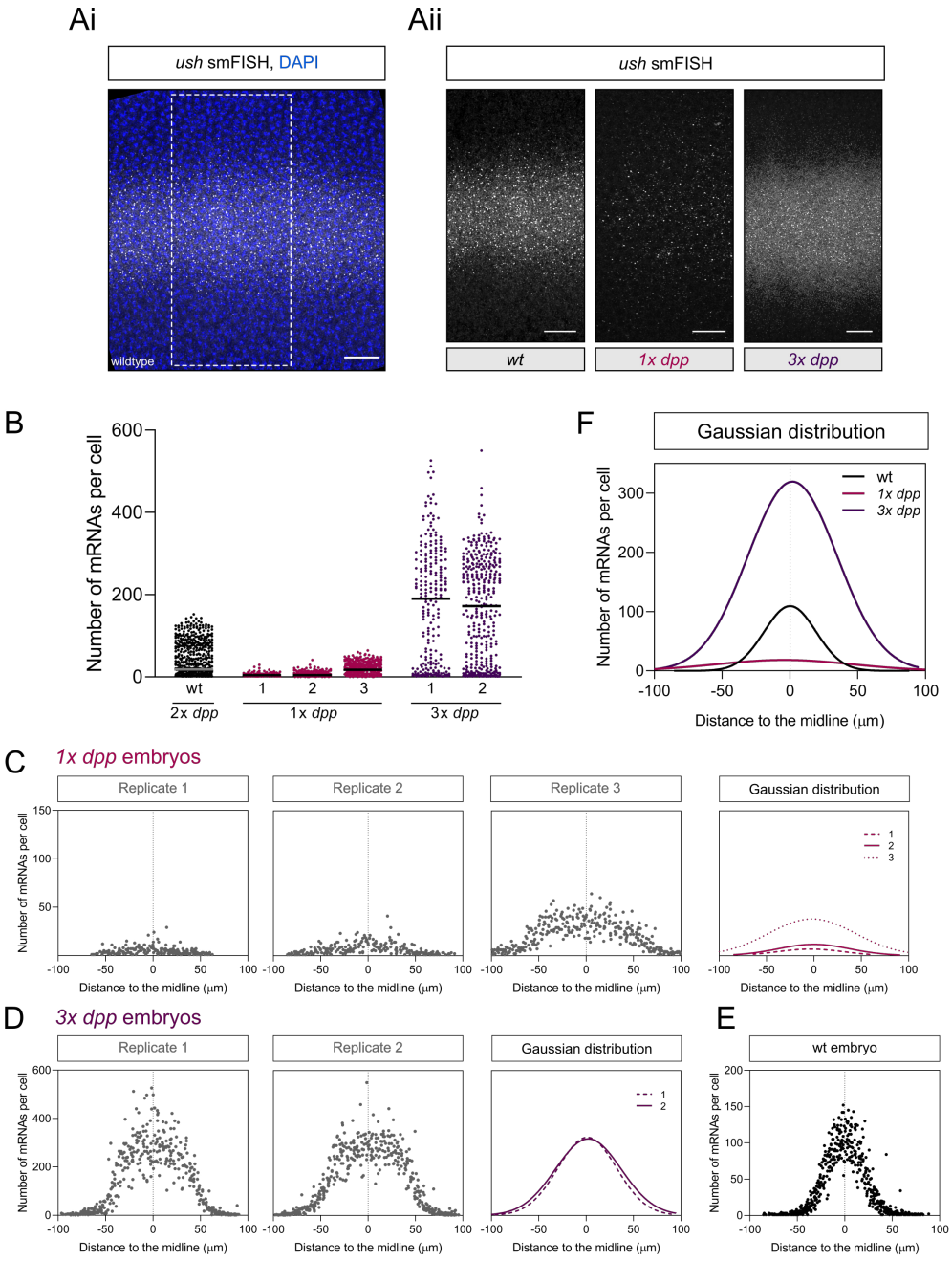



**Figure 3.15: High sensitivity of *ush* transcription to changes in *dpp* dose.** (Ai) Microscopy image of *ush* transcription in a wt embryo, showing exonic smFISH probes (grey), DAPI staining (blue) and the analysis region is highlighted with the white box. (Aii) Enlarged region of the wt embryo (Ai) and representative regions from *1x dpp* and *3x dpp* embryos showing exonic *ush* transcription (grey). (B) The graph shows *ush* transcription output in wt (black), *1x dpp* (pink) and *3x dpp* embryos (purple). Three biological replicates of *1x dpp* (C) and two replicates of *3x dpp* (D) embryos were analysed and their mRNA counts are shown here across the midline. (E) A representative wt embryo is shown for comparison. Data obtained from *ush* time course experiment. (F) Pooled transcription profiles from all genotypes are shown as Gaussian curves for direct comparison.
Scale bar in (Ai, Aii) = 20 μm. Median shown in (B), n = 3 biological replicates for *1x dpp*, n = 2 for *3x dpp* and n = 1 for wt. Dorsal views of the embryos are shown in (Ai, Aii) with anterior to the left.

that BMP signalling levels are not directly proportional to *dpp* dosage. Instead, this observation suggests that an additional feedback mechanism exists that increases Dpp signalling. A positive feedback mechanism during BMP signalling has previously been shown, where future Dpp binding to receptors is promoted by the previous signalling strength (Wang & Ferguson 2005). The effect of increased BMP signalling levels on the proportion of monoallelic and biallelic transcription will be investigated later in this chapter in response to increased ectopic *dpp* expression (see *st2-dpp* Section).

Next, this study investigated how reduction in Smad activator concentration alters *ush* transcription. *ush* transcription was analysed in embryos from a cross of females heterozygous for the $Med^{13}$ null allele (nonsense point mutation in *Med* locus) (Hudson et al. 1998), with wt males. These embryos (here referred to as $Med^{13}$) received reduced levels of maternally deposited Med, which leads to reduced BMP signalling transduction (Hudson et al. 1998).

Analysis of a section of the *ush* expression domain in $Med^{13}$ embryos of different ages (Figure 3.16 A) shows a decrease in the maximum *ush* output per cell compared to wt embryos (Figure 3.16 B). The embryos representing early and middle time points of *ush* transcription in nc14 show reduced maximum transcription counts of 43 and 112 mRNAs per cell, compared to wt embryos of similar age which showed maximum outputs of 86 and 152 mRNAs per cell (Figure 3.16 B). The reduction in *ush* transcription suggests that transcription output is sensitive to the reduction in BMP signalling levels. Interestingly, the oldest embryos in the wt and $Med^{13}$ time course show similar levels of *ush* transcription output, suggesting compensation at this stage during nc14, (Figure 3.16 B).

Next, the *ush* transcription profiles of all embryos were plotted according to the nuclear position along the embryo cross-section. The graph reveals that *ush* transcription in $Med^{13}$ embryos is reduced during the early and middle transcription period in nc14 (Figure 3.16 C). The reduction is most severe in the peak height around the dorsal midline (Figure 3.16 C). As already observed in Figure 3.16 B, the transcription profile of a $Med^{13}$ embryo during late stages of nc14 appears almost identical to the wt embryo of similar age (Figure 3.16 C). These observations suggest that *ush* transcription is most sensitive to Smad levels particularly at the two early time points.

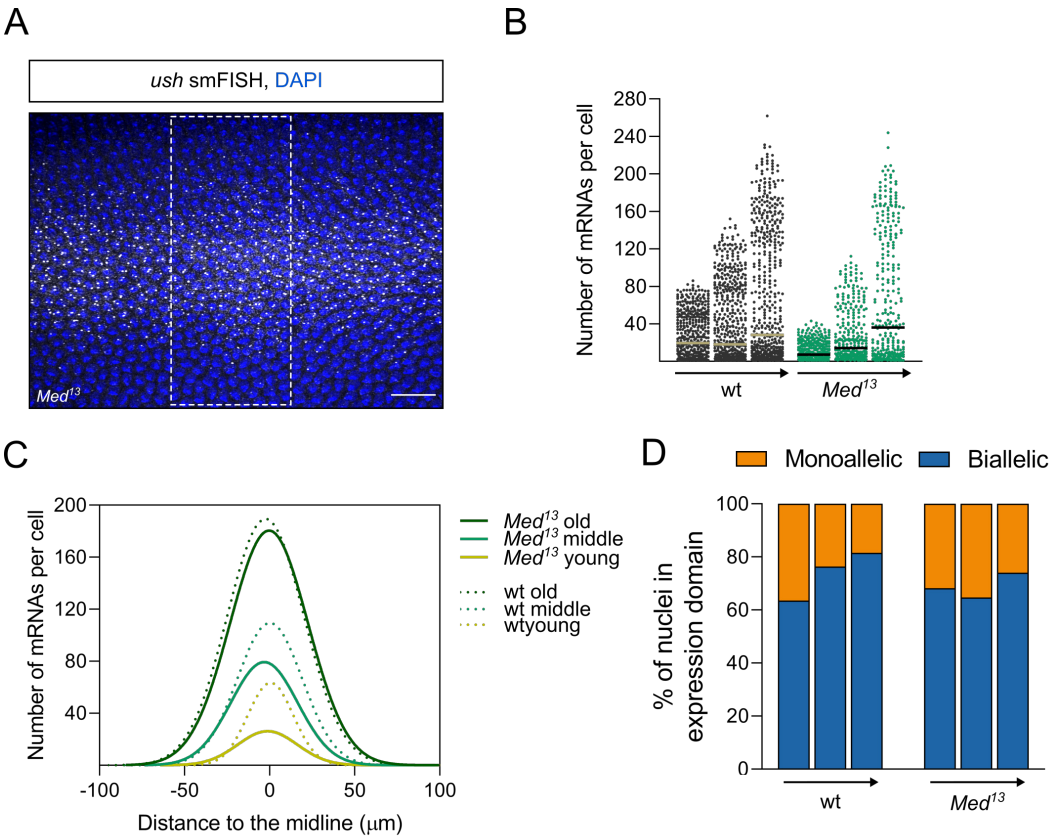



**Figure 3.16: *ush* transcription profile changes during reduced BMP signalling.** (A) Example microscopy image of *ush* transcription in *Med*$^{13}$ heterozygous embryos, showing smFISH probes (white), DAPI staining (blue) and the analysis region is highlighted with the white box. (B) The graph shows *ush* transcription output in wt (black) and *Med*$^{13}$ embryos (green). Three embryos for each genotype were analysed. Each embryo represents a different time point in nc14. Time points were matched between wt and *Med*$^{13}$ embryos. (C) Transcription profiles of all embryos are shown as Gaussian curves according to the nuclear position along the embryo cross-section. *Med*$^{13}$ embryos are shown as solid lines and wt embryos of similar age in dotted lines. (D) Proportion of monoallelic (orange) and biallelic (blue) nuclei in expression domains analysed.
Scale bar in (A) = 20 µm. Median shown in (B), n = 1 biological replicate for each time point. Dorsal view of the embryo is shown in (A) with anterior to the left.

Previous data obtained in this chapter suggested that monoallelic nuclei are a result of reduced activator levels. If this is the case, more monoallelic nuclei would be expected in *Med*$^{13}$ embryos with reduced levels of Med. Apart from during transcription onset, *Med*$^{13}$ maternal heterozygous embryos show higher monoallelic transcription than in wt (32% and 35%, Figure 3.16 D). At the end of nc14 the level of monoallelic transcription decreases slightly to 26%, consistent with the mRNA numbers at this stage being similar to that of wt (Figure 3.16 D). Compensation at the later stage could potentially be due to accumulation of zygotic Med protein (see discussion). The data are consistent with reduced Med levels but to confirm these preliminary results more biological replicates will have to be analysed in the future.

Previous experiments in this chapter have suggested that reduced levels of nuclear Smad activator at the expression domain edge lead to increased proportions of monoallelic transcription. To further investigate the relationship between monoallelic transcription and BMP signalling levels, ectopic *dpp* was introduced into the embryo. Ectopic expression of a *st2-dpp* transgene (expressing *dpp* under the control of the *even skipped stripe 2* enhancer) increases the width of *ush* transcription (Ashe et al. 2000) (Figure 3.17 Ai). If monoallelic transcription results from low levels of activator, increasing BMP signalling levels, and thereby enhancing signal, should result in more biallelic transcription at a position equivalent to the wt domain border. The *st2-dpp* transgene expands the whole BMP signalling domain but we focused on the area of greatest expansion corresponding to the st2 expression domain (Ashe et al. 2000). The hypothetical wt edge was identified using the typical expression domain width centered at the midline. The transcription state of nuclei in the area that represents the wt edge (orange rectangle) was compared to nuclei at the "new" outer edge of the *st2-dpp* expression region (red rectangle) (Figure 3.17 Aii). When quantified, the percentage of monoallelic nuclei in the region that represents the wt edge of *ush* expression is significantly lower than the actual expression domain edge generated by *st2-dpp* expression (Figure 3.17 B). The wt domain showed little monoallelic expression at 13% and therefore lower than a wt embryo, suggesting that the increased levels of BMP signalling in this domain increased biallelic transcription. This supports the conclusion that monoallelic transcription observed in wt embryos is most likely a result of low levels of nuclear activator, which are insufficient to initiate transcription on both alleles with high frequency.

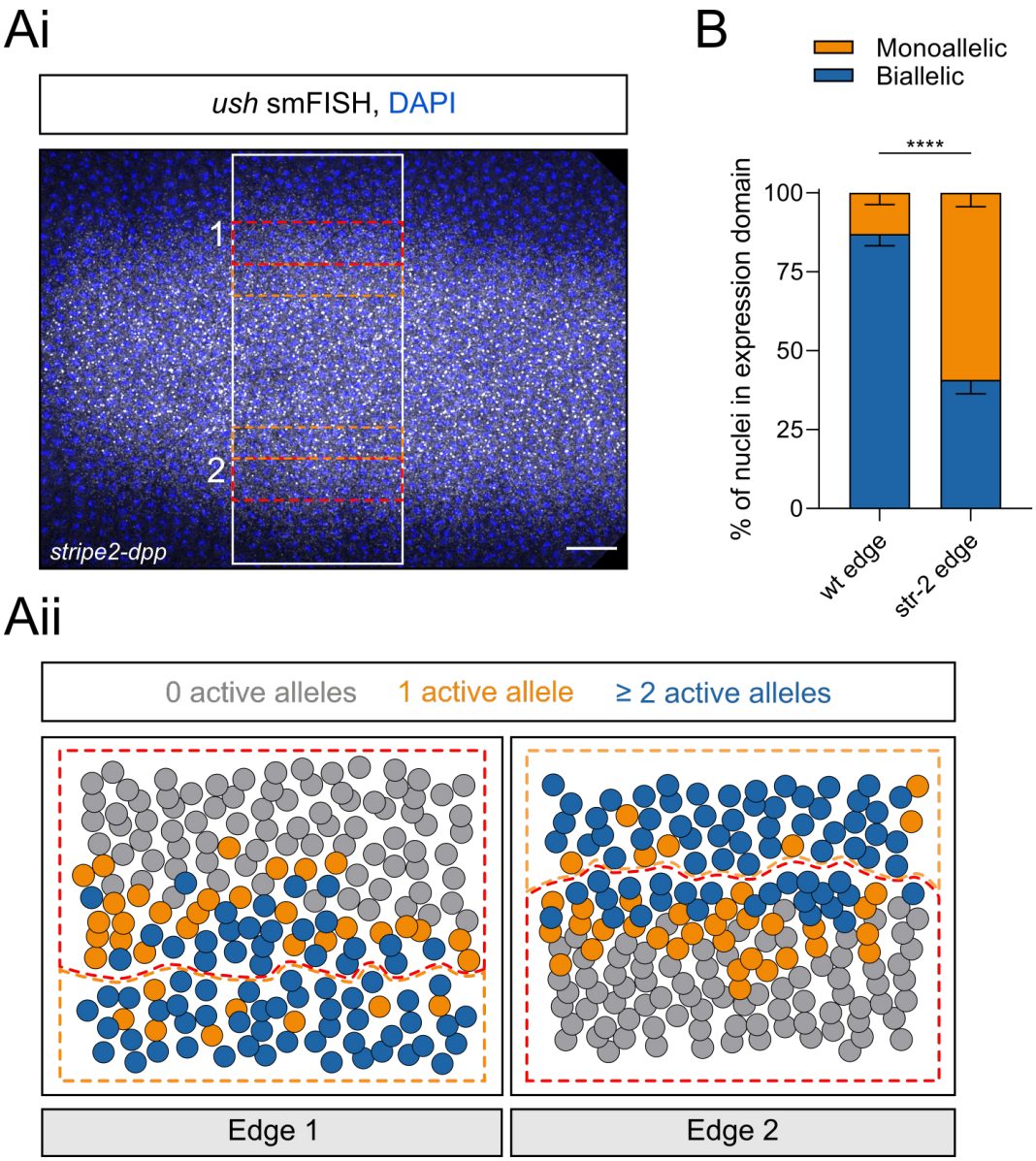



**Figure 3.17: Monoallelic *ush* transcription results from low activator concentration.** (Ai) Example microscopy image of *ush* transcription in a *st2-dpp* embryo, showing smFISH probes (white) and DAPI staining (blue). The white rectangle highlights the analysis region, the orange rectangles show the wt expression domain edges while the red rectangles show the expression domain edges in the *st2-dpp* expression domain. (Aii) *ush* expression patterns of edge regions highlighted in (Ai) are shown. Nuclei were coloured depending on their transcription state and the wt (orange) and widened *st2 dpp* (red) expression domains are indicated. (B) Proportion of monoallelic (orange) and biallelic (blue) nuclei in expression domains analysed. Data from wt edge (orange) and *st2 dpp* edge (red) in (Ai).
Scale bar in (Ai) = 20 μm. Mean and SD (B) shown, n = 3 biological replicates with 2 expression domain edges analysed per embryo. Significance tested with an unpaired Student's t-test between the proportions of monoallelic nuclei (B). ****p< 0.0001. Dorsal view of the embryo is shown in (A) with anterior to the left.

## 3.5 Discussion

Experiments described in this chapter aimed to investigate the transcription profiles of BMP target genes and revealed the following:

1. Transcription of BMP target genes occurs monoallelically at the outer edges of the expression domains.

2. Monoallelic transcription appears to be a general phenomenon during gradient interpretation.

3. Transcription profiles of *hnt*, *ush* and *tup* mirror the BMP signalling gradient in their transcriptional output.

4. Nuclei positioned in a five cell-wide region around the dorsal midline contribute to more than 80% of the overall transcriptional output of the BMP target genes *hnt*, *ush* and *tup*.

**Monoallelic transcription in response to limiting activator concentration**

Using RNA-FISH this study identified a subset of nuclei that show monoallelic transcription for all the BMP target genes that were analysed. In these nuclei, only one active transcription site was identified, compared to the two nascent transcription sites observed during biallelic expression. This observation is unexpected as gene expression in diploid organisms usually occurs through equal expression of the two homologous gene alleles. In some cases however, as observed in this study, transcription can occur monoallelically.

Imprinting is a prominent example of monoallelic gene expression, where either the maternal or paternal gene allele is expressed throughout a tissue or the whole organism (Reik & Walter 2001). A broader class of monoallelic gene expression is random monoallelic gene expression, where a random allelic choice occurs. During X-chromosome inactivation in mammalian females, one X-chromosome is transcriptionally silenced at random in order to regulate gene dosage (Erwin & Lee 2008). In contrast, random autosomal monoallelic gene expression can be found throughout the genome, and in some cases results from a single nucleotide polymorphism which can lead to disease (Bjornsson et al. 2008). Random monoallelic expression can be fixed, where allelic expression is conserved in the daughter nuclei after mi-

totic division (Gimelbrant et al. 2007), or dynamic, where allelic expression is temporal and not mitotically transmitted (Deng et al. 2014). Classic examples in the literature for monoallelic transcription include the expression of immunoglobulin-domain proteins (Pernis et al. 1965) and olfactory receptors (Chess et al. 1994). In both cases, allele-specific expression contributes to allotype variability in gene expression.

The form of monoallelic gene expression observed here for BMP target genes is most likely random and temporal since both biallelic and monoallelic transcription are observed together in the same expression domain. This study shows that nuclei with a single transcription site are found significantly further away from the embryo midline, and therefore closer to the expression domain edge, than nuclei with biallelic transcription. This asymmetric distribution suggests an underlying cause for the increased occurrence of monoallelic transcription in nuclei at the edge of the expression domain. Since BMP target genes directly respond to BMP signalling levels, it can be hypothesised that monoallelic transcription is a direct result of reduced activator concentrations in nuclei at expression domain edges. This hypothesis is further supported by the reduced proportion of monoallelic nuclei at the wt *ush* expression border in the *even skipped* st2 area with ectopic *dpp* expression. Ectopic *dpp* increased the BMP signalling levels and *ush* expression domain width, thereby moving the area of monoallelic nuclei further away from the midline.

Most research on monoallelic gene expression has been performed in human and murine cells (Reinius & Sandberg 2015), but some examples exist in *in vivo* models (Gimelbrant et al. 2007, Deng et al. 2014, Gendrel et al. 2014). This study proposes that monoallelic transcription not only exists in BMP target genes in *Drosophila* but that it is a global phenomenon during gradient interpretation. Monoallelic transcription was shown here to also occur in the Dl target genes *sog, brk* and *sna*. These genes were found to exhibit similar levels of monoallelic gene expression in edge domains as BMP targets. Since Dl acts as a morphogen analogous to Dpp, this observation suggests it is a general feature of signalling gradient readout. If this is the case monoallelic transcription should be present in other genes responding to gradients, in domains where there are low levels of activator.

Even though monoallelic transcription in *Drosophila* has not been investigated in detail, examples can be identified in the literature. Several studies have investigated the expression pattern of zygotic *hunchback* transcription using smFISH or RNA-FISH. Monoallelic expression of *hunchback* is visible in all these studies at the

expression domain border, similar to the monoallelic gene expression observed for BMP and Dl target genes in this study (Porcher et al. 2010, Little et al. 2013, Zoller et al. 2018, Little & Gregor 2018). These publications, however, do not discuss this observation. As zygotically expressed *hunchback* is activated in response to a morphogen gradient of Bicoid protein (Struhl et al. 1989), the presence of monoallelic gene expression in these studies further support the hypothesis that monoallelic transcription and hence low initiation frequency is a consequence of reduced activator levels.

Based on the results from this study, and observations from the literature, similar transcriptional behaviour for other genes activated by concentration gradients would be predicted. Consistent with this prediction, monoallelic expression was also observed at the edges of Dl target genes *sog* and *brk* and an increase in the proportion of monoallelic nuclei was found at the dorsal border of both *sog* and *brk* expression domains where lower levels of Dl activator are present. Using a live imaging approach it was found that transcription of the two *sna* alleles can be repressed individually (Esposito et al. 2016). Hence, repression of *sna* alleles at different times would produce a more interspersed pattern of monoallelic expression, as can be observed in this study. Furthermore, the presence of monoallelic nuclei at the ventral border of *sog* and *brk* expression domains, in the presence of high levels of Dl, can be explained by allele by allele repression.

**BMP signalling in mutant backgrounds**

In its dynamic form, monoallelic expression has been suggested to be a result of independent stochastic activation of alleles (Eckersley-Maslin & Spector 2014). Accordingly, reduced levels of nuclear Mad/Med activator complexes at the expression domain border would result in a reduced probability of initiating transcription or activation of shorter bursts and conversely the increased occurrence of monoallelic transcription. In agreement, this study shows that embryos with reduced levels of Med protein show increased levels of monoallelic transcription in the *ush* expression domain during early stages. The *ush* mRNA numbers per cell produced in embryos from heterozygous $Med^{13}$ mothers are reduced shortly after transcription initiation but reach wt levels at later stages, suggesting compensation. It is possible that compensation occurs through increased Med levels that arise from accumulation of zygotically produced protein. Zygotic production of Med could thereby overcome the reduction in the maternal Med load. The transcription output of *ush* in embryos heterozygous for *dpp* was also shown to be reduced. *ush* transcription in

these embryos was identified to be reduced to an average of 10 transcripts per cell. It can be predicted that transcription of other BMP target genes would be affected with similar severity. The production of mRNA numbers that are so low suggests a molecular explanation for the haploinsufficiency observed in these embryos (Irish & Gelbart 1987).

BMP signalling levels in the absence of Sog protein are uniformly low due to the loss of Dpp shuttling to the dorsal midline (Francois et al. 1994, Biehs et al. 1996). Quantification of *tup* transcription in *sog*$^{-}$ embryos revealed low numbers of mRNAs per cell and a broader expression domain. These data visualise a quantitative BMP target gene transcription output in a *sog*$^{-}$ background for the first time. The broadening of the *tup* expression domain is caused by the loss of Sog mediated repression of BMP signalling in lateral domains (Ray et al. 1991, Francois et al. 1994). Modelling of the BMP signalling gradient in *sog*$^{-}$ embryos predicted uniformly low levels of signalling in the dorsal ectoderm (Mizutani et al. 2005). In agreement, a classical non-quantitative ISH study showed uniform and broadened expression of *dpp* in *sog*$^{-}$ embryos (Francois et al. 1994). Therefore, according to the literature, uniform levels of *tup* expression would be predicted. This study, however, observes a shallow peak with slightly increased levels of *tup* transcription at the dorsal midline in the absence of Sog. This observation could be explained through an inverse Brk repression gradient. Brk has been shown to repress *dpp* expression in lateral domains following Dl repression of *dpp* earlier in embryogenesis (Jaźwińska et al. 1999). A shallow inverse gradient of Brk has been proposed by Ashe et al. (2000), which would reduce Dpp levels and therefore BMP signalling slightly in lateral domains, even in the absence of Sog. Brk has also been suggested to repress *tup* transcription (Ashe et al. 2000). Therefore, repression of *dpp* and *tup* by low Brk levels in the dorsal epidermis adjacent to the neuroectoderm could explain the lower *tup* mRNA number per cell at the expression domain border in *sog* mutant embryos.

Besides the decrease in the overall *tup* mRNA number in *sog*$^{-}$ embryos, an increase in monoallelic expression is identified. This observation is consistent with *sog* mutants having low levels of BMP signalling, similar to that near the expression domain edge in wt embryos. An area of biallelic expression is identified within the broadened *tup* expression domain. It can be hypothesised that biallelic nuclei are a result of a positive Tsg function that acts independently of Sog (Wang & Ferguson 2005). Data obtained from epistasis analysis in mutant embryos suggests that Tsg promotes Dpp- receptor binding in the absence of Sog and thereby enhances

signalling independently of Sog (Wang & Ferguson 2005). The area of biallelic nuclei appears to coincide with the spatial Tsg pattern and therefore could explain the occurrence of biallelic transcription in the low BMP concentration environment through enhanced binding to receptors.

**Heterogeneous transcription profiles of BMP target genes mirror BMP signalling**

This study revealed considerable heterogeneity in the transcriptional output of BMP target genes. Previously, BMP target gene expression domains have been investigated using traditional ISH, which is not quantitative (Ashe & Levine 1999, Ashe et al. 2000). Therefore, heterogeneity in transcriptional output had not been observed. Results from this study reveal that the domains of the BMP target genes *hnt, ush*, and *tup* display sharp gradients of mature cytoplasmic transcripts, with peak levels at the dorsal midline. Additionally, time course data showed that the maximum level of accumulated cytoplasmic transcripts increases with age. The mRNA half-life of *hnt, ush* and *tup* has not been experimentally determined but Lagha et al. (2013) proposed that half-lives of mRNAs expressed in the early embryo are likely to fall in the range of 5–15 min. With a proposed short half-life, it can be estimated that most of the cytoplasmic RNA profile results from *de novo* RNA synthesis.

Strikingly, the number of mRNAs produced in the centre of the respective BMP target gene expression domains is 10-12 fold greater than at the edges. These results extends evidence from an imaging study which investigated the transcriptional output of gap genes during *Drosophila* embryogenesis. Using smFISH, this study highlighted spatial gradients of transcription along the AP axis of the developing embryo. They found that for example *hunchback, krüppel* and *giant* transcription shows similar absolute levels of cytoplasmic mRNAs to each other and that their transcription output occurs in gradients similar to the observed BMP target gene transcription (Zoller et al. 2018). Together, these data indicate that transcription occurs in gradients in response to AP or DV input signals, leading to heterogeneous expression patterns and cell-cell variability.

Examples of cell-cell variability have been reported in other systems, including mice (Itzkovitz et al. 2012), yeast (Zenklusen et al. 2008, To & Maheshri 2010), *Drosophila* (Paré et al. 2009) and mammalian cell culture (Raj et al. 2006). Together these studies reported a transcription heterogeneity of 60% –300%. It is important to note, however, that not all genes during early embryogenesis adhere

to this mode of transcription. *sna* transcription, initiated by peak levels of Dl, was shown to be homogeneous throughout the expression domain with a reported variation between 10% –12% (Boettiger & Levine 2013). This study concluded that the small variation in gene expression of *sna* could be explained by a high RNA synthesis rate and the short time the promoter spends in the "OFF–position" (Boettiger & Levine 2013). It is thought that longer periods of promoter inactivity increase the level of noise and hence cell-cell variability (Raser & O'Shea 2005, Raj et al. 2006, Munsky et al. 2012). Additionally, weak autorepression was suggested to occur during *sna* transcription. This would mean that cells with high levels of cytoplasmic mRNA experience negative feedback to reduce mRNA levels, and that cells with fewer mRNA numbers are de-repressed to increase transcription. This way all cells within the expression domain would produce similar mRNA levels (Boettiger & Levine 2013).

The BMP target gene outputs mirror the BMP signalling gradient, with peak transcription levels at the dorsal midline. As such, it can be suggested that single cells read out BMP signalling levels, not in a digital fashion but in an analogue one. A digital mode of transcription would suggest that nuclei respond to a threshold and once activated, behave similarly. An analogue mode of transcription would suggest that cells can read out the signal concentration and react in different ways (Falo-Sanjuan et al. 2019). A large variation in transcriptional output and the differences in nuclear dynamics across the midline of BMP target genes suggests analogue control of gene expression.

This study also investigated the contribution of individual cell rows to the total transcriptional output. Here, the gradient of BMP signalling is visible. By combining data from multiple biological replicates, a plateau of transcript number in a 4-cell radius around the midline was discovered. This area corresponds to peak BMP signalling, inferred from pMad levels, before a significant drop in signalling further away from the midline (Dorfman & Shilo 2001, Xu et al. 2005, Sawala et al. 2015). When quantified the five-cell radius around the midline contributed to more than 60% of the total transcriptional output of BMP target genes. The large mRNA counts in the middle of the expression domain could produce a buffering effect, and has previously been suggested to protect against fluctuations in parameters over a physiological range to ensure correct cell fate decisions (Little et al. 2013). Overproduction of mRNA could also buffer against fluctuations in translation that have been observed in other systems (Bar-Even et al. 2006, Newman et al. 2006, Taniguchi et al. 2010).

Transcript numbers observed in this study showing peak outputs of 222, 262 and 189 transcripts per cell for *hnt*, *ush* and *tup* respectively, are within the dynamic range of transcript numbers determined by other studies. The homogeneous expression domain of *sna* was reported to generate an average of 200 transcripts in late nc13 and 150 transcripts in nc14 (Boettiger & Levine 2013), whereas *hunchback* was described to contain a range of 100 –500 transcripts per cell (Little et al. 2013). Average mRNA counts for *Sex comb reduced*, *Deformed* and *Ultrabithorax* were shown to be 94, 92 and 74 respectively (Paré et al. 2009). All studies together suggest a dynamic range of 50 –500 transcripts per cell for early patterning genes.

An alternative method to study single-cell transcriptomics is Nano-String. The Nano-String system quantifies the number of mRNAs directly without reverse transcriptase or the need to fragment (Geiss et al. 2008). Using Nano-String, a study by Sandler & Stathopoulos (2016) generated single embryo time course data, quantifying 68 genes during *Drosophila* embryogenesis. The relative order of expression profiles for *hnt, ush*, and *tup* is consistent with data generated in this study. Accordingly, *hnt* produces the lowest and *ush* the highest number of transcripts when the whole expression domain is combined. The sections of the *ush* expression domains that were analysed in this study encompassed approximately one third of the total expression domain. Therefore, the approximated absolute number of *ush* mRNAs produced in the oldest time course embryo is ~ 112,000. The absolute number of detected *ush* transcripts detected using Nano-String sequencing, however, is approximately six-fold higher than the estimated values from this study. Similarly, Sandler & Stathopoulos (2016) report that their detection of *sna* transcription is two–to five-fold higher than the smFISH data generated by Boettiger & Levine (2013). Sandler & Stathopoulos (2016) suggest that the differences in absolute number arise from microscopy and image analysis errors, as their transcription counts correlate to qPCR data which is more accurate. They further suggest that thresholds and the dense fluorescence signals observed in smFISH might complicate the analysis (Sandler & Stathopoulos 2016). However, it is also possible that calibration of the Nano-String data is not entirely accurate.

The possible underestimation of mRNA number by quantitative FISH techniques was also shown by expansion microscopy. During expansion microscopy samples are physically enlarged using acrydite integration into a polymer network which, after further treatment, expands. The physical expansion of samples results in an increased resolution and enables the visualisation of particles that were previously

below the diffraction limit (Chen et al. 2015). When expansion microscopy was combined with smiFISH probes targeting CRM1 transcripts in HeLa cells, the nuclear area was shown to be expanded 14-fold and single molecule counts were shown to be increased two-fold (Tsanov et al. 2016). Use of expansion microscopy further demonstrates that mRNA detection by smFISH or smiFISH is likely to underestimate the actual number of mRNAs present per cell. However, even the two-fold increase in mRNA number observed after expansion microscopy does not explain the much higher estimate by Nano-String.

Despite the possibility that smFISH likely underestimates the transcriptional output it still is a valid experimental technique to investigate spatial and temporal aspects of transcription. All spatial information from tissues is lost during genomic approaches, such as Nano-String, and therefore smFISH has an advantage over these techniques. Furthermore, as the relative transcription levels between genes are similar when comparing Nano-String and smFISH data both techniques are valid to study transcriptional output.

**Transcription dynamics and stochasticity**

As discussed above, monoallelic transcription most likely corresponds to reduced transcription initiation frequency in nuclei experiencing lower levels of activator. Hence, transcription occurs dynamically. A further indication of dynamic transcription of BMP target genes was the detection of a broad range of mRNAs in transcription sites. The differences in fluorescence suggest variation in the number of polymerases transcribing along genes, potentially resulting from changes in transcriptional burst kinetics. It can be hypothesised that nuclei positioned at the expression domain border would contain less intense fluorescent transcription sites and therefore, fewer mRNAs. This hypothesis could be tested in the future by correlating the nuclear fluorescence intensity to the nuclear position.

The calculation of total nascent mRNA number in this study lacks absolute accuracy and therefore can only be used for mRNA number estimation. This is due to the presence of incomplete mRNAs within the transcription site. All smFISH probes might not be able to bind unfinished transcripts, resulting in a lower fluorescence than a mature RNA. As an estimate, the number of mRNAs present at the active transcription site should be indicative of the number of elongating polymerases. SmFISH probes detect mRNAs that are stably associated with Pol II molecules, and it is assumed that terminated transcripts diffuse away from the site of transcrip-

tion. The footprint of an engaged Pol II complex is ~80 nt (Boettiger & Levine 2013). However, Pol II kinetics, including gene body pausing and loading rate will vary the Pol II density on a gene (Neuman et al. 2003). Transcript lengths for BMP target genes are *hnt* ~8 kb, *ush* ~16 kb and *tup* ~20 kb. Hence, more Pol II molecules should be able to fit on the genomic region of *tup* compared to *hnt*. This prediction is consistent with experimental data from this study which identified that the maximum number of mRNAs detected in nascent transcription sites was 105 for *tup*, compared to 61 for *hnt*.

In summary, this chapter has identified dynamic transcriptional behaviour for the BMP target genes *Race, hnt, ush* and *tup*. Monoallelic transcription, predominantly at the expression domain edges, suggests dynamic transcription of BMP target genes in those regions. The transcriptional output of cells has been found to mirror the BMP signalling gradient and this suggests an analogue mode of transcription. The large cell-cell variation in transcript number that has been observed, summarised by the fano factor, suggests that BMP target genes exhibit transcriptional bursting. Transcriptional bursts are known to produce a greater variability of transcripts compared to continuous transcription (Munsky et al. 2012). The calculated fano factors for *hnt*, *ush* and *tup* have an average value of ~40 and the high fano factor therefore suggests that BMP target genes experience transcriptional bursting. The experimental methods used in this chapter provide snap-shot data at the time of embryo fixation and therefore are only indicative of transcription dynamics at that point in time. In order to investigate transcriptional burst kinetics further, and to investigate changes that occur based on BMP signalling levels, live imaging approaches are needed and this will be the focus of the next chapter.

# 4 | Live imaging reveals differences in burst kinetics

## 4.1 Introduction

The previous chapter has identified that BMP target gene transcription is heterogeneous and occurs in a gradient. Furthermore, the dynamic nature of BMP target gene transcription was shown by identifying monoallelic nuclei at the edge of expression domains where low levels of activator are present. Furthermore, large fano factors were observed for BMP target genes. Fano factors higher than one indicate transcriptional bursting. Bursting describes dynamic transcription that occurs in non-continuous burst that are separated by periods of inactivity (Elowitz et al. 2002, Chubb et al. 2006, Fukaya et al. 2016).

Static image analysis reflects snapshot data from when the tissue was fixed. *Drosophila* embryos develop rapidly, and cellular transcription profiles might change on the order of seconds. Therefore, *in situ* data should be complemented with live imaging to encompass the full transcriptional process and activation dynamics. Live imaging systems offer powerful non-invasive tools to study gene expression on a molecular level. Adding a dynamic dimension to imaging enables the holistic view of transcription and substantially increases the density of information that is collected. The MS2/MCP imaging system has been successfully implemented in the *Drosophila* embryo, and previous studies have highlighted the importance of collecting live data. For example, transcription regulated by the *even-skipped* stripe 2 enhancer was visualised in *Drosophila* embryos and the burst activity was found to be much shorter than predicted based on preceding fixed embryo data (Bothma et al. 2014).

This chapter will investigate the transcription dynamics of BMP target genes *hnt* and *ush* using the MS2/MCP live imaging system. By characterising the mRNA production kinetics, insight will be gained into how genes respond to different BMP thresholds and modulate transcription output according to their position within the gradient. Furthermore, differences in signalling readout between the high threshold response gene *hnt* and the low threshold response gene *ush* will be investigated.

## 4.2 Optimisation of MS2 live imaging

One major challenge using the MS2/MCP live imaging system is the high level of background fluorescence that originates from unbound MCP-GFP molecules. The low signal-to-noise ratio complicates data analysis and limits the detection threshold of low expressing nuclei. Furthermore, no standardised analysis pipeline exists in the field of *Drosophila* research to extract data from live imaging time-lapses that were generated using different microscopy setups. Therefore, image analysis challenges are addressed first and the strategies used to overcome the problems described.

### 4.2.1 Image analysis pipeline

No image analysis pipeline existed that was compatible with the time-lapse data produced by this study. Hence the following pipeline was generated and used for all further analysis. The analysis relies on semi-automated segmentation of transcription foci (based on MCP-GFP signal) and nuclei (based on His-RFP signal) (Figure 4.1 A). Segmented nuclei are then tracked through time and analysed by subsequent computational methods using custom python scrips (Figure 4.1 B). This analysis pipeline will be transferable and can be used to extract quantitative information for different genes and using different microscopy setups.

### 4.2.2 Reduction of background fluorescence

Attempts to optimise the signal detection revealed that the background fluorescence from diffuse MCP-GFP, present in the cytoplasm and the nucleoplasm, complicates image analysis. Semi-automated identification of active transcription sites lacks accuracy when background levels are too high. In an attempt to reduce background levels, embryos were imaged with different maternal loads of MCP-GFP. All MCP-GFP protein present in the early *Drosophila* embryo is maternally deposited. The MCP-GFP fusion protein is expressed as a transgene under the control of the *nanos* promoter without a nuclear localisation signal (Garcia et al. 2013). Zygotically transcribed MCP-GFP does not add to the background fluorescence as GFP maturation time is approximately 40 min in *Drosophila* embryos (Hazelrigg et al. 1998, Little et al. 2011, Bothma et al. 2018). Hence, newly synthesised MCP-GFP does not interfere with live imaging during nc14, and tagged transcripts recruit MCP-GFP from the maternal pool.

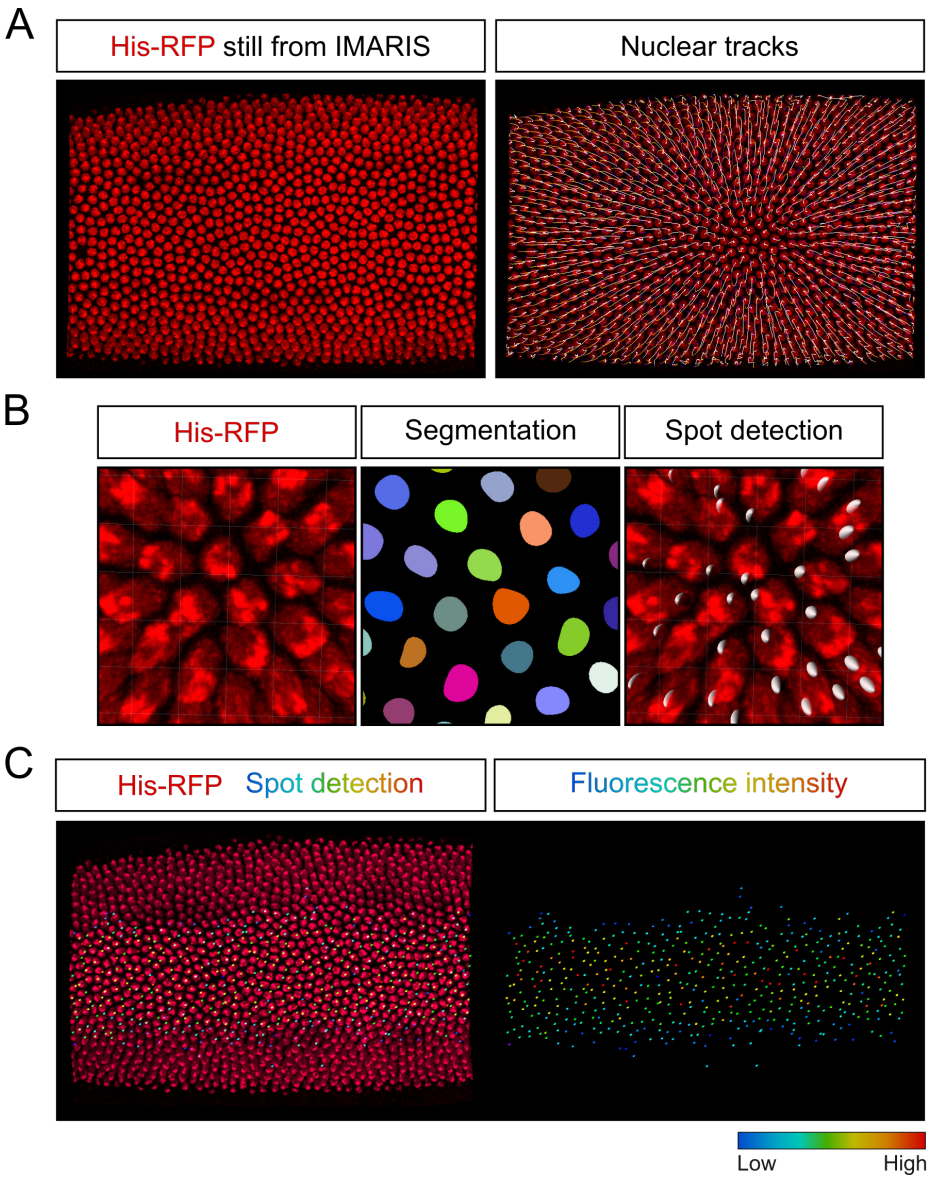



**Figure 4.1: MS2 live imaging analysis pipeline.** (A) Stills from a live imaging data set showing *ush* transcription in a dorsally placed embryo. Nuclei (red) were tracked through time shown here by dragon tails (white) superimposed onto individual nuclei on the right. (B) Enlarged region of a live imaging data set, focussing on a few nuclei visualised by His-RFP fluorescence (red). Nuclei were segmented and coloured based on their nuclear ID. Next, transcription sites were segmented into volumes of equal size (white spheres) based on the MCP-GFP fluorescence (channel not shown). (C) Still from a video showing transcription site detection in IMARIS. Spheres represent transcription sites and are colour-coded based on their fluorescence intensity. The full video can be found at: `https://www.tiny.cc/thesis_videos`
Embryos oriented with anterior to the left.

It can be predicted that lower background fluorescence levels would improve image analysis and spot detection by increasing the signal-to-noise ratio. However, if the available pool of MCP-GFP is too low it might impair imaging results due to insufficient loop binding. In previous studies female virgin flies, homozygous for both *MCP-GFP* and *His-RFP*, were mated with males carrying the MS2 stem-loop locus (for example, Garcia et al. 2013, Bothma et al. 2014). The resulting embryos were used for microscopy. Genetically, these embryos were heterozygous for the imaging locus, and therefore, transcription was observed from one allele. In this scheme, maternal MCP-GFP was produced from two alleles resulting in the maximum deposited load (Figure 4.2 Ai). In an effort to reduce background fluorescence, *;His-RFP;MCP-GFP* virgin flies were first crossed to wildtype flies (*+/+*) and virgins resulting from this cross, heterozygous for the *MCP-GFP* locus, were then crossed to males carrying the imaging MS2 locus in the *ush* 5'UTR (Figure 4.2 Aii). Hence, embryos imaged from this second generation cross carried half the load of deposited maternal MCP-GFP protein.

One embryo each from the first or second generation cross was imaged with subsequent data analysis. Males used in these crosses carried the *ush* imaging locus, which will be described in more detail in Section 4.3. Background fluorescence levels were measured in four randomly selected positions (outside of the active expression domain) at every third time point after *ush* transcription was first observed (Figure 4.2 B). Analysis of the results shows that the background fluorescence was almost twice as high in the embryo with maximal maternal MCP-GFP load. A downward trend in overall background fluorescence is visible, which is most likely explained by photobleaching (Figure 4.2 B).

Next, a linear regression was performed and using the line equation, background levels were determined for each time point. Subsequently, calculated background levels were subtracted from fluorescence values recorded for *ush* transcription sites. The resulting background corrected fluorescent values of *ush* transcription sites are shown in Figure 4.2 C. Background subtraction causes a few fluorescent values to fall below 0 (red data points). The fluorescence value of those nuclei was set to 0 after background correction was implemented as part of the analysis pipeline. Transcription sites below the background level could represent false-positive transcription sites. The range of fluorescent values that were recorded for *ush* transcription sites is similar between the embryo that received maximum and the embryo that received half the dose of maternal MCP-GFP (Figure 4.2 C). Therefore it can be concluded that the reduced load of MCP-GFP deposited by *;His-RFP;MCP-GFP/+*

females almost halved the background levels of unbound MCP-GFP and at the same time did not limit the overall fluorescent intensity of MCP-GFP bound transcription sites. All subsequent imaging experiments were performed using embryos from a second-generation cross to reduce the maternally deposited MCP-GFP load. Additionally, background correction was performed by subtracting the background value from fluorescent readings and then dividing the resulting fluorescence by the background. With the improved signal-to-noise ratio and after optimising the analysis pipeline, the image analysis of time-lapse movies still requires significant manual detection and correction, and therefore is extremely time consuming.

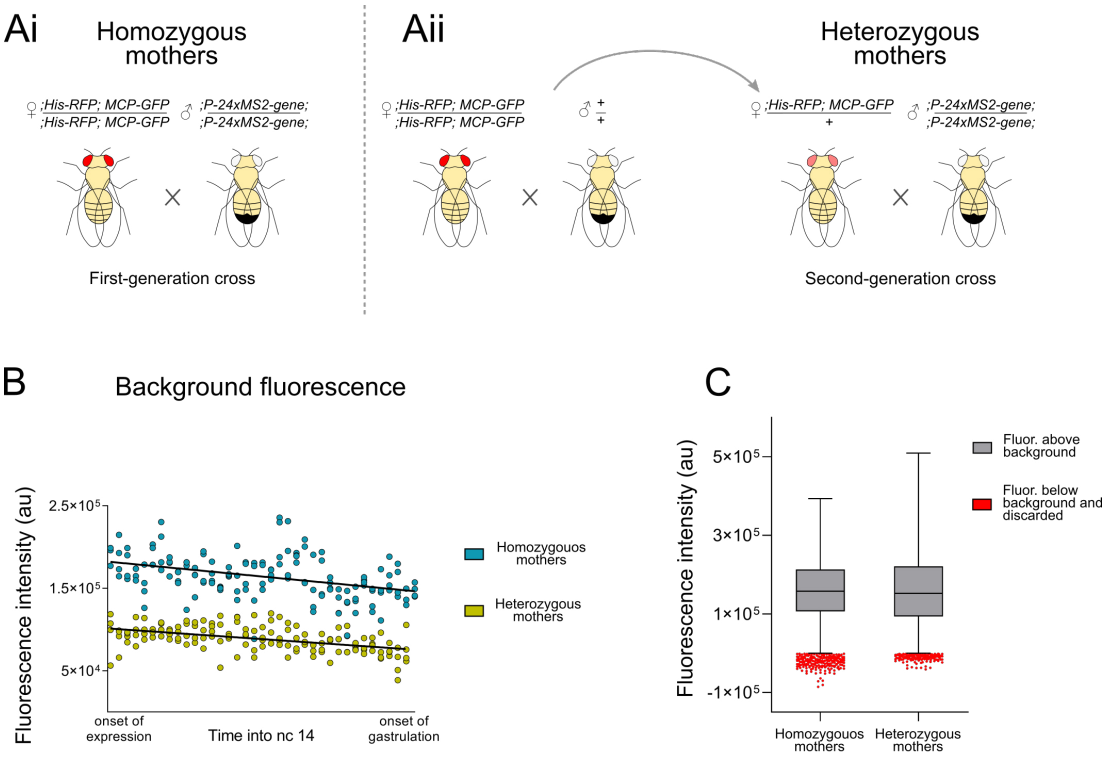



**Figure 4.2: Improved background fluorescence levels of the MS2/MCP system**. Crossing schemes to obtain first (Ai) or second-generation (Aii) offspring for imaging studies. Maximal maternal load of MCP-GFP was obtained by crossing homozygous virgin females of the *;His-RFP;MCP-GFP* genotype to homozygous males with the genotype *ush-24xMS2*. Reduction of the maternal load was achieved by a second generation cross, crossing heterozygous *;His-RFP;MCP-GFP/+* virgins to homozygous *ush-24xMS2* males. (B) Background fluorescence readings from embryos imaged during nc14 with different loads of maternal MCP-GFP. Embryo offspring from homozygous (blue) and heterozygous (green) mothers. Regression line equations were used for background subtraction from fluorescent transcription site values. (C) Fluorescence intensity values after background correction. Fluorescent values below (red) and above (grey) zero after background subtraction. Boxes show 25$^{th}$ to 75$^{th}$ percentile, line shows median and whiskers show range.

## 4.3 Transcription dynamics of BMP target genes *hnt* and *ush*

### 4.3.1 Design and characterisation of endogenous MS2 lines

To characterise the transcription dynamics of the BMP target genes *hnt* and *ush*, endogenous MS2-loop knock-in fly lines were created. The *hnt* gene is located on the X-chromosome and has two alternative isoforms that differ in their 3'UTR and therefore share the promoter and 5'UTR region. The *hnt* promoter region contains a TATA box. 24 copies of the MS2 stem loop cassette were inserted into the endogenous 5'UTR downstream of the TSS using CRISPR/Cas9 genome engineering and ΦC31 mediated reintegration (see methods, summary in Figure 4.3 A). The resulting *MS2-hnt* line was used for live imaging 4D experiments. The genome region encoding the *ush* gene is located on the left arm of the second chromosome and has five isoforms (Figure 4.3 B). A GRO-seq study identified isoform C (RC) to be primarily expressed during early *Drosophila* development. Little activity was found for isoform A (RA) (Saunders et al. 2013). The promoters of RA and RC contain an Inr sequence at the TSS according to the consensus sequence TCAGTY (Smale & Baltimore 1989, Haberle & Stark 2018). MS2-loops were inserted into the endogenous gene locus downstream of the TSS in the 5'UTR (see methods, summary in Figure 4.3 B).

Colourimetric ISH was performed to evaluate whether the insertion of MS2-loops changed the expression pattern of *hnt* and *ush* transcription. A comparison between the *hnt* expression domain in a wt and *MS2-hnt* embryos shows the *hnt* expression pattern in the same relative dorsal location with a similar width (Figure 4.4 A). Similar to *hnt* expression, the overall transcription pattern of *ush* is equivalent between wt and *MS2-ush* embryos (Figure 4.4 A). These data suggest that the expression pattern during nc14 is not altered by the addition of MS2-loops into the genome. For a quantitative comparison between wt and embryos containing MS2-loops, the number of transcripts would need to be quantified by either smFISH or qRT-PCR.

Next, amnioserosa cells were counted to investigate the influence of MS2-loops in the genome on later stages of development. The number of Amnioserosa cells is a direct readout of the correct development of dorsal tissues. Amnioserosa cells were visualised by Hnt antibody staining (Figure 4.4 B). The mean number of amnioserosa cells counted in wt embryos is 131 cells, which is consistent with previ-

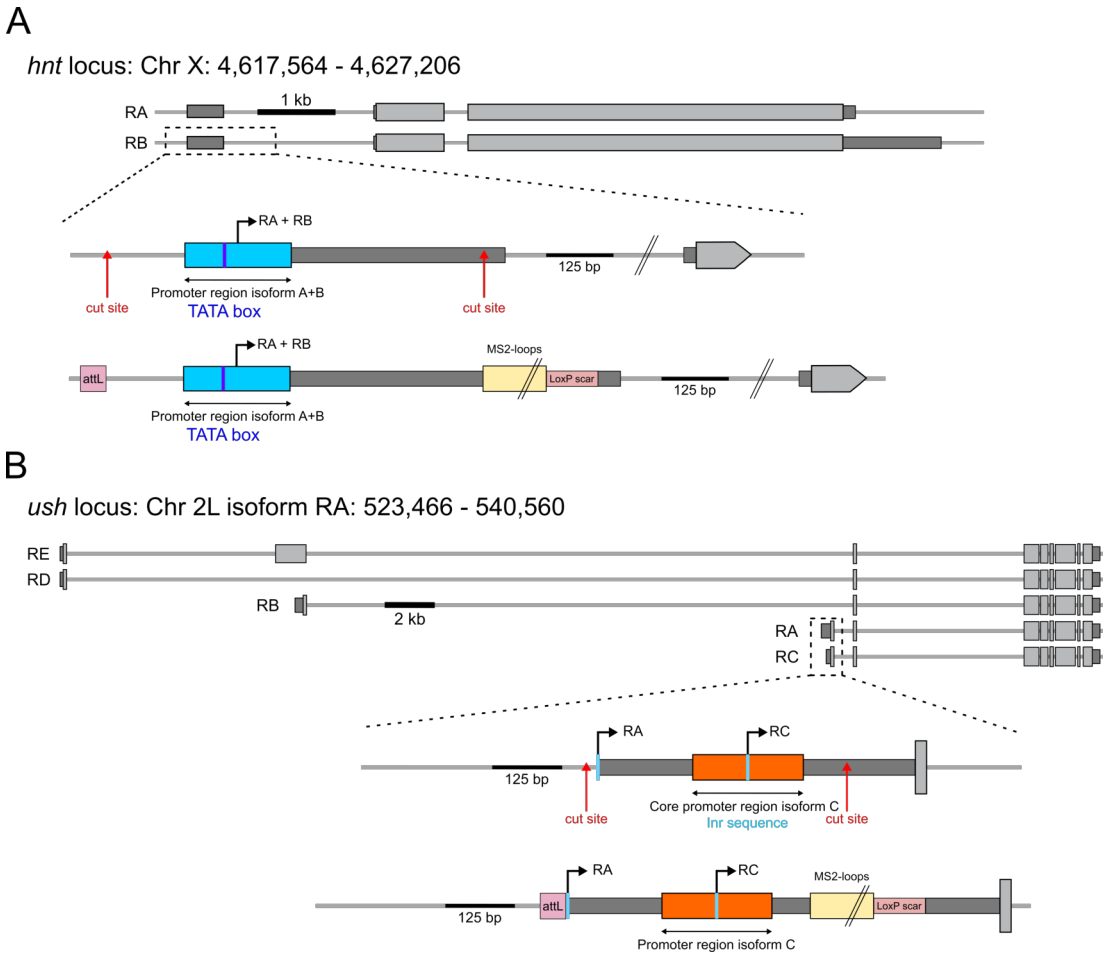



**Figure 4.3: Generation of *hnt* and *ush* MS2 live imaging lines.** (A) A schematic showing the genomic region encoding *hnt* on the X-chromosome. Two isoforms share the 5'UTR and promoter region (RA, RB) and contain a TATA box. 24xMS2 cassettes were inserted into the endogenous *hnt* 5'UTR. (B) A schematic showing the genomic region encoding *ush* on the second chromosome. Five isoforms exist of which the shorter two are actively transcribed with predominant contribution by isoform (RC) during early *Drosophila* development. The isoforms contain an Inr sequence at the TSS's. 24xMS2 cassettes were inserted into the endogenous *ush* 5'UTRs using CRISPR genome editing. All chromosome coordinates were obtained from the *Drosophila* dm6 genome.

ous studies (Figure 4.4 B) (Wharton et al. 1993, Newton et al. 2015, Deignan et al. 2016). Embryos homozygous for the *MS2-ush* locus showed no statistical difference to wt embryos with a mean of 137 cells (Figure 4.4 B). The number of amnioserosa cells in homozygous *MS2-hnt* embryos is shown to be significantly reduced compared to wt numbers (mean = 90, Figure 4.4 B). Hence, these data suggest that the addition of MS2-loops or the general changes in the genome introduced by CRISPR editing led to changed *hnt* expression that could have occurred during transcription or potentially post-transcriptionally. Possibly amnioserosa counts were influenced more by changes in *hnt* expression than *ush*, as Hnt is a direct marker for amnioserosa cells and vital for its maintenance (Frank & Rushlow 1996, Yip et al. 1997). Ideally, a different amnioserosa marker would have been used, since compromised levels of Hnt in *hnt-MS2* lines could lead to detection threshold problems. At this time no other antibody detecting an amnioserosa marker is commercially available.

*Drosophila* lines were generated to investigate whether the addition of MS2-loops or the changes caused by CRISPR engineering were responsible for the changes in amnioserosa number. The modified lines underwent the same genome engineering procedure, but instead of MS2-loops, they had only wt sequences reintegrated. After CRISPR genome editing both the MS2 imaging lines and the CRISPR control fly lines contain small scars on either side of the double strand breaks. The scar lengths in flies with modified *ush* locus are 82bps and 103 bps and with modified *hnt* locus 100 and 95 bps (see methods). The *ush* line without loops but with genome editing (*CRISPR-ush*) shows the same number of amnioserosa cells as wt and the *MS2-ush* line (Figure 4.4 B). Unfortunately, the more important *CRISPR-hnt* line was not available due to reintegration difficulties.

The wildtype-like expression patterns of *hnt* and *ush* RNA during nc14 suggest that transcription occurs in similar regions as in wt embryos, but that posttranscriptional changes might lead to developmental defects in the case of Hnt. Since this study focuses on transcription kinetics, the investigation into protein production is beyond the scope of this work. When embryos were imaged using the MS2/MCP system to investigate transcription kinetics, active fluorescence was recorded from one allele since these embryos were heterozygous for the respective imaging locus (Figure 4.5). The following experiments will investigate the paternal MS2 tagged allele based on the crossing scheme outlined in Figure 4.2 Aii.

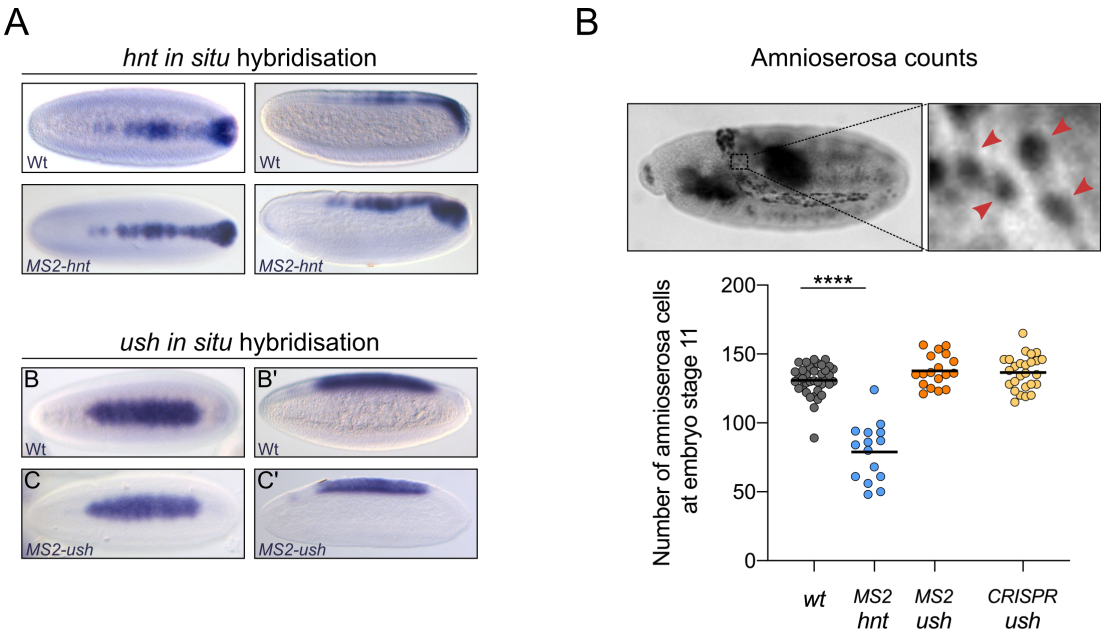



**Figure 4.4: Expression patterns of live imaging lines.** (A) Colourimetric ISH detecting *hnt* (left) or *ush* (right) transcripts in wt embryos or embryos homozygous for the MS2-loop tagged gene locus. (B) The top shows a microscopy image of a stage 11 *Drosophila* embryo with amnioserosa cells (red arrowheads) stained with Hnt antibody. The graph underneath shows amnioserosa cell counts. Each data point represents one embryo. *CRISPR-ush* embryos were used as a control as they were treated the same way as MS2-loop lines during genome modification but without loop insertion.
Line in scatter plots (B) depicts mean. Significance was tested with a One-Way ANOVA test with a Holm-Sidak's multiple comparisons test (B), ****$p < 0.0001$. Technical replicates for (B) *wt* n = 5, *MS2-hnt* n = 2, *MS2-ush* n = 3, *CRISPR-ush* n = 3. Embryos in (A) shown in dorsal and lateral position, anterior to the left.

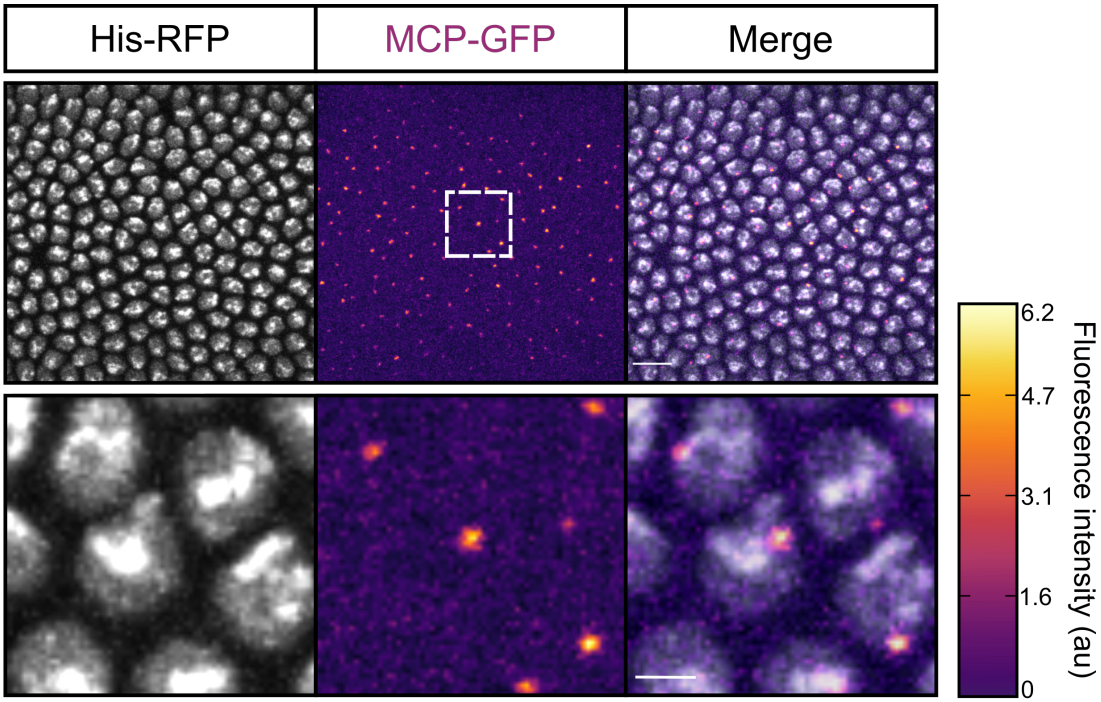



**Figure 4.5: MS2/MCP system produces bright fluorescent signal of nascent transcription**. Maximum projected still from a microscopy time-lapse dataset. Transcription of the *MS2-ush* locus, where MS2-loops are bound by fusion MCP-GFP protein, was detected by bright fluorescence. Enlarged region (bottom) shows one active transcription site per nucleus.
Scale bars = 10 μm (top) and 3 μm (bottom)

### 4.3.2 Wildtype *ush* and *hnt* transcription profiles

*Drosophila* embryos were imaged from the onset of cleavage cycle 14 to generate time-lapse datasets of *hnt* and *ush* transcription (Figure 4.6 A). The imaging regions, for dorsally placed embryos, are shown in Figure 4.6 Bi. To capture *ush* transcription, the central domain of the embryo was imaged, whereas the centre to posterior region was imaged to capture *hnt* transcription as *hnt* has less anterior expression pattern (Figure 4.6 Bi). Maximum projection images at 37.5 min into nc14 illustrate the dynamic behaviour of *hnt* and *ush* transcription. A small subset of nuclei that are part of the *hnt* and *ush* core expression domain (surrounded by expressing cells) did not show active transcription (grey nuclei, Figure 4.6 Bi). It is plausible that these inactive nuclei were in between two transcriptional bursts and therefore show dynamic transcription. Analysis of the imaging domains results in single cell expression data. Representative single cell traces show differences in transcriptional onset, where *ush* precedes that of *hnt* (Figure 4.6 Bii).

All analysis was performed using three-dimensional datasets. For time-lapse movies, maximum projected images are shown. Stills from movies of embryos with hnt and ush transcription are shown in Figure 4.7 and 4.8.

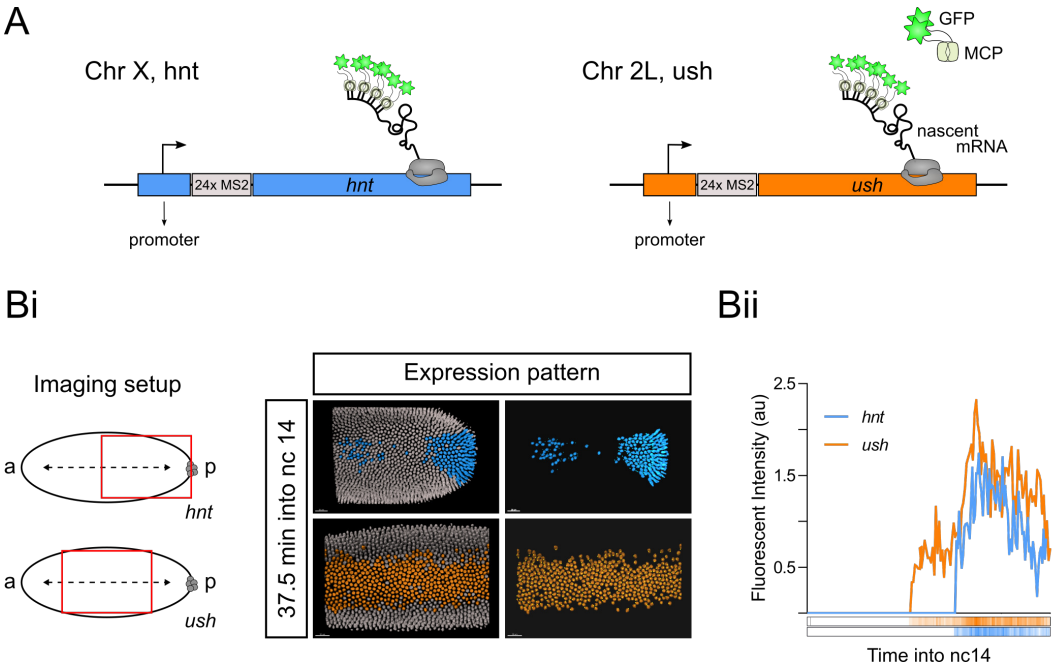



**Figure 4.6: Live imaging setup of *hnt* and *ush* transcription.** (A) Summary of the genomic imaging locus region for *hnt* (blue) and *ush* (orange). (Bi) Imaging setup to visualise BMP target gene transcription. *hnt* transcription was captured in the posterior and central domain, whereas *ush* transcription was imaged in the central dorsal domain. Falsely coloured stills from time-lapse data sets show the expression pattern at 37.5 min into nc14. (Bii) Example single nuclear fluorescent trace of *hnt* (blue) and *ush* (orange) transcription. Heatmap of imaging traces below graph.

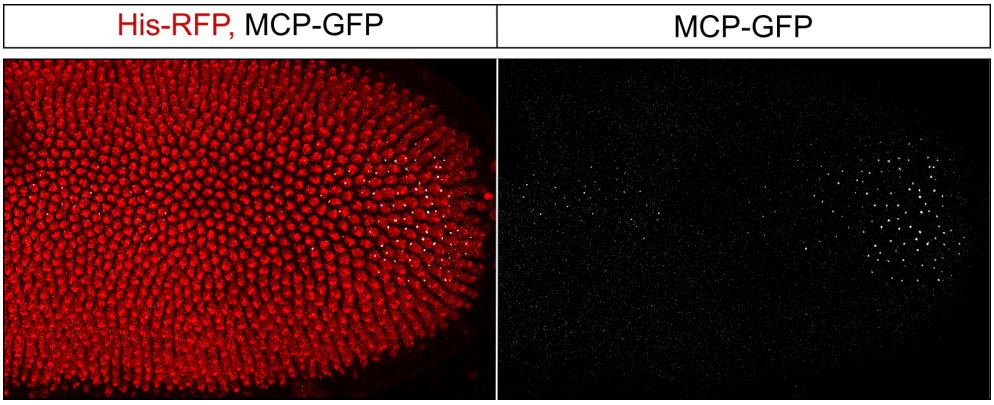



**Figure 4.7: Time-lapse microscopy detecting *hnt* transcription.** Still of a maximum intensity projection (54x 1 μm stacks) during nc14. Nuclei visualised by His-RFP (red) and *hnt* transcription by MCP-GFP fluorescence (grey). Time resolution 20 sec/frame. Embryo was imaged in a dorsal position with anterior to the left. Full movie can be found at `https://www.tiny.cc/thesis_videos`.

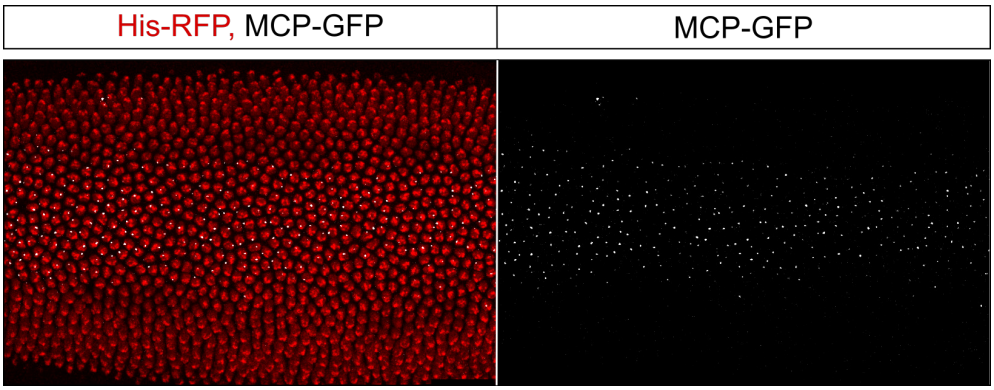



**Figure 4.8: Time-lapse microscopy detecting *ush* transcription.** Still of a maximum intensity projection (54x 1 μm stacks) during nc14. Nuclei visualised by His-RFP (red) and *ush* transcription by MCP-GFP fluorescence (grey). Time resolution 20 sec/frame. Embryo was imaged in a dorsal position with anterior to the left. Full movie can be found at `https://www.tiny.cc/thesis_videos`.

## 4.4 Characterisation of *hnt* and *ush* expression domains along the AP axis

For initial characterisation of the full imaging domain, one embryo visualising *hnt* and one embryo visualising *ush* transcription was analysed. Hence, data were collected from one biological replicate containing n = 918 and n = 1013 nuclei in the field of view, respectively.

### 4.4.1 Characterisation of the *hnt* expression domain

To investigate the *hnt* transcription domain of a full imaging dataset, the single-cell transcription profiles were analysed first concerning the onset time of transcription. The *hnt* core expression domain of this embryo contains 368 transcriptionally active nuclei. All of these nuclei were identified to show active *hnt* transcription for at least one time point within the time-lapse dataset. First, the onset time of each nucleus was plotted against the nuclear AP position at the time of transcription initiation. The onset time of transcription shows a curved shape along the AP axis and a correlation between onset time and position along the AP axis is found ($r^2$ = 0.31, Figure 4.9 Ai). The time at which *hnt* transcription was initiated for the first time is significantly lower in the posterior domain of the embryo compared to the central domain (Figure 4.9 Aii). These data suggest differences in the transcriptional behaviour of *hnt*. It has been reported that the posterior expression domain of *hnt* is less responsive to BMP signalling. In experiments where BMP signalling was reduced due to a reduction in Collagen IV levels, *hnt* expression was lost in the central domain (presumptive amnioserosa) but persisted in the posterior region (Wang et al. 2008). Similarly, *hnt* expression in the presumptive amnioserosa is lost in *nejire* mutant embryos but remains active in the posterior pole. Hence, in can be concluded that the central domain is responsive to BMP signalling, and the posterior domain might be regulated by other signalling pathways in addition to BMP (Ashe et al. 2000). Here, the differences in regulation manifest in significantly different transcriptional onset times.

As this study is interested in the transcription behaviour in response to BMP signalling, the central domain was analysed in more detail and separated into three sub-domains. The median onset time of transcription was found to be lowest at the most anterior end and increased towards the posterior end of the central domain (medians = 38.59 min, 39.67 min, 42.93 min, Figure 4.9 Aiii). No significant cor-

relation was found between the onset time and nuclear position along the cross-section of the embryo (Figure 4.9 B).

Next, the transcriptional output of single-cell traces was investigated over time. The mean fluorescent intensity for each section was plotted at each time point. Even though the time at which the maximal transcriptional output was reached is similar (central = 47.3 min, posterior = 44.5 min), the absolute fluorescence values are very different (Figure 4.9 C). Since MCP-GFP emitted fluorescence is directly proportional to the number of transcripts produced, it can be concluded that the posterior region produced approximately 3x as much RNA as the central region (Figure 4.9 C). This difference in overall output can also be visualised by calculating the sum fluorescence that a nucleus produces throughout the complete nc14 (Figure 4.9 D). Again, nuclei positioned posteriorly are shown to produce a greater amount of total fluorescence compared to nuclei positioned in the central domain (Figure 4.9 D).

These data show different transcriptional behaviour between the BMP responsive and unresponsive domain of *hnt*. All further analysis will be focused on the central section of the *hnt* expression domain, which is BMP responsive.

### 4.4.2 Characterisation of the *ush* expression domain

Similar to the *hnt* transcription profile, *ush* transcription was characterised first in the whole imaging domain. Out of the 1013 single cell traces obtained from one time-lapse experiment, 660 nuclei showed active *ush* transcription. The nuclear AP position was split into three equally sized analysis areas: anterior, centre and posterior. The onset time of *ush* transcription is shown to be less regulated along the AP position (no correlation), compared to *hnt* transcription onset ($r^2 = 0.04$, Figure 4.10 Ai).

Transcription is first initiated around 20 min into nc14 and new nuclei still initiate transcription towards the end of nc14 (Figure 4.10 Ai). Quantification shows that the onset time of transcription is significantly lower in the anterior domain compared to the central or posterior areas (Figure 4.10 Aii). This finding suggests that the anterior region might be regulated by other signalling inputs besides BMP, as for example Bicoid. If *ush* is regulated by another anterior signal, it potentially does so in an AP gradient. To test this hypothesis, the anterior region of the *ush* embryo was further divided into four bins. The most anterior bin shows the ear-

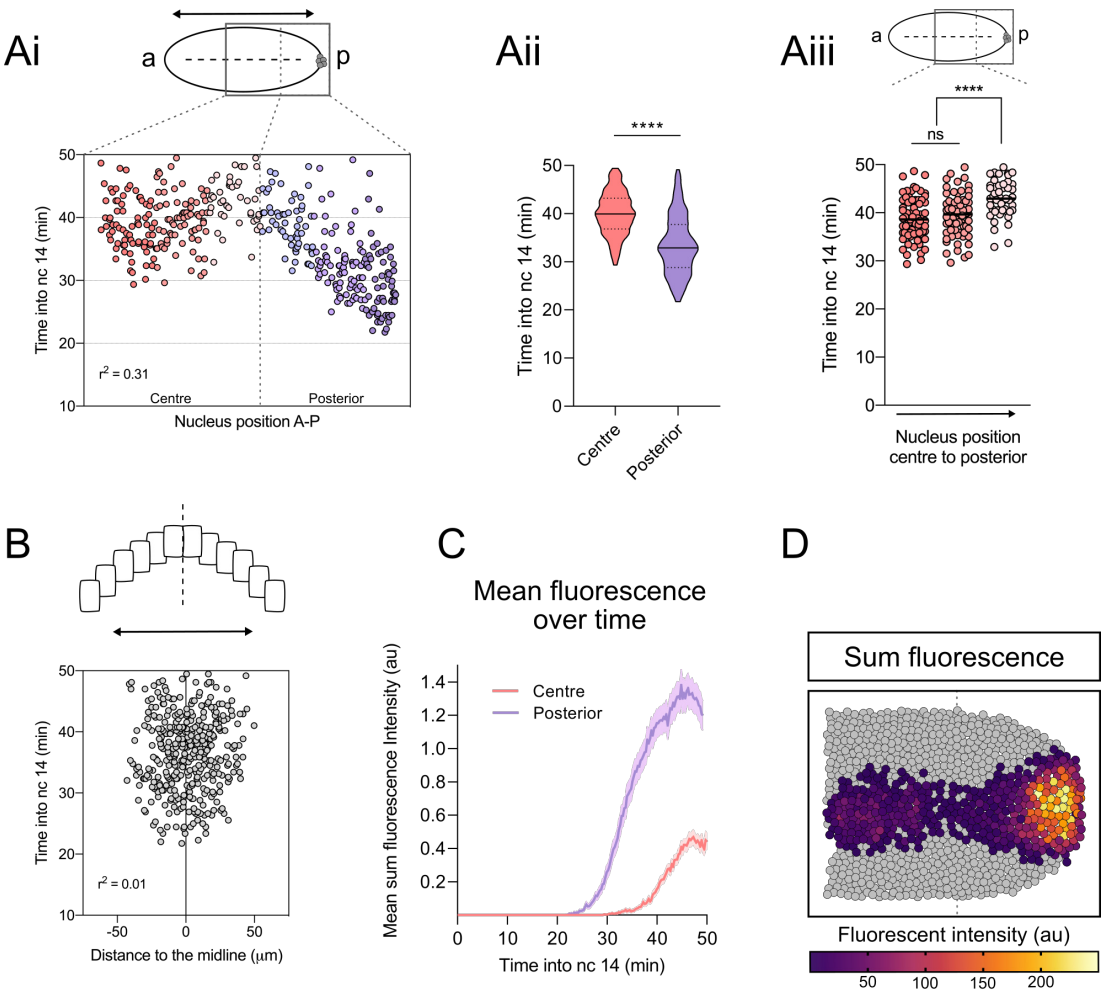



**Figure 4.9: Characterisation of the *hnt* expression domain.** (Ai) Transcription onset time is plotted according to the nuclear position along the AP axis. Each dot represents a nucleus and nuclei are coloured according to AP position bins (Bins = 5 μm) (Aii) Quantification of transcription onset time between the central (pink) and the posterior (purple) region. (Aiii) Time to transcription onset in subsections of the central domain. (B) Transcription onset time plotted according to the nuclear position across the dorsal midline. (C) Mean fluorescence of central and posterior domain. Maximum transcriptional output reached in the central domain = 47.3 min, posterior domain = 44.5 min. (D) Representative image of expression domain. Nuclei coloured depending on total fluorescence value.
Lines in violin plots (Aii) depict median and 95% confidence intervals. Line in scatter plot (Aiii) depicts median. Line plots (C) depict mean and SEM. Significance was tested with a Mann-Whitney test (Aii) and a Kruskal-Wallis test with a Dunn's multiple comparisons test (Aiii), ****$p < 0.0001$, ns = not significant. Spearman correlation coefficient (Ai, B) calculated based on nuclear position versus onset of transcription. Biological replicate = 1 with n = 368 nuclei. Embryo depictions in dorsal position, anterior to the left.

liest *ush* transcription onset (Figure 4.10 Aiii). From there the median onset time increases with nuclear position towards the central domain (Figure 4.10 Aiii).

Contrary to the *hnt* transcription domain, *ush* transcription onset is shown to be correlated with the nuclear position across the midline ($r^2 = 0.44$, Figure 4.10 B). Nuclei that are positioned around the dorsal midline initiate *ush* transcription earlier than nuclei in the periphery (Figure 4.10 B). The observation that *ush* transcription onset is regulated across the midline over time is consistent with previously obtained smFISH results, where static images showed that the transcription domain of *ush* increased in width over time (see, Figure 3.10 Bi).

The analysis of different AP domains in regards to their mean fluorescence output reveals no difference between the centre and the posterior domains (Figure 4.10 C). The anterior domain shows slightly increased mean fluorescence levels (Figure 4.10 C). Furthermore, the anterior domain curve initiates transcription earlier and reaches its maximum level quicker (Figure 4.10 C), consistent with an earlier average onset time (Figure 4.10 Aii). The slight increase in fluorescence of the anterior region is also visible when the sum fluorescence of nuclei is falsely coloured in a heatmap (Figure 4.10 D). Here, the fluorescence produced at each time point is summed and the nuclear position visualised in the embryo (Figure 4.10 D). The increase in total fluorescence indicates increased transcription output in the anterior domain. This increase is less severe than observed for *hnt* transcription in the posterior *hnt* domain (compare, Figure 4.9 D).

All further analysis will be focused on the anterior section of the *ush* expression domain as no differences in BMP responsiveness of *ush* have been described previously. The anterior *ush* domain is representative of the overall *ush* expression domain, as little differences in mean and total fluorescence were found between the domains.

### 4.4.3 Comparison of *hnt* and *ush* expression domains

After characterising the overall expression domains of *hnt* and *ush*, their transcription behaviour will be compared. To obtain complete data sets two more time-lapse experiments were analysed for *ush* and one more experiment was analysed for *hnt* transcription. One *hnt* dataset was discarded as developmental defects were observed during cleavage cycle 14. Together with the data described in the previous sections, the number of biological replicates for *ush* is 3 and for *hnt* is 2. The cen-

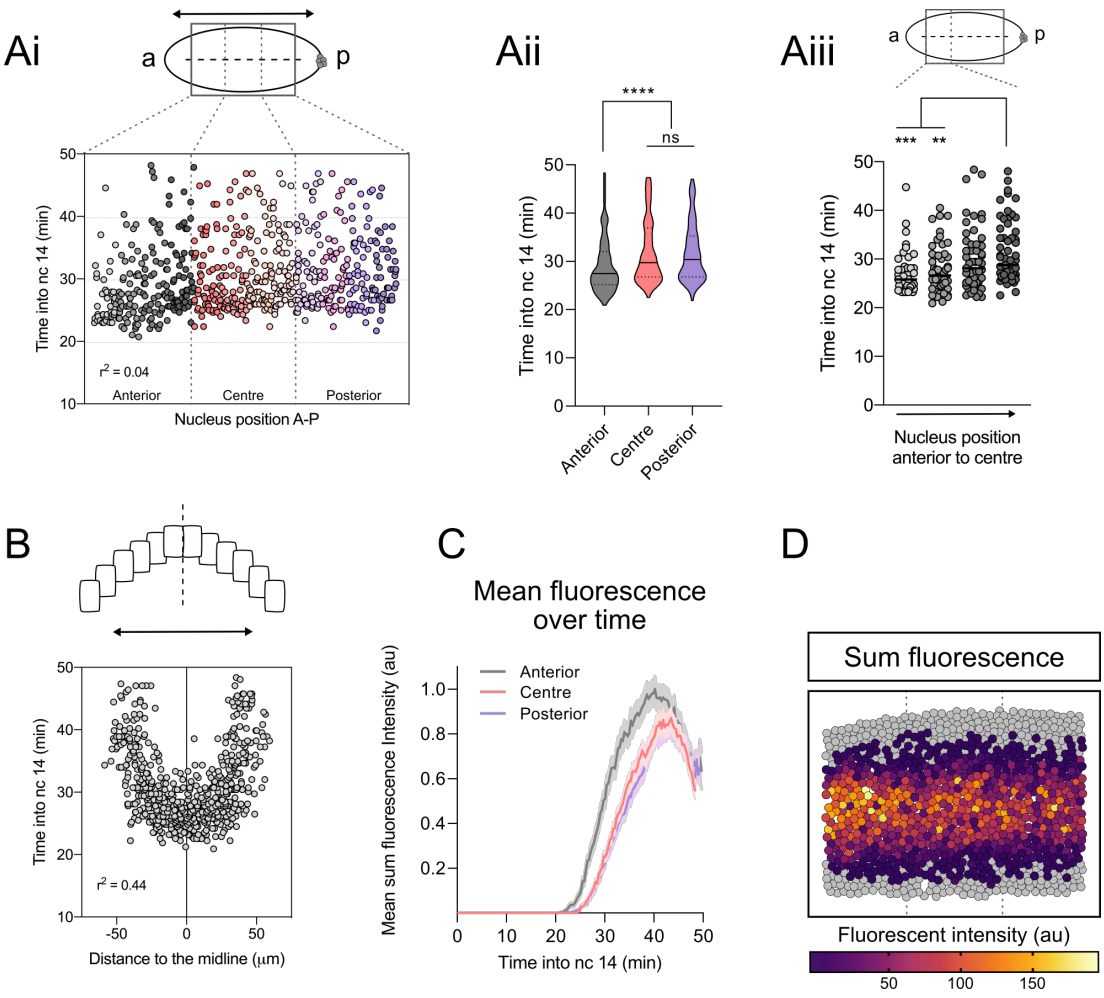



**Figure 4.10: Characterisation of the *ush* expression domain.** (Ai) The transcription onset time of *ush* is plotted according to the nuclear position along the AP axis. Each dot represents a nucleus and nuclei are coloured according to AP position bins. (Aii) Quantification of transcription onset time between three equally sized domains: anterior (grey), central (pink), and posterior (purple) region. (Aiii) Time to transcription onset in subsections of the anterior domain. (B) Transcription onset time plotted according to the nuclear position across the dorsal midline. (C) Mean fluorescence of sub-domains. Maximum transcriptional output reached at anterior = 40.2 min, central = 43.5 min, posterior = 43.5 min. (D) Representative image of expression domain. Nuclei coloured depending on total fluorescence value.
Lines in violin plots (Aii) depict median and 95% confidence intervals and in scatter plots (Aiii) the median. Line plots (C) depict mean and SEM. Significance was tested with a Kruskal-Wallis test with Dunn's multiple comparisons test (Aii, Aiii). ****p < 0.0001, ***p < 0.001, **p < 0.01, ns = not significant. Spearman coefficient (Ai, B) calculated based on nuclear position versus onset of transcription. Biological replicate = 1 with n = 660 nuclei. Embryo depictions in dorsal position, anterior to the left.

tre domain was analysed for all *hnt* datasets and the anterior domain for all *ush* datasets. The *hnt* dataset, analysed in Figure 4.9, will be called *hnt* 1 and the *ush* dataset analysed in Figure 4.10 will be referred to as *ush* 1.

The differences in imaging and analysis domains are visualised in embryo schematics (Figure 4.11 A). The full cumulative expression domains of *hnt* and *ush* correspond to the data analysed in Figures 4.9 and 4.10. The analysis domains show the regions selected for all further analysis and the resulting expression patterns (Figure 4.11 A). Single-cell traces of biological replicates were analysed and their transcription onset times are shown in Figure 4.11 B. The transcription onset of *hnt* and *ush* is consistent between biological replicates and not statistically different. *ush* transcription is initiated earlier (median = 28.11 min) than *hnt* transcription (median = 39.5 min) during nc14 (Figure 4.11 B).

The mean fluorescence intensity curves of *hnt* replicates are very similar to each other. Their overall shape, the range of fluorescence values and time of maximum fluorescence (transcriptional output) is consistent. The maximum fluorescence intensity is reached at 47.3 min for *hnt* replicate 1 and 46.9 min for *hnt* replicate 2 (Figure 4.11 Ci). *ush* transcription displays a similar consistency in mean fluorescence over time between biological replicates. Here, the curve shapes are similar and the maximum fluorescence intensity is reached at 40.2 min for *ush* 1, 41.3 min for *ush* 2 and at 42.8 min for *ush* 3 (Figure 4.11 Cii).

## 4.5 Transcription levels depend on nuclear position within the BMP gradient

Having detected onset times for *hnt* and *ush* transcription and variations in the mean expression levels the variation in total transcriptional activity was investigated.

False colouring of nuclei in the analysis domains, based on the total fluorescence, reveals increased mRNA output in the regions adjacent to the dorsal midline for the representative embryos shown (Figure 4.12 A). The *ush* expression domain is shown to produce higher fluorescence output than the *hnt* expression domain (note different scales on heatmaps, Figure 4.12 A). Next, the transcriptional output of *hnt* and *ush* domains was analysed over time and in relation to the nuclear position across the dorsal midline. The *hnt* transcription profile of pooled biological

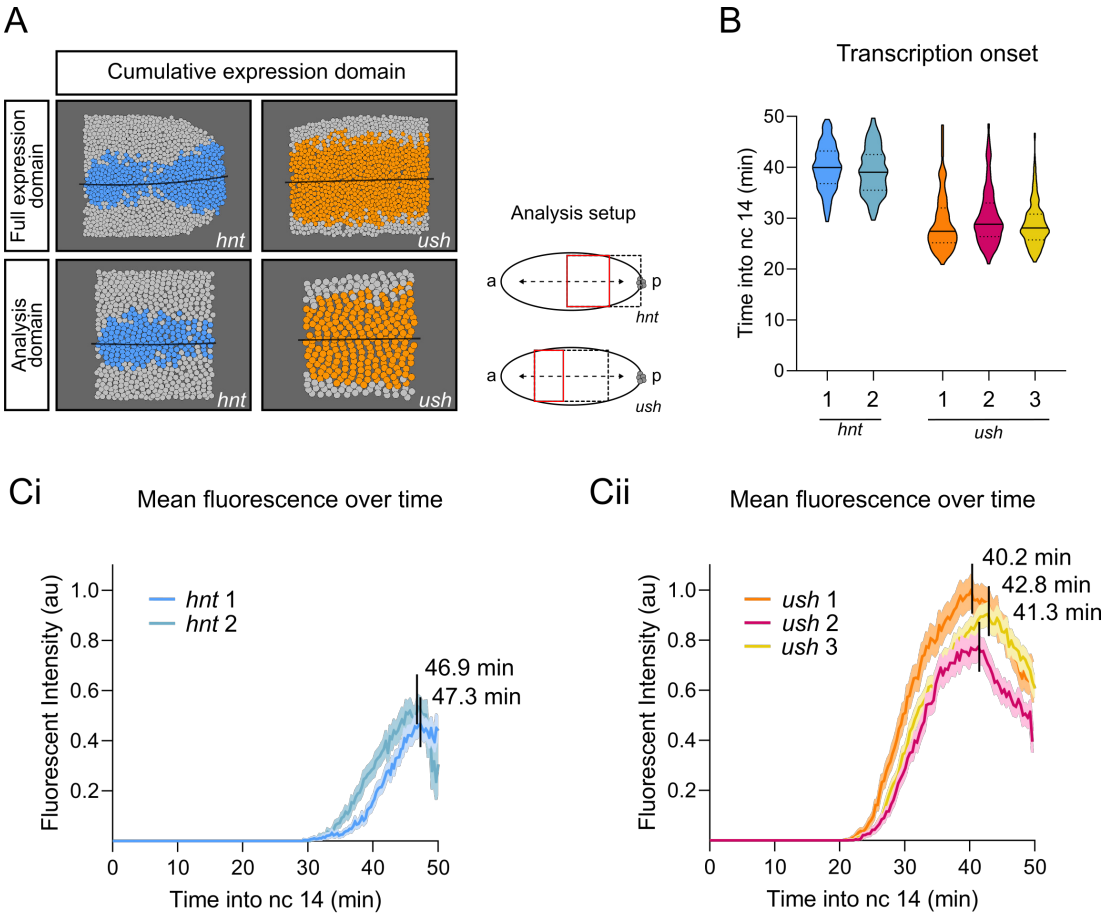



**Figure 4.11: Comparison of *hnt* and *ush* transcription profiles using live imaging.** (A) Cumulative expression domains of *hnt* and *ush* transcription are depicted for the full imaging domains and the analysis domains. Schematic of dorsally positioned embryos highlights the chosen analysis domains. (B) Transcription onset of *hnt* and *ush* transcription. Mean fluorescence values of *hnt* (Ci) and *ush* (Cii) transcription over time. Time point of maximum fluorescence next to curve peaks. Lines in violin plots (B) depict median and 95% confidence intervals. Line plots (Ci, Cii) depict mean and SEM. Biological replicates n = 2 for *hnt* (192, 188 nuclei) and n = 3 for *ush* (209, 186, 223 nuclei).

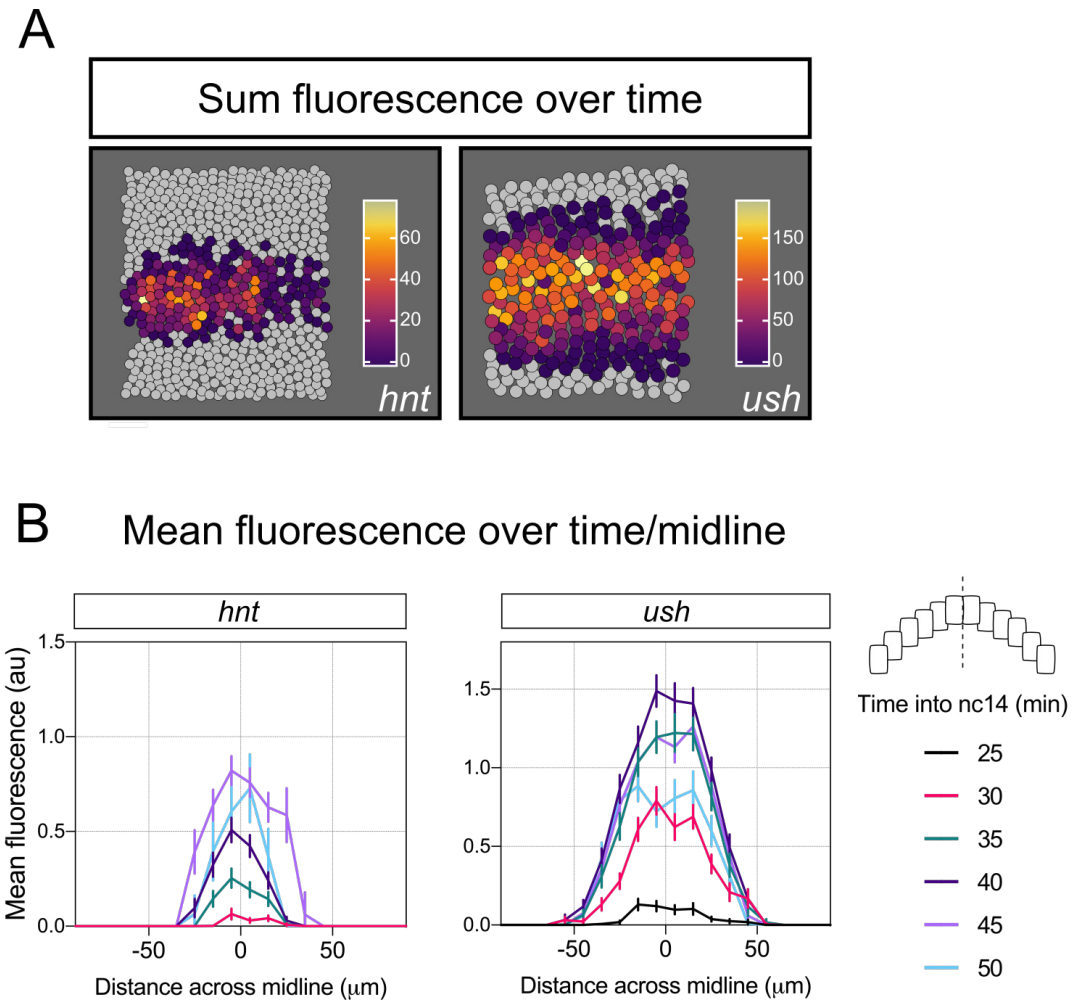



**Figure 4.12: Live imaging reveals increased mRNA output at dorsal midline.** (A) Representative images of *hnt* (central) and *ush* (anterior) analysis domains. Nuclei were coloured depending on their total sum fluorescence, and maximum values were chosen according to the gene's maximum fluorescence respectively (note different scales). (B) Mean fluorescence of *hnt* and *ush* transcription domains were determined every 5 min after transcription initiation and divided into positional bins along the embryo cross-section. Bin width = 10 μm. Graphs show mean and 95 % confidence interval. Data were pooled from all biological replicates.

replicates is shown to increase in mean fluorescence over time with peak levels at approximately 45 min into nc14 (Figure 4.12 B). Additionally, the expression domain widens over time, as was previously observed using smFISH. A steep gradient of fluorescence is shown, which starts to form at 35 min into nc14 (Figure 4.12 B). The transcription profile of *ush* is shown to increase rapidly between 25 and 35 min into nc14 before it peaks at 40 min (Figure 4.12 B). Again, the expression domain of *ush* increases in width over time and the highest mean fluorescence is greater than that of *hnt* transcription (Figure 4.12 B). These data highlight the changes of *hnt* and *ush* expression domains over time. The mRNA output gradient was shown in the previous chapter by smFISH probes. However, the gradient and its formation are shown here with a higher temporal resolution of 5 min.

### 4.5.1 K-Means clustering identifies clusters of transcription behaviour

So far, the fluorescence intensity of single-cell traces was summarised in mean intensities or according to the nuclear position. Next, individual traces were investigated in detail using the K-Means clustering algorithm. During K-Means clustering, traces are divided into groups based on feature similarity. The K-Means algorithm aims to partition observations into sets by alternating between two steps until they converge. First is the assignment of fluorescence traces to clusters based a more similar mean value, followed by calculation of a new mean within clusters (Lloyd & S. 1982). Clustering revealed the existence of two clusters for *hnt* transcription (Figure 4.13) and three clusters for *ush* transcription (Figure 4.14). The number of clusters was determined using an elbow graph (data not shown).

When nuclei of the *hnt* core expression domain are coloured according to the cluster they belong to, it is observed that nuclei near the dorsal midline mostly belong to cluster 1 (yellow). This middle cluster is enriched in highly expressing nuclei (Figure 4.13). Conversely, nuclear traces from the broader intermediate/edge domain belong to cluster two (green). The edge cluster contains traces from lower expressing nuclei (Figure 4.13). Visualisation of the individual nuclear traces as heatmaps shows the clear separation of high and low expressing nuclei. The mean fluorescence output of nuclei in the middle cluster is greater than the output of the edge cluster (right, Figure 4.13). Furthermore, individual traces show periods of increased and decreased fluorescence. These periods represent dynamic transcription fluctuating in the number of elongating polymerases (Figure 4.13). In the

heatmaps, individual nuclear traces were ordered according to their transcriptional onset time. The transcriptional onset front of the middle cluster appears steeper than the onset front of the edge cluster (Figure 4.13).

For *ush* transcription, K-Means clustering grouped fluorescent traces into three groups which map to the centre (yellow), intermediate (green) and the edge (orange) location of the transcription domain (Figure 4.14). Similar to the *hnt* expression domain, the central cluster of *ush* contains nuclei that show high fluorescence (average mean fluorescence ∼1). The intermediate cluster produces fluorescence levels that are half that of the central cluster (average mean fluorescence ∼0.5, Figure 4.14) and there is an even further reduction in fluorescence in the edge domain (Figure 4.14).

In summary, clustering of *hnt* and *ush* transcription added independent evidence for a relationship between nuclear output and position. It can be speculated that this relationship is the direct result of different BMP signalling levels. The heatmaps have illustrated that the fluorescence intensity and therefore mRNA production varies greatly between nuclei in different positions. It can be hypothesised that those changes are caused by different transcription burst kinetics. Furthermore, periods of increased and decreased fluorescence within single traces have been observed, .which may represent transcriptional bursts. The off periods of transcriptional bursts during *hnt* and *ush* transcription would rarely be visible due to the positioning of MS2-loops in the gene's 5'UTR. When the promoter switches into an off-state, the time resolution is limited by the time it takes already initiated Pol II molecules to complete transcription. A Pol II molecule with an elongation speed of 3 kb/min (Fukaya et al. 2017) would take approximately 2.5 min to complete one *hnt* transcript and 5.5 min to transcribe one *ush* transcript at constant elongation speed. Therefore, off-periods that are shorter than 2.5 or 5.5 min for *hnt* and *ush* respectively cannot be resolved visually. Furthermore, the underlying transcription kinetics that generate the differences observed in the heatmaps, cannot be determined visually and computational modelling approaches are needed.

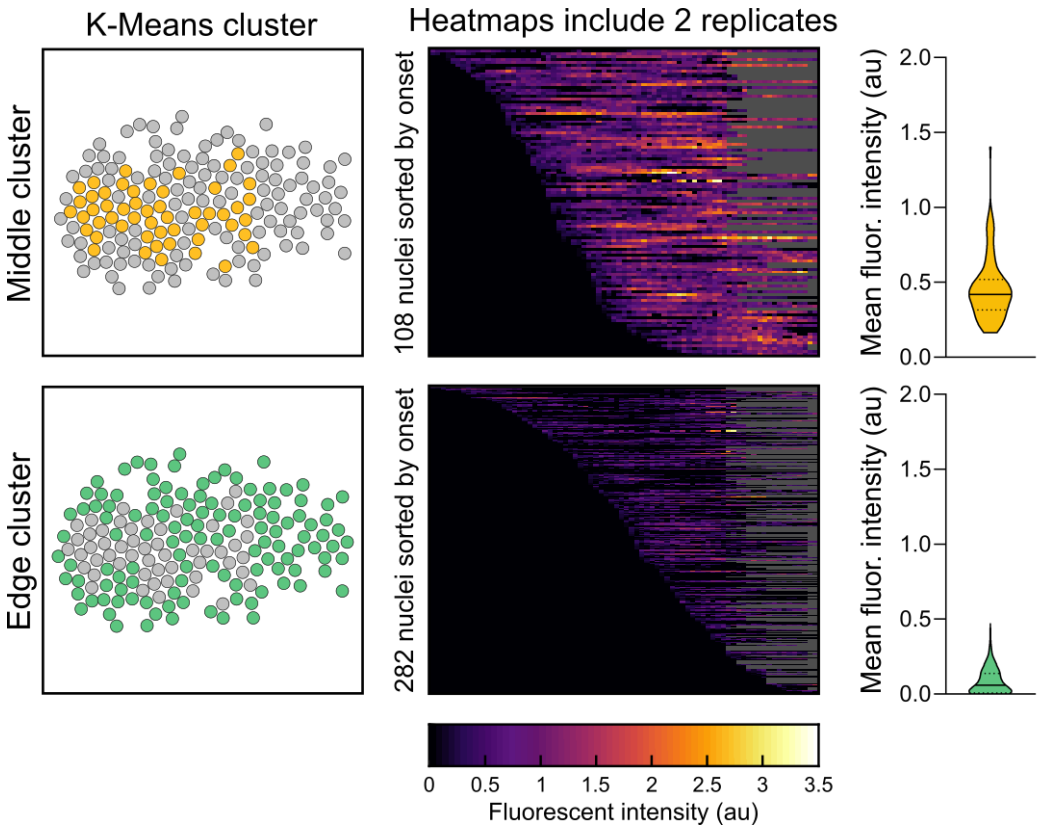



**Figure 4.13: K-Means clustering identifies *hnt* transcription domains.** Single cell *hnt* transcription profiles were grouped into two clusters (middle and edge), using the K-Means algorithm. (Left) Nuclei are coloured by their cluster and plotted according to their position within one representative embryo (*hnt 1*. (Middle) Heatmaps visualising transcriptional profiles of single-cell traces, pooling all biological replicates. Traces were sorted according to transcriptional onset (scale as indicated, black = no expression, yellow = high expression; grey indicates periods where nuclei were not tracked due to temperature related lengths of time-lapses). (Right) Graph shows mean fluorescence values of nuclear traces in each cluster.
Lines in violin plots depict median and 95% confidence intervals. Nuclear traces were cut at transcriptional onset time of first nucleus.

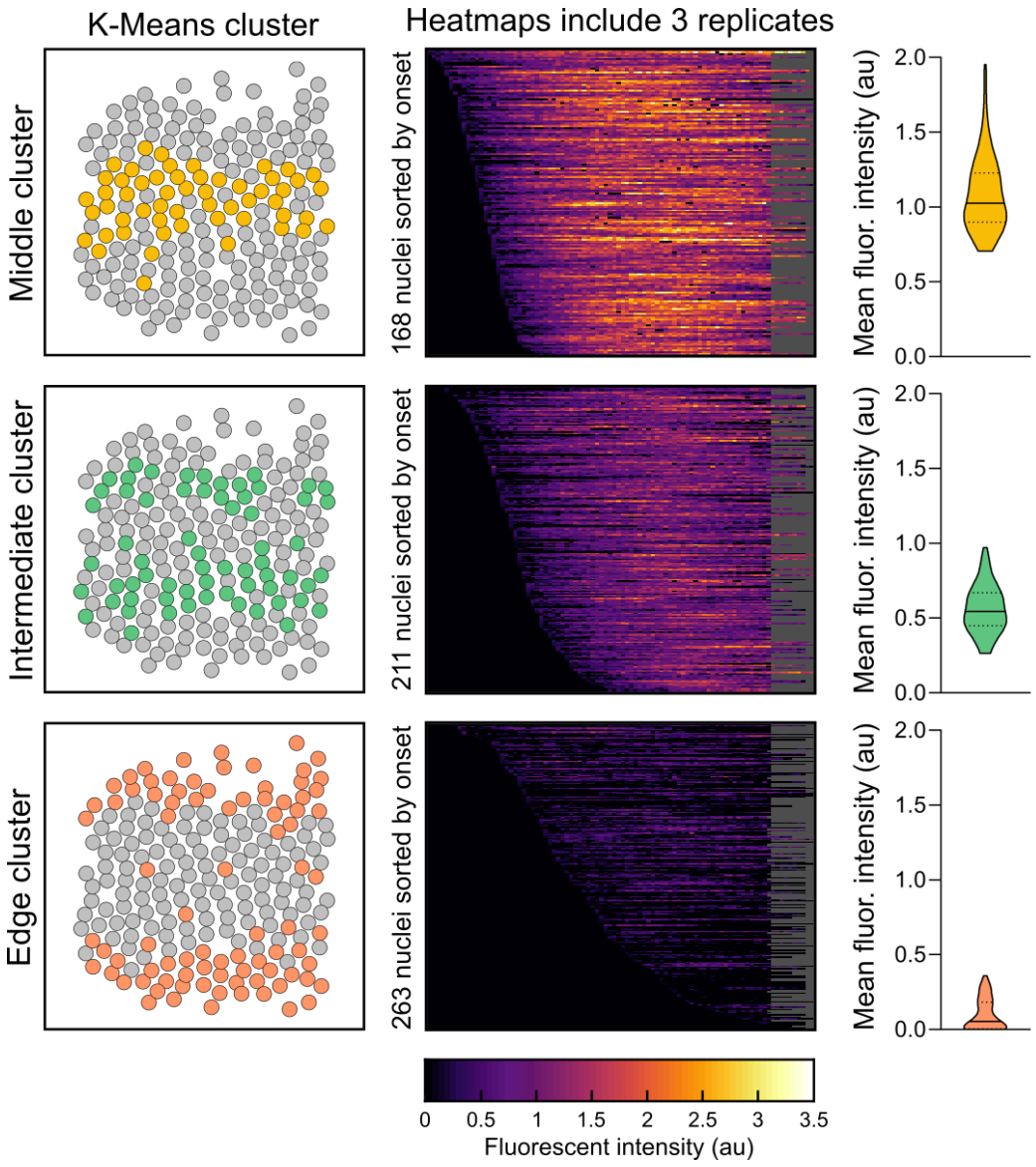



**Figure 4.14: K-Means clustering identifies *ush* transcription domains.** Single cell *ush* transcription profiles were grouped into three clusters (middle, intermediate and edge), using the K-Means algorithm. (Left) Nuclei are coloured by their cluster and plotted according to their position within one representative embryo (*ush 1*). (Middle) Heatmaps visualising transcriptional profiles of single-cell traces, pooling all biological replicates. Traces are sorted according to the transcriptional onset (scale as indicated, black = no expression, yellow = high expression; grey indicates periods where nuclei were not tracked due to temperature related lengths of time-lapses). (Right) Mean fluorescence values of each nuclear trace belonging to a cluster. Lines in violin plots depict median and 95% confidence intervals.

## 4.6 Transcription kinetics of *ush* and *hnt* target genes

### 4.6.1 Modelling of transcription kinetics

The parameters that control transcription can be inferred using a memory-adjusted HMM (Lammers et al. 2019). This computational model is based on a two-state promoter model, where the promoter can be in an ON or OFF state (Figure 4.15 Ai). The promoter transitions between the two states with rates of Kon and Koff (Figure 4.15 Ai). Promoter states can be modelled from single-cell nuclear fluorescence traces (Figure 4.15 Aii) and burst kinetics can be inferred based on the promoter behaviour. Definitions of bursting parameters are shown in Figure 4.15 Aiii (Zoller et al. 2018, Lammers et al. 2019). The model does not include RNA stability or degradation measurements/terms, since the MS2 system captures only nascent transcription sites and therefore RNA production. *hnt* and *ush* transcription kinetics were modelled for different regions of the expression domain individually to investigate the changes in burst parameters in response to different levels of BMP signalling.

#### Burst kinetics of *ush* transcription

Computational modelling of promoter states (performed by Jonathan Bowles) enables the analysis of burst kinetics, which dictate the changes in transcription output that were observed using the MS2/MCP fluorescence traces. Transcription parameters were analysed for the three clusters of *ush* transcription that were previously identified using the K-Means clustering algorithm (see Figure 4.14).

The promoter states and transition rates between them were calculated for every nuclear trace at every time point. The promoter states are shown as "ON" or "OFF" in heatmaps for the different *ush* expression clusters (Figure 4.16 A). All three clusters showed large fractions of promoter activity (red) (Figure 4.16 A). Furthermore, the promoter state profiles of the central cluster and intermediate cluster appear similar, whereas the edge cluster shows a more sparse ON-state pattern (Figure 4.16 A).

One representative fluorescence trace from each *ush* cluster and its inferred promoter states are shown in Figure 4.16 B. A greater number of transcriptional bursts is visible for the central cluster (5) compared to the edge cluster (2), highlighting different burst frequencies (Figure 4.16 B). Since the frequency of bursts is defined by the rate with which the promoter switches between the ON and the OFF state

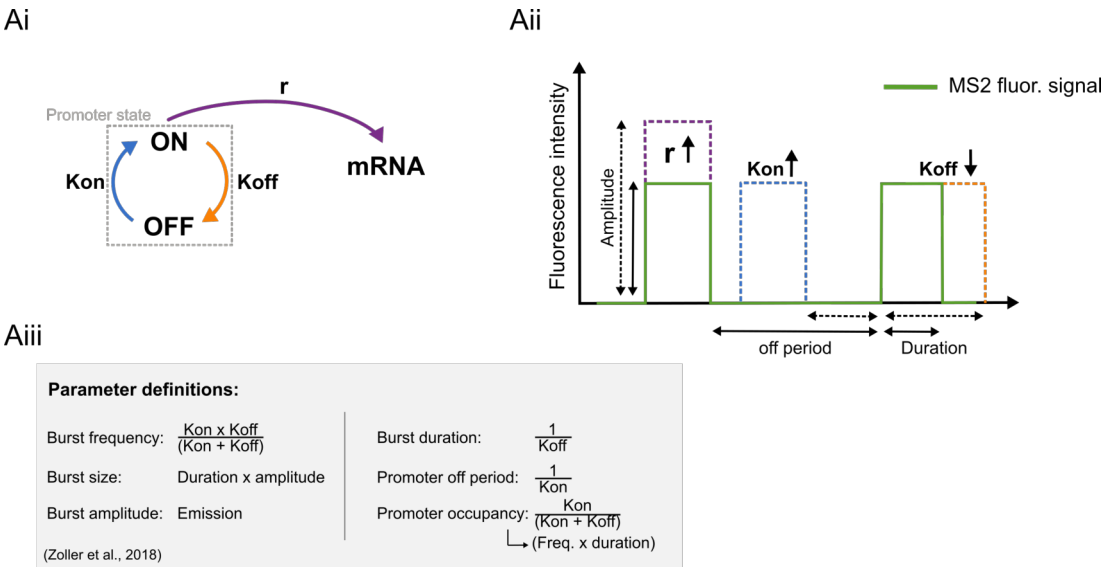



**Figure 4.15: Schematic of two-state promoter architecture and burst kinetics.** (Ai) Schematic of a two-state promoter model where the promoter switches between an active ON and and inactive OFF state. When in the ON state it produces mRNA with the rate of r (purple). The probability of switching between the two states is denoted in Kon (blue) and Koff (orange). (Aii) Nuclear MS2 fluorescence traces can visualise changes in the burst kinetic parameters. An increase in r could lead to increased burst amplitude and therefore greater fluorescence. An increase in Kon could lead to an increase in burst frequency. A decrease in Koff could lead to a longer burst duration and a more persistent fluorescence trace. (Aiii) The burst parameter definitions (Zoller et al. 2018) that will be used to investigate transcriptional changes are shown. This figure was adapted from Falo-Sanjuan et al. (2019).

(Figure 4.15 Aiii), it can be predicted that the edge cluster shows lower Kon and/or higher Koff values.

A global analysis of all burst parameters was performed to examine changes between the spatial domains. All graphs in Figure 4.16 C show mean values of burst parameters for all three *ush* biological replicates. The global analysis of transcription parameters for *ush* clusters will first be discussed between the central and intermediate clusters, and then for the edge cluster. First, changes in the transcriptional dynamics between the central and intermediate cluster will be described. Kon determines the rate with which the promoter switches into the ON state and therefore directly informs the time period that the promoter spends in the OFF position (1/Kon) (Figure 4.15 Aiii). The rate with which the promoter initiates transcription is significantly reduced in the intermediate cluster (Kon, Figure 4.16 C) and the dependent promoter OFF period is significantly increased (Off period, Figure 4.16 C). The frequency of transcriptional bursts is influenced both by Kon and Koff (Figure 4.15 Aiii). Analysis of burst frequency revealed that *ush* transcription

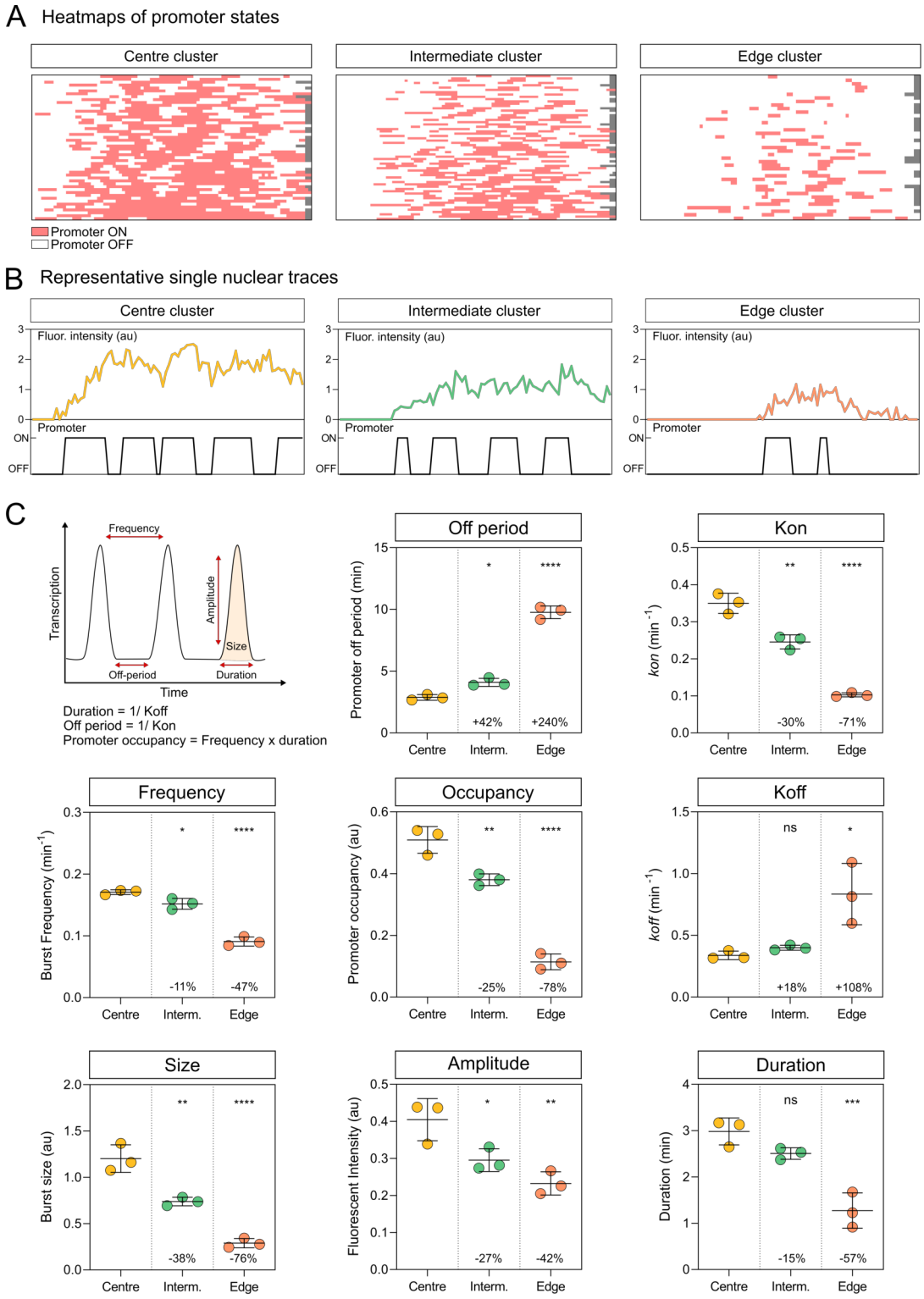



**Figure 4.16: Bursting parameters *ush* transcription.** (A) Heatmaps of promoter states that were inferred using a memory-adjusted HMM. Promoters are shown in active "ON" (red) and inactive "OFF" (white) state. Traces are shown in their clusters. Each row corresponds to a single nuclear trace. Grey boxes indicate periods where nuclei were not tracked. (B) Representative fluorescence traces are plotted from nuclei corresponding to the centre, intermediate and edge cluster and their inferred promoter states are shown below. (C) Schematic shows burst parameters. Global analysis of burst parameters was performed for *ush* transcription in different spatial domains.
*(continued on next page)*

*(continued from previous page)*
Data points are coloured according to clusters. Percentages denote changes of the mean relative to the central domain. Lines in (C) show mean and SD. Significance was tested in (C) using a one way ANOVA test with a Dunnett's multiple comparisons test and denotes the difference to the central cluster. *p < 0.05, **p < 0.01, ***p < 0.001, ****p < 0.0001, ns = not significant. Number of biological replicates n = 3.

bursts on average 0.17 times per minute in the central cluster (Figure 4.16 C). A small but significant reduction occurs between the central and intermediate cluster (Figure 4.16 C). The reduction in burst frequency can be observed in the representative example where the nuclear trace from the intermediate cluster shows four and the central cluster shows five bursts (Figure 4.16 B). A reduction in promoter occupancy is also detected between the central and the intermediate cluster. The fractional mean promoter occupancy describes the time a promoter spends in the ON state as a probability and therefore can adopt values between 0 and 1 (Lucas et al. 2013, Zoller et al. 2018). The promoter occupancy level is dependent on the Kon and Koff rate (Figure 4.15 Aiii). No significant difference is found between the Koff values of the central and intermediate clusters, and therefore, the change in mean promoter occupancy is largely due to changes in Kon (Figure 4.16 C).

Additionally, a difference between the central and intermediate cluster is observed in burst size (Figure 4.16 C). The burst size is a product of burst amplitude and duration and therefore can be influenced by both (Figure 4.15 Aiii). The difference in burst size between the two clusters is shown to be caused mostly by a reduction in burst amplitude, as there is no significant change in burst duration (Figure 4.16 C). Together these observations suggest that the reduced BMP signalling levels between the central and intermediate clusters is decoded transcriptionally as a reduction in burst size and the promoter occupancy. Changes in burst size are shown to be dominated by changes in amplitude and changes in promoter occupancy by changes in Kon values.

Cells in the edge cluster receive lowest BMP signalling levels and shows significant differences in all burst parameters compared to the central and intermediate cluster (Figure 4.16 C). Burst frequency, promoter occupancy and Kon are reduced and there is an increase in the duration that the promoter spends in the OFF position (Figure 4.16 C). Additionally, burst size is low due to low burst amplitude and duration with the latter reflecting an increased Koff rate (Figure 4.16 C). These data suggest that low levels of BMP signalling are interpreted by edge cells as infrequent bursts of small size.

After analysing the changes in burst kinetics on a global level, a more detailed analysis was performed to determine the main BMP signalling responsive parameters. The resolution was increased by calculating burst parameters for every single nucleus. The results are shown here for one biological replicate (*ush 1*). First, it was determined which parameter is most responsive to BMP signalling levels. Results from fixed and live imaging experiments have shown during this study that the BMP signalling gradient leads to changes in the transcriptional output of *ush*, which is quantified here by fluorescence. In order to evaluate which parameter is most likely to account for changes in BMP signalling levels, a correlation analysis was performed between mean fluorescence values (transcriptional output) and burst parameters.

The mean expression profile of cells are shown across the dorsal midline of the embryo (Figure 4.17 A). Single-cell analysis revealed the distribution of promoter occupancy values for individual cells to closely repeat that of mean expression, suggesting that occupancy is likely a key determinant of BMP signalling integration and therefore controls transcription output (Figure 4.16 B). The mean fluorescence and mean promoter occupancy values are shown to be almost perfectly correlated with a correlation coefficient of 0.97 (Figure 4.17 B). Similar to promoter occupancy, the distribution profile of mean Kon values resembles that of mean fluorescence (Figure 4.17 C). A significant correlation was also found for Kon values consistent with the earlier conclusion that this parameter underpins the changes in promoter occupancy (Figure 4.17 C). The expression domain shape could not be reconstructed using mean Koff parameter values, which also shown a weaker correlation to mean fluorescence levels (Figure 4.17 D). Burst frequency is less likely to control BMP signalling integration as the distribution profile is less similar to that of mean fluorescence and the correlation between the two parameters is reduced, again in agreement with the analysis of expression clusters (Figure 4.17 E). Similar to burst frequency, also burst amplitude recapitulates the mean expression pattern of nuclei well but is less correlated than promoter occupancy and Kon (Figure 4.17 F).

Overall, the analysis of promoter states suggest that single cells integrate BMP signalling levels into *ush* transcription processes through changes in promoter occupancy under the control of Kon and burst amplitude.

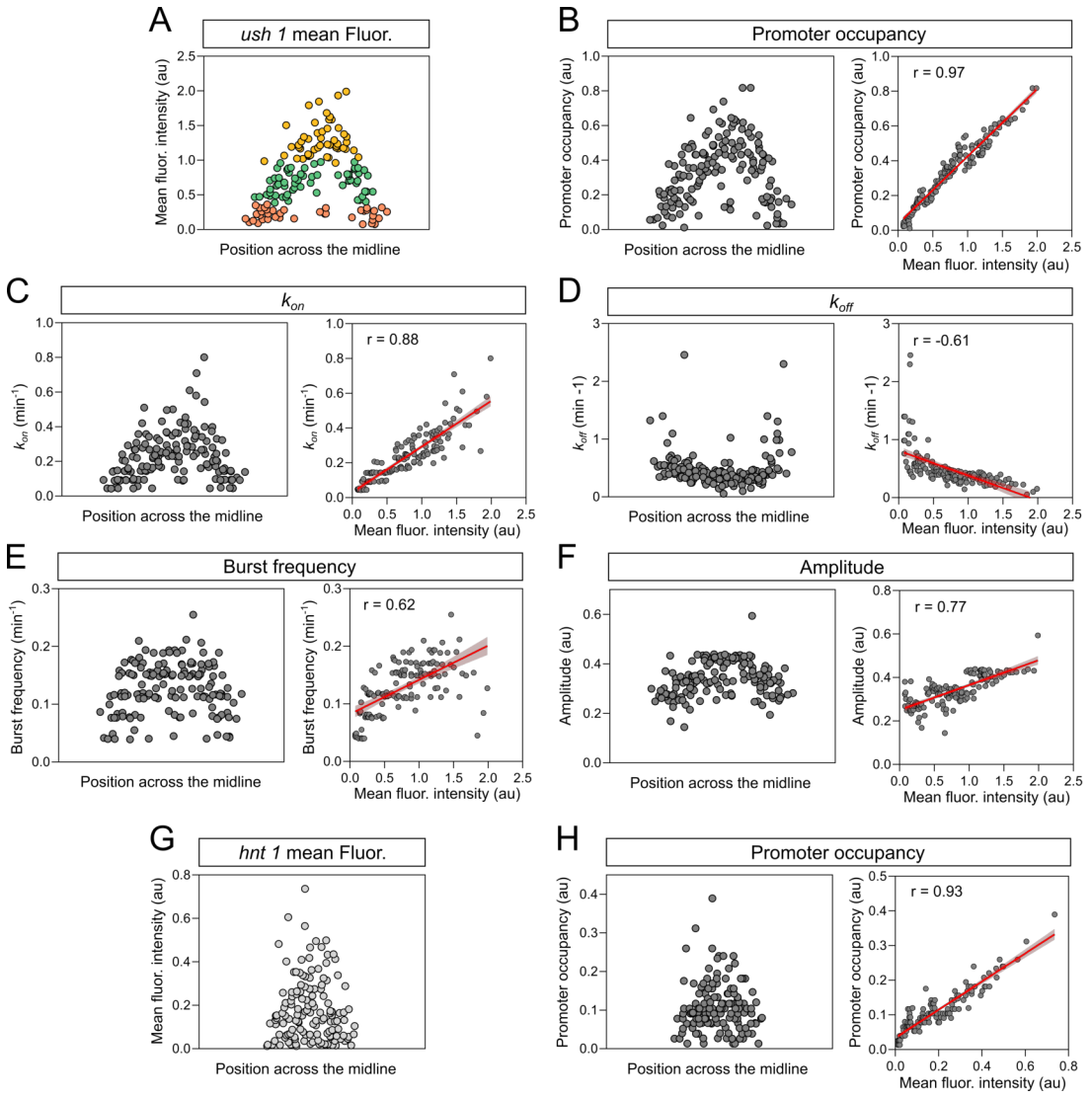



**Figure 4.17: Changes in BMP signalling levels correlate well with promoter occupancy and Kon.** (A) The mean fluorescence profile of nuclei in the *ush* expression domain are plotted. Nuclei are coloured based on the global cluster they belong to. Distribution profiles of mean burst parameters and their correlation to the mean fluorescence intensity are shown according to their nuclear position for promoter occupancy (B), Kon (C), Koff (D), burst frequency (E) and amplitude (F). (G) The graph shows the mean fluorescence profile of nuclei in the *hnt* expression domain. (H) Correlation between mean promoter occupancy and mean fluorescence intensity for *hnt* transcription. Each dot represents one nucleus and lines in (B-F, G) show linear regressions with 95% confidence intervals. Correlation was determined using the Pearson correlation coefficient.

### Burst kinetics of *hnt* transcription

Nuclear traces of *hnt* transcription were clustered into two groups based on the K-Means clustering algorithm (Figure 4.13). In order to investigate how changes in BMP signalling levels in the middle, intermediate and edge of the expression domain affect *hnt* transcription nuclear traces were redistributed into three domains based on their mean fluorescence levels. This separation is valid as it can be noted that the *ush* clustering analysis primarily separates nuclei on mean expression (Figure 4.17 A). The domains are referred to here as centre, intermediate and edge as they broadly reflect nuclei distribution.

The promoter states are shown as "ON" or "OFF" in heatmaps for the different *hnt* transcription domains (Figure 4.18 A). All three domains show shorter periods of promoter activity compared to *ush* transcription clusters, suggesting shorter burst durations (compare Figures 4.18 A and 4.16 A). Overall, the promoter transitions into the ON state are fewer in the edge domain compared to the central domain (Figure 4.18 A). One representative fluorescence trace from each domain is shown together with the inferred promoter traces (Figure 4.18 B). The highest number of transcriptional bursts is observed in the central domain (10) compared to the intermediate (6) and edge domains (3), highlighting different burst frequencies. Additionally, the transcriptional bursts are observed to be short in their duration (Figure 4.18 B).

A global analysis of all burst parameters was performed to examine changes between the spatial domains. All graphs in Figure 4.18 C show mean values of burst parameters for each of two *hnt* biological replicates. Data from the two replicates are in good agreement and show only small differences throughout the analysis (Figure 4.18 C). First changes between the centre and intermediate regions will be described. Promoters of nuclei in the central domain show high Kon levels, meaning that the promoter transitions into the ON state more often resulting in short OFF periods (Figure 4.18 C). The intermediate domain shows a reduction in the Kon rate with which the promoter becomes active and shows an increase in its mean off period length (Figure 4.18 C).

As observed in the single representative traces, the burst frequency is identified to be highest in the central domain with a large reduction in the intermediate region (Figure 4.18 C). The values of Koff and associated duration are largely unchanged between the central and intermediate region (Figure 4.18 C). Therefore, it can be

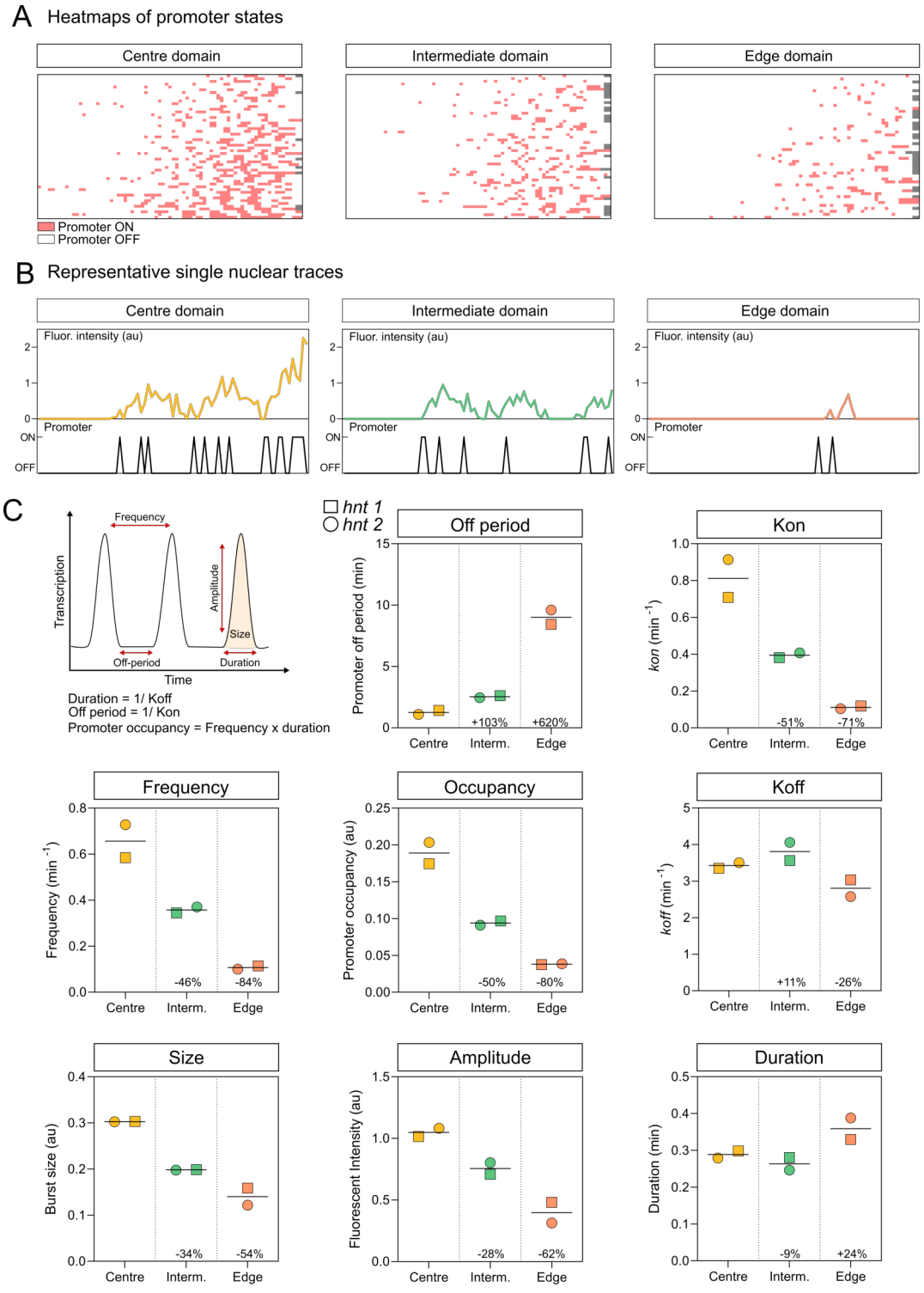



**Figure 4.18: Bursting parameters for *hnt* transcription.** (A) Heatmaps of promoter states that were inferred using a memory-adjusted HMM. Promoters can be in an ON (red) and OFF (white) state and traces were divided into a centre, intermediate and edge domain based on nuclear position. Each row corresponds to a single traces. Grey boxes indicate periods where nuclei were not tracked. (B) Representative fluorescence traces are plotted from nuclei corresponding to the three domains with their inferred promoter states shown below. (C) Schematic shows burst parameters. Global analysis of burst parameters identified for *hnt* transcription in different spatial domains. Percentages denote the mean changes relative to the central domain. Lines in (C) show mean, biological replicates n = 2.

concluded that changes in Kon mostly determine the large decrease in burst frequency in the intermediate domain. Similar to burst frequency, the intermediate region of *hnt* expression shows large reductions in promoter occupancy compared to the centre. As the occupancy value is dependent on Kon and Koff, and Koff is mostly constant between regions, it can be concluded that changes in Kon contribute strongly to the changes observed in occupancy. The burst duration, 1/Koff, is mostly constant between the central and intermediate region. Hence, transcriptional burst durations (1/Koff) should be similar between regions, which can be observed in the representative promoter traces (Figure 4.18 B).

The burst size (number of mRNAs produced per burst) depends on the burst amplitude and the burst duration (Figure 4.15 Aiii). With nearly constant burst duration, the changes in burst size are mostly determined by differences in burst amplitude (Figure 4.18 C). The reduced amplitude in the intermediate region of *hnt* is consistent with differences observed in mRNA production (see Figure 4.12 B). These data suggest that the *hnt* promoter interprets the changes from high to intermediate BMP signalling levels mainly through changes in promoter occupancy but also burst size, which result mostly from changes in amplitude. The change in promoter occupancy is primarily caused by changes in the rate with which the promoter transitions into the ON state (Kon). This interpretation is further supported by a high correlation between the mean promoter occupancy and the mean fluorescence intensity (transcriptional output) when burst parameters were analysed with single-cell resolution (Figure 4.17 G-H). Therefore, the *hnt* promoter seems to respond to changes in BMP signalling levels between the centre and intermediate region in the same way as the *ush* promoter.

The edge domain shows an even further reduction in Kon compared to the intermediate domain with the associated increase in the off-period (Figure 4.18 C). At the edge, the mean frequency is only 0.1 bursts per minute. Since *hnt* transcription is initiated at ~30 min into nc14 (Figure 4.11 B), approximately 2 transcription bursts would be predicted to occur until the end of nc14 at 50 min. This prediction is consistent with two bursts depicted in the representative single nuclear trace (Figure 4.18 B). Promoter occupancy as well as the burst size is further reduced in the edge region (Figure 4.18 C). Again, the decrease in burst size between the intermediate and edge domain is mostly caused by a decrease in amplitude as the burst duration remains largely unchanged (Figure 4.18 C). Overall these data suggest that low levels of BMP signalling result in very infrequent bursts that are characterised by a low amplitude and short duration.

To compare *hnt* and *ush* burst profiles, they were simulated based on mean burst parameters for amplitude, duration and off-period in the central cluster. Bursts were generated for a 30 min period for *ush* and a 20 min period for *hnt* corresponding to the maximum length of transcription that was visualised in this study (see Figure 4.11 B). The burst duration identified for *hnt* clusters is shorter than that of *ush* transcription (central domains 3 min for *ush* vs 0.3 min for *hnt*) (Figure 4.19). At the same time burst amplitude (1 au *hnt* vs 0.4 in *ush*) and frequency (0.6 bursts/ min *hnt* vs 0.17 bursts/ min *ush*) are greater during *hnt* transcription compared to *ush* (Figure 4.19). Therefore, *hnt* and *ush* transcription show different burst kinetics, where *hnt* transcription exhibits high amplitude, short duration but very high frequency bursts, compared to *ush* transcription which occurs with longer, more infrequent and lower amplitude bursts (Figure 4.19). Nevertheless, promoters integrate different levels of BMP signalling through a similar molecular mechanism during both *hnt* and *ush* transcription. These data suggest that a small reduction in signalling levels are modulated through changes in the promoter occupancy and amplitude, although duration is also reduced when signalling levels are very low.

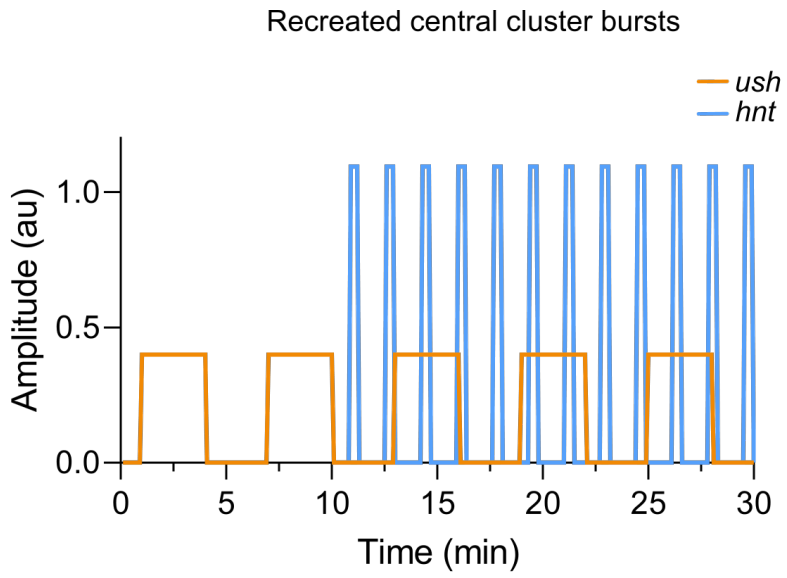



**Figure 4.19: Comparison of *ush* and *hnt* transcriptional burst profiles.** Bursting simulation of *hnt* (blue) and *ush* (orange) transcription. Bursts were calculated based on mean burst parameter values from the central clusters and are shown for a transcription period of 30 min for *ush* and 20 min for *hnt*.
Biological replicates: *hnt* n = 2, *ush* n = 3. Lines in (C) show mean and SD.

## 4.7 Changes in promoter occupancy integrates BMP signalling levels

The previous data predict that increasing BMP signalling levels will lead to greater occupancy values. In order to test this hypothesis, the influence of increased BMP signalling on *ush* transcription dynamics was investigated by introducing ectopic Dpp into the embryo. The concentration of BMP signalling was increased through ectopic expression of *dpp* (Ashe et al. 2000). One embryo was analysed to obtain preliminary data, and a movie still is shown in Figure 4.20. The embryo tested was heterozygous for the *st2-dpp* transgene and shows an expanded *ush* expression domain (Ashe et al. 2000).

A strip of cells spanning the full width of the embryo imaged was analysed in the area corresponding to the stripe 2 expression domain (imaging and analysis domain indicated on embryo schematic, Figure 4.21 A). The expression domain is shown to extend into almost all of the analysed region (Figure 4.21 A). This observation is consistent with results in Chapter 3, where *ush* was expanded in *st2-dpp* embryos (see Figure 3.17) as well as published data (Ashe et al. 2000).

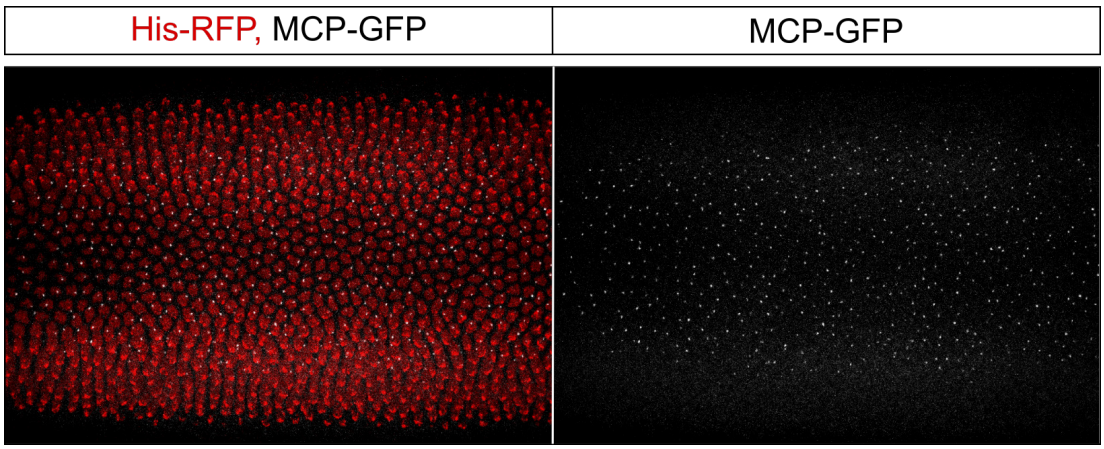



**Figure 4.20: Time-lapse microscopy detecting *ush* transcription in a *st2-dpp* heterozygous background.** Still of a maximum intensity projection (54x 1 μm stacks) during nc14. Nuclei visualised by His-RFP (red) and *ush* transcription by MCP-GFP fluorescence (grey). Time resolution 20 sec/frame. The embryo was imaged in a dorsal position with anterior to the left. Full movie can be found at `https://www.tiny.cc/thesis_videos`.

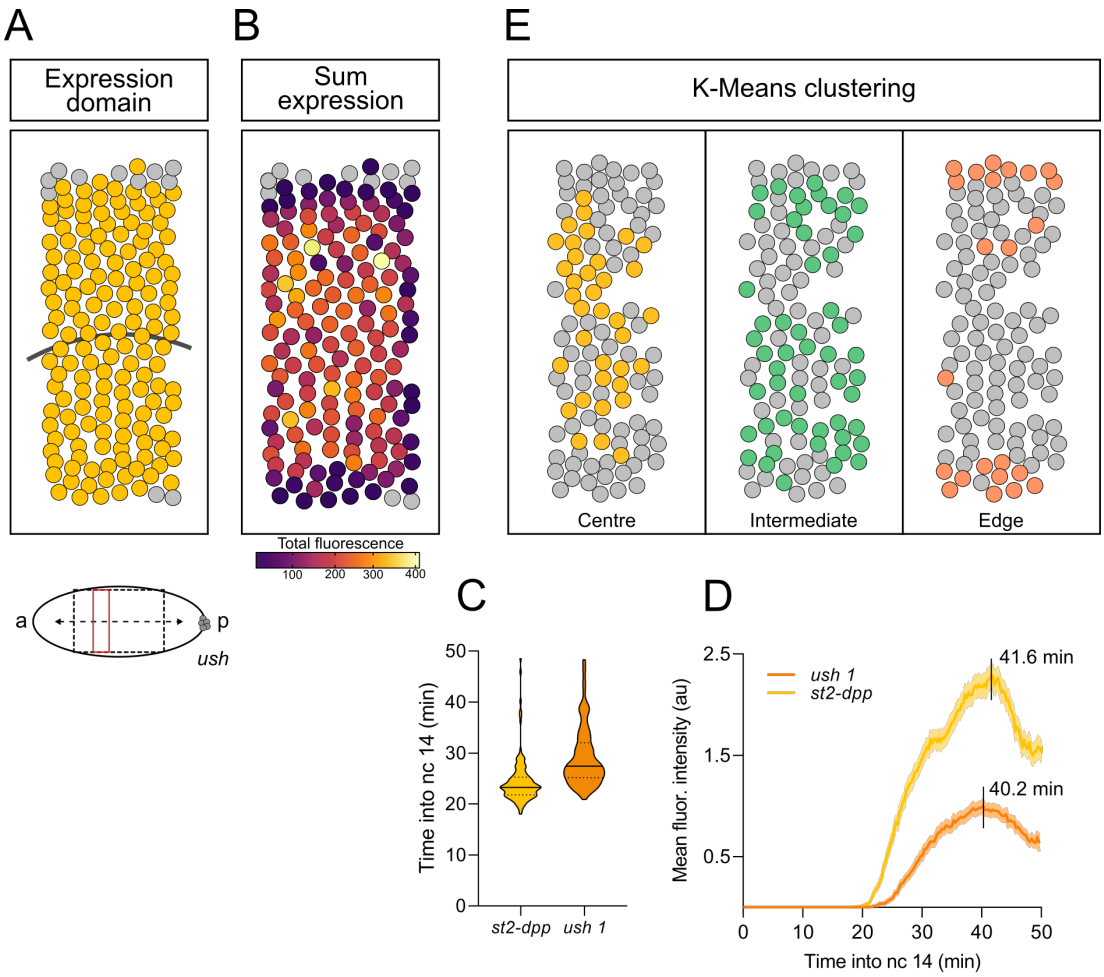



**Figure 4.21: Transription profile of *ush* in *st2-dpp* heterozygous background.** (A) Embryo schematic shows imaging domain (black) and analysis domain (red) indicated as rectangles on embryo schematic. Cumulative expression domain is depicted for the analysis domain. (B) Sum fluorescence shown as a heatmap for each nucleus in the core expression domain. (C) Transcription onset for *ush* in wt (orange) and *st2-dpp* (yellow) background. (D) Graph shows mean fluorescence values over time. (E) Clustering results shown for ush transcription in the *st2-dpp* embryo.
Lines in violin plots (C) depict median and 95% confidence intervals. Line plots (D) depicts mean and SEM. One biological replicate for *st2-dpp* (121 nuclei) and one representative *ush* replicate shown for comparison (*ush 1* = 209 nuclei).

In *st2-dpp* embryos the total fluorescence produced by nuclei is shown to be increased (max = 400) compared to wt *ush* transcription (max wt= 180) (compare Figure 4.12 A and Figure 4.21 B). The increase in BMP signalling level is also shown to result in an earlier transcription onset time (median = 23.3 min) compared to the representative *ush* embryo (*ush 1*, median = 27.5 min) (Figure 4.21 C).

In addition to the increase in sum fluorescence, an increase in the mean fluorescence over time is also detected for *ush* transcription in the *st2-dpp* background (Figure 4.21 D). The peak height of the mean *ush* transcription is 3 fold higher in the *st2-dpp* embryo compared to wt but the peak is reached at almost the same time (40.2 vs 41.6 min, Figure 4.21 D).

In order to evaluate the transcriptional changes resulting from increased BMP signalling levels, the nuclei were distributed into clusters. Similar to *ush* embryos, the *st2-dpp* embryo was grouped into three clusters (Figure 4.21 E). The centre and intermediate clusters are more intermixed and less clearly distinguished by their distance to the dorsal midline compared to *ush* embryos (compare to Figure 4.14). This observation suggests that BMP signalling levels are evenly high around the dorsal midline in the *st2-dpp* embryo with a sharp decrease towards the edge cluster.

**Burst kinetics of *ush* transcription in a heterozygous *st2-dpp* background**

Similar to transcription in *ush* wt embryos, the promoter states of single nuclear traces were inferred using the memory-adjusted HMM. Representative fluorescence traces and corresponding promoter states from each cluster are shown in Figure 4.22 A.

Global analysis of burst kinetics reveals an increase in *ush* burst size in all clusters of the *st2-dpp* embryo compared to wt embryos (Figure 4.22 B). Again, the main contribution to the increase in burst size is shown to occur through an increase in burst amplitude, which is increased by approximately two-fold in all clusters (Figure 4.22 B). The burst duration is considered to be similar between genotypes (high duration in intermediate cluster is likely an outlier due to the analysis being dominated by certain nuclear traces, Figure 4.22 B) suggesting that increased signalling does not affect burst duration.

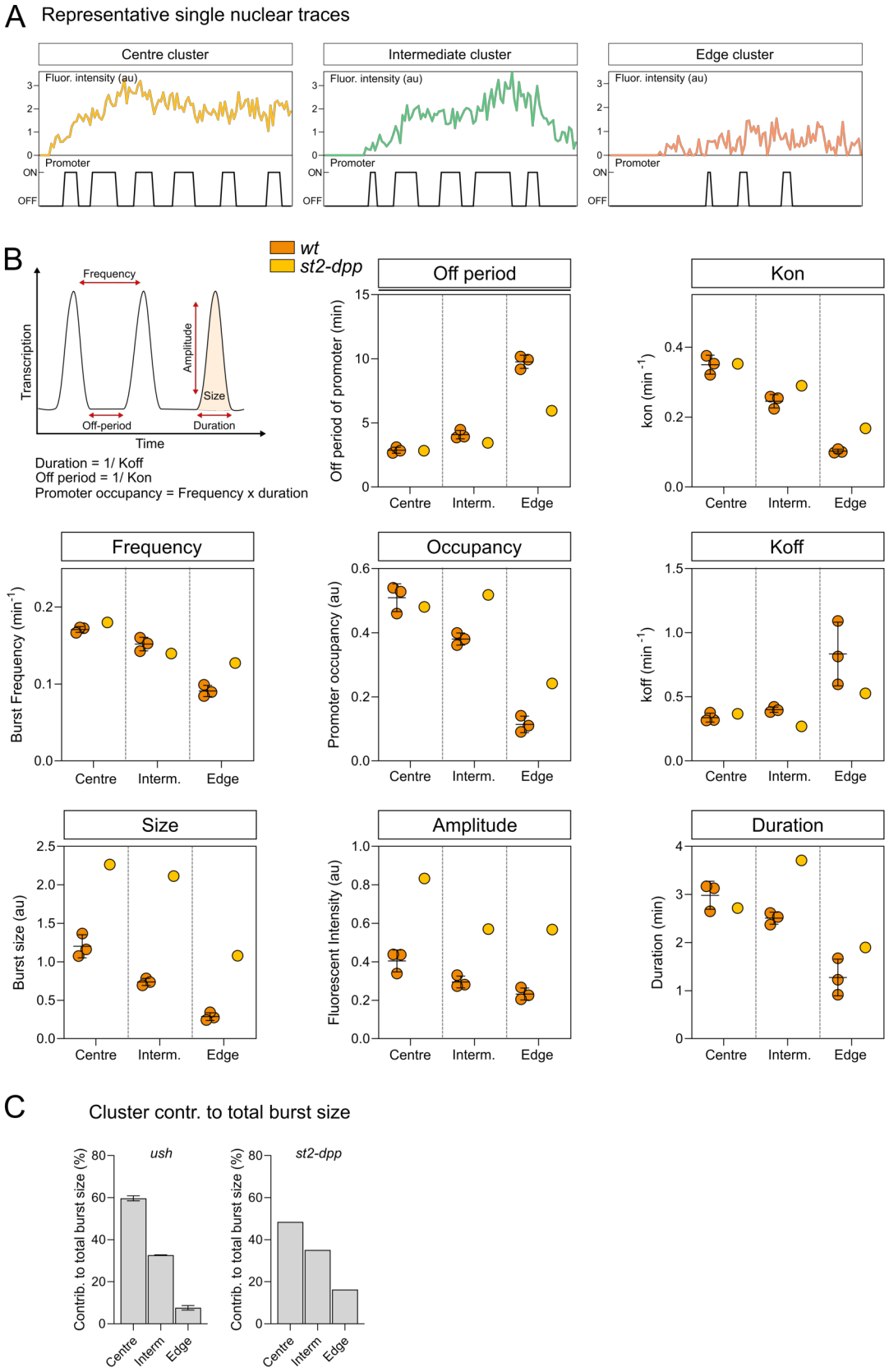



**Figure 4.22: High BMP signalling levels are decoded by *ush* promoter occupancy.** (A) Representative single nuclear fluorescence traces with their corresponding promoter states below. (B) Schematic outlines burst parameters. Global analysis of *ush* burst kinetics in wt (orange) and *st2-dpp* (yellow) backgrounds.
*(continued on next page)*

*(continued from previous page)* (C) Contribution of individual clusters to the total mean burst size, calculated based on mean burst parameters from (B).
Lines in (B,D) show mean and SD. Number of biological replicates, *ush* = 3, *st2-dpp* = 1.

Increased promoter occupancy rates are detected for the intermediate and the edge cluster in the *st2-dpp* embryo. The intermediate cluster now shows a similar level to the centre cluster of a wt and *st2-dpp* (Figure 4.22 B). Even the edge cluster, which in wt embryos shows poor transcription output and initiation, can be improved for promoter occupancy by increasing the levels of BMP (Figure 4.22 B). The promoter occupancy is not further enriched in centre of the *st2-dpp* embryo (Figure 4.22 B). This observation suggests that *ush* promoter occupancy levels are saturated in the centre and therefore cannot be improved further by the increased signal. These data provide further evidence that BMP responsive promoters decode BMP signalling levels through adjusting promoter occupancy. Based on previous results in wt embryos, it would be predicted that the change in promoter occupancy mainly occurs through increased levels of Kon (see Figure 4.16 C). This appears to be the case as Kon decreases with reduced signalling levels although there is a small decrease in Koff also 4.22 B). The burst frequency remains unchanged between genotypes in the central and intermediate cluster, likely due to changes in both Kon and Koff (Figure 4.22 B). While this preliminary analysis supports the idea that BMP signalling effects promoter occupancy, additional biological repeats will be needed further investigate how the parameters change.

Expanding the *ush* expression domain in the *st2-dpp* embryo and increasing amplitude and promoter occupancy in the intermediate and edge cluster changed the contribution of clusters to the total expression output (total mRNA produced by field of cells) (Figure 4.22 C). In wt embryos the central cluster of the *ush* expression domain shows approximately 60% of the transcriptional activity, whereas the edge cluster contributes little (Figure 4.22 C). Addition of extra Dpp results in the increased contribution from the intermediate and edge clusters to the total mRNA production (Figure 4.22 C).

Overall, results obtained in this section propose that signal integration occurs through changes in target gene promoter occupancy, primarily due to increasing Kon, the rate with which the promoter switches on. Increased BMP signalling also increases burst size, due to the effect on amplitude. The next chapter will investigate the role of promoter during signal integration further.

## 4.8 Live imaging of both *ush* alleles

The previous chapter identified nuclei with monoallelic transcription, where only one homologous gene allele was active. In the live imaging setup used until now, transcription of one allele was visible due to the chosen crossing scheme. Further investigation of monoallelic transcription was possible by visualising both alleles using the MS2/MCP system.

Here, one embryo was analysed to generate preliminary results. A still from the time-lapse experiment can be found in Figure 4.23. This embryo was positioned dorsolaterally to focus on the edge of the *ush* expression domain. Still images from the movie reveal monoallelic transcription in edge nuclei (Figure 4.24 Ai). Nuclei are shown (yellow arrowheads) at both 26 and 45 min into nc14 where one transcription site is active compared to more dorsal nuclei where biallelic transcription is visible (Figure 4.24 Ai). The position of monoallelic nuclei at the expression domain border corresponds to earlier observations using RNA-FISH and smFISH probes. Many monoallelic nuclei show transcription from both alleles at later or earlier time points (examples below). However, a small subset of nuclei was identified where *ush* transcription occurs only through monoallelic transcription and a representative nucleus is shown in Figure 4.24 Aii. This nucleus, selected from the edge domain (white outline in Figure 4.24 Ai), showed *ush* transcription from one allele during the transcriptional active period in nc14 (Figure 4.24 Aii). Stills from the time-lapse movie show *ush* transcription initiation at 34 min into nc14 (note, onset much later than median of the *ush* expression domain Figure 4.11 B) and until its termination at approximately 50 min, transcription is visible from the same allele (Figure 4.24 Aii). Further, the fluorescence produced by this nucleus reaches its maximum at approximately 40 min into nc14 and therefore is consistent with previous observations of the *ush* locus (Figure 4.24 Aiii). These data indicated that monoallelic transcription in the *ush* expression domain is often dynamic but can also occur in a "sustained" fashion where one allele is actively transcribed and the second allele never initiates transcription.

The single transcription site was tracked in three-dimensional space to ensure that it was not lost and that the second allele did not take over transcription at any point. A still from a three-dimensional representation of the transcription site tracking and corresponding maximum projected images can be found in Figure 4.25.

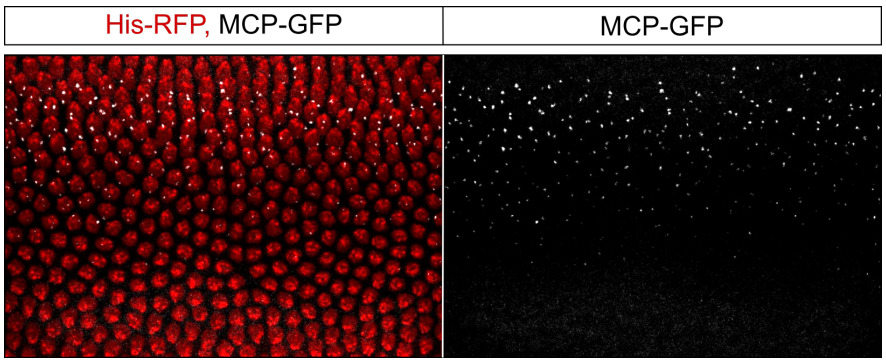



**Figure 4.23: Time-lapse microscopy investigating *ush* transcription on both alleles.** Still of a maximum intensity projection (40x 1 μm stacks) during nc14. Nuclei visualised by His-RFP (red) and *ush* transcription by MCP-GFP fluorescence (grey). Time resolution 15.2 sec/frame and images obtained with 2.15 x confocal zoom. The embryo was imaged dorsolaterally with anterior to the left and dorsal at the top. The full movie can be found at `https://www.tiny.cc/thesis_videos`.

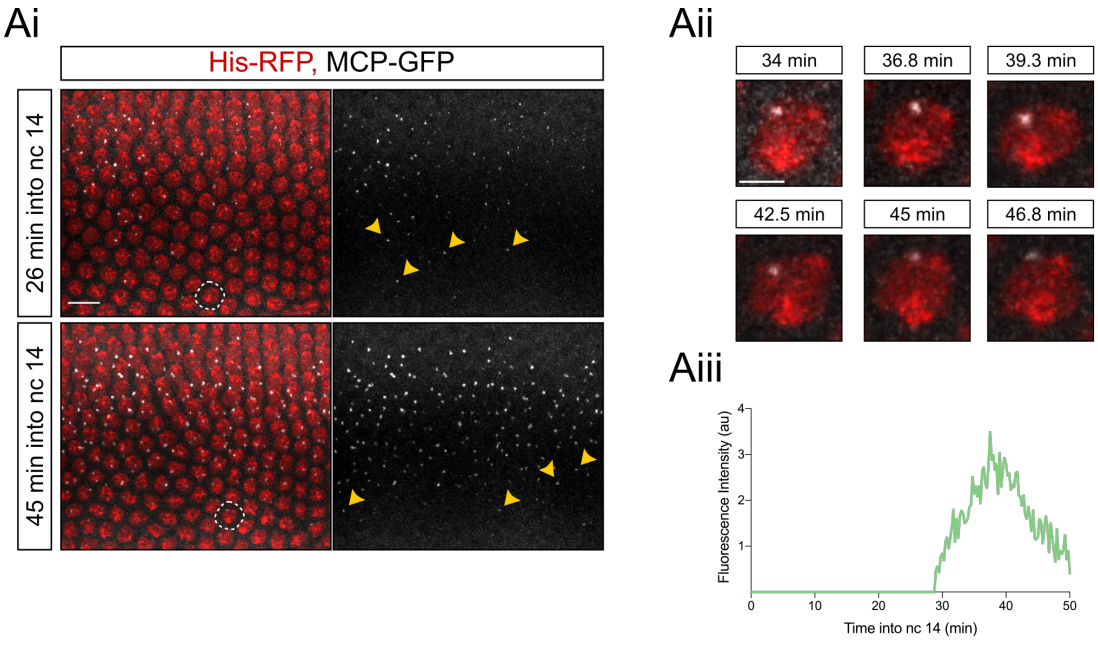



**Figure 4.24: Homozygous embryos for the *MS2-ush*.** (Ai) Still maximum projected images from a time-lapse experiment at 26 min and 45 min into nc14. Nuclei are visualised by His-RFP (red) and active transcription by MCP-GFP (grey). Monoallelic nuclei are highlighted by yellow arrowheads. (Aii) Time course of a single nucleus showing monoallelic transcription (white outline in Ai). (Aiii) Fluorescence trace over time from the nucleus in Aii.
Scale bar = 10 μm (A) and 2 μm (Aii). The embryo was positioned dorsolaterally with anterior to the left and the dorsal midline at the top.

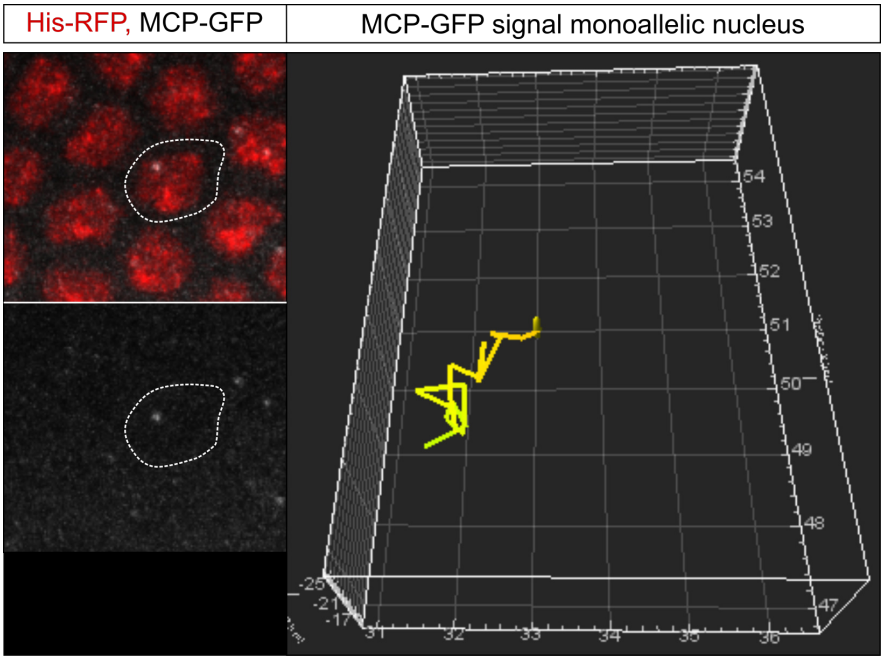



**Figure 4.25: Sustained monoallelic transcription of *ush* throughout nc14.** Still of a maximum intensity projected image (40x 1 μm stacks) during nc14 on the left, where His-RFP visualises nuclei (red) and *ush* transcription is shown by MCP-GFP fluorescence (grey). Sustained monoallelic transcription is shown in a single nucleus (white outline) and the corresponding three-dimensional tracking is shown on the right side. Colour of dragon tail visualises time. The full video of this analysis can be found at `https://www.tiny.cc/thesis_videos`.

It is possible that what appears to be monoallelic transcription corresponds to both alleles sitting in the same axial plane. In that case, it would be difficult to resolve both alleles. However, if this was the case, the combined large transcription site containing two alleles would be predicted to produce higher fluorescence. Since the observed fluorescence for monoallelic transcription sites is similar to fluorescence observed by independent alleles from biallelic nuclei at the same relative position in the expression domain (compare Figure 4.24 Aiii with 4.26), it can be concluded that two alleles rarely sit in the same axial plane.

### 4.8.1 Coordinated dynamic transcription between alleles

This section will investigate the behaviour of individual *ush* alleles further. Nuclei that showed biallelic transcription for at least one-third of the transcriptionally active period were chosen for this analysis. All nuclei analysed were located in the intermediate to edge region of the *ush* expression domain. This region was chosen as nuclei at the extreme edge are predicted to show poor transcription output, whereas nuclei in the centre are predicted to show little variation in fluores-

cence traces due to high levels of transcription. Transcription from the independent loci was detected as distinct sites of fluorescence and was tracked independently through manual tracking. Three representative examples of allele transcription are shown in Figure 4.26 Ai-Aiii. Although the transcription behaviour of alleles in the same nucleus produces similar overall profiles, their detailed behaviour was not synchronous. The small variations in peaks and troughs were not synchronised (Figure 4.26 Ai-Aiii). This behaviour is predicted from transcription loci that function via an uncorrelated manner (Harper et al. 2011, Little et al. 2013, Fritzsch et al. 2018, Falo-Sanjuan et al. 2019). The fluorescence produced by the representative nuclei varied (highest in Ai and lowest in Aiii) but the two alleles within a nucleus produced approximately the same maximal output (Figure 4.26 Ai-Aiii). The time-lapse data of individual allele transcription show periods where the same nucleus displays monoallelic and biallelic transcription (monoallelic periods shaded in grey, Figure 4.26 Ai-Aiii). Fixation of embryos during any of those periods would result in the observation of monoallelic transcription and explain the observations made in the previous chapter.

Given the initial observation that transcription output of alleles seem coordinated in their intensity, their transcription output was analysed in more detail. The fluorescence produced by allele A and allele B at each time point was plotted and is shown to be correlated (Figure 4.27 A), therefore suggesting synchronicity in the transcription output between alleles (Figure 4.27 A).

Since the fluorescent output is correlated the synchronicity of transcriptional onset time was investigated. When the allele data was pooled, the mean onset time is shown to be 26.5 min and little variation is observed (Figure 4.27 Bi). This time corresponds to onset times observed earlier for nuclei positioned in the intermediate region of the *ush* expression domain (compare, Figure 4.10 B). The time difference between the transcriptional onset of allele A and allele B is shown to be small with a median difference of 2 min (Figure 4.27 Bii), suggesting high synchronicity even if the allele onset is not directly correlated.

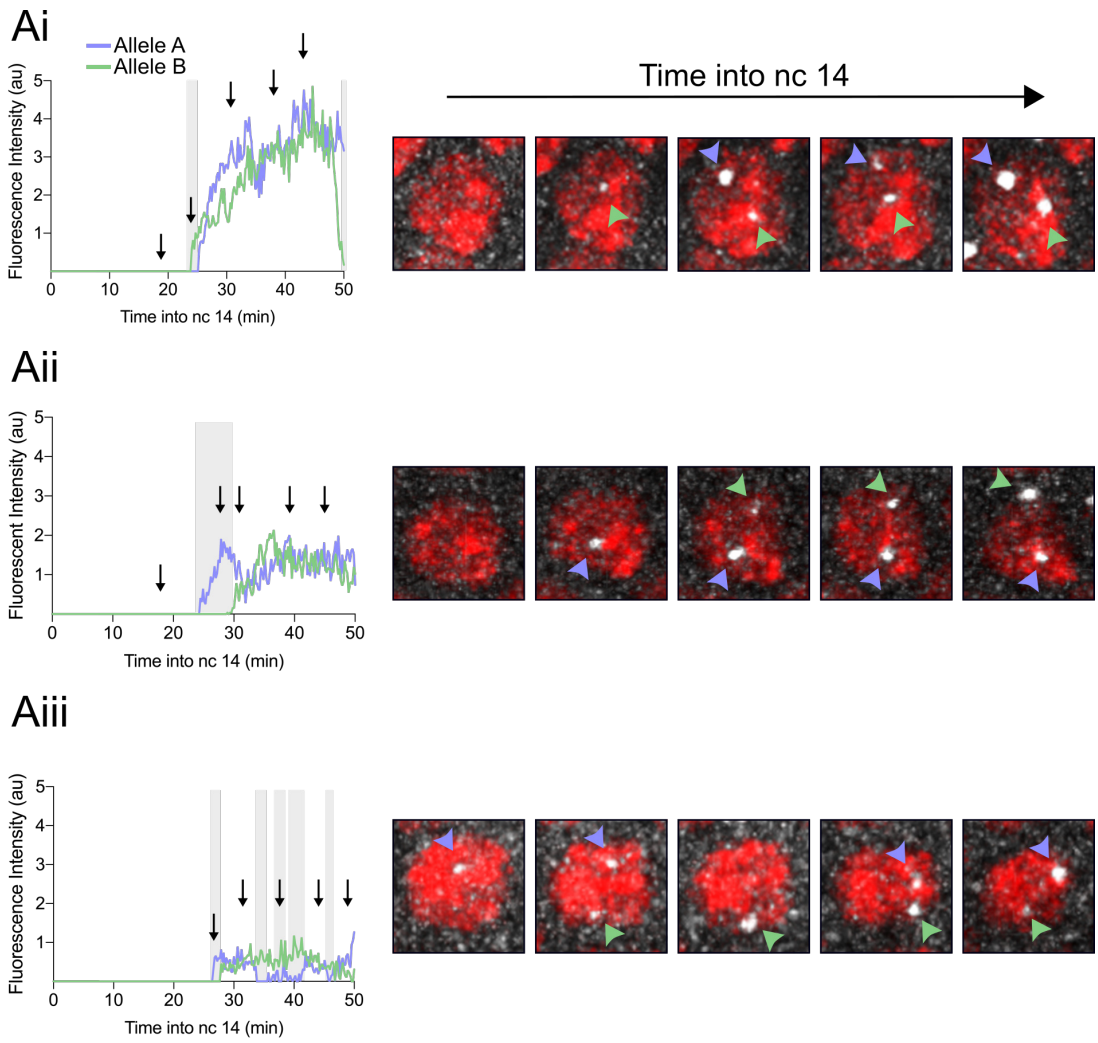



**Figure 4.26: Dynamic coordinated transcription of two homologous *ush* alleles.** Transcription profiles from three nuclei (Ai-Aiii) with different transcription profiles. Absolute transcription was highest in (Ai) and lowest in (Aiii). Alleles were assigned randomly (allele A in blue and allele B in green) and grey shading indicates periods of monoallelic transcription. Maximum projected images from the time-lapse experiment show nuclear transcription states at time points indicated by black arrows on transcription profiles. Nuclei are labelled with His-RFP (red) and *ush* transcription with MCP-GFP (grey).

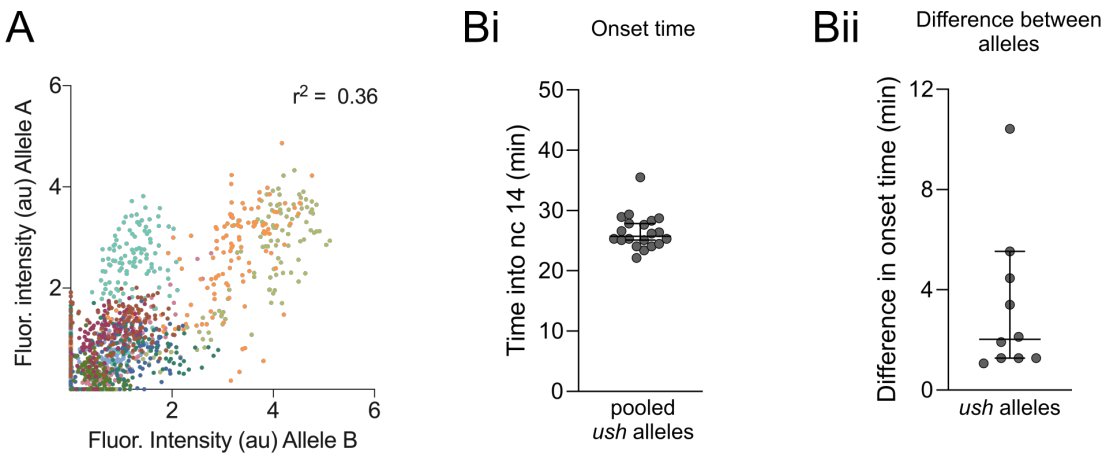



**Figure 4.27: Coordinated transcription output by homologous *ush* alleles.** (A) The graph shows the fluorescence intensity measurements of alleles A and B. Each dot represents a fluorescence measurement at a single time point and colours represent the nuclear identity, correlation coefficient $r^2 = 0.36$. (Bi) Onset time of alleles were pooled and plotted. Mean = 26.5 min. (Bii) The difference in onset time between allele A and allele B was calculated. Median = 2 min.
Lines in (Bi) represents Mean and SD, and in (Bii) depict median and 95% confidence intervals. Correlation between the fluorescence of allele A and allele B was calculated using the Spearman correlation coefficient.

## 4.9 Discussion

Experiments described in this chapter aimed to investigate the transcriptional burst kinetics that underlie BMP target gene expression and revealed the following:

1. The transcription onset of *ush* precedes that of *hnt*.

2. K-Means clustering identified two clusters of nuclear transcription signatures in *hnt* and three clusters in *ush* expression data.

3. The BMP signalling gradient is interpreted by single cells through changes in burst size and promoter occupancy.

4. Dynamic and sustained monoallelic transcription occurs during *ush* transcription.

**Improvements of the MS2/MCP imaging system and in image analysis**

The snapshot data that were obtained by smFISH in fixed tissues in Chapter 3 lack the temporal resolution that is needed to characterise transcription kinetics and distinguish between transcriptional heterogeneity (bursting) and inherent noise (Boettiger & Levine 2009, Porcher et al. 2010). The MS2 live imaging system offers scientists the possibility to investigate molecular mechanisms and production kinetics of transcription in real-time and with single-cell resolution. Scientific questions can now be addressed that were previously impossible, due to limitations in microscopy techniques. High-speed microscopy setups generate data sets with a final time resolution of seconds in the z-direction (for example Lim et al. (2018)). These setups allow data collection of molecular processes that occur in the range of seconds to minutes.

The MS2 imaging system, as it is used currently in the *Drosophila* field, presents a few limitations and three considerations and areas for improvement will be discussed here. Firstly, the low signal-to-noise ratio limits signal detection, which is mainly due to the fluorescent background that is generated by unbound MCP-GFP. Background levels can be reduced for studies of nascent transcription by targeting any unbound MCP-GFP to the cytoplasm through removing a nuclear localisation signal. This study had used MCP-GFP with the nuclear localisation signal removed. The nuclear localisation signal was used previously to sequester unbound MCP-GFP

in the nucleus and reduce background fluorescence in the cytoplasm for RNA localisation studies (Bertrand et al. 1998). To optimise the signal-to-noise ratio further, this study identified that using MCP-GFP/+ reduced the background fluorescence almost by half. MS2/MCP studies in *Drosophila* embryos rely on a pool of MCP-GFP protein that is deposited by the mother (Garcia et al. 2013). This study has shown that mothers heterozygous for the MCP-GFP transgene locus, deposit lower amounts of MCP-GFP which led to a 50% reduction in background fluorescence.

Increasing the signal-to-noise ratio by reducing background fluorescence is essential for image analysis. The fluorescence of active transcription sites is, on average, only twice as high as the background fluorescence. Previous studies have acknowledged this fact by stating that the errors made in fluorescent background estimation dominate the overall imaging error (Garcia et al. 2013, Bothma et al. 2015). By reducing the background fluorescence, this study improves the detection threshold during semi-automated image analysis and therefore, should improve the sensitivity of overall data collection during analysis.

In other approaches to improve the signal-to-noise ratio, multiples of the 24 MS2-loop repeats have been incorporated into test genes. 128 loops have been shown to improve detection thresholds, prevent rapid photobleaching and visualise cytoplasmic single mRNA molecules (Tantale et al. 2016). However, the increased number of MS2-loops would introduce a much larger DNA sequence into the genome, which could lead to artefacts.

Furthermore, no consensus analysis pipeline exists for MS2 live imaging data sets within the *Drosophila* community. This study developed an analysis framework that offers transferable applications. The created analysis pipeline can be applied to images that are obtained using different microscopy systems and therefore, could be applied to a wide range of applications to extract quantitative information from different tissues or organisms.

A second consideration when using the MS2 system is the lack of MS2-loop saturation with MCP-GFP molecules. Quantification of the MS2 signal assumes a proportional relationship between the fluorescence intensity and the number of mRNA molecules that are produced. The MCP-GFP fusion protein binds to MS2-loops as a dimer (LeCuyer et al. 1996). Hence, 24 copies of the MS2 stem-loops should be bound by 48 MCP molecules resulting in 48 fluorophores at maximum occupancy. However, it was observed in cell culture, that coat proteins did not occupy all MS2-

loops and that MS2 labelling was not uniform between cells (Fusco et al. 2003, Wu et al. 2012). The *in vitro* study showed that on average, only 40% of MS2-loops were bound by MCP-GFP. To circumvent this problem, Wu et al. designed a single chain tandem dimer of MCP, which they showed increased the uniformity of MS2 labelling and occupancy to close to 100% (Wu et al. 2012). Therefore, assuming full occupancy binding by MCP-GFP (without tandem dimer) in *Drosophila* studies, including this study, introduces an analysis error. So far, no *Drosophila* study has reported the use of the tandem dimer MCP construct, and therefore, the results obtained in this study are directly comparable to other studies in the *Drosophila* field. The Ashe lab has addressed the lack of a MCP tandem dimer and has generated a fly line expressing it. However, due to time constrains of this study the tandem dimer fly line was not introduced as a new variable into the experiments in this study. All future experiments carrying on from this study will use the new to ensure full MS2-loop occupancy.

The third consideration when using the MS2 system is the choice of MS2-loops. A recent article suggested that MS2 tagging in yeast prevents RNA degradation and forms decay fragments, which can still bind MCP and have generated false-positive signal (Garcia & Parker 2015). The major problem was identified to be the high affinity of MCP to the MS2-loops (Carey & Uhlenbeck 1983), which led to the decay artefacts. To produce the initial MS2-loops, used in many experiments (and in this study), the bacteriophage MS2 sequence was mutated. A single Uracil to Cytidine substitution within the stem-loop sequence has been shown to increase the affinity to MCP by 10-fold and to decrease the dissociation kinetics by approximately 90-fold over that of wildtype (Lowary & Uhlenbeck 1987, Valegård et al. 1997). This higher affinity aids visualisation but prevents RNA degradation (Garcia & Parker 2015). To circumvent degradation fragments, the Singer lab generated new MS2-loop sequences where the initial mutation of U to C was reversed (Tutucci et al. 2018). *Drosophila* studies have yet to make use of the new version of MS2-loops and therefore, no data on their practicality and effects on the signal affinity exist. Furthermore, the problems resulting from the high MS2/MCP-GFP bond stability will occur post-transcriptionally and therefore, will not alter the results obtained in this study. Although MS2 imaging encounters limitations such as the ones discussed above, it is still a valuable system to obtain single-cell data, as it provides direct access to the temporal dimension of transcription while preserving the tissue morphology and microenvironment.

This study used the CRISPR/Cas9 genome editing system for the targeted insertion of MS2-loops into endogenous gene loci of the BMP target genes *hnt* and *ush*. Together with Lim et al. (2018), this study is the first *Drosophila* study that has investigated endogenous gene loci using the MS2 system. Most previous studies in the *Drosophila* embryo have addressed scientific questions using short reporter transgenes that often contain minimal enhancer and promoter sequences to drive transcription (for example Garcia et al. (2013), Falo-Sanjuan et al. (2019)). However, transgenes may lack critical unknown regulatory regions and therefore might not recapitulate endogenous transcription levels or general transcription behaviour.

**Transcription dynamics of *hnt* and *ush* show differences**

This study used *Drosophila* lines that contain MS2-loops in the gene's endogenous 5'UTR. Studies in *Drosophila* embryos have inserted MS2-loops into transgenic constructs and positioned them in different genomic locations. Reports of MS2-loops in a gene's 5'UTR, 3'UTR and intronic sequences have been published (Martin et al. 2013, Fukaya et al. 2016). The different placements correspond to different scientific questions aimed to be answered. Positioning in the 5'UTR leads to a brighter fluorescent signal as the fluorescence is present during the complete transcription process of the coding region (Fukaya et al. 2017).

A possible disadvantage to MS2-loop positioning in the 5'UTR is induced defects in translation. It has been shown that a RNA stem loop, similar to the MS2-loops, positioned in the 5'UTR leads to the stalling of scanning ribosomes (Vattem & Wek 2004, Palam et al. 2011). It is plausible that the MS2 stem loop placement in the 5'UTR elicits a similar response, where the scanning ribosome cannot displace the MCP-GFP from the stem loop, thereby reducing RNA translation. An inability to displace MCP-GFP would likely be due to the high affinity of MCP-GFP to MS2-loops and use of the newly generated low-affinity MS2-loops could improve translation efficiency. The stalling of polymerases can result in RNA decay (Doma & Parker 2006) and therefore, reduced levels of the mature protein. Reduced protein levels in MS2-loop fly lines could explain the significant reduction in amnioserosa cell number observed in *MS2-hnt* flies, as Hnt protein is essential for amnioserosa maintenance (Frank & Rushlow 1996). Alternatively, the number of amnioserosa cells could have been underestimated due to low levels of Hnt protein present in the cells, which could not be detected in the assay. Nevertheless, problems during translation are not relevant to this study.

Using the MS2/MCP system, single-cell fluorescence traces were generated to investigate *hnt* and *ush* transcription dynamics. This study showed that *ush* transcription onset precedes that of *hnt* transcription, which was previously unknown. The differences in transcription onset between BMP target genes provide new insight into the responsiveness of threshold target genes to BMP signalling levels. The transcription of the low threshold response gene *ush* was shown to initiate at ~20 min into nc14, whereas *hnt* transcription was initiated 10 min later at ~30 min into nc14.

Modelling of the BMP receptor complex binding showed that the concentration of bound BMP receptor complexes doubles between 20 min and 30 min into nc14 at the dorsal midline peak (Mizutani et al. 2005, O'Connor et al. 2006, Umulis et al. 2010) and only *ush* transcription can respond to the initial low levels of signalling. Based on the results from this study, it can be suggested that the increase in signalling that is elicited during this 10 min time window is sufficient to activate *hnt* transcription. Furthermore, the maximum peak concentration of BMP receptor binding is found at 40 min into nc14 using computational modelling (Mizutani et al. 2005, O'Connor et al. 2006). This timing coincides with the observed maximum fluorescence output during this study. The maximum expression of *hnt* was observed at ~46 min whereas maximum *ush* transcription was observed at 41 min. At the end of nc14 when *ush* transcription decreases, embryos enter stage 6 and gastrulation, which coincides with refinement of the pMad gradient (Dorfman & Shilo 2001, Mizutani et al. 2005).

Additionally, this study found that the maximum mean fluorescence output of *ush* transcription foci is approximately twice as high as that of *hnt* (see Figure 3.10 D). This observation is consistent with smFISH data shown in Chapter 3, where the number of cytoplasmic mRNAs per cell produced by *ush* was greater than for *hnt*. Similarly, visualising the fluorescence output of nuclei according to their cross-section position validated the observed gradient in expression output using smFISH. The refined temporal resolution, gained through live imaging, showed at first shallow and then very steep transcription gradient formation at 5 min intervals. The computational model described above, showed an increase of BMP-bound receptor concentration over time (O'Connor et al. 2006). Within 15 minutes of activation, BMP signalling was shown to increase above a threshold level eliciting signalling (O'Connor et al. 2006). This timing would mean that *ush* transcription, which is initiated by low levels of signalling, is induced with an approximate five-minute delay. This lag phase would represent the maximum time needed for the intracel-

lular signalling cascade to relay the BMP signal to the transcription machinery and for enough Pol II molecules to be recruited to generate a detectable transcription signal.

**Transcription changes along the AP position**

Differences in transcription output were observed for both *hnt* and *ush* expression along the AP axis. The posterior domain of *hnt* has been previously shown to be less responsive to BMP signalling (Wang et al. 2008, Ashe et al. 2000). Similarly, the anterior domain of *ush* showed slightly different transcriptional onset behaviour compared to the central and posterior domains of the *ush* expression pattern. Additionally, small changes were observed in the mean fluorescence intensity produced by the different regions. Together these changes suggest that *ush* transcription could be responsive to other factors besides BMP. The main patterning genes active during this period and expressed in this region are Bicoid and Hunchback. Their expression patterns form a concentration gradient in the anterior domain of the *Drosophila* embryo (Tautz et al. 1987, Driever & Nüsslein-Volhard 1988). In support of this hypothesis two potential Bicoid (AAATCC) and one potential Hunchback (TTTTTTG) binding motifs (Ho et al. 2009) are present in the *ush* enhancer (enhancer sequence identification in Chapter 5. Whether *ush* expression is responsive to Bicoid or Hunchback levels would be interesting to to test by using smFISH and live imaging in mutant embryos.

**Individual alleles show coordinated transcription profiles**

Transient monoallelic expression observed in single cells can correspond to transcriptional bursting (Deng et al. 2014). Bursting is not always coordinated between alleles, and it is possible to have only one homologous allele actively transcribed at any given time. Allele contribution to transcription has been addressed in a few studies, but it is challenging to visualise and analyse this highly dynamic process using single time point analysis approaches (Eckersley-Maslin & Spector 2014, Larsson et al. 2019). In this study, the monoallelic expression of the BMP target gene *ush*, first identified and characterised by static FISH techniques in Chapter 3, was verified using live imaging data. Preliminary evidence from the *ush* homozygous movie shows dynamic monoallelic expression where the two homologous alleles were not entirely synchronous in their bursting states, and therefore, at times, only one allele was observed to be actively transcribing.

The detailed analysis of ten nuclei, positioned between the middle and the edge of the *ush* expression domain, revealed that fluorescence levels obtained from the individual loci within the same nucleus are correlated. Hence, it can be suggested that the extrinsic BMP concentration a cell receives is interpreted synchronously by alleles in nuclei at intermediate distance to the midline. Transcription onset time of alleles differed on average by less than five minutes further suggesting a coordinated behaviour of alleles. Similar observations were made in a recent study investigating the transcriptional responses to Notch signalling. Using the MS2/MCP system, this study observed that 80% of cells expressing a transgene under the *single minded* enhancer initiated transcription from both alleles within a five-minute time frame (Falo-Sanjuan et al. 2019). Therefore, both BMP and Notch signalling leads to highly correlated expression between alleles which is initiated within a small time window showing a coordinated response by gene enhancers.

Interestingly, this study identified a subset of nuclei in the *ush* expression domain that showed "sustained" monoallelic expression, where the second allele never initiated active transcription. These data suggest that the activator levels in the nucleus could have been very low and insufficient to activate transcription on both alleles. It has been proposed in the literature that monoallelic gene expression is a way of fine-tuning gene dosage, which could be advantageous in development and cell identity (Gendrel et al. 2016).

The occurrence of "sustained" monoallelic transcription could be explained by a positive feedback mechanism where alleles that were actively transcribing are primed for re-initiation and therefore are preferentially transcribed. A proposed priming or memory effect of alleles could potentially occur through readily available transcription factors and machinery in the promoter vicinity (see General Discussion, transcription hubs) or chromosome accessibility. Transcriptional memory has been described for transcription that is sustained through mitosis (Ferraro et al. 2016). Further studies of nuclei with both alleles labelled will be necessary to explore this idea further.

This study cannot differentiate between the contribution of the maternal and paternal allele to transcription levels. Therefore, potential biases cannot be investigated. Since the transcriptional output of homologous alleles was observed to be correlated in intermediate nuclei, it can be hypothesised that no bias exists. To test the contribution of maternal and paternal alleles experimentally, the endogenous loci could be tagged with MS2 and PP7 loops, respectively. Tagging the

endogenous loci with two analogous stem-loop systems would allow to distinguish between them and recognise each using a different fluorophore.

**Burst kinetics observed by live-imaging**

*Drosophila* single-cell studies have highlighted that gene transcription often features intrinsic stochasticity (Boettiger & Levine 2009), cell-to-cell heterogeneity (Bintu et al. 2018) and high levels of transcriptional noise (Gregor et al. 2007). Therefore, the analysis of transcription dynamics can be challenging. It is unknown how single cells readout the BMP signalling gradient and how changes in the activator levels change burst kinetics of single cells and therefore contribute to cell-to-cell heterogeneity within an expression domain. Here, burst kinetics were extracted from single-cell fluorescence traces using a memory-adjusted HMM. During transcriptional bursts, Pol II complexes are released from the promoter to form transcription convoys. These convoys travel along the gene body and the mediator complex has been shown to influence convoy size (Tantale et al. 2016). Active periods of the promoter are followed by inactive refractory periods with little to no activity.

The observation of MS2/MCP fluorescence only indirectly visualises changes in promoter behaviour, as transcribing Pol II molecules continue elongation after the promoter transitions into the OFF state. A newly developed memory-adjusted HMM accounts for the persistence of fluorescence during mRNA elongation (Lammers et al. 2019). The HMM can infer a hidden state within a dataset, corresponding in this case to the promoter states. These can be indirectly observed through the MCP-GFP fluorescence. The traditional HMM does not account for the continuous elongation of Pol II molecules throughout the gene body after the promoter has already switched off. Once the promoter has switched off, the remaining Pol II molecules along the gene will continue to emit fluorescence. Therefore the HMM was adapted by Lammers et al. to include a memory function considering that the fluorescence at time point x conditionally depends on the hidden promoter state as well as the promoter states of previous time points (Lammers et al. 2019). Using this model promoter states can be inferred and transcription parameters for *hnt* and *ush* were extracted.

This study investigated how the promoter and enhancer integrate changes in BMP signalling. Global analysis of transcription kinetics revealed different bursting profiles for *hnt* and *ush* transcription. *hnt* transcription is characterised by short-

duration bursts that occur with high frequency and amplitude, whereas *ush* transcription shows a lower burst amplitude but longer burst periods resulting in a reduced burst frequency compared to *hnt*. The high burst amplitude obtained during *hnt* transcription may reflect high polymerase pausing levels. Even though *hnt* and *ush* transcription feature very different burst profiles, their response to BMP signalling levels is similar. In both cases, the promoter occupancy was identified to be the key feature regulated by the Kon parameter in response to BMP signalling levels. Although either or both Kon and Koff rates can be tuned by the transcription system to modulate the number of mRNAs that are produced (Zoller et al. 2018), in the case of *ush* and *hnt* transcription in wt embryos, this study found Kon rates to be predominantly changed in response to signalling levels across the dorsal midline. In addition, this study identified that high BMP signalling levels around the dorsal midline lead to increased burst sizes that are dominated by increases in amplitude. Hence, changes in BMP signalling are integrated by *hnt* and *ush* promoters through changes in promoter occupancy as well as in burst amplitude.

The promoter occupancy describes the fraction of time the promoter stays in the ON state (Lucas et al. 2013, Zoller et al. 2018) and differs more between spatial clusters than other burst parameters. The promoter occupancy for both *hnt* and *ush* was shown to be dominated by changes in Kon. Previous studies have shown that Kon is regulated by enhancers (Bartman et al. 2016), therefore it is likely that the enhancer decodes the BMP signalling level which results in changes of the promoter occupancy. According to the modelling results, the promoter occupancy parameter and burst amplitude are most responsive to BMP signalling levels, and a possible molecular mechanism through hub formation will be described in the General Discussion.

Further evidence that BMP signal integration occurs through changes in promoter occupancy was obtained by increasing signalling levels. Ectopic *dpp* expression by the *st2-dpp* allele increases BMP signalling levels and expands the *ush* expression domain (Ashe et al. 2000). Preliminary results show that the increase in signalling levels did not change the burst frequency of transcription in the centre or intermediate expression domain cluster. A small increase in burst frequency was observed in the edge cluster that previously showed low burst frequency. Increased levels of BMP signalling through ectopic DPP were shown to increase promoter occupancy in the intermediate cluster to that found in the central cluster of wt embryos. The central cluster in *st2-dpp* embryos, however, did not show a further increase in occupancy levels suggesting that the promoter already is at maximal occupancy

and the rate of switching on cannot be further increased. This may be different for a promoter of a different class containing different promoter elements. Higher BMP signalling levels were shown to result in an overall increase of burst size by approximately two-fold in all spatial clusters, consistent with increased mRNA production being controlled through changes in burst amplitude. However, additional replicates of the *st2-dpp* embryo need to be analysed to investigate transcription kinetics further.

Nuclei that are part of the edge cluster of *ush* and *hnt* expression profiles show overall reduced burst kinetics, suggesting that the level of BMP signalling in these regions is insufficient to drive extensive bursting behaviour. The low burst frequency and increased off-period between bursts in the edge cluster of *hnt* and *ush* transcription offers an explanation for the high proportion of monoallelic nuclei that was observed previously at the expression domain border of BMP target genes (see Chapter 3). As only small short bursts are produced infrequently, the probability of capturing both alleles on at the point of fixation (prior to FISH analysis) is low.

The in-depth analysis of bursting kinetics showed that *hnt* and *ush* transcription display very distinct bursting profiles in response to identical BMP signalling level exposure in the centre cluster. In response to BMP, *ush* transcription occurs in long duration, low amplitude and low-frequency bursts in comparison to high amplitude, higher frequency and short duration bursts for *hnt*. Overall, the longer and more infrequent bursts of *ush* produce a larger burst size and therefore more mRNAs, which was also shown earlier using smFISH. Even though *hnt* and *ush* bursting profiles differ substantially, the parameters that integrate BMP signalling were identified to be the same. The burst frequencies observed for *ush* in the central and intermediate clusters are comparable to previously identified burst frequencies in *Drosophila* live-imaging studies. Transgenes containing the *sna*, *Krüppel*, *rhomboid* and *Abdominal-B* enhancers have been shown to elicit transcription burst frequencies that are similar to *ush* ranging from 0.1-0.2 bursts per min (Fukaya et al. 2016). Interestingly, the burst frequency observed for *hnt* transcription in the central domain (0.7 bursts per min) is higher than any burst frequency that has been observed for *Drosophila* enhancers. The difference in burst behaviour in response to identical BMP signalling levels could be a result of enhancer-promoter architecture or communication.

The results obtained in this chapter suggest that BMP signalling levels are interpreted differently by enhancers and promoters compared to Notch signalling levels. A recent study found that enhancers responsive to the Notch intracellular domain (NICD) integrate activator levels through modulation of burst size and concluded that the main effect of NICD is a reduction in Koff and therefore an increase in the burst duration (Falo-Sanjuan et al. 2019). This study identified variations in Kon as the primary source of control during BMP signal integration and did not identify large changes in burst duration. Consistent with changes in Kon, the burst frequency responds to different BMP signalling levels, again in contrast to the findings with Notch, which suggest that frequency is not modulated (Falo-Sanjuan et al. 2019). Possible changes in promoter occupancy, however, were not investigated in response to increased Notch signalling. Further studies will be necessary to determine whether distinct signals modulate burst kinetics by very different rules or whether Notch or BMP signals are an exception to a general rule.

In summary, this chapter has further characterised the dynamic transcription behaviour of BMP target genes *hnt* and *ush*. The transcription onset time of *ush* was shown to precede that of *hnt*. Furthermore, transcription onset time was found to be correlated to the nuclear position. Furthermore, both *hnt* and *ush* transcription profiles are partitioned into clusters of similar transcription behaviour. These clusters were shown to correlate with nuclear distance to the dorsal midline. Computational modelling revealed that BMP signalling levels are integrated mostly through changes in promoter occupancy and burst amplitude. The role of promoters in burst dynamics and the contribution of shadow enhancers to transcription will be the focus of the next chapter.

# 5 | The role of promoters and enhancers in transcription dynamics

## 5.1 Introduction

The previous chapter showed that BMP target gene expression of *hnt* and *ush* occurs in bursts and that changes in burst kinetics regulate the transcriptional output of both genes. However, the influence of their DNA regulatory elements on transcription is still unknown. Cis-regulatory DNA sequences, for example promoters and enhancers, regulate gene expression through binding of activators or repressors. The first half of this chapter will investigate the influence of the promoter on burst kinetics. The core promoter sequences of *hnt* and *ush* genes were exchanged so that *hnt* transcription is driven by the *ush* promoter and *ush* transcription is driven by the *hnt* core promoter. The resulting changes in burst dynamics will be investigated to advance the understanding of promoter influence on transcription.

The second part of this chapter will focus on the role of shadow enhancers. The presence of multiple enhancers that can regulate the same gene has been shown to increase the robustness of gene transcription during development (Bothma et al. 2015, Frankel et al. 2010, Cannavò et al. 2016). New enhancers in remote (shadow) positions were identified by genome-wide transcription factor occupancy analyses (Hong et al. 2008). Studies of enhancers in isolation and in pairs have shown complex behaviours for example additive, sub-additive and super-additive function (Bothma et al. 2015, Scholes et al. 2019). This chapter will investigate if the *tup* enhancer architecture is important for BMP signalling interpretation.

## 5.2 Promoter swaps change transcription dynamics

### 5.2.1 Design and characterisation of endogenous MS2 promoter swap lines

This study generated promoter swap lines to investigate the influence of the core promoter sequences on transcriptional bursting. Previous studies have shown that the exchange of promoter sequences results in genes exhibiting the degree of PPP encoded within the new promoter (Lagha et al. 2013), which suggested that important regulatory information is contained within the core promoter sequence. The *hnt* and *ush* MS2 lines that have been used in Chapter 2 were modified by exchanging their core promoter sequences, while retaining the MS2-loops in the same position. Core promoter sequence annotation was determined by a GRO-Seq study (Saunders et al. 2013). The newly created *ush*-Promoter-*hnt* (*ushPhnt*) line contains the 200 bp core promoter region (around the TSS) of *ush* isoform RC (Figure 5.1 A). The introduction of the *hnt* promoter sequence into the endogenous *ush* isoform RC 5'UTR resulted in the loss of an Inr sequence and gain of a TATA box (Figure 5.1 B). The newly generated promoter swap line *hnt*-Promoter-*ush* (*hntPush*) now expresses the *ush* gene under the *hnt* promoter.

First, the general expression patterns of live imaging lines with promoter swaps were analysed using ISH, and the effect on development was evaluated by counting the number of amnioserosa cells. The overall expression pattern generated by *ushPhnt* embryos is shown to be similar to the wt *hnt* expression domain and present in the same dorsal region (Figure 5.2 Ai) The number of amnioserosa cells produced in *hnt* embryos was shown to be reduced in Chapter 2 (see Figure 4.4 B). The promoter swap line shows a further reduction in amnioserosa cell number (Figure 5.2 Bi and Bii). The additional reduction in amniosersa cell counts suggests that the *ush* promoter driving *hnt* expression leads to reduced Hnt levels.

The expression pattern of *ush* transcription is shown to be very similar between wt and *hntPush* expressing embryos (Figure 5.2 Aii). The overall dorsal location and width in expression pattern appears very similar (Figure 5.2 Aii). Likewise, the numbers of amnioserosa cells that were counted for *hntPush* embryos are statistically the same as for wt and *ush* embryos (shown in Chapter 4, Figure 4.4 B and replotted here), with a mean of 137 cells (Figure 5.2 Bii). To account for changes introduced through the MS2-loops alone, *Drosophila* lines were generated

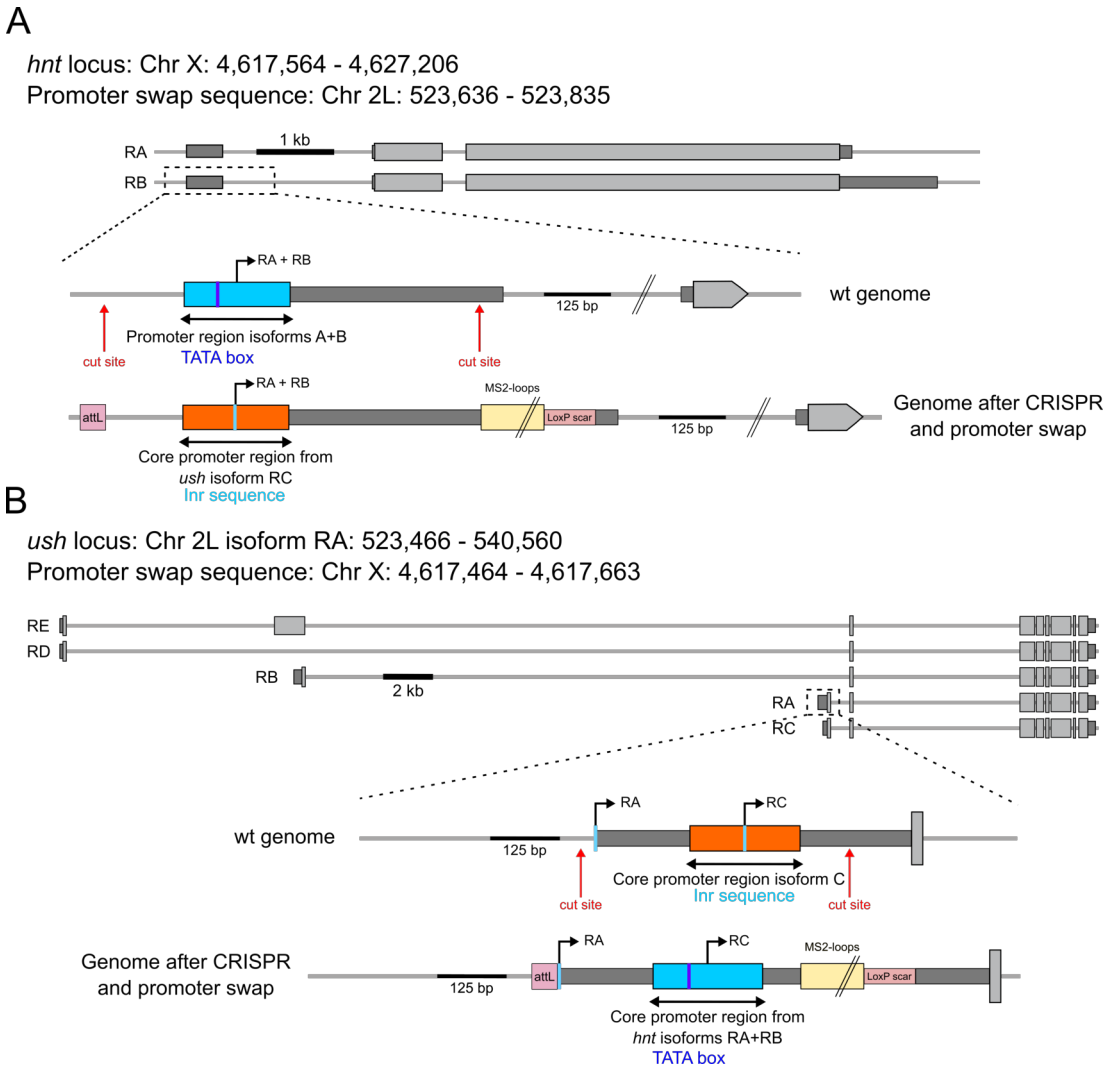



**Figure 5.1: Generation of *hnt* and *ush* MS2 promoter swap lines.** (A) A schematic showing the genomic region encoding *hnt* on the X-chromosome. To create a promoter swap line, the 200 bp core promoter region of *ush* isoform RC was inserted into the endogenous *hnt* 5'UTR using a two step CRISPR approach including attB/P reintegration (see Methods). A TATA box in the *hnt* promoter region was thereby lost. (B) A schematic showing the genomic region encoding *ush* on the 2$^{nd}$ chromosome. To create a promoter swap line, the 200 bp core promoter region of *hnt* was inserted into the endogenous *ush* RC promoter region. The insertion resulted in the loss of an Inr sequence at the TSS and gain of a TATA box. The resulting fly lines are referred to as *ushPhnt* and *hntPush*. All chromosome coordinates obtained from the *Drosophila* dm6 genome.

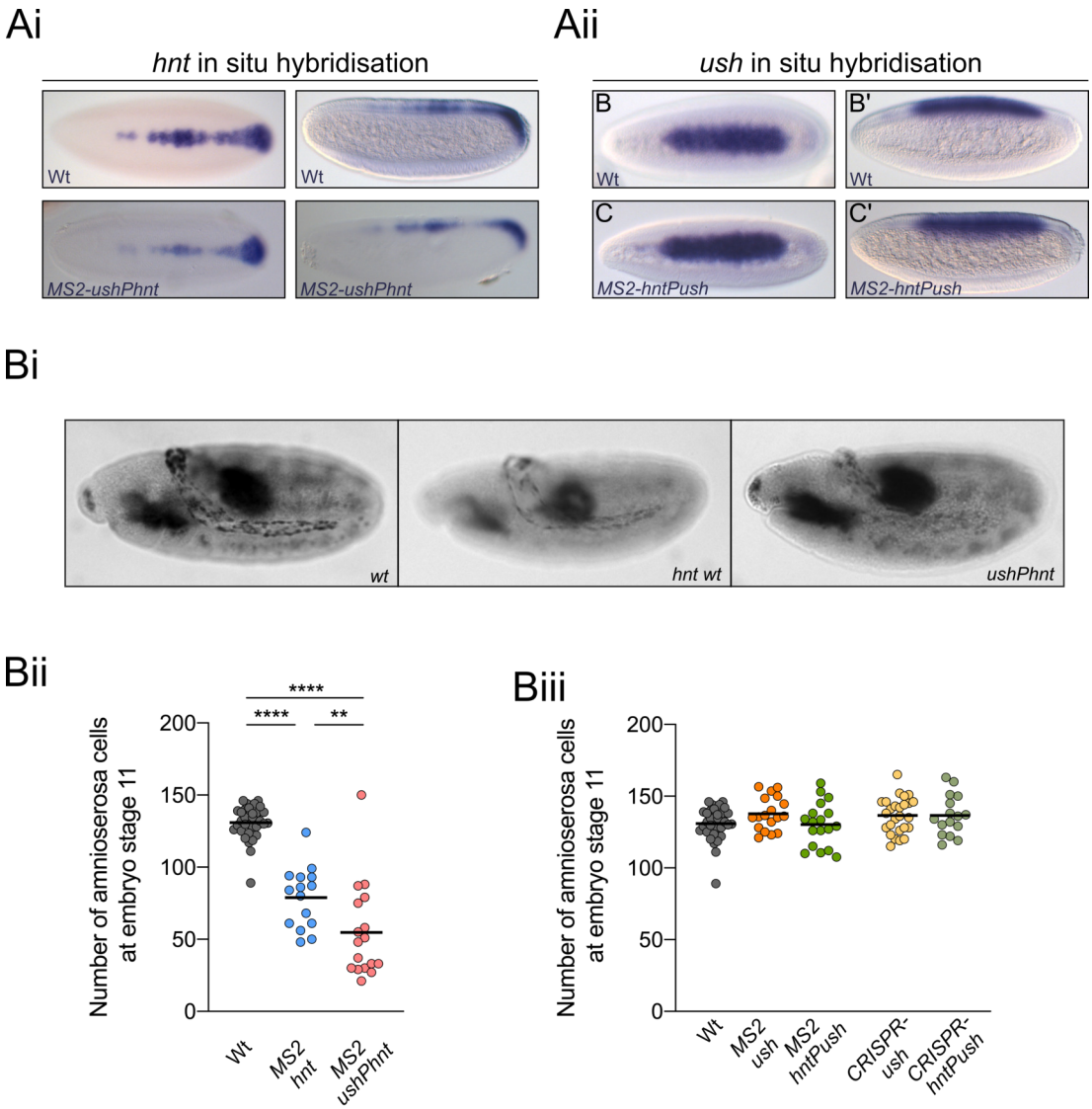



**Figure 5.2: Expression patterns of promoter swap live imaging lines.** Colourimetric ISH visualising *hnt* (Ai) or *ush* (Aii) transcripts in wt or MS2-loop promoter swap lines. (Bi) Representative images of stage 11 embryos showing Hnt antibody stains to detect amnioserosa cells. Amnioserosa counts for *hnt* (Bii) and *ush* (Biii) based live imaging lines in embryo stage 11 embryos. Each data point represents one embryo count. *CRISPR-ush* and *CRISPR-hntPush* embryos were used as a control as they were treated the same way as MS2-loop lines during genome modification but without loop insertion.
Line in scatter plots (Bii, Biii) depicts mean. Significance was tested with a One-Way ANOVA test with a Dunnett's multiple comparisons test (Bi), **p $< 0.01$; ****p $< 0.0001$. Technical replicates *wt* n = 5, *hnt* n = 3, *ushPhnt* n = 3, *ush* n = 3, *hntPush* n = 3, *CRISPR-ush* n = 3, *CRISPR-hntPush* n = 3. Embryos in (Ai, Aii) shown in dorsal and lateral position, embryos in (Bi) shown in lateral position, anterior to the left. Data from wt, *hnt* and *ush* embryos previously shown in Chapter 2.

that underwent the identical CRISPR engineering procedure but had no MS2-loops inserted into their genome. Embryos of the genotype *CRISPR-ush* were analysed in Chapter 3 and show the same number of amniosersa cells as *CRISPR-hntPush*, embryos, suggesting that the insertion of MS2-loops did not affect amnioserosa cell development (Figure 5.2 Bii).

These data suggest that the addition of the *hnt* promoter into the *ush* genomic locus did not influence the *ush* transcription and later developmental cell fate decisions as much as the *ush* promoter in the *hnt* locus. This observation is constrained by the fact that Hnt detection was used as a marker for amniosersa cells, due to the lack of alternative amnioserosa markers.

### 5.2.2 Expression profiles and clustering of *ushPhnt* embryos

Live imaging of the promoter swap lines was performed as previously described in Chapter 4. Here, the results of promoter swap lines will be compared to their "wildtype" counterparts, which were previously analysed in Chapter 4 and their data are re-plotted in this chapter for direct comparison.

#### Characterisation of *hnt* promoter swap line

First, global differences on a cell population level are compared to characterise the changes that were caused by the promoter swap in *ushPhnt* embryos. To match the *hnt* datasets, one full imaging domain of an *ushPhnt* embryo was analysed and a second biological replicate was analysed in the central domain only. A still image and the corresponding maximum projected movie of *ushPhnt* transcription can be found in Figure 5.3.

The cumulative expression pattern visualises all nuclei that show *hnt* transcription at any point during nc14. The full expression pattern in the *ushPhnt* embryo is similar to the previously analysed *hnt* embryos. However, fewer actively transcribing nuclei are observed in the connecting region between the central and posterior domain of the *ushPhnt* embryo (Figure 5.4 B). The analysis region represents the central embryo domain which was chosen for all further analysis (Figure 5.4 B). Representative single-cell nuclear traces are shown in Figure 5.4 C. The number of active nuclei in the *hnt* expression domain was quantified for all biological replicates and is shown to be reduced in the *ushPhnt* embryos (Mean *hnt* = 190.0 and *ushPhnt* = 139.0, Figure 5.5 A). Next, the time period of active transcription was

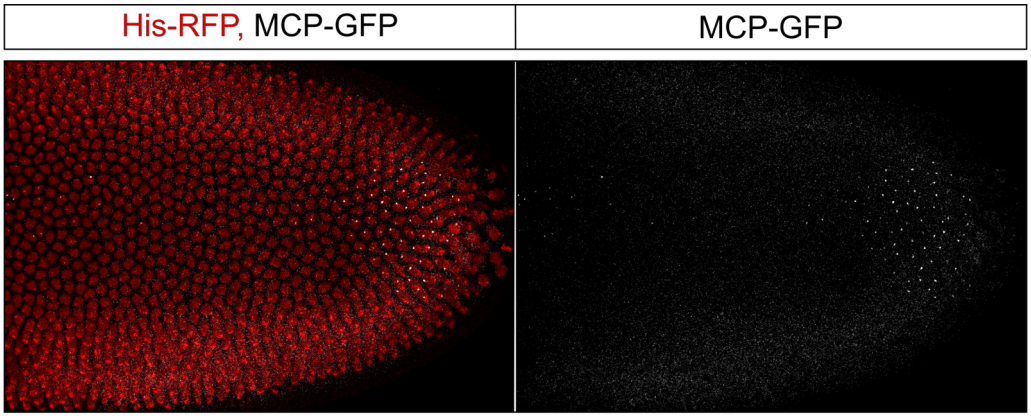



**Figure 5.3: Time-lapse microscopy detecting *ushPhnt* transcription.** Still of a maximum intensity projection (54x 1 μm stacks) during nc14. Nuclei visualised by His-RFP (red) and *hnt* transcription by MCP-GFP fluorescence (grey). Time resolution was 20 sec/frame. Embryo was imaged in a dorsal position with the anterior to the left. Full movie can be found at `https://www.tiny.cc/thesis_videos`.

analysed for all nuclei. The time period between transcription onset and the last time point of transcriptional activity present in the datasets was calculated (Figure 5.5 Bi). This measured time period could contain multiple transcriptional bursts and only determines the overall active period of nuclei.

The median duration of transcriptional activity is shown to be similar between genotypes with 8.8 and 8.3 min, respectively (Figure 5.5 Bii). Besides the length of transcriptional activity, the onset time of transcription was compared between *hnt* and *ushPhnt* embryos. The onset time of nuclei within biological replicates shows a similar distribution (shape of violin plot) and is not statistically different within or between genotypes (Figure 5.5 Ci). The median onset time is shown to be almost identical between genotypes with a mean of *hnt* = 39.5 min and *ushPhnt* = 38.4 min (Figure 5.5 Cii). The onset time for nuclei within an expression domain can be converted into a cumulative onset time which reveals the "time to synchrony".

The "time to synchrony" is defined as the time that is required to achieve coordinated gene activation across a field of cells (Lagha et al. 2013). Coordinated gene expression is measured when 50% of nuclei in the expression domain have initiated transcription. The time is smaller when a gene exhibits synchronous gene activation (Lagha et al. 2013). When the "time to synchrony" is quantified for all biological replicates, it is observed that the curves show similar shapes (Figure 5.6 A). The mean time to synchrony was calculated between replicates and shows that activation of *hnt* transcription across a field of cells was initiated with a very small

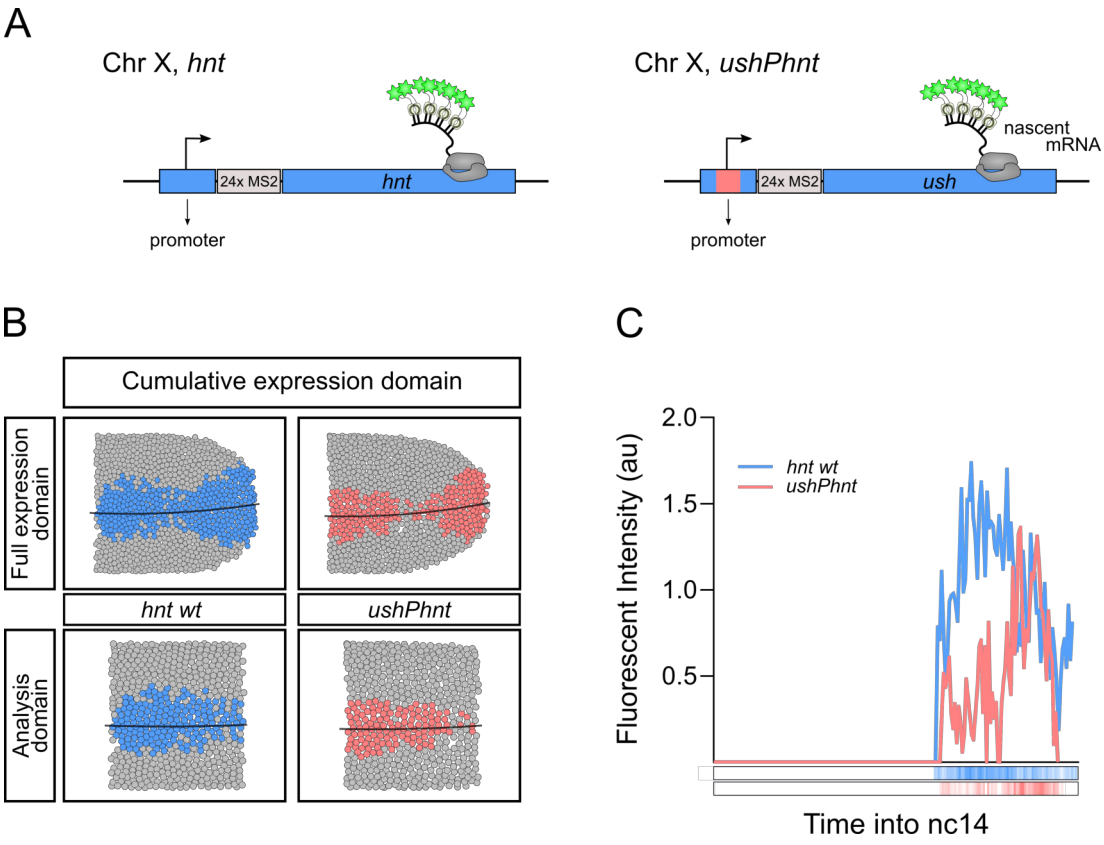



**Figure 5.4: Characterisation of *hnt* promoter swap embryo.** (A) Summary of the genomic imaging regions for embryos with a wt *hnt* promoter (blue) and embryos carrying a *ush* promoter (pink). (B) Cumulative expression domains of *hnt* and *ushPhnt* transcription are shown for the full imaging domains and the analysis domains. (C) Graph shows example single nuclear fluorescent traces of *hnt* (blue) and *ushPhnt* (pink) transcription. A heatmap of the imaging traces is shown below the graph.
Expression domains are oriented with the anterior to the left.

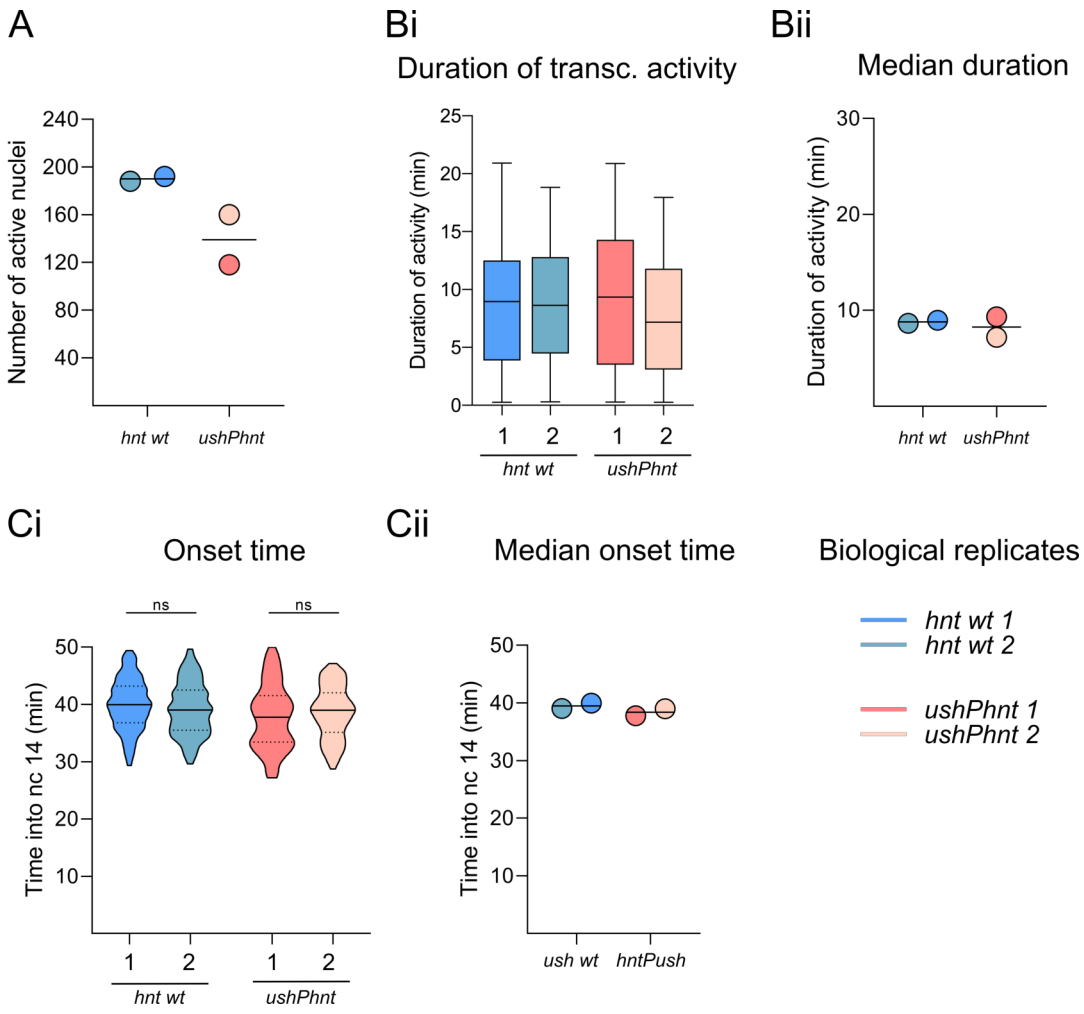
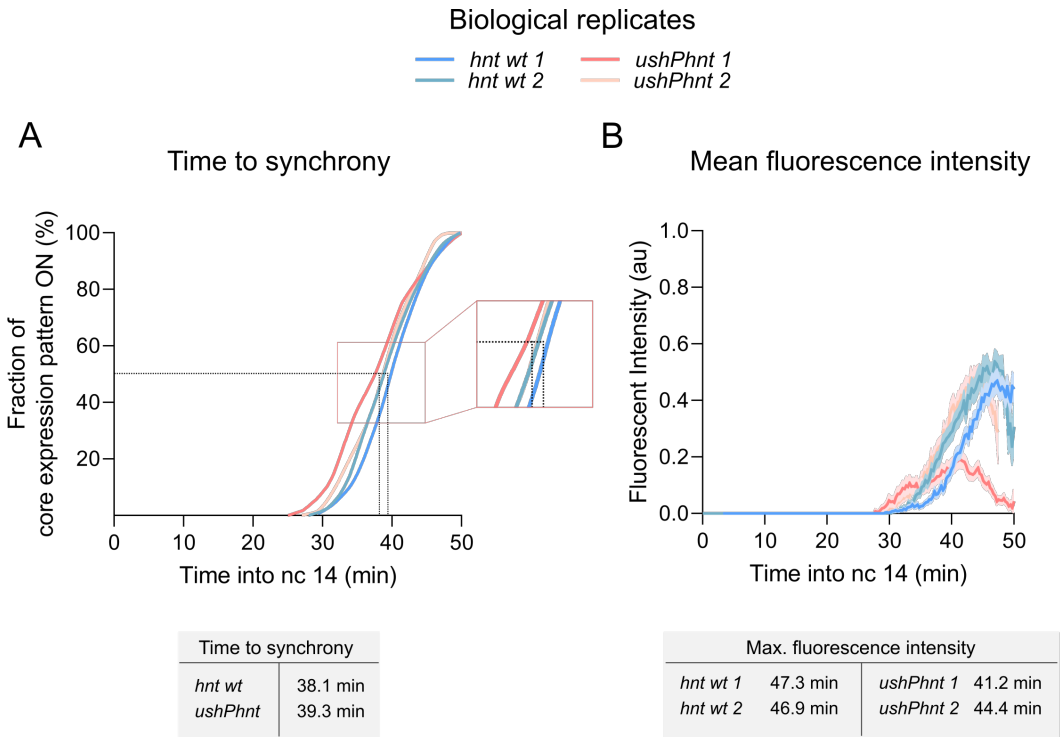



**Figure 5.5: *ush* driven *hnt* transcription results in fewer transcribing nuclei.** (A) The graph shows quantification of transcriptionally active nuclei in biological replicates. The duration (Bi) and median duration (Bii) of transcription is shown in this graph. The time interval was determined based on the transcription onset time and the last frame of detected active transcription in nc14. (Ci) Here, the onset time of individual nuclei is shown and the median onset time (Cii) for individual biological replicates.
Number of biological replicates for *hnt* n = 2 with 192 and 188 nuclei and for *ushPhnt* n = 2 with 118 and 160 nuclei respectively. Means are shown in (A, Bii, Cii) and lines in violin plots (Ci) depict median and 95% confidence intervals. Boxes (Bi) show 25$^{th}$ to 75$^{th}$ percentile, line shows median and whiskers show range.




| Time to synchrony | |
|---|---|
| *hnt wt* | 38.1 min |
| *ushPhnt* | 39.3 min |

| Max. fluorescence intensity | | | |
|---|---|---|---|
| *hnt wt 1* | 47.3 min | *ushPhnt 1* | 41.2 min |
| *hnt wt 2* | 46.9 min | *ushPhnt 2* | 44.4 min |

**Figure 5.6: The *ush* promoter causes greater embryo-to-embryo variability in *hnt* expression.** (A) "Time to synchrony" curves are shown for all biological replicates and measurements are shown representing 50% core expression pattern onset. (B) Mean fluorescence intensity traces are shown in this graph. The time points at which the maximum fluorescence output was detected are shown in the table below.

difference of one minute after the promoter swap (Figure 5.6 A). Significant differences are observed when the mean fluorescence intensity of all biological replicates was quantified. The two replicates of *ushPhnt* embryos show very different fluorescence behaviours. One replicate follows the same shape as *hnt* embryos, while replicate number 1 shows a flattened curve (Figure 5.6 B). Additionally, replicate number 1 reaches only one-third of the maximum mean fluorescence intensity compared to all other replicates, while reaching the maximum at a similar time (Figure 5.6 B).

The above data show minimal changes in the overall transcription behaviour in *ushPhnt* embryos compared to *hnt* embryos. A reduced number of *ushPhnt* expressing nuclei was found. Furthermore, a significant change in mean fluorescence was observed. The large variability in *ushPhnt* embryos could be a result of an outlier embryo, and with two biological replicates, this study is unable to determine if this is the case. However, large variability in gene expression, induced by genetic changes, has been reported to occur in other early *Drosophila* patterning genes. A study analysing *tailless* mutations has shown large embryo-to-embryo variability in gap gene expression in mutant backgrounds, concluding that a loss of canalisation (loss of developmental robustness) can be buffered to different extends in mutant embryos (Janssens et al. 2013). Therefore, it is plausible that the change in promoter sequence results in greater embryo-to-embryo variability in expression levels.

Besides the mean fluorescence intensity, the sum fluorescence produced by individual nuclei can be compared between genotypes. The sum fluorescence produced by nuclei in the *ushPhnt* embryo is shown to be highest in the posterior domain but appears reduced overall compared to the *hnt* embryo (top row, Figure 5.7 Ai). However, after quantification, this trend is not statistically significant (top row, Figure 5.7 Aii). When only comparing the central domains of the embryos, nuclei in the central domain of the *hnt* embryo reach higher levels of sum fluorescence (bottom row, Figure 5.7 Ai). This observation is confirmed statistically as the *hnt* embryo is shown to produce a significantly higher median fluorescence than the *ushPhnt* embryo (bottom row, Figure 5.7 Aii). These data are consistent with the results analysed in Figure 5.6 B, where the mean fluorescence intensity of *ushPhnt* embryo 1 was found to be lower than *hnt* embryos and the second biological replicate of *ushPhnt*. Overall, the results obtained in this section characterise the *hnt* expression in *hnt* and *ushPhnt* embryos on a population level. For further analysis of the changes in transcription dynamics single nuclear traces will be analysed.

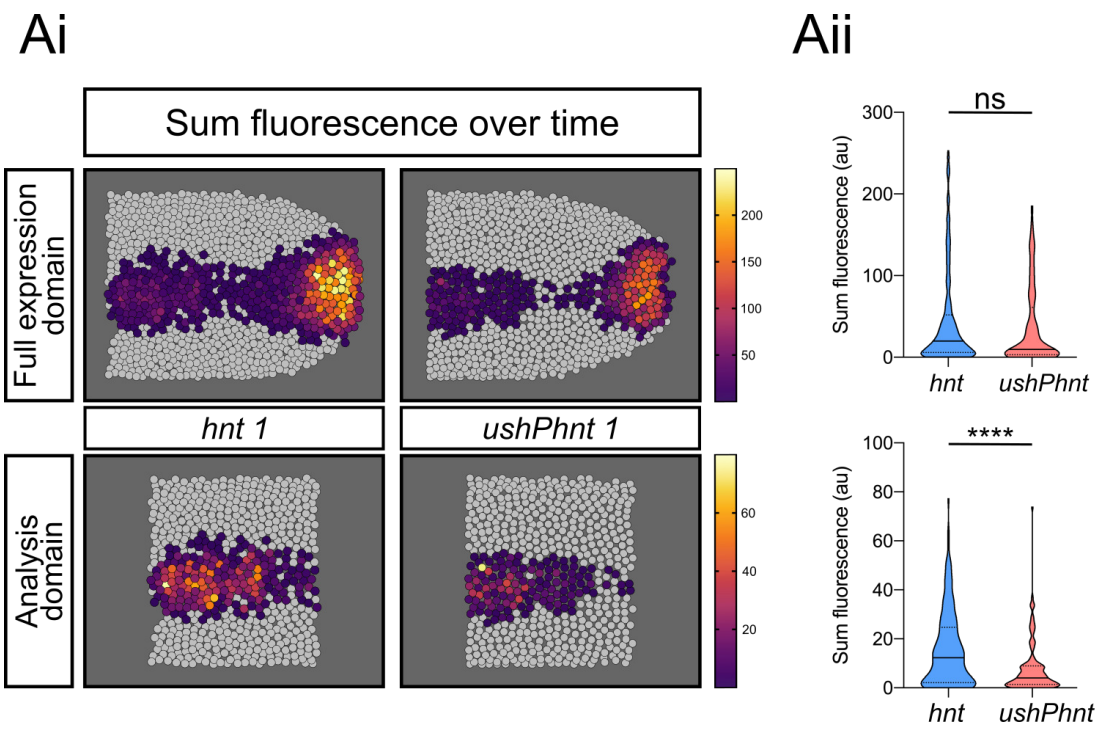



**Figure 5.7: Differences in sum fluorescence between *hnt* and *ushPhnt* transcription.** (Ai) Representative images of *hnt wt 1* and *ushPhnt 1* domains. Nuclei were coloured depending on their total sum fluorescence, and maximum values were chosen according to the maximum fluorescence of *hnt wt*. (Aii) The graphs show the quantification of the sum fluorescence for the individual replicates. Note different scales.
Lines in violin plots depict median and 95% confidence intervals, significance was tested with a Mann-Whitney test, ****p $<$ 0.0001; ns = not significant. Expression domains are oriented with the anterior to the left.

**Clustering results for *hnt* transcription**

The differences in RNA transcription between single nuclei was evaluated using K-Means clustering. The K-Means algorithm divides nuclei into groups of similar mean fluorescence and identified two clusters within the *ushPhnt* data, similar to *hnt* embryos (Figure 5.8). Nuclei positioned around the dorsal midline were grouped into the middle cluster (yellow). Representative schematics are shown here for *hnt* embryo 1 and *ushPhnt* embryo 1. The number of nuclei that were grouped into the middle cluster is much reduced in the *ushPhnt* embryo compared to the *hnt* embryo (52 *hnt* vs 19 *ushPhnt*, Figure 5.8). In addition, the nuclei that are part of the middle cluster of the *ushPhnt* embryo show a significantly reduced mean fluorescence intensity compared to the *hnt* embryo (Figure 5.8). Similarly, nuclei in the edge cluster of *ushPhnt* show reduced mean fluorescence (green) (Figure 5.8). These data suggest that more nuclei in the *ushPhnt* embryo showed lower mean fluorescence which lead to their sorting into the edge cluster.

Heatmaps were generated for each biological replicate individually as the two biological replicates of *ushPhnt* showed great variability in their mean fluorescence (see Figure 5.6 B). Heatmaps of the additional *hnt* and *ushPhnt* replicates can be found in Figure B.1. The heatmaps of replicates *hnt* 1 and *ushPhnt* 1 correspond to the K-Means clusters shown in the previous figure and show highest fluorescence intensity in the middle cluster (top, Figure 5.9). A few nuclear traces, especially in the edge cluster, show periods of inactivity, where fluorescence values first decrease to 0 and then increase again (Figure 5.9). These periods most likely represent refractive periods between transcriptional bursts.

Nuclear traces in the heatmaps were ordered by their onset times and the onset front of transcription was traced to investigate differences between clusters of the two genotypes. The onset front of the *hnt* middle cluster was traced and is shown by the white line. The onset line was then superimposed onto the *ushPhnt* middle cluster. A difference is observed between the onset fronts of the two genotypes, with the *ushPhnt* embryo initiating transcription earlier (Figure 5.9). Quantified is this difference by measuring the time at which 50% of nuclei within a cluster have initiated transcription. Given the great embryo-to-embryo variability within the *ushPhnt* genotype, the difference in transcription onset between the middle clusters varied (graph, Figure 5.9). Earlier analysis of the global transcription onset time showed no difference between genotypes, suggesting that differences between clusters were masked (compare Figure 5.5 Cii). The transcription onset time in the

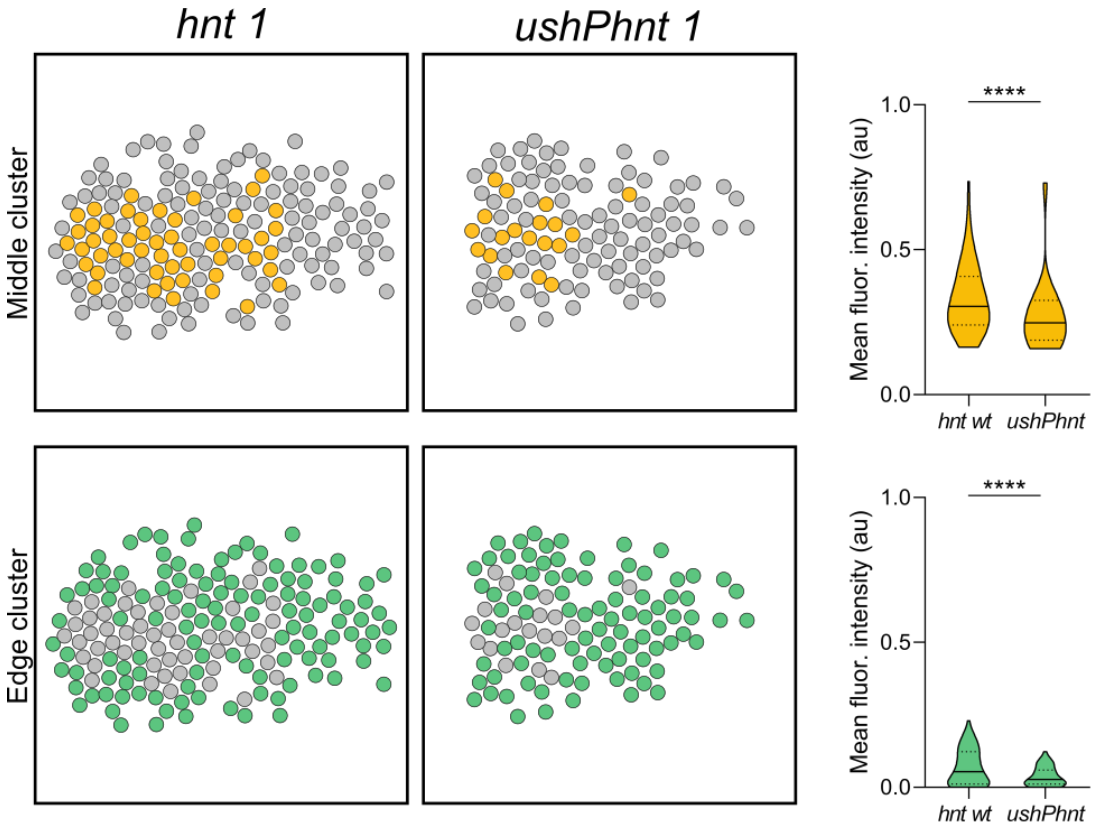



**Figure 5.8: K-mean clustering results for *ushPhnt* embryos.** (Left) Schematics show representative expression domains of *hnt* wt embryo 1 and *ushPhnt* embryo 1. Nuclei are coloured depending on their K-means cluster (yellow=middle cluster and green = edge cluster). (Right) Mean fluorescence values from nuclei pooled according to their cluster.
Lines in violin plots depict median and 95% confidence intervals, ns = not significant. Expression domains are oriented with the anterior to the left.

edge clusters is less variable between *hnt* and *ushPhnt* embryos (Figure 5.9). This observation suggests that the main contribution to embryo-to-embryo variability in *hnt* transcription originates from differences in the middle cluster despite there being fewer cells in this region.

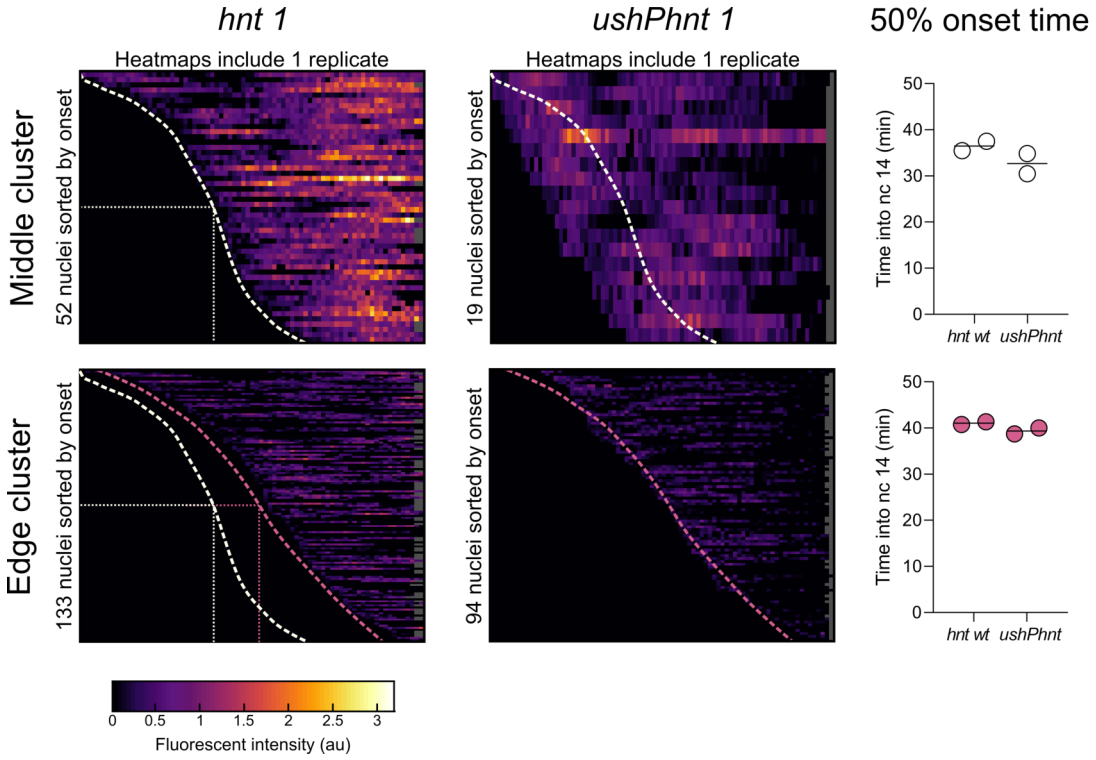



**Figure 5.9: Heatmaps of *ushPhnt* embryo K-Means clusters.** (Left) Heatmaps represent single nuclear traces according to their clusters (middle = top, edge = bottom). Nuclear traces were ordered by their onset times and the onset front of each cluster was traced according to the *hnt* wt onset front and superimposed onto the heatmaps of the *ushPhnt* embryo. Scale as indicated below heatmaps, black = no expression, grey at the end of traces indicates periods where nuclei were not tracked. The time into nc14 at which 50% of nuclei had activated transcription was measured and compared (right). Graphs show 50% onset time for all biological replicates.
Nuclear traces start at the transcriptional onset time of the first nucleus. Lines represent mean. Number of biological replicates n = 1 in heatmaps and n = 2 in graphs for each genotype. Heatmaps of other replicates can be found in Figure B.1. *hnt 1* data was shown previously in Figure 4.13.

### 5.2.3 Expression profiles, clustering and transcription kinetics in *hntPush* embryos

#### Characterisation of *ush* promoter swap line

For an initial characterisation of the *ush* and *hntPush* lines (Figure 5.10 A), their cumulative expression domains were visualised. The cumulative expression pattern shown for *ush* expression in a *hntPush embryo* appears similar to the "wildtype" *ush* expression domain (Figure 5.10 B). The analysis domains show the anterior regions of the full expression domains (Figure 5.10 B). One representative nuclear trace was plotted for each genotype (Figure 5.10 C). A still image and the corresponding maximum projected movie of *hntPush* transcription can be found in Figure 5.11.

Next, the global differences in expression behaviour are compared to characterise the changes that were caused by the insertion of the *hnt* promoter into the *ush* genomic locus. To match the *ush* datasets, one full imaging domain of a *hntPush* embryo was analysed and is shown in Figure 5.10. Two further biological replicats of *hntPush* were analysed in the anterior domain only. The number of nuclei that show active *ush* transcription is statistically the same between genotypes with a downward trend in *hntPush* embryos (Figure 5.12 A). Similarly, when the duration of nuclear activity was calculated (Figure 5.12 Bi) no changes were observed in the median duration (Figure 5.12 Bii). The onset time of nuclei within genotypes is shown to be the same (Figure 5.12 Ci). However, when the median onset time of *ush* transcription was compared between genotypes, a small but significant difference is found. Transcription in *hntPush* embryos is shown to be initiated earlier (Figure 5.12 Cii). The difference in onset times could represent increased PPP levels or suggests that the *hnt* promoter leads to an increased sensitivity to BMP signalling levels.

The "time to synchrony" measures the synchronicity with which a field of cells activate gene expression (Lagha et al. 2013). Embryos that carry the *hntPush* gene locus are shown to reach the 50% mark quicker than *ush* embryos (Figure 5.13 A). On average the *ush* transcription domain reaches 50% activity at 28.5 min into nc14. This measurement is reduced in *hntPush* embryos where *ush* transcription is activated more synchronously reaching 50% activity on average at 25.5 min (Figure 5.13 A).

Besides a change in transcription synchronicity, a large change in mean fluorescence was observed in promoter swap embryos. The three biological replicates of

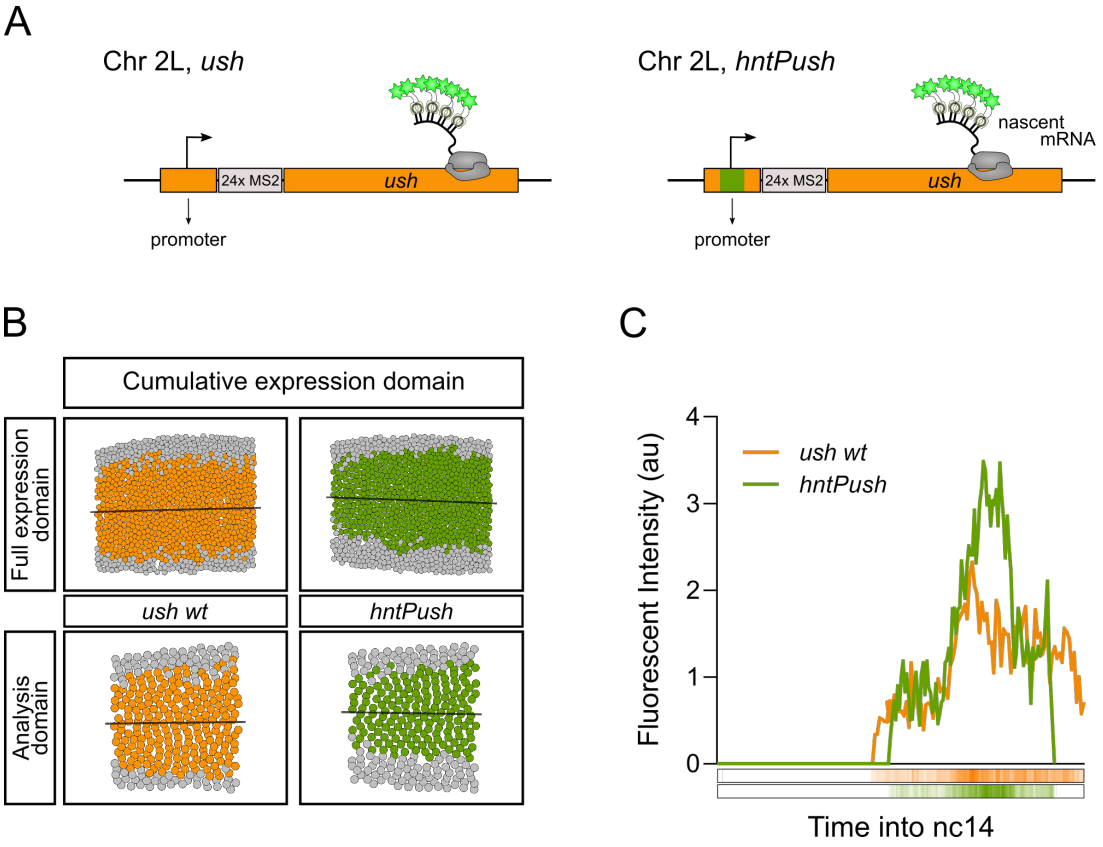



**Figure 5.10: Characterisation of *ush* promoter swap embryo.**(A) Summary of the genomic imaging loci for embryos with a wt *ush* promoter (orange) and embryos carrying a *hnt* promoter (green). (B) Cumulative expression domains of *ush* and *hntPush* transcription are shown for the full imaging domains and the analysis domains. (C) Graph shows example single nuclear fluorescent traces of *ush* (orange) and *hntPush* (green) transcription. A heatmap of the imaging traces is shown below the graph.
Expression domains are oriented with the anterior to the left.

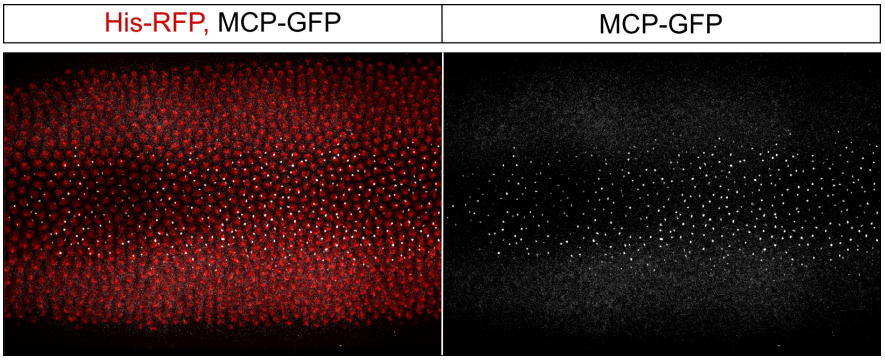



**Figure 5.11: Time-lapse microscopy detecting *hntPush* transcription.** Still of a maximum intensity projection (54x 1 μm stacks) during nc14. Nuclei visualised by His-RFP (red) and *hnt* transcription by MCP-GFP fluorescence (grey). Time resolution was 20 sec/frame. Embryo was imaged in a dorsal position with the anterior to the left. Full movie can be found at `https://www.tiny.cc/thesis_videos`.

*hntPush* show great variation in their levels of *ush* transcription (Figure 5.13 B). One replicate (*hntPush 1*) shows similarity in the maximum fluorescence value and the curve shape to *ush* embryos. The other two replicates show maximum mean fluorescence values that are approximately 50% and 100% greater than *ush* wt transcription (Figure 5.13 B). The observed variation could be caused by promoter-enhancer incompatibility or robustness issues (see Discussion). Consistent with an earlier onset time and the increased synchronicity, *hntPush* transcription reached its maximum fluorescence value earlier (mean = 38.1 min) than *ush* wt transcription (mean = 41.4 min) (Figure 5.13 B). The observation of increased embryo-to-embryo variability in *hntPush* embryos is similar to *ushPhnt* embryos. Therefore, it is plausible that promoter changes cause a greater embryo-to-embryo variation that represents different buffering capacities.

After analysing the mean fluorescence intensity, the overall sum fluorescence produced by single nuclei was analysed (Figure 5.14 Ai). It can be observed that, similar to the *ush* embryo, the *hntPush* embryo produced the highest levels of fluorescence in the middle of the expression domain along the dorsal midline (Figure 5.14 Ai). Quantification of the total fluorescence produced by nuclei shows that the *hntPush* embryo produced significantly higher levels in its full as well as in its anterior expression domain (Figure 5.14 Aii). The results from this section highlight the embryo's wide differences in the expression between *ush* and *hntPush* embryos. In order to analyse these datasets further, the individual nuclear traces will be investigated.

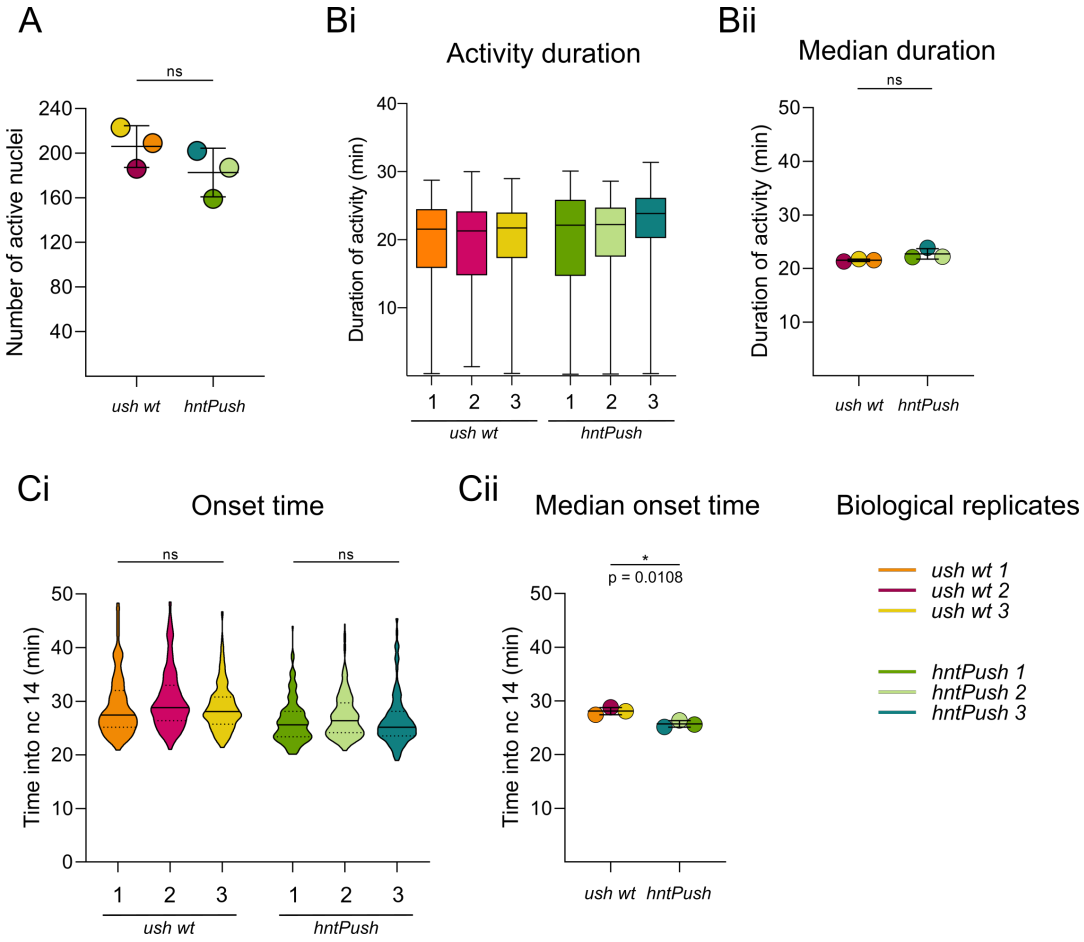
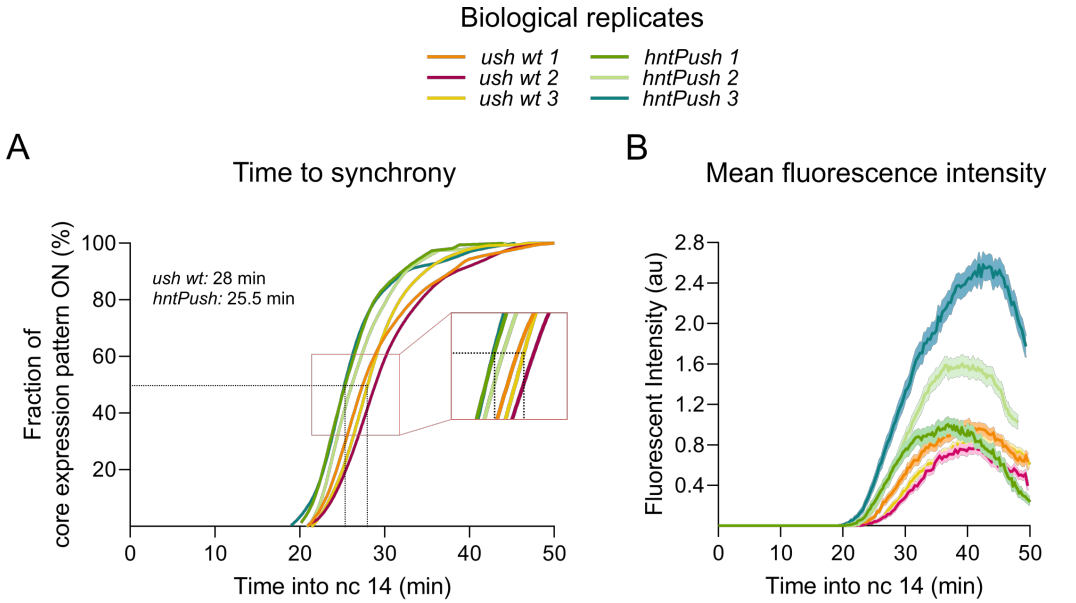



**Figure 5.12: *hnt* driven *ush* transcription onset precedes *ush* wt.** (A) The graph shows quantification of transcriptionally active nuclei in biological replicates. The duration (Bi) and median duration (Bii) of transcription is plotted here. The time interval was determined based on the transcription onset time and the last frame of detected active transcription in nc14. Here, the onset time of individual nuclei (Ci) and the median onset time (Cii) is shown for individual biological replicates.
Number of biological replicates for *ush* n = 3 with 186, 209 and 222 nuclei and for *hntPush* n = 3 with 159, 187 and 202 nuclei, respectively. Mean and SD are shown in (A, Bii, Cii) and lines in violin plots (Ci) depict median and 95% confidence intervals. Boxes (Bi) show 25$^{th}$ to 75$^{th}$ percentile, line shows median and whiskers show range. Significance (Cii) was tested using a Student's t-test; *p $< 0.05$.




| Time to synchrony | |
|---|---|
| *ush wt* | 28 min |
| *hntPush* | 25.5 min |

| Max. fluorescence intensity | | | |
|---|---|---|---|
| *ush wt 1* | 40.2 min | *hntPush 1* | 36.8 min |
| *ush wt 2* | 41.3 min | *hntPush 2* | 36.4 min |
| *ush wt 3* | 42.8 min | *hntPush 3* | 41.2 min |

**Figure 5.13: The *hnt* promoter causes greater embryo-to-embryo variability and greater *ush* synchronicity.** (A) "Time to synchrony" curves are shown for all biological replicates and measurements represent the time at which 50% of the core expression pattern has initiated transcription. (B) Mean fluorescence intensity traces are shown in this graph. The time points at which the maximum fluorescence output was detected are shown in the table below.

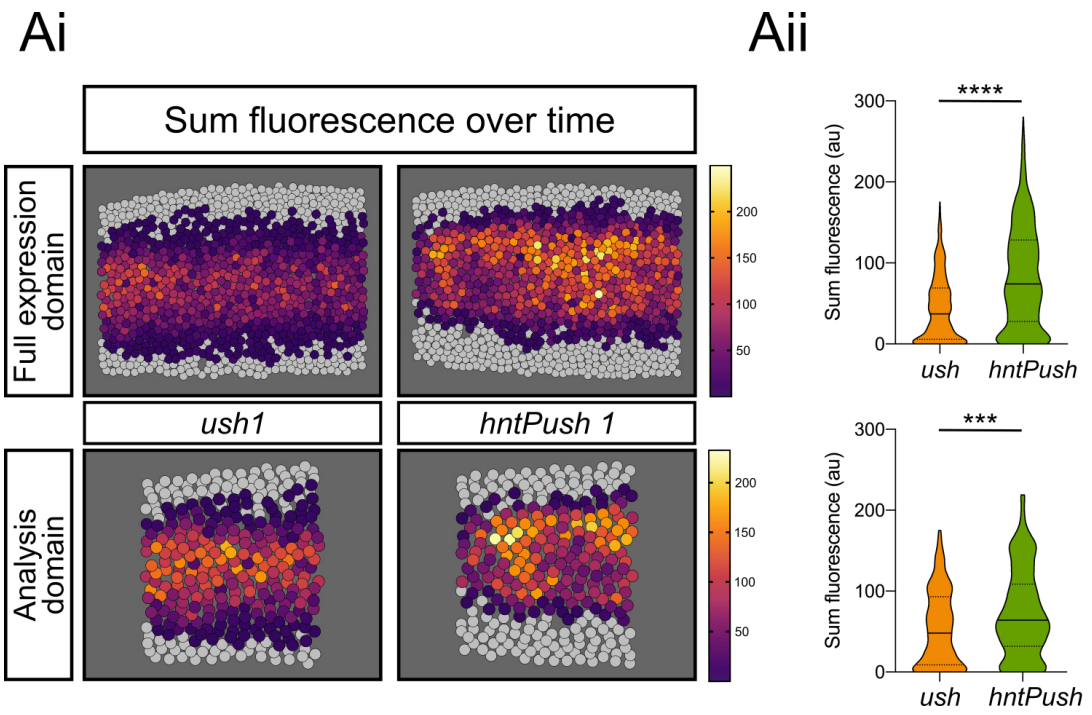



**Figure 5.14: Differences in sum fluorescence between *ush* and *hntPush* transcription.** (Ai) Representative images of *ush wt 1* and *hntPush 1* domains. Nuclei were coloured depending on their total sum fluorescence, and maximum values were chosen according to the maximum fluorescence of *hntPush*. (Aii) The graph shows the quantification of the the sum fluorescence for the individual replicates.
Lines in violin plots depict median and 95% confidence intervals, significance was tested with a Mann-Whitney test, ***p $<$ 0.001; ****p $<$ 0.0001. Expression domains are oriented with the anterior to the left.

### 5.2.4 Possible transcription biases in MS2 lines

Before investigating the *ush* transcription output of single-cells in detail, the influence of MS2-loops on transcription was further investigated. This section uses static images and smFISH probes to directly compare the transcriptional output of MS2 imaging lines (*ush* and *hntPush*) as well as CRISPR control lines. The CRISPR control lines were used earlier to compare amnioserosa cell counts and underwent the same CRISPR editing procedure (with identical scarring as imaging lines) but do not contain MS2 imaging loops.

First, the number of cytoplasmic mRNAs produced by CRISPR control lines was compared to wt embryos using exonic *ush* smFISH probes (Figure 5.15 A). The median number of *ush* transcripts produced by all genotypes is similar (Figure 5.15 A). The maximum number of transcripts produced, however, varies. Embryos homozygous for the *hntPush*Δ*MS2* locus are shown to produce larger numbers of maximum transcripts per cell (wt = 152, *hntPush*Δ*MS2* = 346, 256, 324; Figure 5.15 A). This observation is consistent with the increased mean nascent fluorescence that was shown in *hntPush* embryos during live imaging. Hence, the smFISH analysis of fixed embryos contributes additional evidence for increased *ush* transcription after the change in promoter. Next, the proportion of monoallelic nuclei within the expression domain was quantified. Both, *ush*Δ*MS2* and *hntPush*Δ*MS2* embryos show monoallelically expressing nuclei at level of wt embryos (compare, Figure 3.9 Aii and Figure 5.15 B). Embryos of the genotype *hntPush*Δ*MS2* show a slight but not significant reduction in monoallelic nuclei compared to *ush*Δ*MS2* embryos (Figure 5.15 B).

Finally, the cross-section distribution of *ush* transcript numbers per cell was visualised. The peak of mRNA number per cell in *hntPush*Δ*MS2* embryos is highest at the dorsal midline and greater than *ush*Δ*MS2* (Figure 5.15 C). Less embryo-to-embryo variation is observed between replicates of *hntPush*Δ*MS2* embryos than would have been predicted based on the great variability in nascent mean fluorescence levels observed in Figure 5.13 B. These data suggest that cytoplasmic mRNA numbers could be regulated post-transcriptionally to control transcription heterogeneity in promoter swap embryos. Additionally, the expression domains of *ush*Δ*MS2* and *hntPush*Δ*MS2* embryos appear wider than the representative wt domain (Figure 5.15 C). Overall, these data suggest that embryos with CRISPR edited *ush*Δ*MS2* loci produce similar numbers of cytoplasmic mRNAs and their transcription profile across the midline is comparable to wt embryos. Hence, MS2 studies

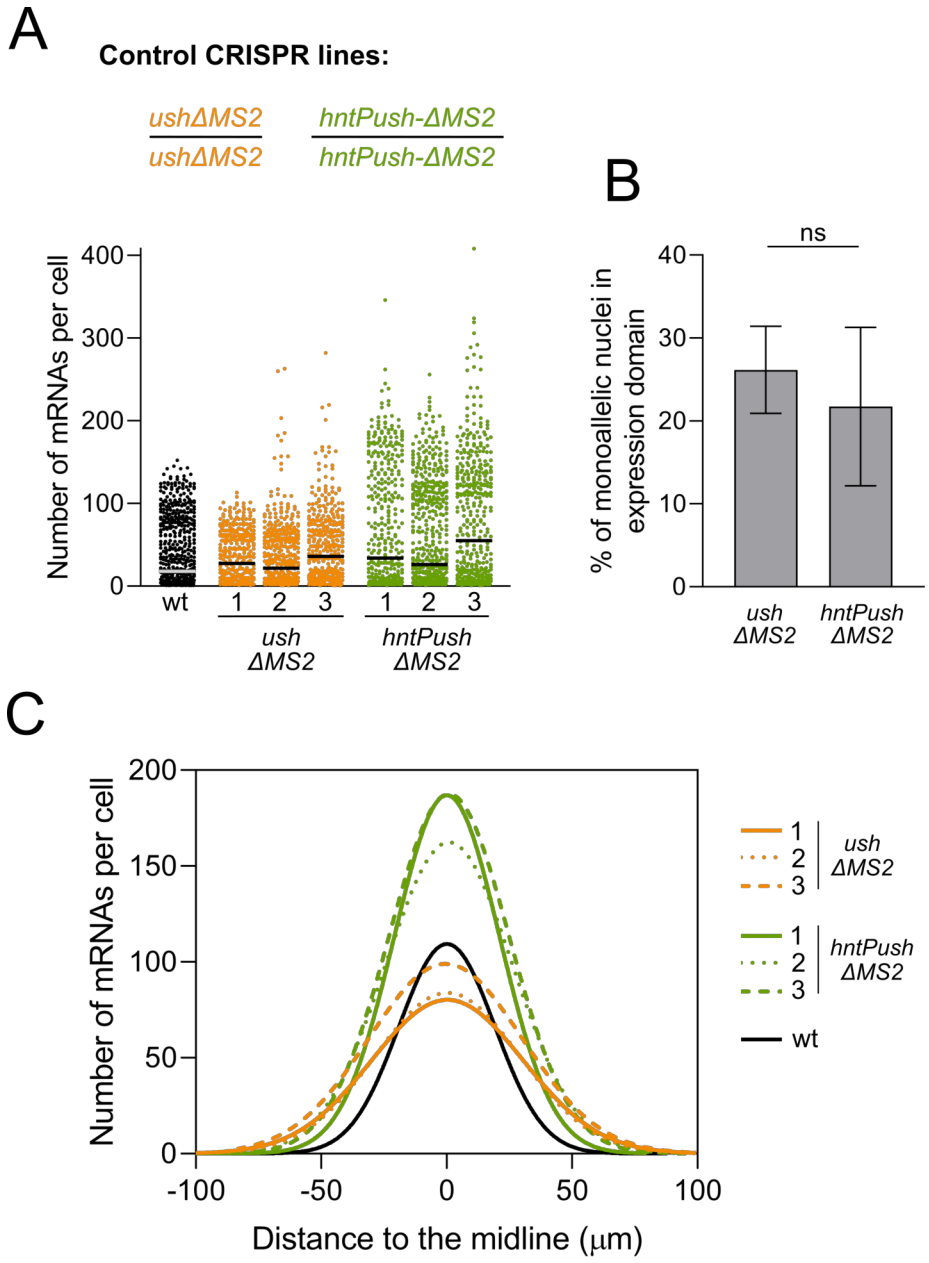



**Figure 5.15: Effects of CRISPR genome editing on *ush* transcription output.** (A) Control CRISPR lines were generated to investigate the effect of CRISPR engineering around the promoter region on *ush* transcription. smFISH was used to quantify cytoplasmic *ush* transcripts. The graph shows the number of mRNAs per cell in wt (black), *ushΔMS2* (orange) and *hntPushΔMS2* (green) embryos. (B) Proportion of monoallelic nuclei identified in *ushΔMS2* and *hntPushΔMS2* expression domains. (C) Gaussian curves representing biological replicates and the number of mRNAs per cell shown according to nuclear location across the midline.
Lines in (A) depict median and in (B) mean and SD. Number of biological replicates: wt = 1 representative, *ushΔMS2* = 3, *hntPushΔMS2* = 3; ns = not significant.

of *ush* transcription are likely to reflect that of the wt locus. To gain further understanding about the influence of MS2-loops on *ush* transcription, the fluorescence of nascent transcription sites was analysed.

Fixed embryos from crosses that were used for live imaging experiments were used to determine the number of nascent mRNA molecules (Figure 5.16 A). Embryos from imaging crosses contain one wt *ush* and one *ush-MS2* or *hntPush-MS2* allele. The fluorescence of the MS2 and the wt *ush* allele were recorded separately, and the MS2 allele was identified by a second set of smFISH probes targeting MS2-loop regions (Figure 5.16 A). The fluorescence of nascent transcription sites in wt embryos was quantified, to generate control values. Homologous alleles in wt embryos were assigned Allele A and Allele B at random. The mean number of mRNAs present in nascent transcription sites of wt Alleles A and B are shown to be similar at 25 and 26 (Figure 5.16 Bi). Next, the ratio of Allele A and Allele B within a nucleus was calculated to determine the contribution of the alleles relative to each other (Figure 5.16 Bii). Equal contribution of both alleles within a nucleus should result in a ratio of 1. Here, the median ratio of wt alleles is shown to be 0.96 (Figure 5.16 Bii). After defining the number of nascent *ush* mRNAs in wt embryos, the same analysis is performed on embryos resulting from the *ushMS2/+* imaging cross.

When the number of *ush* mRNAs present in nascent transcription sites is compared, it is observed that the *ush-MS2* allele contains significantly fewer mRNAs than the wt allele (Figure 5.16 Ci). The number of mRNAs is equivalent to the number of Pol II molecules that were transcribing the *ush* locus at the time of fixation. Therefore, the reduced number of mRNAs in transcription foci corresponds to under-transcription of the *ush-MS2* allele. This is also shown in the ratio of wt mRNA number to that produced by the *ush-MS2* allele (Figure 5.16 Cii). The observed median ratio of 1.4 indicates an asymmetric contribution to the overall mRNA output by alleles with a bias towards the wt allele (orange bar, Figure 5.16 Cii).

Based on the increased mean fluorescence in *hntPush* embryos, observed during live imaging, it can be predicted that more mRNAs will be present within nascent transcription foci in fixed embryos carrying a *hntPush-MS2* allele. Consistent with this prediction, significantly more mRNAs are found in *hntPush-MS2* foci compared to the corresponding wt alleles in the same embryo (Figure 5.16 Di). The higher transcription of the *hntPush-MS2* allele results in a mean ratio of 0.5 (wt vs MS2; green bar, Figure 5.16 Dii). Interestingly, the number of mRNAs in wt allele foci in

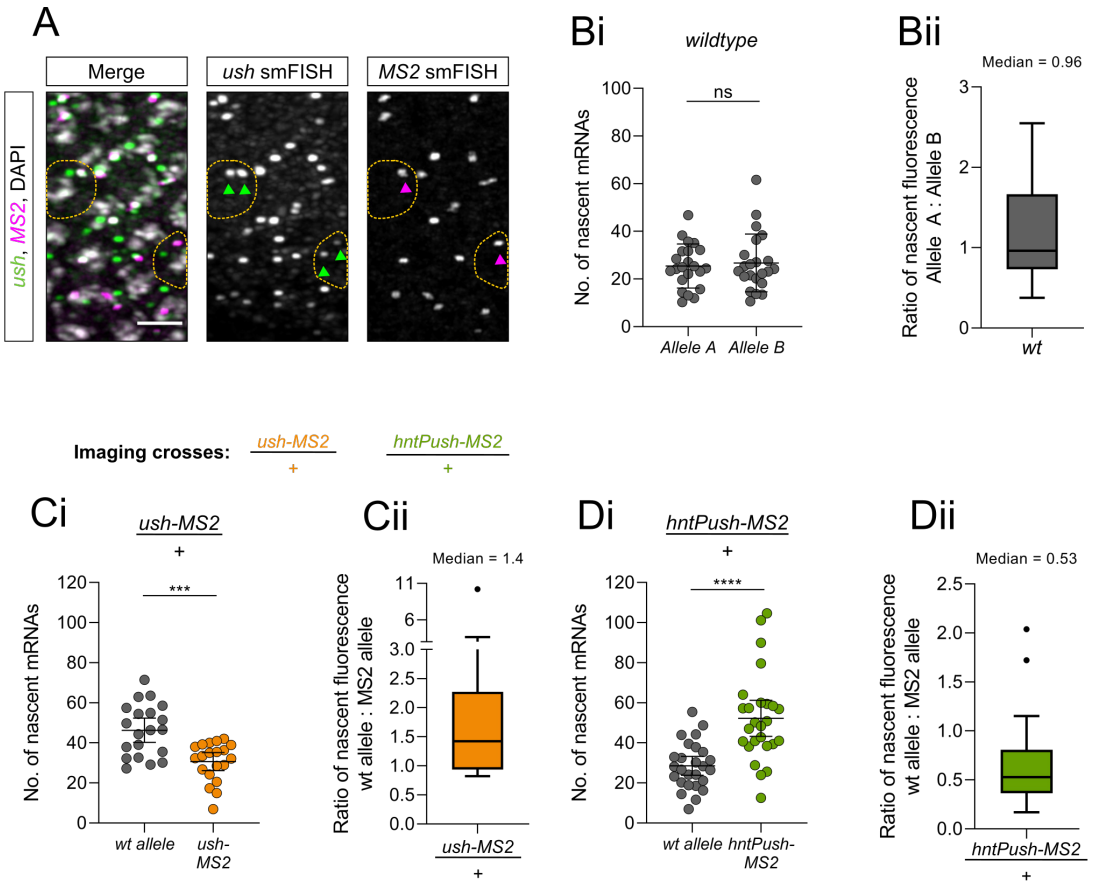



**Figure 5.16: Quantification of nascent mRNAs in wt and MS2 imaging lines using smFISH probes.** (A) The fluorescence intensity of fixed embryos (here *ush-MS2/+*) from imaging crosses carrying one wt and one *ush-MS2* allele were quantified using *ush* smFISH (green) probes. MS2 alleles were identified with a second smFISH probe set (pink). (Bi) The graph shows mRNA number in transcription foci of *ush* wt alleles. Alleles were assigned randomly. (Bii) Ratio of allele mRNA number is shown for wt. (Ci) Here, the allele contribution of wt (dark grey) and *ush-MS2* alleles (orange) is compared in *ush-MS2/+* embryos. (Cii) The graph shows the ratio of wt to MS2 allele mRNA number in *ush-MS2/+* embryos. (Di) The allele contribution of wt (dark grey) and *hntPush-MS2* alleles (green) is shown in *hntPush-MS2/+* nuclei. (Dii) Graph shows the ratio of wt to MS2 allele mRNA number in *hntPush-MS2/+* embryos.
Number of biological replicates: n = 22 (wt), 20 (*ush-MS2/+*), 26 (*hntPush-MS2/+*) nuclei from 1 representative embryo. Lines in (Bi, Ci, Di) show mean and SD. Boxplots (Bii, Cii, Dii) show median with interquartile range ($25^{th}$ - $75^{th}$ percentile); error bars show Tukey's whiskers. Significance was tested with unpaired Student's t-test (Bi, Ci, Di); ***$p < 0.001$; ****$p < 0.0001$, ns = not significant.

the *hntPush-MS2* embryo are lower than in the *ush-MS2* embryo (means = 46 and 29, respectively) (compare Figure 5.16 Ci and Di).

These data suggest that transcription biases exist between wt and MS2 tagged alleles. This limitation to the live imaging setup has to be taken into consideration but cannot currently be circumvented due to the need of crossing flies to the His-RFP;MCP-GFP line. It can be hypothesised that allele biases do not exist in homozygous *ush-MS2* embryos as live imaging of both homologous alleles in Chapter 4 showed correlated fluorescence between the two alleles (Figure 4.27 A). Ideally, flies homozygous for the MS2 imaging locus would be crossed to CRISPR control flies (containing CRISPR scar). That way the difference in transcription caused by the MS2-loops alone could be investigated in the future.

**Clustering results for *ush* transcription**

In *hntPush* embryos the clustering algorithm identified three groups of nuclear traces (Figure 5.17). As for *ush* embryos, they show a clear pattern in their position along the embryo cross-section and were therefore labelled middle (yellow), intermediate (green) and edge (orange) cluster (Figure 5.17). For this analysis *hntPush* embryo 2 was chosen, as the previous characterisation of *ush* expression in *hntPush* embryos showed large variability in fluorescence values between biological replicates (see Figure 5.13 B). Hence, the embryo with intermediate fluorescence intensity was used for this representative figure.

The mean fluorescence of nuclear traces is shown to be significantly greater in all clusters of the *hntPush* embryo (Figure 5.17). This observation is consistent with the previous observation of much increased mean fluorescence levels in *hntPush* embryos compared to *ush* embryos (compare, Figure 5.13 B).

Heatmaps were generated for each biological replicate individually as *hntPush* replicates showed embryo-to-embryo variability. Replicates two and three of *ush* can be found in Figure B.2 and replicate one and three for *hntPush* embryos can be found in Figure B.3. The heatmaps of replicates *ush* 1 and *hntPush* 2 correspond to the K-Means clusters and show highest fluorescence intensity in the middle cluster (top) and lowest intensity in the edge cluster (bottom, Figure 5.18).

The onset front of transcription was visualised to investigate differences between the clusters of the two genotypes. The onset front of the *ush* middle cluster was

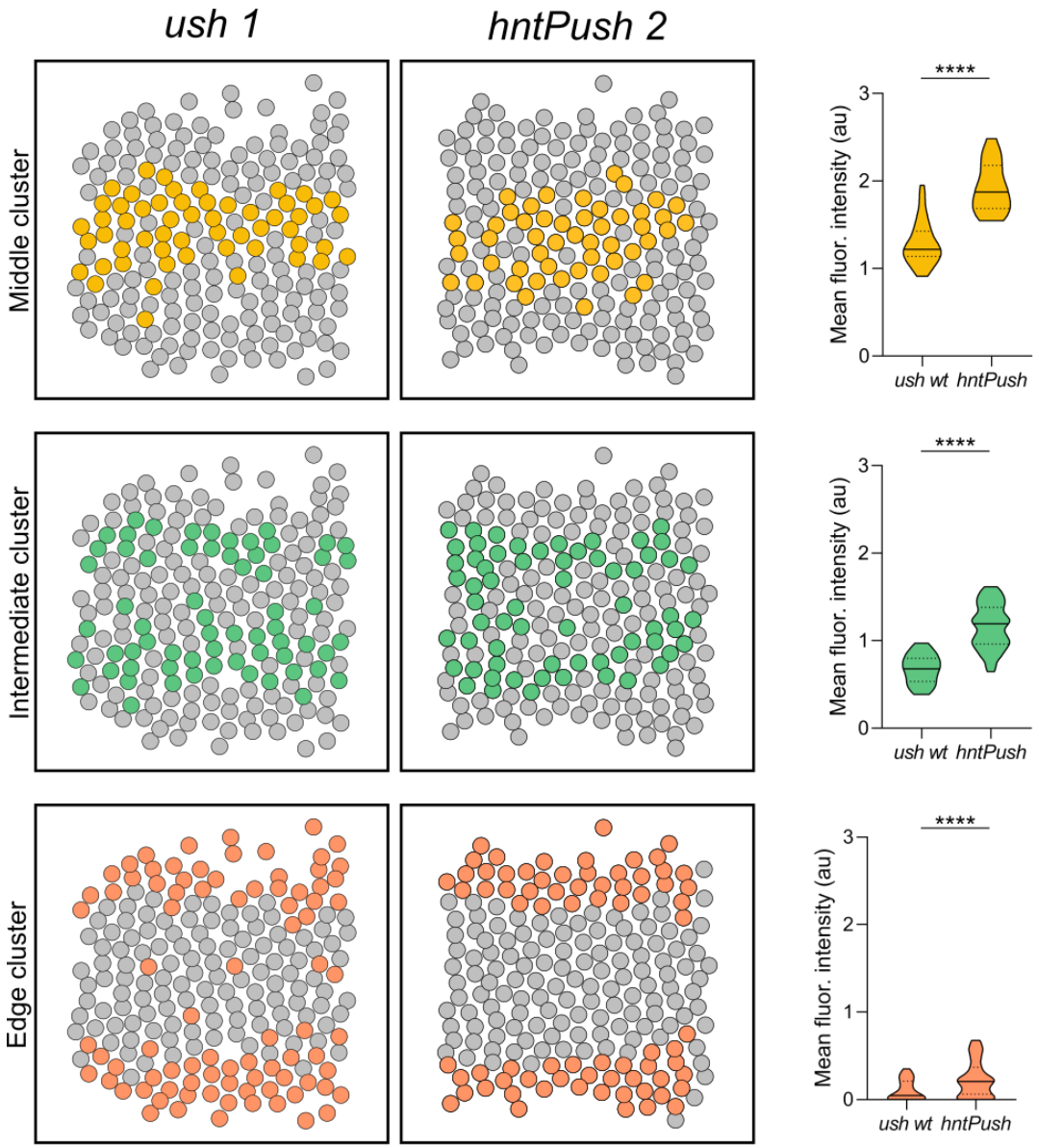



**Figure 5.17: K-mean clustering results for *hntPush* embryos.** (Left) Schematics show representative expression domains of *ush* wt embryo 1 and *hntPush* embryo 2. Nuclei are coloured depending on their K-means cluster (yellow=middle cluster, green = intermediate and orange = edge cluster). (Right) Mean fluorescence values from nuclei pooled according to their cluster.
Lines in violin plots depict median and 95% confidence intervals and significance was tested using a Mann-Whitney test; ****p<0.0001. Expression domains are oriented with the anterior to the left.

traced and is shown by the white line (Figure 5.18). The onset line was then superimposed onto the *hntPush* middle cluster and no difference is observed (Figure 5.18). Next, the time at which 50% of nuclei within a cluster had initiated transcription was determined, to further compare the clusters quantitatively between genotypes. This measurement was taken for all biological replicates and shows no significant difference in onset time between genotypes in the middle cluster (Figure 5.18). Likewise, other clusters were investigated by tracing the onset front of *ush* clusters and then superimposing them onto the equivalent *hntPush* clusters (Figure 5.18). The intermediate clusters showed a small but significant difference in onset time when traced and quantified (pink line, Figure 5.18). The *hntPush* intermediate clusters initiates transcription quicker than the comparable *ush* cluster (Figure 5.18). Similarly, the edge cluster of *hntPush* embryos reaches the 50% onset time significantly faster than the *ush* edge cluster (red line, Figure 5.18). These observations are consistent with an increased synchronicity within the field of cells, conferred by the *hnt* promoter sequence.

These results suggest that the previously observed earlier onset time of the *hntPush* expression domain is mostly due to the more responsive nuclei in the intermediate and edge region (see Figure 5.12 Cii). These data further predict that the change in promoter influences transcription dynamics most significantly in those domains of *hntPush* embryos. To investigate the changes in transcription dynamics, the memory-adjusted Hidden Markov Model will be used.

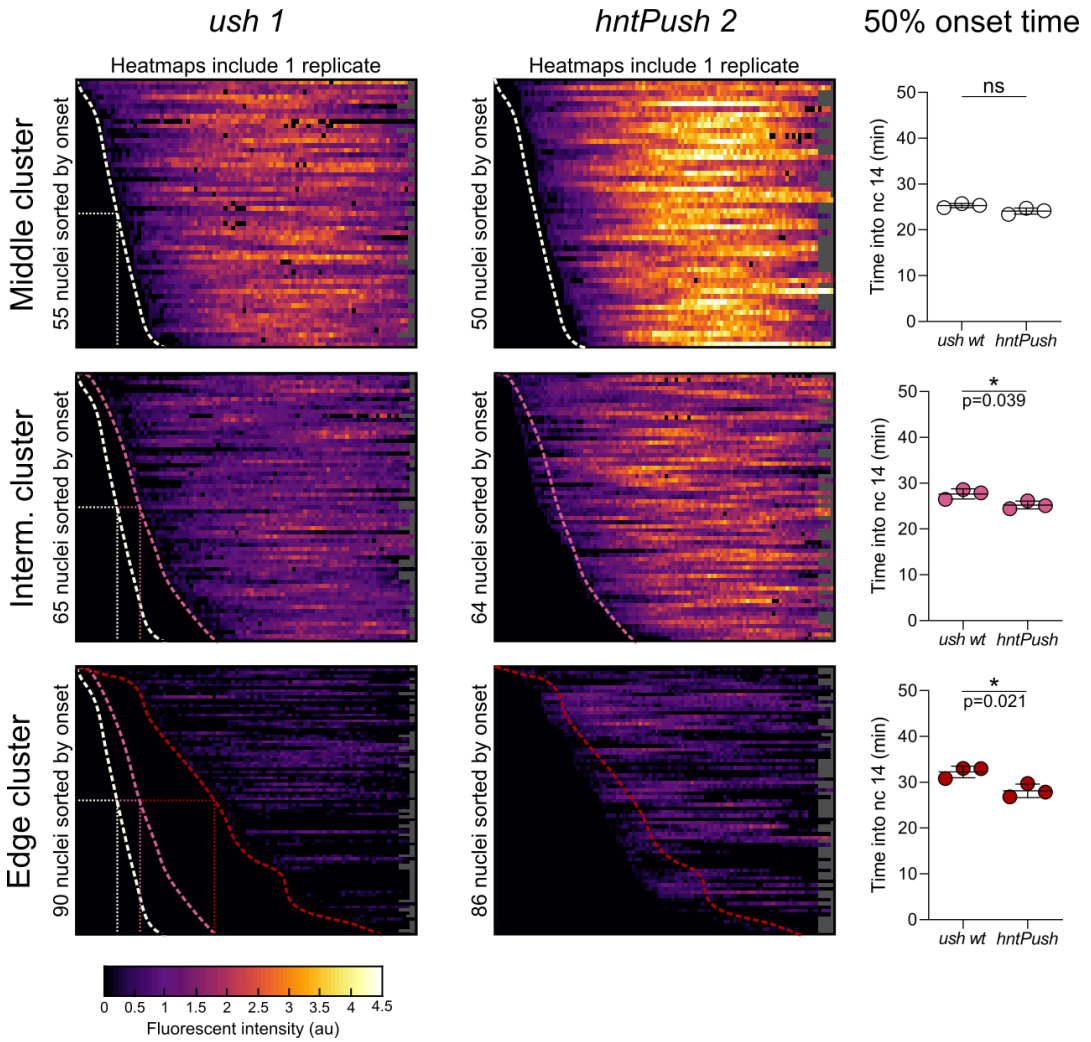



**Figure 5.18: Heatmaps of *hntPush* embryo K-Means clusters.** (Left) Heatmaps represent single nuclear traces according to their clusters (middle = top, edge = bottom). Nuclear traces were ordered by their onset times and the onset front of each cluster was traced according to the *ush* wt onset front and superimposed onto the heatmaps of the *hntPush* embryo. Scale as indicated below heatmaps, black = no expression, grey at the end of traces indicates periods where nuclei were not tracked. The time into nc14 at which 50% of nuclei had activated transcription was measured and compared (right). Graphs show 50% onset time for all biological replicates.
Nuclear traces start at the transcriptional onset time of the first nucleus. Lines represent mean and SD and significance was tested using a Student's t-test; *$p < 0.05$, ns = not significant. Number of biological replicates n = 1 in heatmaps and n = 3 in graphs for each genotype. Heatmaps of other replicates can be found in Figure B.2 and B.3. *ush 1* data was shown previously in Figure 4.13.

### 5.2.5 Changes in burst kinetics resulting from promoter swaps

**Burst kinetics for *hntPush* embryos**

The burst kinetics defining the transcriptional output of *ush* in *hntPush* embryos were investigated and compared to *ush* and *hnt* embryos to determine which burst parameters were affected by exchanging the promoter. The promoter states and their transition rates were determined using the memory-adjusted HMM (Lammers et al. 2019) and parameter definitions can be found in Figure 4.15 Aiii. Transcription parameters were analysed for the three clusters of *hntPush* transcription that were identified using the K-Means clustering algorithm (see Figure 5.17). Analysing these regions allows comparison between *hntPush* and *ush* as well as *hnt* embryos.

One representative single nuclear fluorescence trace is shown for each *hntPush* cluster together with their respective promoter states in Figure 5.19 A. The highest number of transcriptional bursts is shown in the central cluster with fewer bursts in the intermediate and edge clusters (Figure 5.19 A). This observation suggests different burst frequencies between the three clusters as previously shown for *ush* transcription in Chapter 4.

A global analysis of all burst parameters was performed for the three *hntPush* replicates and compared to the *ush* and *hnt* parameters. Changes between genotypes in relation to the centre and intermediate clusters will be focused on as the edge region was shown in the previous chapter to experience very low burst activity with poor values for every parameter. Burst size, which is a product of the burst amplitude and duration, is shown to be slightly increased in the centre and intermediate cluster of *hntPush* embryos compared to *ush* and *hnt* (Figure 5.19 B). This observation is consistent with increased mean and maximum fluorescence levels that were detected in *hntPush* embryos (see Figure 5.14 and 5.13 B). Here, the burst size is shown to be influenced by both parameters working against one another. The burst duration in the central cluster of *hntPush* embryos is decreased compared to *ush* as a result of slightly increased Koff values and thereby forms an intermediate between *ush* and *hnt*. Hence, the *hnt* promoter in the *ush* gene locus transitions into the OFF state with a slightly higher rate (Figure 5.19 B). The slight decrease in burst duration is shown in the representative central cluster trace when compared to the *ush* trace (compare Figures 4.16 B and 5.19 A).

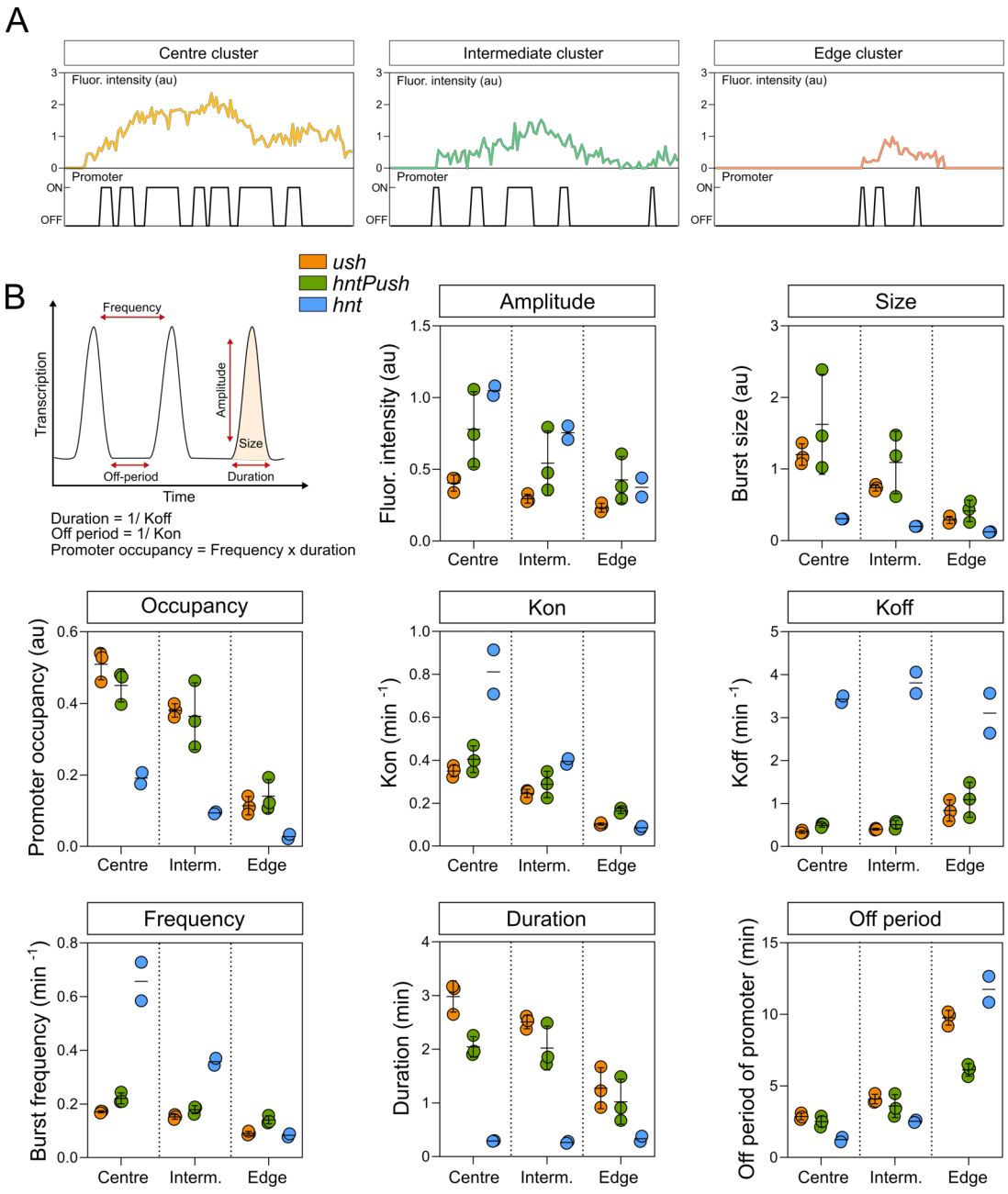



**Figure 5.19: Burst parameters for *hntPush* promoter swap embryos.** (A) Representative fluorescence traces are plotted from nuclei expressing *ush* and corresponding to the three domains with their inferred promoter states shown below. (B) The schematic shows burst parameters. Graphs show global analysis of burst parameters identified for *ush* transcription in *hntPush* (green) and *ush* (orange) embryos as well as *hnt* (blue).
Biological replicates: *ush* n = 3, *hntPush* n = 3, *hnt* n = 2. Lines in (B) show mean and SD.

The burst amplitude is increased in all clusters, resulting in an overall increase in burst size in *hntPush* embryos compared to *ush* (note large SD in amplitude indicating embryo-to-embryo variation, Figure 5.19 B). These data suggest that the addition of the *hnt* promoter changed the burst size of *ush* transcription by modulating the amplitude, and to a lesser extend duration, of *ush* transcription bursts and now forms an intermediate between *ush* and *hnt* transcription.

The change in promoter is shown to have little effect on Kon and Koff values when comparing *hntPush* and *ush* (Figure 5.19 B). Consistent with this, no difference is observed in the mean promoter occupancy between *ush* and *hntPush* embryos in any of the three clusters (Figure 5.19 B). The promoter occupancy was identified in Chapter 4 to be one of the key parameters integrating changes in BMP signalling levels. The observation that the promoter occupancy is unchanged in the promoter swap embryos suggests that occupancy is not regulated by the promoter but instead is a feature controlled by the enhancer. Very little changes are observed in burst frequency between genotypes (Figure 5.19 B). This is consistent with burst frequency being controlled by Kon and Koff, which are largely unchanged (Figure 5.19 B).

In summary, changing the promoter appears to have the biggest effect on *ush* burst amplitude. Furthermore, no change in promoter occupancy, Kon, Koff and little change in burst frequency parameters suggests that these burst features are regulated not by the promoter but most likely by the enhancer. When compared to hnt, the hntPush promoter parameters form an intermediate between ush and hnt in terms of the burst amplitude and duration.

Both genotypes *ush* and *hntPush* show consistent changes between the centre and intermediate cluster, which receive peak and intermediate levels of BMP signalling, respectively. The decrease in burst size between the centre and intermediate cluster in *hntPush* is mainly controlled through changes in amplitude, similar to *ush* wt (Figure 5.19 B). Additionally, both genotypes show a decrease in promoter occupancy between the central and intermediate cluster, which is controlled by decreasing levels of Kon while Koff stays constant (Figure 5.19 B). These data are consistent with previous observations and suggest that BMP signalling affects promoter occupancy, regardless of the nature of the promoter tested.

Next, the changes in burst kinetics resulting from the promoter swap were investigated in more detail. Transcriptional bursts were simulated for a 30 min period, corresponding to the maximum length of *ush* transcription that was followed in this

study (see Figure 4.11 B), using mean burst parameters of amplitude, duration, and off-period in the central cluster (Figure 5.20 A). It can be observed that the burst amplitude in *hntPush* embryos is much higher and the burst frequency is slightly shorter than in *ush* embryos (Figure 5.20 A). The slightly shorter burst frequency in *hntPush* is shown to result in six full bursts during the 30 min window, whereas *ush* bursts five time in the comparable time frame (Figure 5.20 A).

After characterising the burst profiles of *ush* and *hntPush* they were compared to the transcription profile of *hnt* (data from Figure 4.18). The *hnt* transcription profile was simulated for 20 min as *ush* transcription precedes that of *hnt* by approximately 10 min during nc14 (see Figure 4.11 B). Based on the transcription profiles it appears that the *hntPush* promoter drives bursts that are hybrids between that of *hnt* and *ush* (Figure 5.20 B). The *hnt* burst profile shows very short duration bursts with short off-periods and high amplitude (Figure 5.20 B). Conversely, *ush* transcription is characterised by longer duration bursts with lower amplitude (Figure 5.20 B). The *hnt* promoter inserted in the *ush* gene produces burst profiles that increase the amplitude of *ush* and reduce the duration to create a profile that becomes an intermediate between the *hnt* and *ush* profiles (Figure 5.20 B). Additionally, the slightly reduced burst duration identified in *hntPush* embryos suggests a change towards the *hnt* short-duration profile (Figure 5.20 B). Overall, the hybrid profile of *hntPush* embryos is still closer to *ush* than to *hnt* transcription.

The contribution of individual clusters to the total *ush* expression output remains unchanged in promoter swap embryos despite changes in the kinetics (Figure 5.20 C). Even though *hntPush* embryos experience different burst kinetics, the increase is proportional between *ush* and *hntPush* embryos and within clusters (Figure 5.19 B) and therefore the overall contribution by clusters is not changed (Figure 5.20 C). Besides the increase in burst size, no change was identified in the promoter occupancy value between genotypes. The occupancy parameter was previously shown to be responsive to BMP signalling. The experiments in this section suggest that the promoter occupancy parameter is not regulated by the promoter and therefore suggests its regulation instead is through the enhancer.

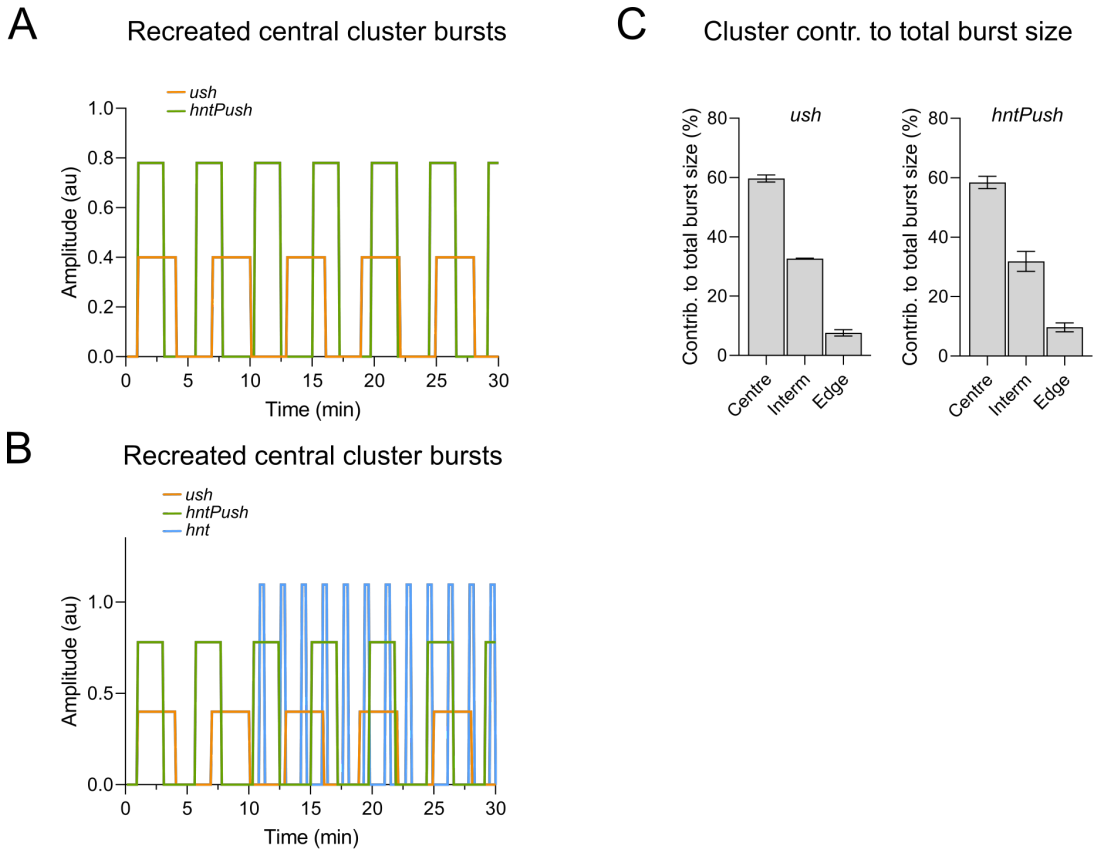



**Figure 5.20: The *hntPush* promoter swap forms a hybrid burst profile between *hnt* and *ush* transcription.** Bursting simulation of *ush* transcription in *ush* wt (orange) and *hntPush* (green) embryos (A) and compared to *hnt* transcription (blue) in (B). Bursts were calculated based on mean burst parameter values from the central clusters and are shown for a transcription period of 30 min for *ush* and 20 min for *hnt*. (C) Contribution of individual clusters to the total transcription output in *ush* and *hntPush* embryos. Percentages calculated based on mean burst parameters from (A).
Biological replicates: *ush* n = 3, *hntPush* n = 3. Lines in (C) show mean and SD.

**Burst kinetics in *ushPhnt* embryos**

After characterising the changes in burst kinetics from *hntPush* promoter swap embryos, the transcription kinetics defining the transcriptional output of *hnt* in *ushPhnt* embryos was investigated. Similar to the *hnt* wt embryos analysed in Chapter 4, the nuclei in *ushPhnt* embryos were separated into three domains based on their mean expression levels. This was done to allow comparison to the *hnt* wt data described previously. The new nuclear domains are referred to here as central, intermediate and edge as clustering shows a general agreement with nuclear position across the midline.

Transcription parameters were analysed for the three domains in two biological replicates of *ushPhnt* embryos. Analysis of mean fluorescence intensity highlighted the large embryo-to-embryo variability in *ushPhnt* embryos (see Figure 5.6 B). In order to be able to draw preliminary conclusions from the analysis only the burst parameters of the *ushPhnt* 1 replicate were plotted in Figure 5.21 B, as there are currently technical issues associated with defining all the parameters for replicate 2. Additionally, here the mean parameter values of the two *hnt* replicates (see Figure 4.18) as well as *ush* parameters were plotted for comparison.

Representative single nuclear fluorescence traces are shown for each *ushPhnt* domain together with their respective promoter states in Figure 4.18 A. The comparison of transcription kinetic parameters will focus on the centre and intermediate cluster as well as differences between the two genotypes as the edge cluster was shown before to be much more variable with poor burst activity.

A large difference between *ushPhnt* and *hnt* embryos is observed in burst size. The burst size in *hntPush* is reduced by approximately half compared to *hnt* embryos and results primarily from a reduction in amplitude, with a smaller contribution from reduced duration (Figure 5.21 B). The changes in burst duration arise from a small increase in Koff in the central cluster but appear unchanged between genotypes in the intermediate cluster (Figure 5.21 B). Compared to *ush*, both *hnt* and *ushPhnt* have a much lower burst size and amplitude.

A slight reduction in burst frequency is detected in the centre and intermediate region of the *ushPhnt* embryo, placing it as an intermediate between *hnt* and *ush* (Figure 5.21 B). The decrease of *ushPhnt* in relation to *hnt* is caused by a minor increase in Koff and decrease in Kon values (Figure 5.21 B). Consistent with little

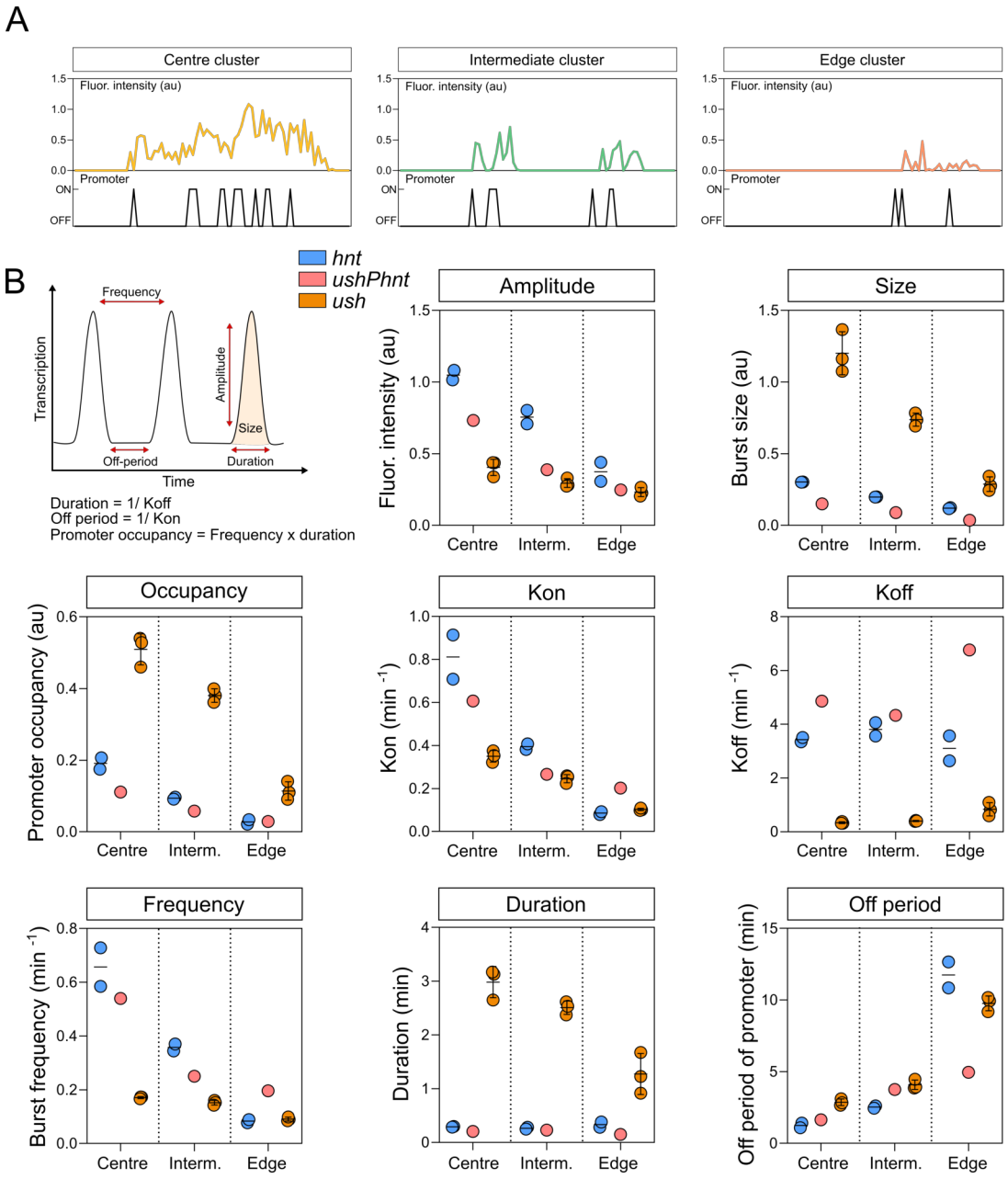



**Figure 5.21: Burst parameters for *ushPhnt* promoter swap embryos.** (A) Representative fluorescence traces are plotted from nuclei corresponding to the three domains with their inferred promoter states shown below. (B) The schematic shows burst parameters. Graphs show global analysis of burst parameters identified for *hnt* transcription in *ushPhnt* (red) and *hnt* (blue) embryos as well as *ush* (orange) transcription.
Biological replicates: *hnt* n = 2, *ushPhnt* n = 1. Symbols for *hnt* show mean and for *ush* show mean and SD in (B). Symbols for *ushPhnt* show one biological replicate (*hnt 1*) in (B).

difference in the Kon values, the off-period of *ushPhnt* is similar compared to *hnt* embryos (Figure 5.21 B).

The fraction of promoter activity, defined by the promoter occupancy value, shows a small decrease in occupancy in the centre and intermediate clusters of the *ushPhnt* embryo (Figure 5.19 B). This is in contrast to the unchanged promoter occupancy detected during *ush* transcription in *hntPush* embryos (Figure 5.19 B). The previous observation in *hntPush* embryos suggested that the occupancy is independent of the promoter as it is unchanged in the promoter swap. More biological replicates will need to be analysed to determine how reproducible this change in occupancy is in the central cluster. Additional replicates will also allow the significance of small changes in other parameters to be determined.

Consistent with the *hntPush* promoter swap embryos, the burst parameter that was identified to be affected the most by the *hnt* promoter swap is burst size. This observation suggests that the burst size is most likely controlled by the promoter (Figure 5.19 B). The decrease in burst size in *ushPhnt* compared to *hnt* is caused by a decrease in amplitude and to a lesser extend burst duration (Figure 5.19 B).

The differences in *ushPhnt* burst kinetics are most notable in the central cluster (Figure 5.21 B). This is consistent with previous observations using heatmaps of K-Means clusters that showed greater variability between *hnt* and *ushPhnt* embryos in the middle cluster (see Figure 5.9). Consistent with the larger differences in burst parameters between genotypes observed in the central region, the overall contribution to the expression domain is found to be altered. In *hnt* embryos the central domain contributes approximately 60% of the overall expression output (Figure 5.22 A). This fraction is increased in *ushPhnt* embryos further indicating a change in overall transcription behaviour in this embryo (Figure 5.22 A).

Using the mean bursting parameters from the central and intermediate clusters, the burst profiles of *ushPhnt* and *hnt* transcription can be compared. Visualisation of the mean burst profile in *hntPush* embryos showed an intermediate burst behaviour between *ush* and *hnt* transcription (Figure 5.20 B). Accordingly, it can be predicted that the *ushPhnt* promoter would also generate an intermediate burst profile with a burst amplitude smaller than that of *hnt* and a lower frequency. Consistent with this prediction, bursts in the central domain of the *ushPhnt* 1 embryo showed a reduction in burst amplitude as well as a slight decrease in burst frequency (Figure 5.22 B). Direct comparison to the *ush* promoter profile reveals that the promoter swap

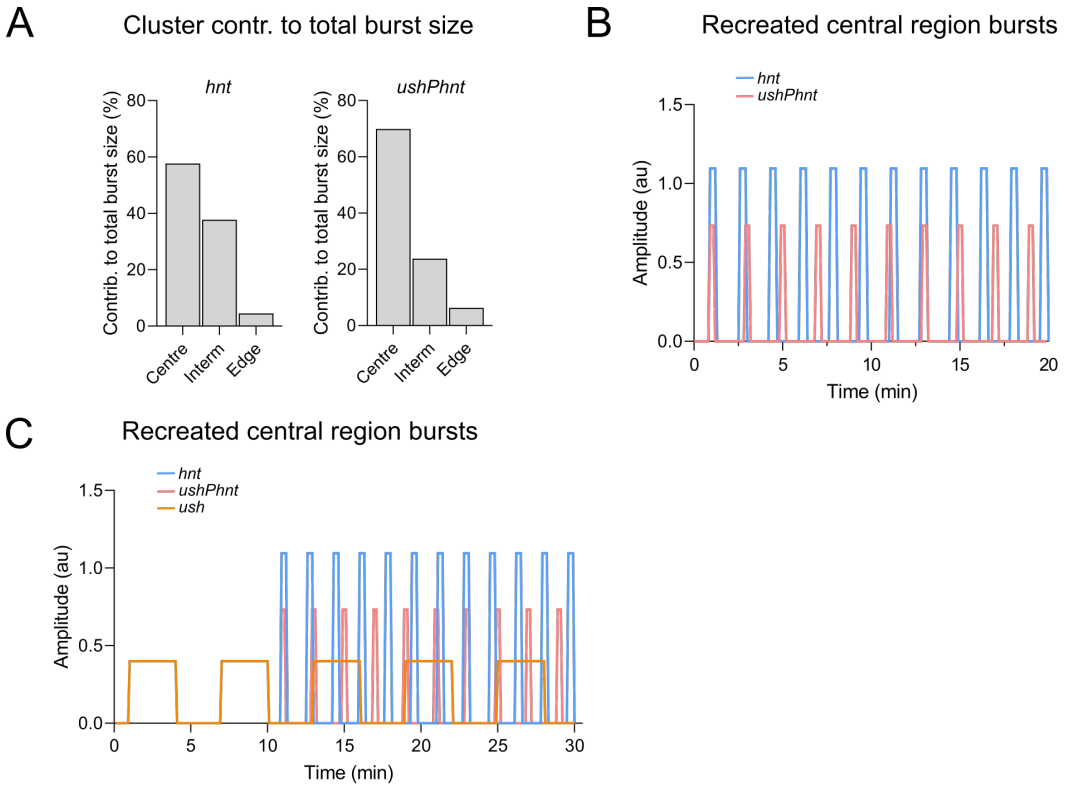



**Figure 5.22: The *ushPhnt* promoter swap forms a hybrid burst profile between *hnt* and *ush* transcription.** (A) Contribution of individual clusters to the total transcription output in *hnt* and *ushPhnt* embryos. Percentages calculated based on mean burst parameters. (B) Bursting simulation of *hnt* transcription in *hnt* wt (blue) and *ushPhnt* (red) embryos in the central domain. Bursts were simulated for a 20 min transcription window based on mean burst parameter values. (C) Burst simulation of *hnt* transcription in wt and the *ushPhnt* embryo compared to *ush* (orange) transcription. Biological replicates: *hnt* n = 2, *ushPhnt* n = 1.

results in an intermediate burst amplitude between *hnt* and *ush* (Figure 5.22 C). Likewise, the burst frequency in *ushPhnt* decreases, although a high burst frequency is retained that is much more similar to that of *hnt* (Figure 5.22 C). Burst duration shows little change with possibly a small decrease, again retaining features similar to that observed for *hnt*. Therefore, both promoter swaps result in burst profiles in between that of the two wildtype promoters, although this is more obvious for introducing the *hnt* promoter into *ush*. The converse swap (*ush* promoter into *hnt*) results in an intermediate profile that retains more characteristics associated with the endogenous *hnt* locus.

## 5.3 Contribution of proximal and shadow enhancers expression patterns

After investigating the influence of promoter regions on transcription, this study will now investigate the contribution of individual enhancers to gene expression profiles. Proximal and shadow enhancers have been shown to confer robustness of gene expression, and therefore, are important to ensure correct tissue formation during embryonic development (Frankel et al. 2010, Perry et al. 2010). Novel enhancers are identified by examining genome regions, which are bound by TFs using ChIP-Seq experiments (Zeitlinger et al. 2007, Deignan et al. 2016). The genomic region of *ush* is shown together with the analysis of pMad (Deignan et al. 2016) and Zld (Harrison et al. 2011) ChIP-Seq tracks obtained from *Drosophila* embryo *in vivo* experiments (Figure 5.23 A). This study identified three putative *ush* enhancer regions based on called peaks (Figure 5.23 A). The identified putative enhancer regions will be tested in isolation to investigate if they can drive transcription and recapitulate the *ush* expression domain.

Three enhancers have been described to drive expression patterns that are similar to the endogenous *tup* expression (Zeitlinger et al. 2007, Deignan et al. 2016). Two of these enhancer regions contain ChIP-Seq peaks for pMad and Zld, and the third enhancer only binds pMad (Figure 5.23 B). These three enhancers have been validated in other studies (Zeitlinger et al. 2007, Deignan et al. 2016) but will be investigated in this study using higher resolution imaging.

### 5.3.1 Investigation of putative *ush* enhancers

Three potential enhancer regions of *ush* were identified using ChIP-Seq data. One region, here called proximal enhancer, is located directly upstream of *ush* isoforms RA and RC (Figure 5.24 Ai). Shadow enhancer 1 is located ~4 kb upstream of the shorter two *ush* transcripts and the shadow enhancer 2 region is located within an *ush* intron and is located furthest away from *ush* isoforms RA and RC, which are expressed during early embryogenesis (Figure 5.24 Ai).

Transgenes were generated to study these enhancers in isolation and determine if each region is sufficient to drive transcription in a pattern similar to endogenous *ush*. Each enhancer region was amplified from genomic DNA and inserted in the same location upstream of a *yellow* reporter gene. The three resulting reporter

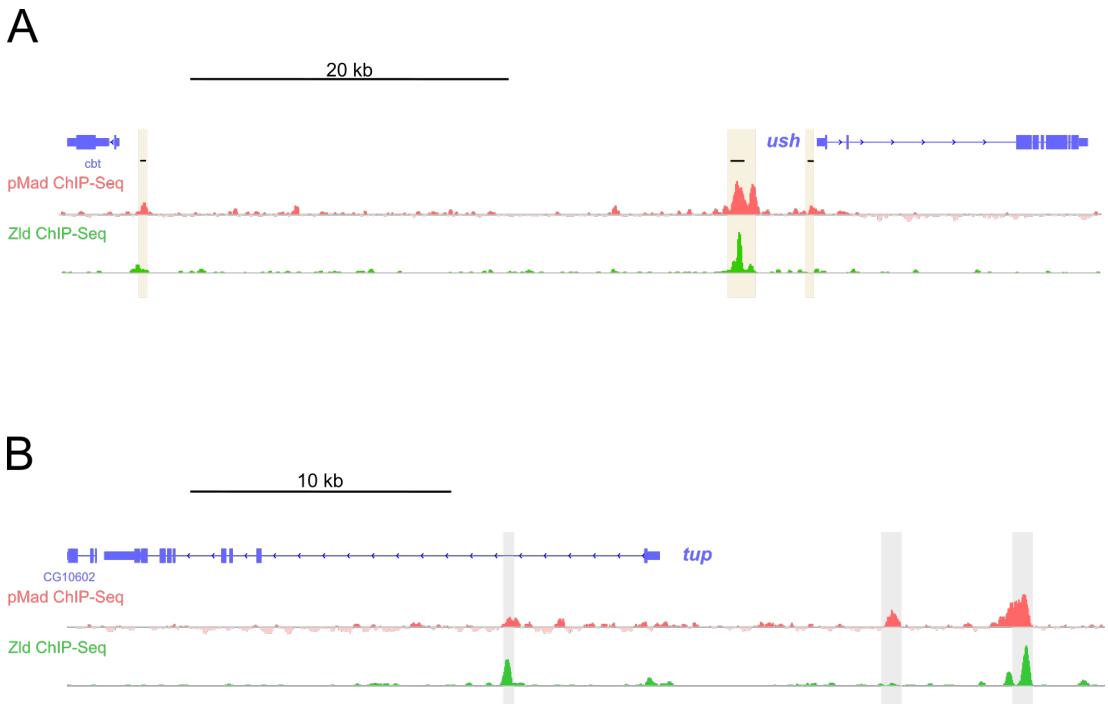



**Figure 5.23: Identification and visualisation of shadow enhancers.** Genome browser view of *ush* (A) and *tup* (B) genomic loci. Annotated gene regions (blue) are overlaid with ChIP-Seq tracks of pMad (red, Deignan et al. (2016)) and Zld (green, Harrison et al. (2011)) showing protein binding sites. Experimentally verified enhancer sequences for *tup* are highlighted in grey. Called peaks shown by black lines and potential enhancer sequences for *ush* are highlighted in yellow.

constructs were integrated into the same genomic region on the third chromosome (Figure 5.24 Aii). The resulting fly lines express endogenous *ush* on the second and at the same time *yellow* on the third chromosome under the control of the test *ush* enhancer sequence.

Next, the expression pattern of *yellow* was analysed using colourimetric ISH and compared to the expression pattern of *ush* in wt embryos (Figure 5.24 B). *yellow* transcription driven by the *shadow enhancer 1* sequence shows an expression domain that is very similar to endogenous *ush* in early embryos (stages 5-6) (Figure 5.24 B). The other two putative enhancer sequences did not drive *yellow* transcription in the pattern of *ush*. The proximal enhancer sequence shows *yellow* transcription in the posterior end of the embryo (Figure 5.24 B). This region does not correspond to the wildtype *ush* expression domain and possibly shows ectopic activation of *yellow*. Together, these data suggest that there is one dominant early enhancer of the *ush* gene. This enhancer sequence contains binding motifs for pMad and is the only sequence identified to contain a canonical Zld site (Figure 5.23 A). The other identified enhancer sequences may become active later during *Drosophila* development or adult life. Furthermore, it is possible that the sequences could not drive *yellow* expression because the test sequence was too short. Since this study aims to investigate the contribution and behaviour of multiple enhancers driving gene expression, the *ush* locus is not suitable for this analysis. Hence, this study will focus on the previously characterised *tup* enhancers.

### 5.3.2 Investigation of *tup* enhancers in isolation

The proximal *tup* enhancer was identified in the gene's first large intron (Zeitlinger et al. 2007) (Figure 5.25 Ai). Two shadow enhancers were identified in a subsequent study located 8.7 kb and 14.3 kb downstream of *tup* (Deignan et al. 2016) (Figure 5.25 Ai). All enhancer sequences were placed upstream of the *yellow* reporter construct in isolation (Figure 5.25 Aii). These constructs differ from previous studies as the lacZ reporter gene was replaced with *yellow*. The *yellow* gene contains an intron and therefore intronic FISH probe sequences could be used. Additionally, a triple enhancer construct was made, where all three enhancer sequences were placed upstream of the *yellow* gene to mimic the wt gene architecture albeit without the correct spacing. All reporter constructs were integrated into the fly genome on the 3$^{rd}$ chromosome. Hence, these flies express endogenous *tup* in addition to *yellow* under the control of *tup* enhancers. The expression pattern of *yellow* was

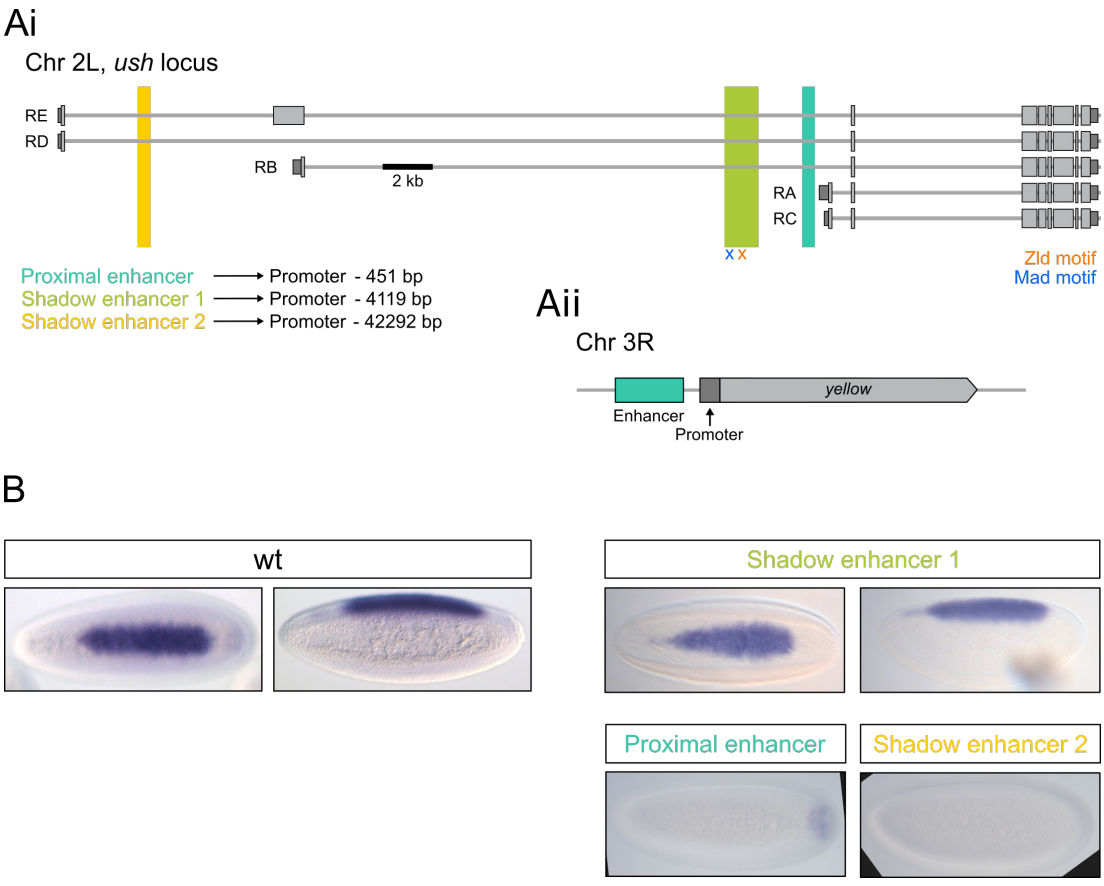



**Figure 5.24: Identification of a novel *ush* enhancer.** (Ai) Schematic of the endogenous *ush* locus with its five isoforms. Putative enhancer sequences are highlighted, proximal enhancer (dark green), shadow enhancer 1 (light green) and shadow enhancer 2 (yellow). Zld (orange) and pMad (blue) binding motifs are indicated. (Aii) This schematic shows the reporter transgene for testing enhancer sequences in isolation. Putative enhancer sequences were inserted upstream of a *yellow* reporter construct. Reporters were integrated into the same landing site on the 3$^{rd}$ chromosome. (B) ISH was performed to determine expression patterns of *yellow*. Wt control shows ISH using a *ush* probe and enhancer transgenes show ISH using a *yellow* probe. Shadow enhancer 1 was able to recapitulate the *ush* expression pattern.
Images show dorsal and lateral view for wt and shadow enhancer 1. Only dorsal views shown for the proximal enhancer and shadow enhancer 2. All embryos were oriented with the anterior to the left.

visualised using ISH probes targeting *yellow* sequences comparing the expression pattern to that of wt *tup* (Figure 5.25 B). The proximal enhancer recapitulated the *tup* expression pattern along the full AP axis (Figure 5.25 B). Shadow enhancers 1 and 2 show *yellow* expression resembling endogenous *tup*, but with reduced anterior expression compared to endogenous *tup* (Figure 5.25 B). The triple enhancer construct drives *yellow* expression that recapitulates the full AP *tup* expression domain. These data show that all transgenes that were generated are able to drive *yellow* transcription in response to individual *tup* enhancers. These fly lines will be used to study *tup* expression through analysis of monoallelic and biallelic transcription.

### 5.3.3 Identification of *tup* enhancer strength

Ideally, changes in transcription due to enhancer isolation would be investigated live using the MS2/MCP system. Due to the time constrains of this study live imaging lines could not be generated and instead a preliminary investigation of transcription states was performed using FISH. In Chapter 3 this study hypothesised that monoallelic transcription in response to BMP signalling levels is indicative of reduced transcription burst frequency or size. This hypothesis was further supported in Chapter 4 using live imaging coupled with computational modelling. Therefore, the transcription state of nuclei will be used here to characterise and compare *yellow* and *tup* transcription.

Embryos homozygous for the reporter transgene were stained using RNA-FISH and smFISH probes against *yellow* and *tup* sequences, respectively (Figure 5.26). Since the transgenes were inserted on the 3$^{rd}$ chromosome in addition to endogenous *tup*, both *tup* and *yellow* expression can be visualised at the same time. Exerts from full embryos show parts of the active expression domains that overlap (Figure 5.26 A). Nascent transcription sites were quantified for both genes and shown in representative images that highlight monoallelic (orange) and biallelic (blue) transcription (Figure 5.26 A). *yellow* transcription driven by the proximal *tup* enhancer generated an expression domain that looks very similar to endogenous *tup* in both its shape and composition of mono- and biallelic nuclei (Figure 5.26 A).

The *yellow* expression domain generated by shadow enhancer 1 resembles that of *tup* but appears to be much thinner and shows almost exclusively monoallelic transcription (Figure 5.26 A). The transcription domain that was driven by shadow enhancer 2 again resembles that of endogenous *tup* but with extended spikes in

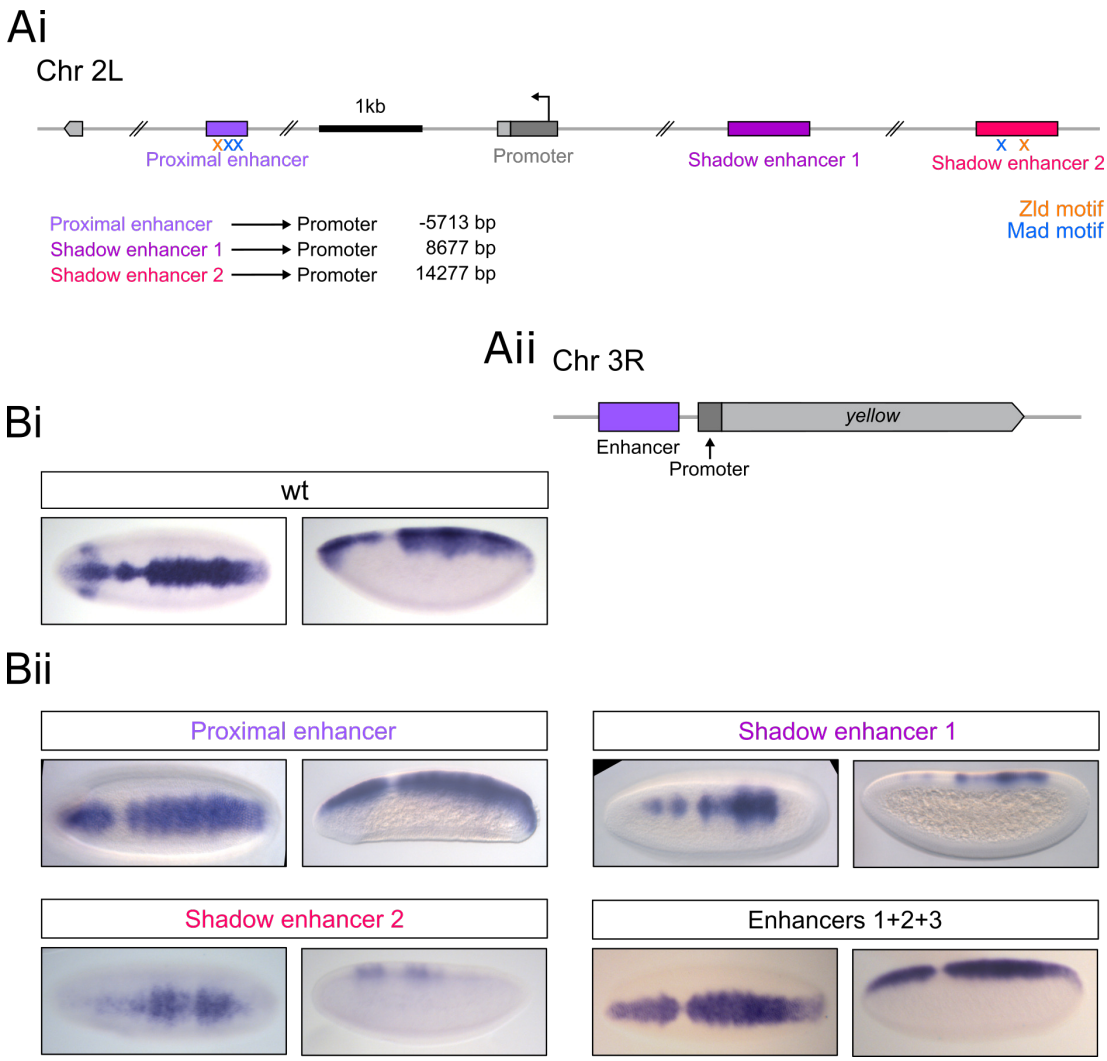



**Figure 5.25: *tup* transcription profile changes in different genotypes.**(Ai) Schematic of the endogenous *tup* locus. Validated enhancer sequences are highlighted, proximal enhancer (purple), shadow enhancer 1 (light purple) and shadow enhancer 2 (pink). Binding motifs of pMad (light blue) and Zld (orange) are indicated. (Aii) This schematic shows the reporter transgene for testing enhancer sequences in isolation. *tup* enhancer sequences were inserted upstream of a *yellow* reporter construct. Reporters were then integrated into the same landing site on the 3$^{rd}$ chromosome. (Bi) Wt control shows ISH using a *tup* probe. (Bii) ISH was performed to determine expression patterns of *yellow*. A triple reporter construct was also made containing all three enhancer sequences, in the correct order, upstream of the *yellow* gene. All enhancer sequences drive *yellow* expression patterns that are similar to endogenous *tup*.
Images show dorsal and lateral views of embryos. All embryos were oriented with the anterior to the left.

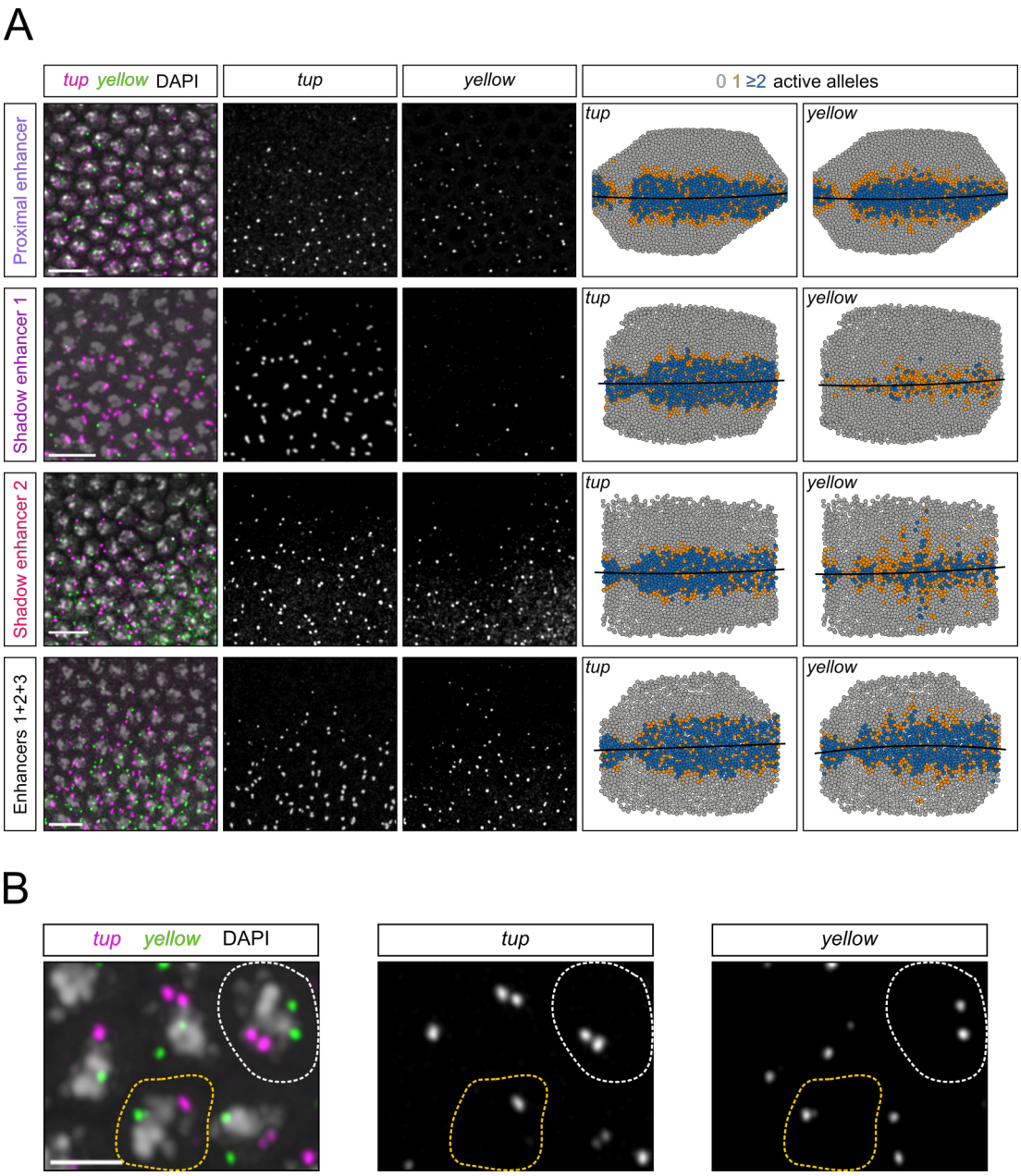



**Figure 5.26: Investigation of nuclear transcription states driven by *tup* enhancers.** (A) Exerts are shown from full embryos stained for *tup* (pink) and *yellow* (green) transcription. Merged images are shown in the left column with nuclei stained with DAPI (grey). Individual channels of *tup* and *yellow* probes are also shown. Representative images of the full embryos are shown in the 2 columns furthest to the right with nuclei false coloured according to their transcription state (Monoallelic = orange, biallelic = blue). Representative images left (*tup* probe) and right (*yellow* probe) correspond to data from the same embryo. (B) Enlarged section from triple enhancer line showing individual nuclei labelled with DAPI, tup (magenta) and yellow (green) probes. Nucleus with biallelic (white) and monoallelic (orange) expression of both genes highlighted.
All embryos were oriented with the anterior to the left. Scale bar = 10 μm (A) and 2 μm (B).

the middle of the expression domain, which span almost the complete dorsal width (Figure 5.26 A). Finally, the analysis of the *yellow* expression domain that was generated by the triple enhancer transgene appears similar to *tup* and reveals allelic activation that appears similar to the endogenous *tup* expression (Figure 5.26 A). However, the *yellow* transcription pattern appears to contain more spikes at the outer edges than *tup* expression, likely due to *shadow enhancer 2* (Figure 5.26 A). An enlarged image from the triple enhancer line shows both biallelic and monoallelic transcription for both *tup* and *yellow* (Figure 5.26 B). After initial observations gained from the expression domains, the proportion of monoallelic nuclei was quantified for each transgene.

The proportion of monoallelic nuclei was calculated based on the combined expression domain, which was defined to include only nuclei that showed both *tup* and *yellow* transcription. Consistent with initial observations, the *yellow* expression domain driven by the proximal enhancer shows high similarity to endogenous *tup*. No significant difference is found in the proportion of monoallelic nuclei (Figure 5.27 A). This suggests that the proximal enhancer contributes significantly to the *tup* expression domain and under permissive conditions is sufficient to drive an expression pattern almost equivalent to endogenous *tup*. Similarly to the identified *ush* enhancer, the proximal *tup* enhancer sequence contains both pMad and Zld binding motifs. Therefore, it can be suggested that the presence of Zld binding motifs could be indicative of stronger enhancer activity.

Analysis of *yellow* transcription driven by the shadow enhancer 1 sequence reveals a significant increase in the proportion of monoallelic nuclei compared to endogenous *tup* expression (Figure 5.27 B). At 80% monoallelic transcription, the expression domain driven by shadow enhancer 1 is not only more narrow (Figure 5.26 A), but also less able to activate biallelic transcription, suggesting that the enhancer sequence is less responsive.

Likewise the expression pattern driven by shadow enhancer 2 shows a smaller but still significant increase in monoallelic transcription at 60% (Figure 5.27 C). In its expression pattern it resembles the wt pattern more closely and together with the smaller increase in monoallelic transcription, it can be hypothesised that this enhancer sequence is more responsive in isolation than shadow enhancer 1.

Finally, the proportion of monoallelic nuclei in the joint expression domain is shown to be equal between the endogenous *tup* and the *yellow* domain generated

by the triple enhancer construct (Figure 5.27 D). This observation suggests that all three enhancer sequences together generate an expression domain similar to the endogenous *tup* expression.

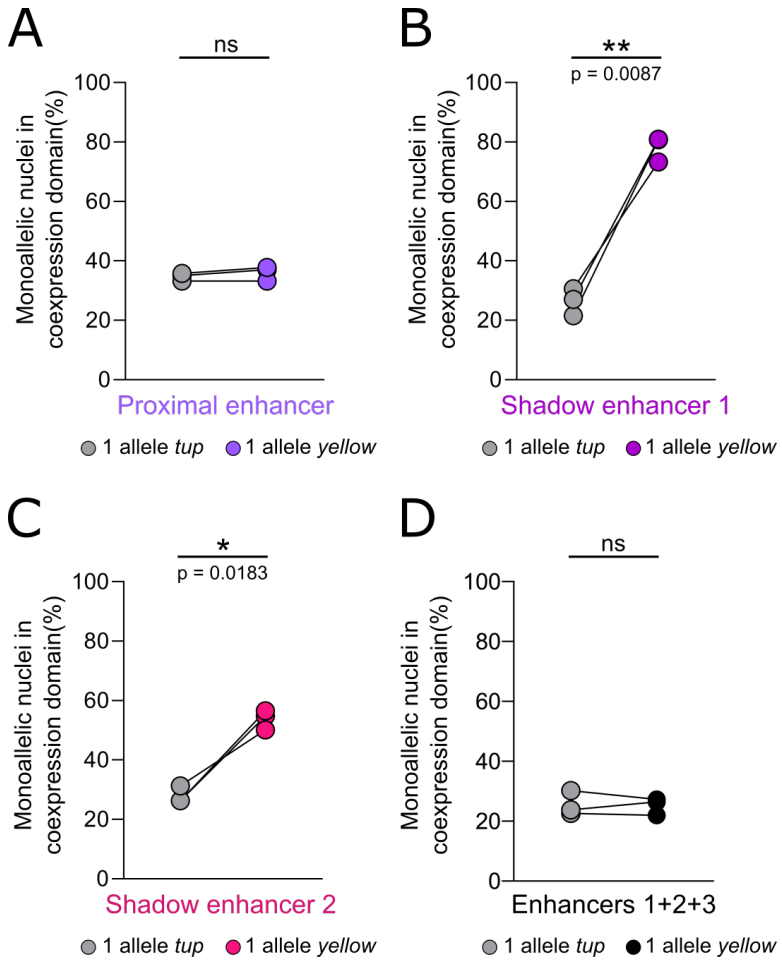



**Figure 5.27: Higher proportion of monoallelic nuclei observed during shadow enhancer driven transcription.** The proportion of monoallelic nuclei within the co-expression domains of *yellow* and *tup* were quantified. The graphs show the monoallelic proportion of *tup* (grey) and monoallelic *yellow* transcription driven by the proximal enhancer (A), shadow enhancer 1 (B), shadow enhancer 2 (C), and the triple enhancer line (D).
Significance tested with a paired Student's t-test. *$p < 0.05$; **$p < 0.01$; ns = not significant. Biological replicates $n = 3$.

## 5.4 Discussion

Experiments described in this chapter aimed to investigate the alterations in transcriptional output in response to changing the promoter and testing the activity of enhancers in isolation. The results presented in this chapter reveal the following:

1. The *hnt* promoter sequence promotes more synchronous *ush* transcription than that of *ush*.

2. Promoter swaps lead to increased embryo-to-embryo variability in transcription output.

3. Burst size but not promoter occupancy is controlled by promoters.

4. The promoter swaps produce hybrid burst profiles in between that of *ush* and *hnt* transcription.

5. A novel enhancer sequence is sufficient to drive an *ush* shaped transcription profile.

6. *tup* enhancer sequences show different strengths and levels of activity in isolation.

### 5.4.1 Changes in transcription kinetics through promoter swaps

#### Increased synchronicity in *ush* transcription

After the addition of the *hnt* promoter into the endogenous *ush* gene locus, increased synchronicity in transcription onset was detected in *hntPush* embryos. The time to synchrony describes how coordinated gene expression is within a field of cells (Lagha et al. 2013). Changing the gene promoter has been previously shown to change the level of gene expression synchronicity through changes in PPP. That the correct temporal expression of patterning genes is important during embryogenesis has been demonstrated by exchanging the highly paused (synchronous) *sna* promoter for a less paused promoter. The reduction in PPP led to stochastic gene activation, which was shown to cause progressive loss of mesoderm invagination during gastrulation (Lagha et al. 2013). Thereby, the authors showed that the temporal changes in the *sna* transcription profile are sufficient to cause developmental

defects and highlight the importance of correct gene expression timing in development, especially during short time windows of mitotic cycles.

The spatio-temporal expression dynamics of BMP target genes *hnt* and *ush* were examined in this chapter by changing their temporal expression patterns through promoter swaps. Experimental data from a GRO-seq study revealed that *hnt* is highly paused, whereas *ush* transcription is weakly paused (Saunders et al. 2013). Based on the results from the study discussed above, the *hnt* promoter with increased pausing is predicted to lead to an increase in *ush* transcription synchronicity. Consistent with this, there is a decrease in time to synchrony by 3 min in the *hntPush* promoter swap line. Hence, *hnt* promoter driven *ush* transcription initiates earlier and more synchronously within the core expression domain. To link this directly to PPP, it will be important to directly measure the degree of PPP in the promoter swap versus wt embryos.

This study did not identify developmental changes caused by the promoter swap as the number of amnioserosa cells was not altered in the *hntPush* promoter swap line. It is possible that the transcriptional network buffers the change in temporal *ush* expression or that the change does not affect the amnioserosa cell fate but instead other developmental mechanisms. *ush* expression is important for dorsal closure events and the communication between the amnioserosa and the dorsal epidermis tissue (Lada et al. 2012). Therefore, it is possible that changes in *ush* transcription caused by the promoter swap could lead to dorsal closure defects, which could be investigated in the future. Additionally, the robustness of *ush* transcription driven by the *hnt* promoter sequence could also be investigated under environmental stress conditions, such as an increase in temperature. Lastly, it is also possible that changing gene transcription to become more synchronous is tolerated better by the gene regulatory network than delaying gene activation as it was the case for *sna* (Lagha et al. 2013).

Interestingly, only a very small change in transcription synchronicity was observed in *ushPhnt* embryos. Based on previous findings, the *ush* promoter addition should reduce the degree of PPP and caused a more stochastic transcription onset. However, no change in onset time and only a slight decrease in time to synchrony was observed in *ushPhnt* embryos. This could be related to the fact that wt *hnt* transcription is already initiated very late during nc14 with an overall short transcription time window.

The exchange of promoters was shown to cause great variability in the transcriptional output based on mean and total fluorescence values for both *hntPush* and *ushPhnt* embryos. This variability in transcription levels could be explained by early signs of compromised transcriptional robustness. For reasons outside of out control, the imaging conditions for time-lapse experiments were variable in temperature, which caused slightly different development speeds (see methods for developmental time adjustments). Therefore, the variability in expression levels could be caused by a lack of transcriptional robustness due to elevated temperature. This hypothesis is possibly supported by cytoplasmic mRNA counts using smFISH. A reduced variability was observed for cytoplasmic mRNA counts in promoter swap lines when compared to the large differences in nascent transcription levels obtained during live imaging. Alternatively, the heterologous promoter might not be perfectly compatible with the endogenous enhancer and therefore produce more varying transcription output.

**Transcription bias in live imaging lines**

To investigate whether the introduction of MS2-loops into the *Drosophila* genome alters the transcription process, the contribution of alleles was compared using smFISH. Interestingly, this study identified under-transcription of the *ush-MS2* allele and over-transcription of the *hntPush-MS2* allele compared to the *ush* wt allele using smFISH probes. This result should be confirmed in the future using RT-qPCR. Furthermore, the contribution of alleles should be investigated in embryos carrying one MS2 tagged allele and one CRISPR control allele (containing the CRISPR scars but no MS2-loops) to determine if the bias is caused by the MS2-loop sequence. Similar analysis of nascent mRNA number in an *in vitro* study showed no bias between wt and MS2 tagged alleles (Sheinberger et al. 2017). That study, however, did not insert the MS2-loops into the gene's 5' or 3'UTR but created an ectopic exon carrying the MS2-loop tagged sequences.

Here, we chose to insert MS2-loops into the endogenous 5'UTR of *ush* and *hnt* to increase the fluorescent signal through long residence time of the fluorophores during transcription and thereby increasing the signal-to-noise ratio for signal detection (Ferraro et al. 2016). It is possible that the MS2 location in the 5'UTR leads to slightly preferential transcription of the wt allele and that this bias is overcome by inserting the stronger *hnt* promoter. Nascent mRNA production between one MS2 and one wt allele has not been analysed in *Drosophila* studies using MS2-loops in a similar location (or in the 3'UTR). It is likely, however, that similar biases

existed during other *Drosophila* MS2/MCP imaging studies, as the experimental setup in the field mostly analyses embryos with one MS2 tagged locus. Despite slightly lower output from the MS2 allele, the dynamics that were observed during live imaging are still likely to be representative of the wt allele with potentially a slight reduction in the loading rate of polymerases.

**Transcriptional burst kinetics in promoter swaps**

The previous chapter suggested that BMP target gene transcription integrates signalling levels through changes in burst size and promoter occupancy. In support of this, increased levels of BMP signalling through ectopic Dpp increased burst size and improved promoter occupancy in the intermediate and edge cluster during *ush* transcription. This chapter investigated how the gene's promoter influences burst parameters as the influence of the promoters on transcriptional burst kinetics has not been addressed directly *in vivo*.

The burst parameter that is changed most strongly by the promoter swap is burst amplitude, the rate with which Pol II is loaded (Zoller et al. 2018). As described in the previous chapter, *hnt* is transcribed in short, high amplitude bursts whereas *ush* bursts have lower amplitude but longer duration. The burst amplitude is increased in the *hntPush* promoter swap embryos and decreased in the *ushPhnt* embryo, suggesting that burst amplitude is regulated directly by the promoter. For *hntPush*, the duration is slightly shortened in the centre and intermediate regions, again changing the *ush* burst duration to one more similar to that observed for endogenous *hnt* transcription. However, this small effect on duration means that burst size for *hntPush* relative to *ush* increases, in line with the larger change in amplitude. Likewise, for *ushPhnt*, there is little change in duration in the centre and intermediate regions so that the reduced amplitude leads to a reduced burst size. Together these results suggest that the promoter primarily alters burst amplitude with only small effects on burst duration in the case of the *ush* promoter (Figure 5.28). In relation to amplitude, the rate of Pol II loading at the promoters can occur in waves where Pol II molecules are loaded onto the gene in discrete bursts, forming polymerase convoys (Tantale et al. 2016). Therefore, the results here suggest that the *hnt* promoter can load larger Pol II convoys than the *ush* promoter.

The finding that the promoter alters burst amplitude is consistent with previous data showing that burst amplitude is not modulated by the enhancer. Testing the effect of different enhancer sequences on burst kinetics while using the same pro-

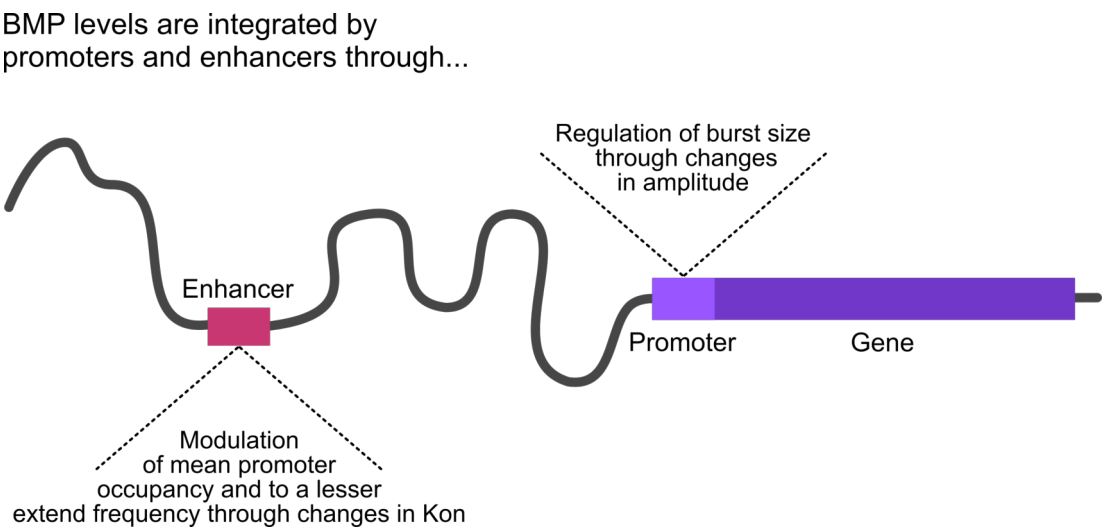



**Figure 5.28: Summary of the role of promoters and enhancers during BMP signalling level integration.** This study has highlighted that the major determinant of transcriptional output is mean promoter occupancy, the time that the promoter is active. Additionally, different levels of BMP signalling alter burst frequency, which together with promoter occupancy seems to be regulated through changes in the Kon rate. The introduction of a heterologous promoter changes the burst size but not the promoter occupancy suggesting that promoter occupancy is dictated by the enhancer and that the burst size is dependent on the promoter sequence.

moter identified changes in frequency but not in burst amplitude, suggesting that different enhancers do not modulate burst amplitude (Fukaya et al. 2016). A recent study investigating the expression of genes in primary mouse fibroblasts by single-cell RNA sequencing also suggests that promoter elements affect burst size (Larsson et al. 2019). Genes that contain TATA boxes or TATA boxes and Inr sequences were shown on average to produce larger burst sizes than genes containing only an Inr sequence or no promoter elements at all with the latter two showing no difference in burst sizes (Larsson et al. 2019). These data are difficult to reconcile with the data presented here. The *hnt* promoter contains a TATA box and *ush* contains an Inr sequence. Hence, *hnt* would be expected to show a larger burst size than *ush*. Even though the amplitude is larger, the burst size of *hnt* was identified here to be smaller than for *ush* and, therefore, does not adhere to the study mentioned above. Additionally, in this study changing the promoter also resulted in small effects on burst frequency. The reason for this is not currently understood. Further experiments with systematic addition/deletion of consensus promoter elements will be necessary to fully characterise the links between promoter features and effects on burst size and other kinetics.

Changing the promoter does not change the mean promoter occupancy in *hntPush* embryos and only small changes were detected in *ushPhnt* embryos compared to their wt counterparts. This observation suggests that the mean promoter occupancy is regulated by the enhancer irrespective of the promoter present (Figure 5.28). Moreover, the introduced promoters respond to different BMP signalling levels by changing occupancy in the same way as the endogenous promoter, as described in Chapter 4. Together these data strongly suggest that BMP signalling levels are decoded mainly by the enhancer, which alters promoter occupancy, regardless of the nature of the promoter present (Figure 5.28). A model outlining BMP signalling integration on a molecular level will be introduced in the General Discussion.

### 5.4.2 Enhancer architecture is important for transcription robustness

#### Identification of *ush* enhancer

Enhancer sequences control the spatial and temporal aspect of gene transcription and are often positioned in distal locations to the target gene promoter. Furthermore, many genes were identified to be regulated by multiple enhancers (Yokoshi & Fukaya 2019). Shadow enhancers seemed redundant initially under optimal conditions, but their importance for transcription robustness and buffering against environmental changes has been shown during development (Bothma et al. 2015, Frankel et al. 2010, Cannavò et al. 2016). However, in some cases enhancers can act non-redundantly even under optimal conditions as has been shown for *sna* (Dunipace et al. 2011).

Here, this study also identified the first *ush* enhancer sequence. This sequence contains both canonical pMad and Zld binding sites and ChIP-Seq data showed binding of both. Similarly, the strong *tup* proximal enhancer sequence also contains both pMad and Zld binding sites, which alone were shown to direct intermediate BMP expression (Deignan et al. 2016). Regions bound by Zld are associated with accessible chromatin and therefore act as a predictor for future TF binding (Li et al. 2008, MacArthur et al. 2009, Harrison et al. 2011). Initially, this study set out to test *ush* enhancers in isolation. However, as only one *ush* enhancer with early activity was identified, this study instead focused on the *tup* enhancers.

### *tup* enhancers in isolation

Three known *tup* enhancers were tested in isolation. Embryos homozygous for a transgene driven by the proximal *tup* enhancer showed wt *tup* expression patterns of yellow. This suggests that the enhancer alone is sufficiently strong to generate an expression domain almost equivalent to wt *tup* under permissive conditions. The two tested shadow enhancers produced expression domains in the general shape of wt *tup* when tested in isolation but showed a significant increase in the proportion of monoalellic expression. This suggests, that even though transcription is activated in the correct spatial location the transcription frequency or duration is vastly reduced, leading to increased observance of monoallelic nuclei. Therefore, it can be predicted that the number of mRNAs produced by these transgenes would be lower than the *tup* levels generated when all three enhancers are present. Unfortunately, the number of cytoplasmic mRNAs produced by the transgenes could not be quantified accurately as Yellow protein is secreted and therefore the RNA is targeted to the secretory pathway leading to clustering of mRNAs which prevents quantification of single transcripts (Kornezos & Chia 1992, Wittkopp et al. 2002).

Additionally, the shadow enhancer 1 and 2 transgene expression domains showed regions without active expression in the wt *tup* expression domain. These "holes" in shadow enhancer expression domains were identified mostly at the edges with reduced BMP signalling levels. In these regions it is likely that the compromised enhancer architecture is not able to initiate transgene expression in response to low levels of BMP. One possibility is that shadow enhancer 1 and 2 contain fewer or lower affinity binding sites for Mad/Med and other TFs. This hypothesis will need to be tested in the future.

Overall, it can be concluded that a wt enhancer architecture is important for BMP signalling interpretation at the *tup* locus and for other BMP target genes. Furthermore, this study finds that the proximal enhancer contributes most strongly to the wt *tup* expression pattern. Based on evidence from experiments testing *sna* shadown enhancer robustness (Dunipace et al. 2011), it would be interesting to test whether the proximal enhancer in isolation can recreate the *tup* wt expression domain and support development under stress conditions, such as an increase in temperature. Furthermore, shadow enhancers 1 and 2 would be predicted to accumulate even less mRNA when exposed to environmental stress conditions. In future experiments this theory will be investigated. If shadow enhancers are needed to buffer *tup* transcription, the triple enhancer line, generated during this study,

should produce a robust transcription profile even under stress conditions. The analysis of the triple enhancer construct will also reveal whether correct enhancer spacing is important under wt or adverse conditions. Furthermore, future experiments could investigate the behaviour of *tup* enhancers in isolation and in combination. Analysing the mRNA output of individual enhancer sequences will reveal if they work in an additive, super- or sub-additive fashion to control the wt *tup* expression pattern. Previous studies have highlighted the often complex behaviour and integration of multiple enhancers (Bothma et al. 2015, Scholes et al. 2019).

In summary, this chapter has investigated the changes in transcription kinetics that are controlled by promoter sequences and identified burst amplitude to be the main promoter responsive burst parameter. Additionally the role of enhancers toward the robustness of *tup* transcription was investigated. The increased proportion of monoallelic transcription in transgene expression driven by shadow enhancers 1 and 2 respectively suggests that a wt enhancer architecture is necessary for optimal burst kinetics.

# 6 | General Discussion

Developmental processes and decisions may occur in a matter of minutes to establish cell fates and set-up tissues during embryogenesis. Transcription events that underlie these developmental decisions have rarely been investigated *in vivo* and with a temporal resolution adequate to capture their dynamic nature. This study has investigated the spatio-temporal response of single cells to the BMP signalling gradient. Static image analysis gained high spatial resolution information on the transcription output of BMP target genes and was combined with live imaging of *hnt* and *ush* transcription. Live imaging and quantification of *hnt* and *ush* transcription profiles at 20 second resolution enabled the identification of transcriptional changes. These data were complemented with computational modelling to extract the changes in burst kinetics that are caused by different levels of BMP signalling. Therefore, this study is the first to identify how single cells read out BMP signalling levels and translate them into a graded mRNA output through modulation of burst kinetics.

Combining the data obtained in the three results chapters, this study proposes two models. The first proposes how changes in BMP signalling levels lead to changes in transcription on a molecular level and the second predicts how mRNA levels define cell fate decisions.

## 6.1 Response to changing Smad levels coordinated by transcription hubs

The role of developmental signalling pathways during embryogenesis has been studied extensively but how signals, such as BMP, are received and integrated by single cells into cell specific transcription kinetics is largely unknown. This study identified the transcription parameters that decode the BMP signalling gradient at single-cell resolution. Based on the results presented in this study, a molecular mechanism is proposed that explains BMP signal integration through transcription hub formation.

Using MS2/MCP live imaging this study has shown that BMP signalling levels are predominantly integrated through promoter occupancy (burst frequency times

duration) and modulation of burst amplitude. Nuclei positioned close to the dorsal midline experience peak levels of BMP signalling leading to high promoter occupancy. The changes in promoter occupancy were shown to mainly result from changes in the rate with which the promoter transitions into the ON state (Kon), which is most likely controlled by the enhancer. Enhancers have been shown previously to control changes in Kon (Fukaya et al. 2016, Lammers et al. 2019, Fritzsch et al. 2018) and further support was gathered in this study by showing that promoter occupancy is not regulated by the promoter sequence. Compatible with the observation of greater mean occupancy and increased burst amplitude in areas of high BMP signalling is a model where signalling activators form a larger/more stable transcription hub.

A growing collection of evidence supporting the formation of transcriptional hubs has emerged recently. These three-dimensional membrane-less nuclear compartments are proposed to form based on the chemical properties of participating molecules, including the presence of intrinsically disordered regions in proteins such as subunits of Mediator co-activator. Thereby, enhancers, promoters, TFs, co-regulators and Pol II form a dynamic microenvironment, which is sufficient for gene activation and increases the local concentration of TFs and the transcription machinery (Mir et al. 2018, Chen et al. 2018, Lim et al. 2018, Heist et al. 2019, Mir et al. 2019).

A key component for the formation of transcription hubs is the Mediator co-activator, which has been shown to co-localise with Pol II in phase-separated condensates (Fan et al. 2006, Cho et al. 2018, Sabari et al. 2018). Transcription hubs were shown in vitro to exhibit both stable long-term or transient short-term dynamics (Cho et al. 2018). FRAP analysis, however, showed that the turnover of Mediator protein and Pol II components was very fast (60% and 90% within 10 sec, respectively) even in more stable transcription hubs (Cho et al. 2018). In regions of high BMP signalling, we propose that the increased concentration of activated Smad molecules present leads to an increased probability of hub formation (Figure 6.1 A). Smads bind to enhancer sequences to initiate transcription by recruiting co-activators, including Mediator (Zhao et al. 2013). Therefore, with high levels of BMP signalling/ activated Smads, a high concentration of Mediator is recruited enhancing hub formation, which allows rapid recruitment of many Pol II molecules. As a result, the rate with which the promoter switches on is high. Pol II convoys, which contain multiple closely spaced polymerase molecules that transverse through the gene body as one unit, have been described, with the number of Pol II molecules in a convoy dependent on Mediator concentration (Tantale et al.

2016). In this model, the high concentration of Pol II would also allow loading of large Pol II convoys, leading to a high burst amplitude during transcription (Figure 6.1 A) (Spitz & Furlong 2012).

In areas with intermediate BMP signalling levels lower concentrations of activated Smad at the enhancer recruit less Mediator so that the proposed transcription hub would contain fewer molecules and therefore would take longer to form and potentially be more transient (Figure 6.1 B). As a result, the rate of promoter switching on is reduced. A lower concentration of Mediator recruited to the hub would lead to smaller polymerase convoys, as was observed by knockdown of a Mediator subunit in cells (Tantale et al. 2016), and therefore reduced burst amplitude (Figure 6.1 B).

In nuclei that are positioned in regions with low levels of BMP signalling, there is low promoter occupancy and small, infrequent bursts are observed. This could be explained by very transient hub formation due to low concentrations of Smad and Mediator leading to infrequent transcription initiation events and loading of only small Pol II convoys (Figure 6.1 C). The short duration of bursts observed in nuclei on the edge of the expression domain suggests that the hub may be unstable so that there is a high probability of the promoter switching off. It will be interesting to further test this model in the future, for example by using super-resolution microscopy to visualise Smads and Mediator in the nucleus, as well as testing the sensitivity of burst kinetics to reduction in Mediator subunit levels.

The data obtained in this study suggest Smad concentration-dependent regulation of bursting kinetics, primarily occupancy and amplitude. Accordingly, an increase in Smad concentration would be predicted to increase hub formation and Mediator recruitment. In support of this model, increased Smad concentration through ectopic *dpp* expression increased the burst amplitude during *ush* transcription in all expression clusters. This observation is consistent with a higher concentration of activated Smad recruiting more Mediator to the hub, resulting in loading of larger Pol II convoys. Higher levels of Dpp increased the mean promoter occupancy in the intermediate and edge clusters suggesting that the promoter occupancy is already saturated in the central domain. Perhaps, there is already full Smad occupancy at the *ush* enhancer in wt embryos in regions of peak signalling, so that hub formation is already occurring at maximal efficiency. In wildtype embryos,

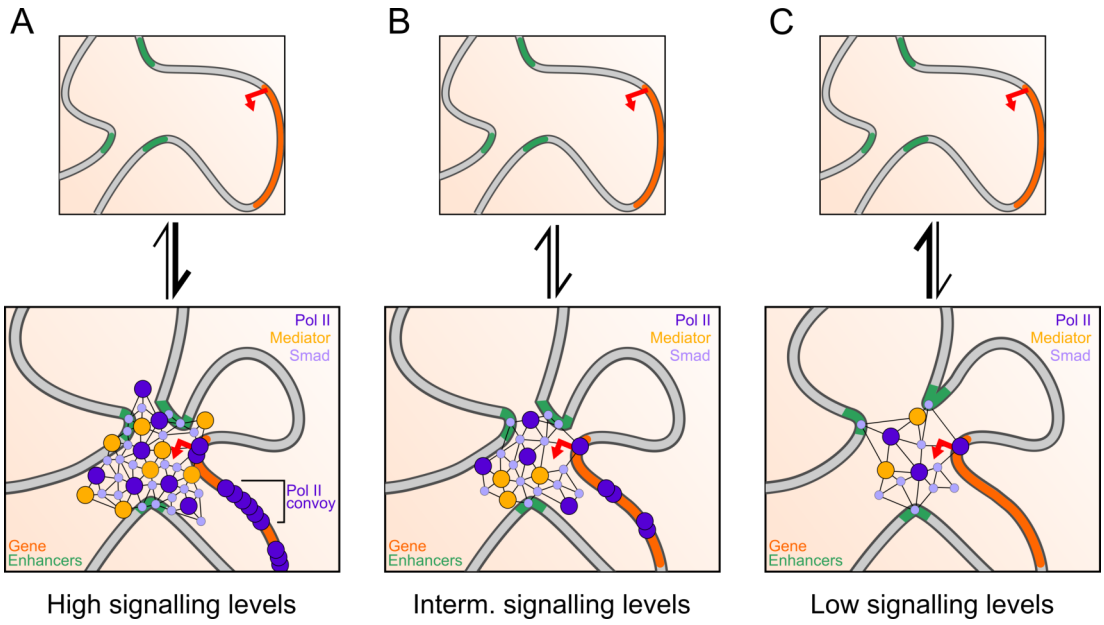



**Figure 6.1: The formation of transcription hubs could explain changes in burst kinetics in response to BMP signalling.** (A) In this model, high levels of activated Smad protein are present in nuclei that are positioned in the centre of the BMP signalling gradient, which leads to increased recruitment of Mediator complexes to the enhancer. High Mediator concentrations and interaction with Pol II increases Pol II convoys and burst amplitude. (B) Transcription hubs are formed more transiently with fewer molecules in nuclei that experience intermediate levels of BMP signalling. (C) In regions with low BMP signalling transcription hubs are unstable and infrequent resulting in reduced transcription initiation and low-amplitude bursts.

the mean promoter occupancy for *hnt* transcription was found to be lower than for *ush* in the central, intermediate and edge region. Hence, it would be interesting to test whether ectopic *dpp* expression increases *hnt* promoter occupancy even in the centre domain where there are peak BMP signalling levels. Although evidence is provided that the enhancer decodes BMP signalling levels to control promoter occupancy, it is currently unclear as to what features of the enhancer (other than Smad binding sites) dictate the degree of promoter occupancy.

Exchanging the promoter sequences between *ush* and *hnt* genes resulted in hybrid bursting profiles that contained features of both wildtype profiles. The promoter occupancy in promoter swap embryos remained at the wildtype level, suggesting that the promoter does not regulate occupancy. This is in agreement with the above hub model that proposes that occupancy is regulated, at least in part, by the concentration of Smad at the enhancer.

The promoter swap experiments showed that the burst parameter most influenced by the promoter is amplitude. The burst amplitude is approximately two-fold higher during *hnt* transcription compared to *ush* transcription. One possibility to explain the higher *hnt* amplitude is that the *hnt* promoter shows a high degree of Pol II pausing (Saunders et al. 2013). Pausing has been associated with displacement of the +1 nucleosome creating a nucleosome free region (Gilchrist et al. 2010). Perhaps the nucleosome free region allows for more efficient Pol II convoy loading at the promoter, which could result in a higher amplitude. This hypothesis could be investigated in the future by testing promoter elements in isolation and specifically the contribution of paused Pol II as part of a well-defined promoter.

## 6.2 Transcript numbers define cell fates

The gene regulatory network that defines DV axis patterning is highly complex, and the information from many genes is integrated to define cell fates in the dorsal ectoderm (Levine & Davidson 2005). This study has shown a 10-fold difference in the mRNA number across the expression domain for the three BMP target genes *hnt*, *ush* and *tup*. Based on the differences in mRNA number, we propose that at the single-cell level a minimum number of mRNA transcripts needs to be produced to ensure sufficient amount of protein present and the correct cell fate to occur (Figure 6.2 A). Under optimal conditions, cells are predicted to produce much greater levels of mRNA than are needed to cross the threshold of minimal mRNA number (left, Figure 6.2 A). By producing more mRNAs than required, these cells can buffer fluctuations in transcription and still ensure the correct cell fate decisions.

This study has collected four pieces of evidence that together suggest a threshold model underlying cell fate decisions. Firstly, analysis of BMP target genes revealed that monoallelic expression is associated with reduced transcript numbers, as a result of poor burst kinetics in these cells. Secondly, this study has shown that the transcriptional response can be challenged by disrupting the enhancer architecture. Testing enhancers in isolation showed that two of the *tup* shadow enhancers could not support biallelic transcription well, but instead there is significant monoallelic transcription. Based on the findings of mRNA number in monoallelic cells it is likely that the monoallelic transcription elicited by enhancers in isolation (or when the enhancer architecture is disrupted) also leads to a reduced number of mRNAs (middle column, Figure 6.2 A).

Thirdly, this study has revealed that monoallelic transcription at the edge of an expression domain is also observed for target genes of the Dl gradient in the early *Drosophila* embryo. Therefore, monoallelic transcription arising from poor bursting kinetics appears to be a general feature of morphogen gradient interpretation.

Finally, data is presented that genetic perturbations can modulate the number of transcripts that are produced. Embryos that contain three copies of *dpp* showed an increase in the numbers of *ush* transcripts that were produced as well as an increase in expression domain width. Conversely, a reduction in BMP signalling, here investigated in embryos with one copy of *dpp* or by using $Med^{13}$ maternal heterozygous embryos, reduced the number of *ush* mRNAs produced by cells (Figure 6.2 Bi).

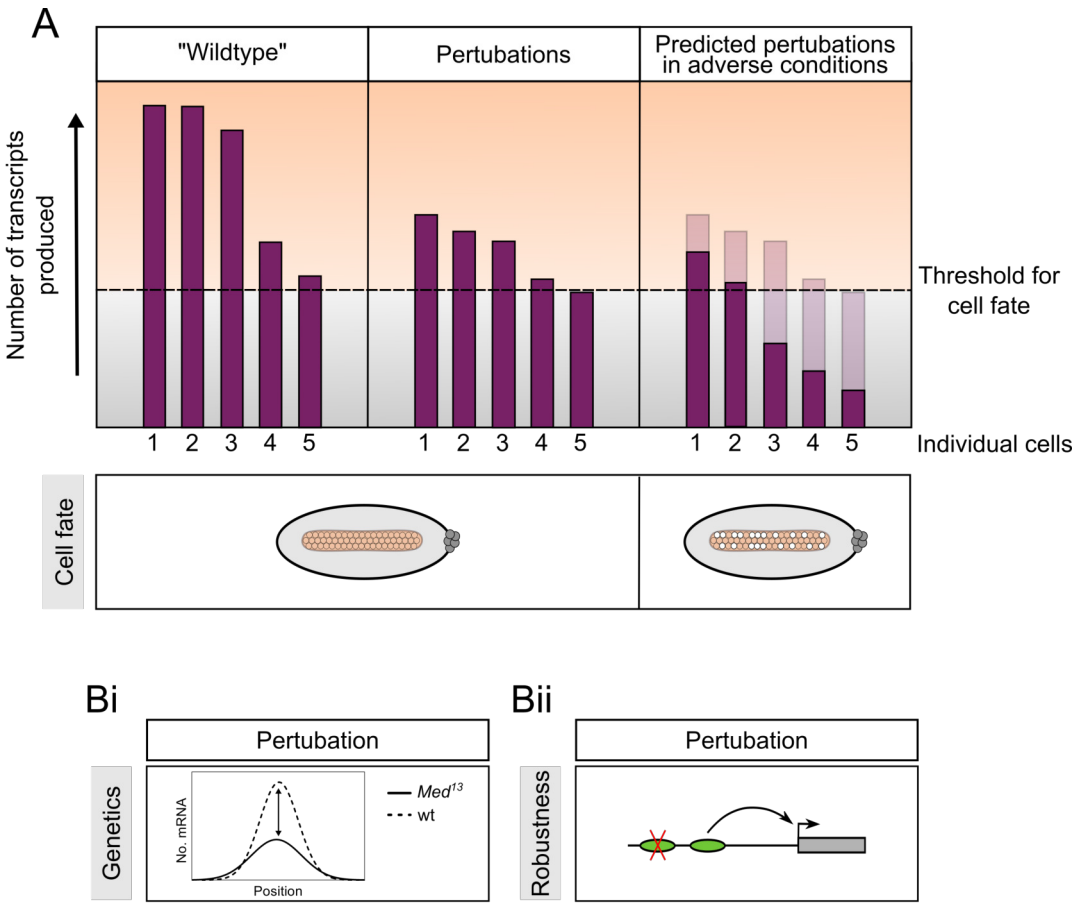



**Figure 6.2: Transcript numbers define cell fate.** (A) Schematic of a threshold model where a minimum number of mRNA transcripts is needed to specify the correct cell fate. Under optimal conditions cells produce more than the required minimum number of transcripts. Perturbations in the system lead to lower numbers of produced transcripts which still result in the correct cell fate. However, under adverse conditions transcription cannot be buffered in embryos with perturbations and mRNA numbers drop below the threshold resulting in incorrect cell fate. (Bi) Genetic perturbations such as reduced Smad concentrations in maternal heterozygous *Med*$^{13}$ embryos reduces the mRNA transcript number per cell. (Bii) Similarly, reducing the robustness of gene expression by deleting an enhancer is predicted to result in lower mRNA numbers and than fall below the threshold in adverse conditions.

Based on the above observations, the mRNA threshold model for cell fate provides an explanation for the lack of robustness identified in other studies when shadow enhancers for key developmental genes were removed and development was challenged under adverse conditions. For example, a loss of robustness has been observed previously in response to deletion of a *sna* enhancer causing defects in mesoderm development (Bothma et al. 2015). Under adverse conditions, such as increased temperature or heterozygosity for a key activator, the model predicts that the mRNA number produced per cell, already reduced by removal of the enhancer or reduced activator (middle column, Figure 6.2 A), is decreased even further (right column, Figure 6.2 A). Therefore, insufficient transcripts are produced by the cell to specify the correct cell fate.

Consistent with this threshold model, we show that the number of BMP target gene mRNAs in embryos with 3x *dpp* is increased, and these embryos adopt extra amnioserosa cells (Wharton et al. 1993). Likewise, the large reduction in *ush* transcript number and likely that of other BMP responsive genes by reducing the *dpp* dose would result in the mRNA number in cells dropping below the threshold, resulting in fewer cells adopting the correct cell fate (here amnioserosa or leading edge fate, right column, Figure 6.2 A). Similarly, previous experiments have shown that embryos with maternal heterozygosity for $Med^{13}$ show a reduced mean and increased variability in amnioserosa number (Gavin-Smyth et al. 2013), whereas here it is shown that these embryos have less *ush* mRNAs per cell, as would be expected for other BMP target gene transcripts.

It would be interesting to test the threshold model by testing the link between robustness and mRNA number directly. One way would be to delete the *tup* proximal enhancer (leaving only the shadow enhancers, Figure 6.2 Bii) and quantitate mRNA number and cell fate during optimal and averse conditions. In addition to amnioserosa cell number for manipulation of peak BMP target gene loci, for intermediate genes such as *tup*, specification of leading edge cells could be quantified. Moreover, an extension to the threshold model is that some cells in the expression domain under optimal conditions fall below the threshold for cell fate and follow a different developmental trajectory to that of other cells in the expression domain. For example, nuclei at the edges of the *ush* expression domain were shown to produce small numbers of mRNA per cell. This raises the possibility that not all expressing cells will contribute to the specification of the leading edge as some would lack sufficient Ush protein. Overall, it will be interesting in the future to fully probe the link between mRNA number and cell fate decisions.

# References


Abrams, J., White, K., Fessler, L. & Steller, H. (1993), 'Programmed cell death during Drosophila embryogenesis', *Development* **117**(1).

Amir-Zilberstein, L., Ainbinder, E., Toube, L., Yamaguchi, Y., Handa, H. & Dikstein, R. (2007), 'Differential regulation of NF-kappaB by elongation factors is determined by core promoter type.', *Molecular and Cellular Biology* **27**(14), 5246–59.

Arora, K., Levine, M. S. & O'Connor, M. B. (1994), 'The screw gene encodes a ubiquitously expressed member of the TGF-beta family required for specification of dorsal cell fates in the Drosophila embryo.', *Genes & development* **8**(21), 2588–601.

Arora, K. & Nüsslein-Volhard, C. (1992), 'Altered mitotic domains reveal fate map changes in Drosophila embryos mutant for zygotic dorsoventral patterning genes.', *Development* **114**(4), 1003–24.

Ashe, H. L. (2005), 'BMP Signalling: Synergy and Feedback Create a Step Gradient', *Current Biology* **15**(10), R375–R377.

Ashe, H. L. & Briscoe, J. (2006), 'The interpretation of morphogen gradients.', *Development* **133**(3), 385–94.

Ashe, H. L. & Levine, M. (1999), 'Local inhibition and long-range enhancement of Dpp signal transduction by Sog.', *Nature* **398**(6726), 427–431.

Ashe, H., Mannervik, M. & Levine, M. (2000), 'Dpp signaling thresholds in the dorsal ectoderm of the Drosophila embryo', *Development* **127**(15), 3305–3312.

Baena-Lopez, L. A., Alexandre, C., Mitchell, A., Pasakarnis, L. & Vincent, J.-P. (2013), 'Accelerated homologous recombination and subsequent genome modification in Drosophila.', *Development* **140**(23), 4818–25.

Bahar Halpern, K. & Itzkovitz, S. (2016), 'Single molecule approaches for quantifying transcription and degradation rates in intact mammalian tissues', *Methods* **98**, 134–142.

Bar-Even, A., Paulsson, J., Maheshri, N., Carmi, M., O'Shea, E., Pilpel, Y. & Barkai, N. (2006), 'Noise in protein expression scales with natural protein abundance', *Nature Genetics* **38**(6), 636–643.

Bartman, C., Hsu, S., Hsiung, C.-S., Raj, A. & Blobel, G. (2016), 'Enhancer Regulation of Transcriptional Bursting Parameters Revealed by Forced Chromatin Looping', *Molecular Cell* **62**(2), 237–247.

Bashirullah, A., Cooperstock, R. L. & Lipshitz, H. D. (2001), 'Spatial and temporal control of RNA stability.', *Proceedings of the National Academy of Sciences of the United States of America* **98**(13), 7025–7028.

Bashirullah, A., Halsell, S. R., Cooperstock, R. L., Kloc, M., Karaiskakis, A., Fisher, W. W., Fu, W., Hamilton, J. K., Etkin, L. D. & Lipshitz, H. D. (1999), 'Joint action of two RNA degradation pathways controls the timing of maternal transcript elimination at the midblastula transition in Drosophila melanogaster', *The EMBO Journal* **18**(9), 2610–2620.

Bertrand, E., Chartrand, P., Schaefer, M., Shenoy, S. M., Singer, R. H. & Long, R. M. (1998), 'Localization of ASH1 mRNA Particles in Living Yeast', *Molecular Cell* **2**(4), 437–445.

Biehs, B., Francois, V. & Bier, E. (1996), 'The Drosophila short gastrulation gene prevents Dpp from autoactivating and suppressing neurogenesis in the neuroectoderm.', *Genes & Development* **10**(22), 2922–2934.

Bintu, B., Mateo, L. J., Su, J.-H., Sinnott-Armstrong, N. A., Parker, M., Kinrot, S., Yamaya, K., Boettiger, A. N. & Zhuang, X. (2018), 'Super-resolution chromatin tracing reveals domains and cooperative interactions in single cells.', *Science* **362**(6413), eaau1783.

Bjornsson, H. T., Albert, T. J., Ladd-Acosta, C. M., Green, R. D., Rongione, M. A., Middle, C. M., Irizarry, R. A., Broman, K. W. & Feinberg, A. P. (2008), 'SNP-specific array-based allele-specific expression analysis.', *Genome Research* **18**(5), 771–9.

Boettiger, A. & Levine, M. (2013), 'Rapid Transcription Fosters Coordinate snail Expression in the Drosophila Embryo', *Cell Reports* **3**(1), 8–15.

Boettiger, A. N. & Levine, M. (2009), 'Synchronous and stochastic patterns of gene activation in the Drosophila embryo.', *Science* **325**(5939), 471–3.

Boettiger, A. N., Ralph, P. L. & Evans, S. N. (2011), 'Transcriptional regulation: effects of promoter proximal pausing on speed, synchrony and reliability.', *PLoS Computational Biology* **7**(5), e1001136.

Bonev, B., Stanley, P. & Papalopulu, N. (2012), 'MicroRNA-9 Modulates Hes1 Ultradian Oscillations by Forming a Double-Negative Feedback Loop', *Cell Reports* **2**(1), 10–18.

Bothma, J. P., Garcia, H. G., Esposito, E., Schlissel, G., Gregor, T. & Levine, M. (2014), 'Dynamic regulation of eve stripe 2 expression reveals transcriptional bursts in living Drosophila embryos.', *Proceedings of the National Academy of Sciences of the United States of America* **111**(29), 10598–603.

Bothma, J. P., Garcia, H. G., Ng, S., Perry, M. W., Gregor, T. & Levine, M. (2015), 'Enhancer additivity and non-additivity are determined by enhancer strength in the Drosophila embryo.', *eLife* **4**.

Bothma, J. P., Norstad, M. R., Alamos, S. & Garcia, H. G. (2018), 'LlamaTags: A Versatile Tool to Image Transcription Factor Dynamics in Live Embryos.', *Cell* **173**(7), 1810–1822.e16.

Bownes, M. (1975), 'A photographic study of development in the living embryo of Drosophila melanogaster', *Development* **33**(3).

Bragdon, B., Moseychuk, O., Saldanha, S., King, D., Julian, J. & Nohe, A. (2011), 'Bone Morphogenetic Proteins: A critical review', *Cellular Signalling* **23**(4), 609–620.

Breathnach, R. & Chambon, P. (1981), 'Organization and Expression of Eucaryotic Split Genes Coding for Proteins', *Annual Review of Biochemistry* **50**(1), 349–383.

Brown, J. D., Lin, C. Y., Duan, Q., Griffin, G., Federation, A., Paranal, R. M., Bair, S., Newton, G., Lichtman, A., Kung, A., Yang, T., Wang, H., Luscinskas, F. W., Croce, K., Bradner, J. E. & Plutzky, J. (2014), 'NF-κB directs dynamic super enhancer formation in inflammation and atherogenesis.', *Molecular Cell* **56**(2), 219–231.

Buckley, M. S., Kwak, H., Zipfel, W. R. & Lis, J. T. (2014), 'Kinetics of promoter Pol II on Hsp70 reveal stable pausing and key insights into its regulation', *Genes & Development* **28**(1), 14–19.

Bulger, M. & Groudine, M. (1999), 'Looping versus linking: toward a model for long-distance gene activation', *Genes & Development* **13**(19), 2465–2477.

Butler, J. E. F. & Kadonaga, J. T. (2002), 'The RNA polymerase II core promoter: a key component in the regulation of gene expression.', *Genes & Development* **16**(20), 2583–92.

Campbell, P. D., Chao, J. A., Singer, R. H. & Marlow, F. L. (2015), 'Dynamic visualization of transcription and RNA subcellular localization in zebrafish.', *Development* **142**(7), 1368–74.

Cannavò, E., Khoueiry, P., Garfield, D. A., Geeleher, P., Zichner, T., Gustafson, E. H., Ciglar, L., Korbel, J. O. & Furlong, E. E. M. (2016), 'Shadow Enhancers Are Pervasive Features of Developmental Regulatory Networks.', *Current biology* **26**(1), 38–51.

Carey, J. & Uhlenbeck, O. C. (1983), 'Kinetic and thermodynamic characterization of the R17 coat protein-ribonucleic acid interaction', *Biochemistry* **22**(11), 2610–2615.

Carter, D., Chakalova, L., Osborne, C. S., Dai, Y.-f. & Fraser, P. (2002), 'Long-range chromatin regulatory interactions in vivo', *Nature Genetics* **32**(4), 623–626.

Chang, S. C., Hoang, B., Thomas, J. T., Vukicevic, S., Luyten, F. P., Ryba, N. J., Kozak, C. A., Reddi, A. H. & Moos, M. (1994), 'Cartilage-derived morphogenetic proteins. New members of the transforming growth factor-beta superfamily predominantly expressed in long bones during human embryonic development.', *The Journal of biological chemistry* **269**(45), 28227–34.

Chen, F., Tillberg, P. W. & Boyden, E. S. (2015), 'Expansion microscopy', *Science* **347**(6221), 543–548.

Chen, H., Levo, M., Barinov, L., Fujioka, M., Jaynes, J. B. & Gregor, T. (2018), 'Dynamic interplay between enhancer-promoter topology and gene activity.', *Nature Genetics* **50**(9), 1296–1303.

Cheng, B., Li, T., Rahl, P. B., Adamson, T. E., Loudas, N. B., Guo, J., Varzavand, K., Cooper, J. J., Hu, X., Gnatt, A., Young, R. A. & Price, D. H. (2012), 'Functional Association of Gdown1 with RNA Polymerase II Poised on Human Genes', *Molecular Cell* **45**(1), 38–50.

Chess, A., Simon, I., Cedar, H. & Axel, R. (1994), 'Allelic inactivation regulates olfactory receptor gene expression.', *Cell* **78**(5), 823–34.

Cho, W.-K., Spille, J.-H., Hecht, M., Lee, C., Li, C., Grube, V. & Cisse, I. I. (2018), 'Mediator and RNA polymerase II clusters associate in transcription-dependent condensates', *Science* **361**(6400), 412–415.

Chubb, J. R., Trcek, T., Shenoy, S. M. & Singer, R. H. (2006), 'Transcriptional Pulsing of a Developmental Gene', *Current Biology* **16**(10), 1018–1025.

Core, L. J., Waterfall, J. J. & Lis, J. T. (2008), 'Nascent RNA Sequencing Reveals Widespread Pausing and Divergent Initiation at Human Promoters', *Science* **322**(5909), 1845–1848.

Corrigan, A. M. & Chubb, J. R. (2014), 'Regulation of transcriptional bursting by a naturally oscillating signal.', *Current Biology* **24**(2), 205–11.

Cubadda, Y., Heitzler, P., Ray, R. P., Bourouis, M., Ramain, P., Gelbart, W., Simpson, P. & Haenlin, M. (1997), 'u-shaped encodes a zinc finger protein that regulates the proneural genes achaete and scute during the formation of bristles in Drosophila.', *Genes & Development* **11**(22), 3083–95.

Dar, R. D., Razooky, B. S., Singh, A., Trimeloni, T. V., McCollum, J. M., Cox, C. D., Simpson, M. L. & Weinberger, L. S. (2012), 'Transcriptional burst frequency and burst size are equally modulated across the human genome.', *Proceedings of the National Academy of Sciences of the United States of America* **109**(43), 17454–9.

Davis, I. & Ish-Horowicz, D. (1991), 'Apical localization of pair-rule transcripts requires 3âĂš sequences and limits protein diffusion in the Drosophila blastoderm embryo', *Cell* **67**(5), 927–940.

De Celis, J. F. (1997), 'Expression and function of decapentaplegic and thick veins during the differentiation of the veins in the Drosophila wing.', *Development* **124**(5), 1007–18.

De Navascués, J. & Modolell, J. (2007), 'tailup, a LIM-HD gene, and Iro-C cooperate in Drosophila dorsal mesothorax specification.', *Development* **134**(9), 1779–88.

De Renzis, S., Elemento, O., Tavazoie, S. & Wieschaus, E. F. (2007), 'Unmasking Activation of the Zygotic Genome Using Chromosomal Deletions in the Drosophila Embryo', *PLoS Biology* **5**(5), e117.

Deignan, L., Pinheiro, M. T., Sutcliffe, C., Saunders, A., Wilcockson, S. G., Zeef, L. A. H., Donaldson, I. J. & Ashe, H. L. (2016), 'Regulation of the BMP Signaling-Responsive Transcriptional Network in the Drosophila Embryo', *PLOS Genetics* **12**(7), e1006164.

Deng, Q., Ramsköld, D., Reinius, B. & Sandberg, R. (2014), 'Single-Cell RNA-Seq Reveals Dynamic, Random Monoallelic Gene Expression in Mammalian Cells', *Science* **343**(6167), 193–6.

Dirks, R. M. & Pierce, N. A. (2004), 'From The Cover: Triggered amplification by hybridization chain reaction', *Proceedings of the National Academy of Sciences* **101**(43), 15275–15278.

Doctor, J. S., Jackson, P., Rashka, K. E., Visalli, M. & Hoffman, F. (1992), 'Sequence, biochemical characterization, and developmental expression of a new member of the TGF-$\beta$ superfamily in Drosophila melanogaster', *Developmental Biology* **151**(2), 491–505.

Doma, M. K. & Parker, R. (2006), 'Endonucleolytic cleavage of eukaryotic mRNAs with stalls in translation elongation', *Nature* **440**(7083), 561–564.

Dorfman, R. & Shilo, B. (2001), 'Biphasic activation of the BMP pathway patterns the Drosophila embryonic dorsal region', *Development* **128**(6), 965–972.

Driever, W. & Nüsslein-Volhard, C. (1988), 'A gradient of bicoid protein in Drosophila embryos', *Cell* **54**(1), 83–93.

Driever, W., Thoma, G. & Nüsslein-Volhard, C. (1989), 'Determination of spatial domains of zygotic gene expression in the Drosophila embryo by the affinity of binding sites for the bicoid morphogen', *Nature* **340**(6232), 363–367.

Dufourt, J., Trullo, A., Hunter, J., Fernandez, C., Lazaro, J., Dejean, M., Morales, L., Nait-Amer, S., Schulz, K. N., Harrison, M. M., Favard, C., Radulescu, O. & Lagha, M. (2018), 'Temporal control of gene expression by the pioneer factor Zelda through transient interactions in hubs', *Nature Communications* **9**(1), 5194.

Dunipace, L., Ozdemir, A., Stathopoulos, A., Dennefeld, C. & Haenlin, M. (2011), 'Complex interactions between cis-regulatory modules in native conformation are critical for Drosophila snail expression.', *Development* **138**(18), 4075–84.

Dvir, A. (2002), 'Promoter escape by RNA polymerase II.', *Biochimica et biophysica acta* **1577**(2), 208–223.

Eckersley-Maslin, M. A. & Spector, D. L. (2014), 'Random monoallelic expression: regulating gene expression one allele at a time.', *Trends in Genetics* **30**(6), 237–44.

Edgar, B. A. & O'Farrell, P. H. (1989), 'Genetic control of cell division patterns in the Drosophila embryo', *Cell* **57**(1), 177–187.

Edgar, B. A. & O'Farrell, P. H. (1990), 'The three postblastoderm cell cycles of Drosophila embryogenesis are regulated in G2 by string.', *Cell* **62**(3), 469–80.

Edgar, B. A. & Schubiger, G. (1986), 'Parameters controlling transcriptional activation during early drosophila development', *Cell* **44**(6), 871–877.

Edgar, B. A., Sprenger, F., Duronio, R. J., Leopold, P. & O'Farrell, P. H. (1994), 'Distinct molecular mechanism regulate cell cycle timing at successive stages of Drosophila embryogenesis.', *Genes & development* **8**(4), 440–52.

Elowitz, M. B., Levine, A. J., Siggia, E. D. & Swain, P. S. (2002), 'Stochastic Gene Expression in a Single Cell', *Science* **297**(5584).

Erwin, J. A. & Lee, J. T. (2008), 'New twists in X-chromosome inactivation.', *Current Opinion in Cell Biology* **20**(3), 349–55.

Esposito, E., Lim, B., Guessous, G., Falahati, H. & Levine, M. (2016), 'Mitosis-associated repression in development.', *Genes & Development* **30**(13), 1503–8.

Falo-Sanjuan, J., Lammers, N. C., Garcia, H. G. & Bray, S. (2019), 'Enhancer priming enables fast and sustained transcriptional responses to Notch signaling', *bioRxiv* p. 497651.

Fan, X., Chou, D. M. & Struhl, K. (2006), 'Activator-specific recruitment of Mediator in vivo', *Nature Structural & Molecular Biology* **13**(2), 117–120.

Femino, A. M., Fay, F. S., Fogarty, K. & Singer, R. H. (1998), 'Visualization of single RNA transcripts in situ.', *Science* **280**(5363), 585–90.

Ferguson, E. L. & Anderson, K. V. (1992), 'decapentaplegic acts as a morphogen to organize dorsal-ventral pattern in the Drosophila embryo', *Cell* **71**(3), 451–461.

Ferraro, T., Esposito, E., Mancini, L., Ng, S., Lucas, T., Coppey, M., Dostatni, N., Walczak, A. M., Levine, M. & Lagha, M. (2016), 'Transcriptional Memory in the Drosophila Embryo.', *Current Biology* **26**(2), 212–218.

FitzGerald, P. C., Sturgill, D., Shyakhtenko, A., Oliver, B. & Vinson, C. (2006), 'Comparative genomics of Drosophila and human core promoters.', *Genome Biology* **7**(7), R53.

Flanagan, P. M., Kelleher, R. J., Sayre, M. H., Tschochner, H. & Kornberg, R. D. (1991), 'A mediator required for activation of RNA polymerase II transcription in vitro', *Nature* **350**(6317), 436–438.

Foe, V. E. (1989), 'Mitotic domains reveal early commitment of cells in Drosophila embryos', *Development* **107**(1).

Foe, V. E. & Alberts, B. M. (1983), 'Studies of nuclear and cytoplasmic behaviour during the five mitotic cycles that precede gastrulation in Drosophila embryogenesis', *Journal of Cell Science* **61**(1).

Forrest, K. M. & Gavis, E. R. (2003), 'Live imaging of endogenous RNA reveals a diffusion and entrapment mechanism for nanos mRNA localization in Drosophila.', *Current Biology* **13**(14), 1159–68.

Fossett, N., Zhang, Q., Gajewski, K., Choi, C. Y., Kim, Y. & Schulz, R. A. (2000), 'The multitype zinc-finger protein U-shaped functions in heart cell specification in the Drosophila embryo.', *Proceedings of the National Academy of Sciences of the United States of America* **97**(13), 7348–53.

Francois, V., Solloway, M., O'Neill, J. W., Emery, J. & Bier, E. (1994), 'Dorsal-ventral patterning of the Drosophila embryo depends on a putative negative growth factor encoded by the short gastrulation gene.', *Genes & Development* **8**(21), 2602–2616.

Frank, L. H. & Rushlow, C. (1996), 'A group of genes required for maintenance of the amnioserosa tissue in Drosophila.', *Development* **122**(5), 1343–52.

Frankel, N., Davis, G. K., Vargas, D., Wang, S., Payre, F. & Stern, D. L. (2010), 'Phenotypic robustness conferred by apparently redundant transcriptional enhancers', *Nature* **466**(7305), 490–493.

Fritzsch, C., Baumgärtner, S., Kuban, M., Steinshorn, D., Reid, G. & Legewie, S. (2018), 'Estrogen-dependent control and cell-to-cell variability of transcriptional bursting.', *Molecular systems biology* **14**(2), e7678.

Fromental-Ramain, C., Vanolst, L., Delaporte, C. & Ramain, P. (2008), 'pannier encodes two structurally related isoforms that are differentially expressed during Drosophila development and display distinct functions during thorax patterning', *Mechanisms of Development* **125**(1), 43–57.

Fujita, K., Iwaki, M. & Yanagida, T. (2016), 'Transcriptional bursting is intrinsically caused by interplay between RNA polymerases on DNA.', *Nature Communications* **7**(1), 13788.

Fukaya, T., Lim, B. & Levine, M. (2016), 'Enhancer Control of Transcriptional Bursting', *Cell* **166**(2), 358–368.

Fukaya, T., Lim, B. & Levine, M. (2017), 'Rapid Rates of Pol II Elongation in the Drosophila Embryo.', *Current Biology* **27**(9), 1387–1391.

Fusco, D., Accornero, N., Lavoie, B., Shenoy, S. M., Blanchard, J. M., Singer, R. H. & Bertrand, E. (2003), 'Single mRNA molecules demonstrate probabilistic movement in living mammalian cells.', *Current Biology* **13**(2), 161–7.

Gaarenstroom, T. & Hill, C. S. (2014), 'TGF-$\beta$ signaling to chromatin: How Smads regulate transcription during self-renewal and differentiation', *Seminars in Cell & Developmental Biology* **32**, 107–118.

Gaertner, B., Johnston, J., Chen, K., Wallaschek, N., Paulson, A., Garruss, A., Gaudenz, K., DeÂăKumar, B., Krumlauf, R. & Zeitlinger, J. (2012), 'Poised RNA Polymerase II Changes over Developmental Time and Prepares Genes for Future Expression', *Cell Reports* **2**(6), 1670–1683.

Gao, S., Steffen, J. & Laughon, A. (2005), 'Dpp-responsive Silencers Are Bound by a Trimeric Mad-Medea Complex', *Journal of Biological Chemistry* **280**(43), 36158–36164.

García-García, M. J., Ramain, P., Simpson, P. & Modolell, J. (1999), 'Different contributions of pannier and wingless to the patterning of the dorsal mesothorax of Drosophila.', *Development* **126**(16), 3523–32.

Garcia, H. G., Tikhonov, M., Lin, A. & Gregor, T. (2013), 'Quantitative imaging of transcription in living Drosophila embryos links polymerase activity to patterning.', *Current Biology* **23**(21), 2140–5.

Garcia, J. F. & Parker, R. (2015), 'MS2 coat proteins bound to yeast mRNAs block 5' to 3' degradation and trap mRNA decay products: implications for the localization of mRNAs by MS2-MCP system.', *RNA* **21**(8), 1393–5.

Gavin-Smyth, J., Wang, Y.-C., Butler, I. & Ferguson, E. L. L. (2013), 'A Genetic Network Conferring Canalization to a Bistable Patterning System in Drosophila', *Current Biology* **23**(22), 2296–2302.

Ge, G. & Greenspan, D. S. (2006), 'BMP1 controls TGFbeta1 activation via cleavage of latent TGFbeta-binding protein.', *The Journal of Cell Biology* **175**(1), 111–20.

Geisler, R., Bergmann, A., Hiromi, Y. & Nüsslein-Volhard, C. (1992), 'cactus, a gene involved in dorsoventral pattern formation of Drosophila, is related to the I kappa B gene family of vertebrates.', *Cell* **71**(4), 613–21.

Geiss, G. K., Bumgarner, R. E., Birditt, B., Dahl, T., Dowidar, N., Dunaway, D. L., Fell, H. P., Ferree, S., George, R. D., Grogan, T., James, J. J., Maysuria, M., Mitton,

J. D., Oliveri, P., Osborn, J. L., Peng, T., Ratcliffe, A. L., Webster, P. J., Davidson, E. H., Hood, L. & Dimitrov, K. (2008), 'Direct multiplexed measurement of gene expression with color-coded probe pairs', *Nature Biotechnology* **26**(3), 317–325.

Gendrel, A.-V., Attia, M., Chen, C.-J., Diabangouaya, P., Servant, N., Barillot, E. & Heard, E. (2014), 'Developmental dynamics and disease potential of random monoallelic gene expression.', *Developmental Cell* **28**(4), 366–80.

Gendrel, A.-V., Marion-Poll, L., Katoh, K. & Heard, E. (2016), 'Random monoallelic expression of genes on autosomes: Parallels with X-chromosome inactivation', *Seminars in Cell & Developmental Biology* **56**, 100–110.

Giardina, C., Pérez-Riba, M. & Lis, J. T. (1992), 'Promoter melting and TFIID complexes on Drosophila genes in vivo.', *Genes & Development* **6**(11), 2190–200.

Gilchrist, D. A., Dos Santos, G., Fargo, D. C., Xie, B., Gao, Y., Li, L. & Adelman, K. (2010), 'Pausing of RNA Polymerase II Disrupts DNA-Specified Nucleosome Organization to Enable Precise Gene Regulation', *Cell* **143**(4), 540–551.

Gimelbrant, A., Hutchinson, J. N., Thompson, B. R. & Chess, A. (2007), 'Widespread Monoallelic Expression on Human Autosomes', *Science* **318**(5853), 1136–1140.

Golding, I., Paulsson, J., Zawilski, S. M. & Cox, E. C. (2005), 'Real-Time Kinetics of Gene Activity in Individual Bacteria', *Cell* **123**(6), 1025–1036.

Goldman-Levi, R., Miller, C., Greenberg, G., Gabai, E. & Zak, N. B. (1996), 'Cellular pathways acting along the germband and in the amnioserosa may participate in germband retraction of the Drosophila melanogaster embryo.', *The International Journal of Developmental Biology* **40**(5), 1043–51.

Gorfinkiel, N., Schamberg, S. & Blanchard, G. B. (2011), 'Integrative approaches to morphogenesis: Lessons from dorsal closure', *genesis* **49**(7), 522–533.

Gray, A. & Mason, A. (1990), 'Requirement for activin A and transforming growth factor–beta 1 pro-regions in homodimer assembly', *Science* **247**(4948), 1328–1330.

Gregor, T., Garcia, H. G. & Little, S. C. (2014), 'The embryo as a laboratory: quantifying transcription in Drosophila', *Trends in Genetics* **30**(8), 364–375.

Gregor, T., Tank, D. W., Wieschaus, E. F. & Bialek, W. (2007), 'Probing the Limits to Positional Information', *Cell* **130**(1), 153–164.

Guenther, M. G., Levine, S. S., Boyer, L. A., Jaenisch, R. & Young, R. A. (2007), 'A Chromatin Landmark and Transcription Initiation at Most Promoters in Human Cells', *Cell* **130**(1), 77–88.

Haberle, V. & Stark, A. (2018), 'Eukaryotic core promoters and the functional basis of transcription initiation.', *Nature Reviews Molecular Cell Biology* **19**(10), 621–637.

Haenlin, M., Cubadda, Y., Blondeau, F., Heitzler, P., Lutz, Y., Simpson, P. & Ramain, P. (1997), 'Transcriptional activity of pannier is regulated negatively by heterodimerization of the GATA DNA-binding domain with a cofactor encoded by the u-shaped gene of Drosophila.', *Genes & Development* **11**(22), 3096–108.

Haerry, T. E., Khalsa, O., O'Connor, M. B. & Wharton, K. A. (1998), 'Synergistic signaling by two BMP ligands through the SAX and TKV receptors controls wing growth and patterning in Drosophila.', *Development* **125**(20), 3977–87.

Hahn, S. (2004), 'Structure and mechanism of the RNA polymerase II transcription machinery', *Nature Structural & Molecular Biology* **11**(5), 394–403.

Hamada, S., Ishiyama, K., Choi, S.-B., Wang, C., Singh, S., Kawai, N., Franceschi, V. R. & Okita, T. W. (2003), 'The transport of prolamine RNAs to prolamine protein bodies in living rice endosperm cells.', *The Plant Cell* **15**(10), 2253–64.

Hao, S. & Baltimore, D. (2013), 'RNA splicing regulates the temporal order of TNF-induced gene expression.', *Proceedings of the National Academy of Sciences of the United States of America* **110**(29), 11934–9.

Harper, C. V., Finkenstädt, B., Woodcock, D. J., Friedrichsen, S., Semprini, S., Ashall, L., Spiller, D. G., Mullins, J. J., Rand, D. A., Davis, J. R. E. & White, M. R. H. (2011), 'Dynamic Analysis of Stochastic Transcription Cycles', *PLoS Biology* **9**(4), e1000607.

Harrison, M. M., Li, X.-Y., Kaplan, T., Botchan, M. R. & Eisen, M. B. (2011), 'Zelda Binding in the Early Drosophila melanogaster Embryo Marks Regions Subsequently Activated at the Maternal-to-Zygotic Transition', *PLoS Genetics* **7**(10), e1002266.

Hartenstein, V. & Campos-Ortega, J. A. (1985), 'Fate-mapping in wild-typeDrosophila melanogaster', *Wilhelm Roux's Archives of Developmental Biology* **194**(4), 181–195.

Hartenstein, V. & Jan, Y. N. (1992), 'Studying Drosophila embryogenesis with P-lacZ enhancer trap lines', *Roux's Archives of Developmental Biology* **201**(4), 194–220.

Hashimoto, C., Hudson, K. L. & Anderson, K. V. (1988), 'The Toll gene of Drosophila, required for dorsal-ventral embryonic polarity, appears to encode a transmembrane protein', *Cell* **52**(2), 269–279.

Hazelrigg, T., Liu, N., Hong, Y. & Wang, S. (1998), 'GFP Expression inDrosophilaTissues: Time Requirements for Formation of a Fluorescent Product', *Developmental Biology* **199**(2), 245–249.

Heisenberg, C.-P. (2009), 'Dorsal closure in Drosophila : cells cannot get out of the tight spot', *BioEssays* **31**(12), 1284–1287.

Heist, T., Fukaya, T. & Levine, M. (2019), 'Large distances separate coregulated genes in living Drosophila embryos.', *Proceedings of the National Academy of Sciences of the United States of America* **116**(30), 15062–15067.

Hendrix, D. A., Hong, J.-W., Zeitlinger, J., Rokhsar, D. S. & Levine, M. S. (2008), 'Promoter elements associated with RNA Pol II stalling in the Drosophila embryo.', *Proceedings of the National Academy of Sciences of the United States of America* **105**(22), 7762–7.

Hill, C. S. (2009), 'Nucleocytoplasmic shuttling of Smad proteins', *Cell Research* **19**(1), 36–46.

Hiraoka, Y., Dernburg, A. F., Parmelee, S. J., Rykowski, M. C., Agard, D. A. & Sedat, J. W. (1993), 'The onset of homologous chromosome pairing during Drosophila melanogaster embryogenesis.', *The Journal of Cell Biology* **120**(3), 591–600.

Hnisz, D., Abraham, B., Lee, T., Lau, A., Saint-André, V., Sigova, A., Hoke, H. & Young, R. (2013), 'Super-Enhancers in the Control of Cell Identity and Disease', *Cell* **155**(4), 934–947.

Hnisz, D., Shrinivas, K., Young, R. A., Chakraborty, A. K. & Sharp, P. A. (2017), 'A Phase Separation Model for Transcriptional Control', *Cell* **169**(1), 13–23.

Ho, M. C. W., Johnsen, H., Goetz, S. E., Schiller, B. J., Bae, E., Tran, D. A., Shur, A. S., Allen, J. M., Rau, C., Bender, W., Fisher, W. W., Celniker, S. E. & Drewell, R. A. (2009), 'Functional Evolution of cis-Regulatory Modules at a Homeotic Gene in Drosophila', *PLoS Genetics* **5**(11), e1000709.

Hogan, B. L. (1996), 'Bone morphogenetic proteins in development.', *Current Opinion in Genetics & Development* **6**(4), 432–8.

Holley, S. A., Neul, J. L., Attisano, L., Wrana, J. L., Sasai, Y., O'Connor, M. B., De Robertis, E. M. & Ferguson, E. L. (1996), 'The Xenopus dorsalizing factor noggin ventralizes Drosophila embryos by preventing DPP from activating its receptor.', *Cell* **86**(4), 607–17.

Hong, J.-W., Hendrix, D. A. & Levine, M. S. (2008), 'Shadow Enhancers as a Source of Evolutionary Novelty', *Science* **321**(5894), 1314–1314.

Huang, J. D., Schwyter, D. H., Shirokawa, J. M. & Courey, A. J. (1993), 'The interplay between multiple enhancer and silencer elements defines the pattern of decapentaplegic expression.', *Genes & Development* **7**(4), 694–704.

Hudson, J. B., Podos, S. D., Keith, K., Simpson, S. L. & Ferguson, E. L. (1998), 'The Drosophila Medea gene is required downstream of dpp and encodes a functional homolog of human Smad4.', *Development* **125**(8), 1407–20.

Hurley, J. M., Loros, J. J. & Dunlap, J. C. (2016), 'Circadian Oscillators: Around the Transcription-Translation Feedback Loop and on to Output.', *Trends in Biochemical Sciences* **41**(10), 834–846.

Hyun, S., Lee, J. H., Jin, H., Nam, J., Namkoong, B., Lee, G., Chung, J. & Kim, V. N. (2009), 'Conserved MicroRNA miR-8/miR-200 and Its Target USH/FOG2 Control Growth by Regulating PI3K', *Cell* **139**(6), 1096–1108.

Inman, G. J. & Hill, C. S. (2002), 'Stoichiometry of active smad-transcription factor complexes on DNA.', *The Journal of Biological Chemistry* **277**(52), 51008–16.

Inman, G. J., Nicolás, F. J. & Hill, C. S. (2002), 'Nucleocytoplasmic shuttling of Smads 2, 3, and 4 permits sensing of TGF-beta receptor activity.', *Molecular Cell* **10**(2), 283–94.

Irish, V. F. & Gelbart, W. M. (1987), 'The decapentaplegic gene is required for dorsal-ventral patterning of the Drosophila embryo.', *Genes & Development* **1**(8), 868–79.

Itzkovitz, S., Lyubimova, A., Blat, I. C., Maynard, M., van Es, J., Lees, J., Jacks, T., Clevers, H. & van Oudenaarden, A. (2012), 'Single-molecule transcript counting of stem-cell markers in the mouse intestine', *Nature Cell Biology* **14**(1), 106–114.

Janssens, H., Crombach, A., Wotton, K. R., Cicin-Sain, D., Surkova, S., Lim, C. L., Samsonova, M., Akam, M. & Jaeger, J. (2013), 'Lack of tailless leads to an increase in expression variability in Drosophila embryos.', *Developmental Biology* **377**(1), 305–17.

Jaźwińska, A., Rushlow, C. & Roth, S. (1999), 'The role of brinker in mediating the graded response to Dpp in early Drosophila embryos.', *Development* **126**(15), 3323–34.

Jin, Y., Tipoe, G. L., Liong, E. C., Lau, T. Y., Fung, P. C. & Leung, K. M. (2001), 'Overexpression of BMP-2/4, -5 and BMPR-IA associated with malignancy of oral epithelium.', *Oral Oncology* **37**(3), 225–33.

Jones, D. L., Brewster, R. C. & Phillips, R. (2014), 'Promoter architecture dictates cell-to-cell variability in gene expression', *Science* **346**(6216), 1533–1536.

Jürgens, G., Wieschaus, E., Nüsslein-Volhard, C. & Kluding, H. (1984), 'Mutations affecting the pattern of the larval cuticle in Drosophila melanogaster: II. Zygotic loci on the third chromosome', *Wilhelm Roux's Archives of Developmental Biology* **193**(5), 283–295.

Kaufmann, B. B. & van Oudenaarden, A. (2007), 'Stochastic gene expression: from single molecules to the proteome', *Current Opinion in Genetics & Development* **17**(2), 107–112.

Kidd, S. (1992), 'Characterization of the Drosophila cactus locus and analysis of interactions between cactus and dorsal proteins', *Cell* **71**(4), 623–635.

Kiehart, D. P., Galbraith, C. G., Edwards, K. A., Rickoll, W. L. & Montague, R. A. (2000), 'Multiple forces contribute to cell sheet morphogenesis for dorsal closure in Drosophila.', *The Journal of cell biology* **149**(2), 471–90.

Kleeff, J., Maruyama, H., Ishiwata, T., Sawhney, H., Friess, H., Büchler, M. W. & Korc, M. (1999), 'Bone morphogenetic protein 2 exerts diverse effects on cell growth in vitro and is expressed in human pancreatic cancer in vivo.', *Gastroenterology* **116**(5), 1202–16.

Kornezos, A. & Chia, W. (1992), 'Apical secretion and association of the Drosophila yellow gene product with developing larval cuticle structures during embryogenesis', *MGG Molecular & General Genetics* **235**(2-3), 397–405.

Kosman, D., Ip, Y., Levine, M. & Arora, K. (1991), 'Establishment of the mesoderm-neuroectoderm boundary in the Drosophila embryo', *Science* **254**(5028), 118–122.

Krumm, A., Hickey, L. B. & Groudine, M. (1995), 'Promoter-proximal pausing of RNA polymerase II defines a general rate-limiting step after transcription initiation.', *Genes & Development* **9**(5), 559–572.

Künnapuu, J., Björkgren, I. & Shimmi, O. (2009), 'The Drosophila DPP signal is produced by cleavage of its proprotein at evolutionary diversified furin-recognition sites.', *Proceedings of the National Academy of Sciences of the United States of America* **106**(21), 8501–6.

Kuntz, S. G. & Eisen, M. B. (2014), 'Drosophila Embryogenesis Scales Uniformly across Temperature in Developmentally Diverse Species', *PLoS Genetics* **10**(4).

Kvon, E. Z., Kazmar, T., Stampfel, G., Yáñez-Cuna, J. O., Pagani, M., Schernhuber, K., Dickson, B. J. & Stark, A. (2014), 'Genome-scale functional characterization of Drosophila developmental enhancers in vivo', *Nature* **512**(7512), 91–95.

Kwak, H., Fuda, N. J., Core, L. J. & Lis, J. T. (2013), 'Precise maps of RNA polymerase reveal how promoters direct initiation and pausing.', *Science* **339**(6122), 950–3.

Lacy, M. E. & Hutson, M. S. (2016), 'Amnioserosa development and function in Drosophila embryogenesis: Critical mechanical roles for an extraembryonic tissue.', *Developmental Dynamics* **245**(5), 558–568.

Lada, K., Gorfinkiel, N. & Martinez Arias, A. (2012), 'Interactions between the amnioserosa and the epidermis revealed by the function of the u-shaped gene.', *Biology open* **1**(4), 353–61.

Lagha, M., Bothma, J. P., Esposito, E., Ng, S., Stefanik, L., Tsui, C., Johnston, J., Chen, K., Gilmour, D. S., Zeitlinger, J. & Levine, M. S. (2013), 'Paused Pol II coordinates tissue morphogenesis in the Drosophila embryo.', *Cell* **153**(5), 976–87.

Lamka, M. L. & Lipshitz, H. D. (1999), 'Role of the Amnioserosa in Germ Band Retraction of the Drosophila melanogaster Embryo', *Developmental Biology* **214**(1), 102–112.

Lammers, N. C., Galstyan, V., Reimer, A., Medin, S. A., Wiggins, C. H. & Garcia, H. G. (2019), 'Multimodal transcriptional control of pattern formation in embryonic development', *bioRxiv* p. 335919.

Langenfeld, E. M., Calvano, S. E., Abou-Nukta, F., Lowry, S. F., Amenta, P. & Langenfeld, J. (2003), 'The mature bone morphogenetic protein-2 is aberrantly expressed in non-small cell lung carcinomas and stimulates tumor growth of A549 cells.', *Carcinogenesis* **24**(9), 1445–54.

Larson, D. R., Singer, R. H. & Zenklusen, D. (2009), 'A single molecule view of gene expression.', *Trends in Cell Biology* **19**(11), 630–7.

Larsson, A. J. M., Johnsson, P., Hagemann-Jensen, M., Hartmanis, L., Faridani, O. R., Reinius, B., Segerstolpe, Å., Rivera, C. M., Ren, B. & Sandberg, R. (2019), 'Genomic encoding of transcriptional burst kinetics', *Nature* **565**(7738), 251–254.

Lecuit, T., Brook, W. J., Ng, M., Calleja, M., Sun, H. & Cohen, S. M. (1996), 'Two distinct mechanisms for long-range patterning by Decapentaplegic in the Drosophila wing', *Nature* **381**(6581), 387–393.

LeCuyer, K. A., Behlen, L. S. & Uhlenbeck, O. C. (1996), 'Mutagenesis of a stacking contact in the MS2 coat protein-RNA complex.', *The EMBO Journal* **15**(24), 6847–53.

Lefebvre, F., Lécuyer, É., Lefebvre, F. A. & Lécuyer, É. (2018), 'Flying the RNA Nest: Drosophila Reveals Novel Insights into the Transcriptome Dynamics of Early Development', *Journal of Developmental Biology* **6**(1), 5.

Lenhard, B., Sandelin, A. & Carninci, P. (2012), 'Metazoan promoters: emerging characteristics and insights into transcriptional regulation', *Nature Reviews Genetics* **13**(4), 233–245.

Letsou, A., Arora, K., Wrana, J. L., Simin, K., Twombly, V., Jamal, J., Staehling-Hampton, K., Hoffmann, F. M., Gelbart, W. M. & Massagué, J. (1995), 'Drosophila Dpp signaling is mediated by the punt gene product: a dual ligand-binding type II receptor of the TGF beta receptor family.', *Cell* **80**(6), 899–908.

Levine, M. (2010), 'Transcriptional Enhancers in Animal Development and Evolution', *Current Biology* **20**(17), R754–R763.

Levine, M. (2011), 'Paused RNA polymerase II as a developmental checkpoint.', *Cell* **145**(4), 502–11.

Levine, M. & Davidson, E. H. (2005), 'Gene regulatory networks for development', *Proceedings of the National Academy of Sciences* **102**(14), 4936–4942.

Li, X.-y., MacArthur, S., Bourgon, R., Nix, D., Pollard, D. A., Iyer, V. N., Hechmer, A., Simirenko, L., Stapleton, M., Hendriks, C. L. L., Chu, H. C., Ogawa, N., Inwood, W., Sementchenko, V., Beaton, A., Weiszmann, R., Celniker, S. E., Knowles, D. W., Gingeras, T., Speed, T. P., Eisen, M. B. & Biggin, M. D. (2008), 'Transcription Factors Bind Thousands of Active and Inactive Regions in the Drosophila Blastoderm', *PLoS Biology* **6**(2), e27.

Liang, H.-L., Nien, C.-Y., Liu, H.-Y., Metzstein, M. M., Kirov, N. & Rushlow, C. (2008), 'The zinc-finger protein Zelda is a key activator of the early zygotic genome in Drosophila', *Nature* **456**(7220), 400–403.

Lim, B., Heist, T., Levine, M. & Fukaya, T. (2018), 'Visualization of Transvection in Living Drosophila Embryos.', *Molecular Cell* **70**(2), 287–296.e6.

Lim, F., Downey, T. P. & Peabody, D. S. (2001), 'Translational Repression and Specific RNA Binding by the Coat Protein of the Pseudomonas Phage PP7', *Journal of Biological Chemistry* **276**(25), 22507–22513.

Lim, F. & Peabody, D. S. (2002), 'RNA recognition site of PP7 coat protein.', *Nucleic Acids Research* **30**(19), 4138–44.

Lionnet, T. & Singer, R. H. (2012), 'Transcription goes digital', *EMBO Reports* **13**(4), 313–321.

Lis, J. (1998), 'Promoter-associated pausing in promoter architecture and postinitiation transcriptional regulation.', *Cold Spring Harbor symposia on quantitative biology* **63**, 347–56.

Little, S. C. & Gregor, T. (2018), Single mRNA molecule detection in drosophila, *in* 'Methods in Molecular Biology', 1 edn, Vol. 1649, Humana Press, New York, NY, pp. 127–142.

Little, S. C., Tkačik, G., Kneeland, T. B., Wieschaus, E. F. & Gregor, T. (2011), 'The Formation of the Bicoid Morphogen Gradient Requires Protein Movement from Anteriorly Localized mRNA', *PLoS Biology* **9**(3), e1000596.

Little, S., Tikhonov, M. & Gregor, T. (2013), 'Precise Developmental Gene Expression Arises from Globally Stochastic Transcriptional Activity', *Cell* **154**(4), 789–800.

Lloyd, S. & S. (1982), 'Least squares quantization in PCM', *IEEE Transactions on Information Theory* **28**(2), 129–137.

Loncar, D. & Singer, S. J. (1995), 'Cell membrane formation during the cellularization of the syncytial blastoderm of Drosophila.', *Proceedings of the National Academy of Sciences of the United States of America* **92**(6), 2199–203.

Lott, S. E., Villalta, J. E., Schroth, G. P., Luo, S., Tonkin, L. A. & Eisen, M. B. (2011), 'Noncanonical Compensation of Zygotic X Transcription in Early Drosophila melanogaster Development Revealed through Single-Embryo RNA-Seq', *PLOS Biology* **9**(2).

Lovén, J., Hoke, H., Lin, C., Lau, A., Orlando, D., Vakoc, C., Bradner, J., Lee, T. & Young, R. (2013), 'Selective Inhibition of Tumor Oncogenes by Disruption of Super-Enhancers', *Cell* **153**(2), 320–334.

Lowary, P. T. & Uhlenbeck, O. C. (1987), 'An RNA mutation that increases the affinity of an RNA-protein interaction.', *Nucleic Acids Research* **15**(24), 10483–93.

Lucas, T., Ferraro, T., Roelens, B., De Las Heras Chanes, J., Walczak, A. M., Coppey, M. & Dostatni, N. (2013), 'Live Imaging of Bicoid-Dependent Transcription in Drosophila Embryos', *Current Biology* **23**(21), 2135–2139.

Lucas, T., Tran, H., Perez Romero, C. A., Guillou, A., Fradin, C., Coppey, M., Walczak, A. M. & Dostatni, N. (2018), '3 minutes to precisely measure morphogen concentration', *PLOS Genetics* **14**(10), e1007676.

MacArthur, S., Li, X.-Y., Li, J., Brown, J. B., Chu, H. C., Zeng, L., Grondona, B. P., Hechmer, A., Simirenko, L., Keränen, S. V., Knowles, D. W., Stapleton, M., Bickel, P., Biggin, M. D. & Eisen, M. B. (2009), 'Developmental roles of 21 Drosophila transcription factors are determined by quantitative differences in binding to an overlapping set of thousands of genomic regions', *Genome Biology* **10**(7), R80.

Malik, S. & Roeder, R. G. (2000), 'Transcriptional regulation through Mediator-like coactivators in yeast and metazoan cells', *Trends in Biochemical Sciences* **25**(6), 277–283.

Mansour, M. R., Abraham, B. J., Anders, L., Berezovskaya, A., Gutierrez, A., Durbin, A. D., Etchin, J., Lawton, L., Sallan, S. E., Silverman, L. B., Loh, M. L., Hunger, S. P., Sanda, T., Young, R. A. & Look, A. T. (2014), 'An oncogenic super-enhancer formed through somatic mutation of a noncoding intergenic element', *Science* **346**(6215), 1373–1377.

Marqués, G., Bao, H., Haerry, T. E., Shimell, M. J., Duchek, P., Zhang, B. & O'Connor, M. B. (2002), 'The Drosophila BMP Type II Receptor Wishful Thinking Regulates Neuromuscular Synapse Morphology and Function', *Neuron* **33**(4), 529–543.

Martin, R., Rino, J., Carvalho, C., Kirchhausen, T. & Carmo-Fonseca, M. (2013), 'Live-Cell Visualization of Pre-mRNA Splicing with Single-Molecule Sensitivity', *Cell Reports* **4**(6), 1144–1155.

Mason, E. D., Konrad, K. D., Webb, C. D. & Marsh, J. L. (1994), 'Dorsal midline fate in Drosophila embryos requires twisted gastrulation, a gene encoding a secreted protein related to human connective tissue growth factor.', *Genes & Development* **8**(13), 1489–501.

Maston, G. A., Evans, S. K. & Green, M. R. (2006), 'Transcriptional Regulatory Elements in the Human Genome', *Annual Review of Genomics and Human Genetics* **7**(1), 29–59.

Miles, W. O., Jaffray, E., Campbell, S. G., Takeda, S., Bayston, L. J., Basu, S. P., Li, M., Raftery, L. A., Ashe, M. P., Hay, R. T. & Ashe, H. L. (2008), 'Medea SUMOylation restricts the signaling range of the Dpp morphogen in the Drosophila embryo.', *Genes & development* **22**(18), 2578–90.

Mir, M., Bickmore, W., Furlong, E. E. M. & Narlikar, G. (2019), 'Chromatin topology, condensates and gene regulation: shifting paradigms or just a phase?', *Development* **146**(19), dev182766.

Mir, M., Stadler, M. R., Ortiz, S. A., Hannon, C. E., Harrison, M. M., Darzacq, X. & Eisen, M. B. (2018), 'Dynamic multifactor hubs interact transiently with sites of active transcription in Drosophila embryos', *eLife* **7**.

Mizutani, C. M., Nie, Q., Wan, F. Y., Zhang, Y.-T., Vilmos, P., Sousa-Neves, R., Bier, E., Marsh, J. L. & Lander, A. D. (2005), 'Formation of the BMP Activity Gradient in the Drosophila Embryo', *Developmental Cell* **8**(6), 915–924.
**URL:** *https://www.sciencedirect.com/science/article/pii/S1534580705001401#fig1*

Montanuy, I., Torremocha, R., Hernández-Munain, C. & Suñé, C. (2008), 'Promoter influences transcription elongation: TATA-box element mediates the assembly of processive transcription complexes responsive to cyclin-dependent kinase 9.', *The Journal of Biological Chemistry* **283**(12), 7368–78.

Moras, D. & Poterszman, A. (1995), 'RNA-protein interactions. Diverse modes of recognition.', *Current Biology* **5**(3), 249–51.

Morisato, D. & Anderson, K. V. (1994), 'The spätzle gene encodes a component of the extracellular signaling pathway establishing the dorsal-ventral pattern of the Drosophila embryo.', *Cell* **76**(4), 677–88.

Munsky, B., Neuert, G. & van Oudenaarden, A. (2012), 'Using gene expression noise to understand gene regulation.', *Science* **336**(6078), 183–7.

Muse, G. W., Gilchrist, D. A., Nechaev, S., Shah, R., Parker, J. S., Grissom, S. F., Zeitlinger, J. & Adelman, K. (2007), 'RNA polymerase is poised for activation across the genome.', *Nature Genetics* **39**(12), 1507–11.

Narita, T., Yamaguchi, Y., Yano, K., Sugimoto, S., Chanarat, S., Wada, T., Kim, D.-k., Hasegawa, J., Omori, M., Inukai, N., Endoh, M., Yamada, T. & Handa, H. (2003), 'Human transcription elongation factor NELF: identification of novel subunits and reconstitution of the functionally active complex.', *Molecular and Cellular Biology* **23**(6), 1863–73.

Nellen, D., Affolter, M. & Basler, K. (1994), 'Receptor serine/threonine kinases implicated in the control of Drosophila body pattern by decapentaplegic.', *Cell* **78**(2), 225–37.

Neuman, K. C., Abbondanzieri, E. A., Landick, R., Gelles, J. & Block, S. M. (2003), 'Ubiquitous transcriptional pausing is independent of RNA polymerase backtracking.', *Cell* **115**(4), 437–47.

Newman, J. R. S., Ghaemmaghami, S., Ihmels, J., Breslow, D. K., Noble, M., DeRisi, J. L. & Weissman, J. S. (2006), 'Single-cell proteomic analysis of S. cerevisiae reveals the architecture of biological noise', *Nature* **441**(7095), 840–846.

Newton, F. G., Harris, R. E., Sutcliffe, C. & Ashe, H. L. (2015), 'Coordinate post-transcriptional repression of Dpp-dependent transcription factors attenuates signal range during development.', *Development* **142**(19), 3362–73.

Nicolas, D., Phillips, N. E. & Naef, F. (2017), 'What shapes eukaryotic transcriptional bursting?', *Molecular BioSystems* **13**(7), 1280–1290.

Nüsslein-Volhard, C., Wieschaus, E. & Kluding, H. (1984), 'Mutations affecting the pattern of the larval cuticle in Drosophila melanogaster: I. Zygotic loci on the second chromosome.', *Wilhelm Roux's Archives of Developmental Biology* **193**(5), 267–282.

O'Connor, M. B., Umulis, D., Othmer, H. G. & Blair, S. S. (2006), 'Shaping BMP morphogen gradients in the Drosophila embryo and pupal wing.', *Development* **133**(2), 183–93.

O'Farrell, P. H. (2001), 'Triggering the all-or-nothing switch into mitosis.', *Trends in cell biology* **11**(12), 512–9.

Ohler, U., Liao, G.-c., Niemann, H. & Rubin, G. M. (2002), 'Computational analysis of core promoters in the Drosophila genome.', *Genome Biology* **3**(12), 87.1.

Oubridge, C., Ito, N., Evans, P. R., Teo, C.-H. & Nagai, K. (1994), 'Crystal structure at 1.92 Å resolution of the RNA-binding domain of the U1A spliceosomal protein complexed with an RNA hairpin', *Nature* **372**(6505), 432–438.

Ozkaynak, E., Schnegelsberg, P. N., Jin, D. F., Clifford, G. M., Warren, F. D., Drier, E. A. & Oppermann, H. (1992), 'Osteogenic protein-2. A new member of the transforming growth factor-beta superfamily expressed early in embryogenesis.', *The Journal of Biological Chemistry* **267**(35), 25220–7.

Padgett, R. W., St. Johnston, R. D. & Gelbart, W. M. (1987), 'A transcript from a Drosophila pattern gene predicts a protein homologous to the transforming growth factor-$\beta$ family', *Nature* **325**(6099), 81–84.

Padovan-Merhar, O., Nair, G. P., Biaesch, A. G., Mayer, A., Scarfone, S., Foley, S. W., Wu, A. R., Churchman, L. S., Singh, A. & Raj, A. (2015), 'Single mammalian cells compensate for differences in cellular volume and DNA copy number through independent global transcriptional mechanisms.', *Molecular Cell* **58**(2), 339–52.

Palam, L. R., Baird, T. D. & Wek, R. C. (2011), 'Phosphorylation of eIF2 Facilitates Ribosomal Bypass of an Inhibitory Upstream ORF to Enhance CHOP Translation', *Journal of Biological Chemistry* **286**(13), 10939–10949.

Paré, A., Lemons, D., Kosman, D., Beaver, W., Freund, Y. & McGinnis, W. (2009), 'Visualization of Individual Scr mRNAs during Drosophila Embryogenesis Yields Evidence for Transcriptional Bursting', *Current Biology* **19**(23), 2037–2042.

Parker, L., Stathakis, D. G. & Arora, K. (2004), 'Regulation of BMP and activin signaling in Drosophila.', *Progress in Molecular and Subcellular Biology* **34**, 73–101.

Peccoud, J. & Ycart, B. (1995), 'Markovian Modeling of Gene-Product Synthesis', *Theoretical Population Biology* **48**(2), 222–234.

Pernis, B., Chiappino, G., Kelus, A. S. & Gell, P. G. (1965), 'Cellular localization of immunoglobulins with different allotypic specificities in rabbit lymphoid tissues.', *The Journal of Experimental Medicine* **122**(5), 853–76.

Perrimon, N., Engstrom, L. & Mahowald, A. P. (1989), 'Zygotic lethals with specific maternal effect phenotypes in Drosophila melanogaster. I. Loci on the X chromosome.', *Genetics* **121**(2), 333–52.

Perry, M. W., Boettiger, A. N., Bothma, J. P. & Levine, M. (2010), 'Shadow Enhancers Foster Robustness of Drosophila Gastrulation', *Current Biology* **20**(17), 1562–1567.

Peters, J.-M., Tedeschi, A. & Schmitz, J. (2008), 'The cohesin complex and its roles in chromosome biology.', *Genes & Development* **22**(22), 3089–114.

Phillips, N. E., Manning, C., Papalopulu, N. & Rattray, M. (2017), 'Identifying stochastic oscillations in single-cell live imaging time series using Gaussian processes.', *PLoS Computational Biology* **13**(5), e1005479.

Pickup, A. T., Lamka, M. L., Sun, Q., Yip, M. L. R. & Lipshitz, H. D. (2002), 'Control of photoreceptor cell morphology, planar polarity and epithelial integrity during Drosophila eye development.', *Development* **129**(9), 2247–58.

Pope, K. L. & Harris, T. J. C. (2008), 'Control of cell flattening and junctional remodeling during squamous epithelial morphogenesis in Drosophila.', *Development* **135**(13), 2227–38.

Porcher, A., Abu-Arish, A., Huart, S., Roelens, B., Fradin, C. & Dostatni, N. (2010), 'The time to measure positional information: maternal Hunchback is required for the synchrony of the Bicoid transcriptional response at the onset of zygotic transcription', *Development* **137**(16), 2795–2804.

Pritchard, D. K. & Schubiger, G. (1996), 'Activation of transcription in Drosophila embryos is a gradual process mediated by the nucleocytoplasmic ratio.', *Genes & Development* **10**(9), 1131–42.

Raftery, L. A. & Sutherland, D. J. (2003), 'Gradients and thresholds: BMP response gradients unveiled in Drosophila embryos', *Trends in Genetics* **19**(12), 701–708.

Raj, A., Peskin, C. S., Tranchina, D., Vargas, D. Y. & Tyagi, S. (2006), 'Stochastic mRNA Synthesis in Mammalian Cells', *PLoS Biology* **4**(10), 309.

Raj, A., van den Bogaard, P., Rifkin, S. A., van Oudenaarden, A. & Tyagi, S. (2008), 'Imaging individual mRNA molecules using multiple singly labeled probes', *Nature Methods* **5**(10), 877–879.

Raj, A. & van Oudenaarden, A. (2008), 'Nature, Nurture, or Chance: Stochastic Gene Expression and Its Consequences', *Cell* **135**(2), 216–226.

Ramírez, F., Bhardwaj, V., Arrigoni, L., Lam, K. C., Grüning, B. A., Villaveces, J., Habermann, B., Akhtar, A. & Manke, T. (2018), 'High-resolution TADs reveal DNA sequences underlying genome organization in flies', *Nature Communications* **9**(1), 189.

Rao, S., Huntley, M., Durand, N., Stamenova, E., Bochkov, I., Robinson, J., Sanborn, A., Machol, I., Omer, A., Lander, E. & Aiden, E. (2014), 'A 3D Map of the Human Genome at Kilobase Resolution Reveals Principles of Chromatin Looping', *Cell* **159**(7), 1665–1680.

Raser, J. M. & O'Shea, E. K. (2005), 'Noise in gene expression: origins, consequences, and control.', *Science* **309**(5743), 2010–3.

Rasmussen, E. B. & Lis, J. T. (1993), 'In vivo transcriptional pausing and cap formation on three Drosophila heat shock genes.', *Proceedings of the National Academy of Sciences of the United States of America* **90**(17), 7923–7.

Ray, R. P., Arora, K., Nüsslein-Volhard, C. & Gelbart, W. M. (1991), 'The control of cell fate along the dorsal-ventral axis of the Drosophila embryo.', *Development* **113**(1), 35–54.

Reed, B. H., Wilk, R. & Lipshitz, H. D. (2001), 'Downregulation of Jun kinase signaling in the amnioserosa is essential for dorsal closure of the Drosophila embryo.', *Current Biology* **11**(14), 1098–108.

Reeves, G. T. & Stathopoulos, A. (2009), 'Graded dorsal and differential gene regulation in the Drosophila embryo.', *Cold Spring Harbor Perspectives in Biology* **1**(4), 836.

Reik, W. & Walter, J. (2001), 'Genomic imprinting: parental influence on the genome.', *Nature Reviews Genetics* **2**(1), 21–32.

Reinius, B. & Sandberg, R. (2015), 'Random monoallelic expression of autosomal genes: stochastic transcription and allele-level regulation', *Nature Reviews Genetics* **16**(11), 653–664.

Reiter, L. T., Potocki, L., Chien, S., Gribskov, M. & Bier, E. (2001), 'A systematic analysis of human disease-associated gene sequences in Drosophila melanogaster.', *Genome research* **11**(6), 1114–25.

Roeder, R. G. (1996), 'The role of general initiation factors in transcription by RNA polymerase II', *Trends in Biochemical Sciences* **21**(9), 327–335.

Rook, M. S., Lu, M. & Kosik, K. S. (2000), 'CaMKIIalpha 3' untranslated region-directed mRNA translocation in living neurons: visualization by GFP linkage.', *The Journal of Neuroscience* **20**(17), 6385–93.

Ross, J. J., Shimmi, O., Vilmos, P., Petryk, A., Kim, H., Gaudenz, K., Hermanson, S., Ekker, S. C., O'Connor, M. B. & Marsh, J. L. (2001), 'Twisted gastrulation is a conserved extracellular BMP antagonist', *Nature* **410**(6827), 479–483.

Roth, S., Hiromi, Y., Godt, D. & Nüsslein-Volhard, C. (1991), 'cactus, a maternal gene required for proper formation of the dorsoventral morphogen gradient in Drosophila embryos', *Development* **112**(2), 371–388.

Roth, S., Stein, D. & Nüsslein-Volhard, C. (1989), A Gradient of Nuclear Localization of the dorsal Protein Determines Dorsoventral Pattern in the Drosophila Embryo, Technical report.

Rothe, M., Pehl, M., Taubert, H. & Jäckle, H. (1992), 'Loss of gene function through rapid mitotic cycles in the Drosophila embryo', *Nature* **359**(6391), 156–159.

Ruberte, E., Marty, T., Nellen, D., Affolter, M. & Basler, K. (1995), 'An absolute requirement for both the type II and type I receptors, punt and thick veins, for dpp signaling in vivo.', *Cell* **80**(6), 889–97.

Rusch, J. & Levine, M. (1996), 'Threshold responses to the dorsal regulatory gradient and the subdivision of primary tissue territories in the Drosophila embryo.', *Current opinion in Genetics & Development* **6**(4), 416–23.

Rusch, J. & Levine, M. (1997), 'Regulation of a dpp target gene in the Drosophila embryo.', *Development* **124**(2), 303–11.

Rushlow, C. A., Han, K., Manley, J. L. & Levine, M. (1989), 'The graded distribution of the dorsal morphogen is initiated by selective nuclear transport in Drosophila', *Cell* **59**(6), 1165–1177.

Rushlow, C. & Roth, S. (1996), 'The Role of the dpp-Group Genes in Dorsoventral Patterning of the Drosophila Embryo', *Advances in Developmental Biology (1992)* **4**, 27–82.

Sabari, B. R., Dall'Agnese, A., Boija, A., Klein, I. A., Coffey, E. L., Shrinivas, K., Abraham, B. J., Hannett, N. M., Zamudio, A. V., Manteiga, J. C., Li, C. H., Guo, Y. E., Day, D. S., Schuijers, J., Vasile, E., Malik, S., Hnisz, D., Lee, T. I., Cisse, I. I., Roeder, R. G., Sharp, P. A., Chakraborty, A. K. & Young, R. A. (2018), 'Coactivator condensation at super-enhancers links phase separation and gene control.', *Science* **361**(6400).

Salazar, V. S., Gamer, L. W. & Rosen, V. (2016), 'BMP signalling in skeletal development, disease and repair', *Nature Reviews Endocrinology* **12**(4), 203–221.

Sandler, J. E. & Stathopoulos, A. (2016), 'Quantitative Single-Embryo Profile of Drosophila Genome Activation and the Dorsal-Ventral Patterning Network.', *Genetics* **202**(4), 1575–84.

Saunders, A., Core, L. J., Sutcliffe, C., Lis, J. T. & Ashe, H. L. (2013), 'Extensive polymerase pausing during Drosophila axis patterning enables high-level and pliable transcription.', *Genes & development* **27**(10), 1146–58.

Sawala, A., Scarcia, M., Sutcliffe, C., Wilcockson, S. G. & Ashe, H. L. (2015), 'Peak BMP Responses in the Drosophila Embryo Are Dependent on the Activation of Integrin Signaling.', *Cell Reports* **12**(10), 1584–93.

Schejter, E. D. & Wieschaus, E. (1993), 'Functional Elements of the Cytoskeleton in the Early Drosophila Embryo', *Annual Review of Cell Biology* **9**(1), 67–99.

Schneider, D. S., Jin, Y., Morisato, D. & Anderson, K. V. (1994), 'A processed form of the Spätzle protein defines dorsal-ventral polarity in the Drosophila embryo.', *Development* **120**(5), 1243–50.

Schöck, F. & Perrimon, N. (2002), 'Cellular Processes Associated with Germ Band Retraction in Drosophila', *Developmental Biology* **248**(1), 29–39.

Schöck, F. & Perrimon, N. (2003), 'Retraction of the Drosophila germ band requires cell-matrix interaction.', *Genes & development* **17**(5), 597–602.

Schoenfelder, S. & Fraser, P. (2019), 'Long-range enhancerâĂŞpromoter contacts in gene expression control', *Nature Reviews Genetics* **20**(8), 437–455.

Scholes, C., Biette, K. M., Harden, T. T. & DePace, A. H. (2019), 'Signal Integration by Shadow Enhancers and Enhancer Duplications Varies across the Drosophila Embryo', *Cell Reports* **26**(9), 2407–2418.e5.

Shao, W., Alcantara, S. G.-M. & Zeitlinger, J. (2019), 'Reporter-ChIP-nexus reveals strong contribution of the Drosophila initiator sequence to RNA polymerase pausing', *eLife* **8**.

Shao, W. & Zeitlinger, J. (2017), 'Paused RNA polymerase II inhibits new transcriptional initiation', *Nature Genetics* **49**(7), 1045–1051.

Shav-Tal, Y., Darzacq, X., Shenoy, S. M., Fusco, D., Janicki, S. M., Spector, D. L. & Singer, R. H. (2004), 'Dynamics of single mRNPs in nuclei of living cells.', *Science* **304**(5678), 1797–800.

Sheinberger, J., Hochberg, H., Lavi, E., Kanter, I., Avivi, S., Reinitz, G., Schwed, A., Aizler, Y., Varon, E., Kinor, N. & Shav-Tal, Y. (2017), 'CD-tagging-MS2: detecting allelic expression of endogenous mRNAs and their protein products in single cells', *Biology Methods and Protocols* **2**(1).

Shermoen, A. W., McCleland, M. L. & O'Farrell, P. H. (2010), 'Developmental Control of Late Replication and S Phase Length', *Current Biology* **20**(23), 2067–2077.

Shi, Y. & Massagué, J. (2003), 'Mechanisms of TGF-beta signaling from cell membrane to the nucleus.', *Cell* **113**(6), 685–700.

Shimell, M. J., Ferguson, E. L., Childs, S. R. & O'Connor, M. B. (1991), 'The Drosophila dorsal-ventral patterning gene tolloid is related to human bone morphogenetic protein 1.', *Cell* **67**(3), 469–81.

Shimmi, O., Umulis, D., Othmer, H. & O'Connor, M. B. (2005), 'Facilitated Transport of a Dpp/Scw Heterodimer by Sog/Tsg Leads to Robust Patterning of the Drosophila Blastoderm Embryo', *Cell* **120**(6), 873–886.

Simon, M., Maresh, J. G., Harris, S. E., Hernandez, J. D., Arar, M., Olson, M. S. & Abboud, H. E. (1999), 'Expression of bone morphogenetic protein-7 mRNA in normal and ischemic adult rat kidney.', *The American Journal of Physiology* **276**(3 Pt 2), F382–9.

Singer, M. A., Penton, A., Twombly, V., Hoffmann, F. M. & Gelbart, W. M. (1997), 'Signaling through both type I DPP receptors is required for anterior-posterior patterning of the entire Drosophila wing.', *Development* **124**(1), 79–89.

Smale, S. T. & Baltimore, D. (1989), 'The âĂIJinitiatorâĂİ as a transcription control element', *Cell* **57**(1), 103–113.

Smale, S. T. & Kadonaga, J. T. (2003), 'The RNA Polymerase II Core Promoter', *Annual Review of Biochemistry* **72**(1), 449–479.

Sopory, S., Kwon, S., Wehrli, M. & Christian, J. L. (2010), 'Regulation of Dpp activity by tissue-specific cleavage of an upstream site within the prodomain', *Developmental Biology* **346**(1), 102–112.

Spitz, F. & Furlong, E. E. M. (2012), 'Transcription factors: from enhancer binding to developmental control', *Nature Reviews Genetics* **13**(9), 613–626.

Srinivasan, S., Rashka, K. E. & Bier, E. (2002), 'Creation of a Sog morphogen gradient in the Drosophila embryo.', *Developmental Cell* **2**(1), 91–101.

St Johnston, D. & Nüsslein-Volhard, C. (1992), 'The origin of pattern and polarity in the Drosophila embryo.', *Cell* **68**(2), 201–19.

St Johnston, R. D., Hoffmann, F. M., Blackman, R. K., Segal, D., Grimaila, R., Padgett, R. W., Irick, H. A. & Gelbart, W. M. (1990), 'Molecular organization of the decapentaplegic gene in Drosophila melanogaster.', *Genes & Development* **4**(7), 1114–1127.

Stapel, L. C., Zechner, C. & Vastenhouw, N. L. (2017), 'Uniform gene expression in embryos is achieved by temporal averaging of transcription noise.', *Genes & Development* **31**(16), 1635–1640.

Stathopoulos, A., Van Drenth, M., Erives, A., Markstein, M. & Levine, M. (2002), 'Whole-genome analysis of dorsal-ventral patterning in the Drosophila embryo.', *Cell* **111**(5), 687–701.

Staudt, N., Fellert, S., Chung, H.-R., Jäckle, H. & Vorbrüggen, G. (2006), 'Mutations of the Drosophila Zinc Finger-encoding Gene vielfältig Impair Mitotic Cell Divisions and Cause Improper Chromosome Segregation', *Molecular Biology of the Cell* **17**(5), 2356–2365.

Stein, D. & Nüsslein-Volhard, C. (1992), 'Multiple extracellular activities in Drosophila egg perivitelline fluid are required for establishment of embryonic dorsal-ventral polarity.', *Cell* **68**(3), 429–40.

Steward, R. (1989), 'Relocalization of the dorsal protein from the cytoplasm to the nucleus correlates with its function.', *Cell* **59**(6), 1179–88.

Struhl, G., Struhl, K. & Macdonald, P. M. (1989), The Gradient Morphogen bicoid Is a Concentration-Dependent Transcriptional Activator, Technical report.

Stumpff, J., Duncan, T., Homola, E., Campbell, S. D. & Su, T. T. (2004), 'Drosophila Wee1 Kinase Regulates Cdk1 and Mitotic Entry during Embryogenesis', *Current Biology* **14**(23), 2143–2148.

Sun, Y., Nien, C.-Y., Chen, K., Liu, H.-Y., Johnston, J., Zeitlinger, J. & Rushlow, C. (2015), 'Zelda overcomes the high intrinsic nucleosome barrier at enhancers during Drosophila zygotic genome activation.', *Genome Research* **25**(11), 1703–14.

Tadros, W., Goldman, A. L., Babak, T., Menzies, F., Vardy, L., Orr-Weaver, T., Hughes, T. R., Westwood, J. T., Smibert, C. A. & Lipshitz, H. D. (2007), 'SMAUG Is a Major Regulator of Maternal mRNA Destabilization in Drosophila and Its Translation Is Activated by the PAN GU Kinase', *Developmental Cell* **12**(1), 143–155.

Tadros, W. & Lipshitz, H. D. (2009), 'The maternal-to-zygotic transition: a play in two acts', *Development* **136**(18), 3033–3042.

Tang, H., Liu, Y., Madabusi, L. & Gilmour, D. S. (2000), 'Promoter-proximal pausing on the hsp70 promoter in Drosophila melanogaster depends on the upstream regulator.', *Molecular and Cellular Biology* **20**(7), 2569–80.

Taniguchi, Y., Choi, P. J., Li, G.-W., Chen, H., Babu, M., Hearn, J., Emili, A. & Xie, X. S. (2010), 'Quantifying E. coli proteome and transcriptome with single-molecule sensitivity in single cells.', *Science* **329**(5991), 533–8.

Tanimoto, H., Itoh, S., ten Dijke, P. & Tabata, T. (2000), 'Hedgehog Creates a Gradient of DPP Activity in Drosophila Wing Imaginal Discs', *Molecular Cell* **5**(1), 59–71.

Tantale, K., Mueller, F., Kozulic-Pirher, A., Lesne, A., Victor, J.-M., Robert, M.-C., Capozi, S., Chouaib, R., Bäcker, V., Mateos-Langerak, J., Darzacq, X., Zimmer, C., Basyuk, E. & Bertrand, E. (2016), 'A single-molecule view of transcription reveals convoys of RNA polymerases and multi-scale bursting', *Nature Communications* **7**(1), 12248.

Tatei, K., Cai, H., Ip, Y. T. & Levine, M. (1995), 'Race: a Drosophila homologue of the angiotensin converting enzyme.', *Mechanisms of Development* **51**(2-3), 157–68.

Tautz, D., Lehmann, R., Schnürch, H., Schuh, R., Seifert, E., Kienlin, A., Jones, K. & Jäckle, H. (1987), 'Finger protein of novel structure encoded by hunchback, a second member of the gap class of Drosophila segmentation genes', *Nature* **327**(6121), 383–389.

Thattai, M. & van Oudenaarden, A. (2001), 'Intrinsic noise in gene regulatory networks.', *Proceedings of the National Academy of Sciences of the United States of America* **98**(15), 8614–9.

Thomas, M. C. & Chiang, C.-M. (2006), 'The General Transcription Machinery and General Cofactors', *Critical Reviews in Biochemistry and Molecular Biology* **41**(3), 105–178.

Thomsen, S., Anders, S., Janga, S. C., Huber, W. & Alonso, C. R. (2010), 'Genome-wide analysis of mRNA decay patterns during early Drosophila development', *Genome Biology* **11**(9), R93.

Thor, S. & Thomas, J. B. (1997), 'The Drosophila islet Gene Governs Axon Pathfinding and Neurotransmitter Identity', *Neuron* **18**(3), 397–409.

To, T.-L. & Maheshri, N. (2010), 'Noise can induce bimodality in positive transcriptional feedback loops without bistability.', *Science* **327**(5969), 1142–5.

Tsanov, N., Samacoits, A., Chouaib, R., Traboulsi, A.-M., Gostan, T., Weber, C., Zimmer, C., Zibara, K., Walter, T., Peter, M., Bertrand, E. & Mueller, F. (2016), 'smiFISH and FISH-quant âĂŞ a flexible single RNA detection approach with super-resolution capability', *Nucleic Acids Research* **44**(22), e165–e165.

Turing, A. M. (1952), 'The Chemical Basis of Morphogenesis', *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences* **237**(641), 37–72.

Tutucci, E., Vera, M., Biswas, J., Garcia, J., Parker, R. & Singer, R. H. (2018), 'An improved MS2 system for accurate reporting of the mRNA life cycle', *Nature Methods* **15**(1), 81–89.

Umulis, D. M., Shimmi, O., O'Connor, M. B. & Othmer, H. G. (2010), 'Organism-scale modeling of early Drosophila patterning via bone morphogenetic proteins.', *Developmental Cell* **18**(2), 260–74.

Urbinati, C. R. & Long, R. M. (2011), 'Techniques for following the movement of single RNAs in living cells', *Wiley Interdisciplinary Reviews: RNA* **2**(4), 601–609.

Urist, M. R. (1965), 'Bone: formation by autoinduction.', *Science* **150**(3698), 893–9.

Valegård, K., Murray, J. B., Stonehouse, N. J., van den Worm, S., Stockley, P. G. & Liljas, L. (1997), 'The three-dimensional structures of two complexes between recombinant MS2 capsids and RNA operator fragments reveal sequence-specific protein-RNA interactions', *Journal of Molecular Biology* **270**(5), 724–738.

Vastenhouw, N. L., Cao, W. X. & Lipshitz, H. D. (2019), 'The maternal-to-zygotic transition revisited', *Development* **146**(11).

Vattem, K. M. & Wek, R. C. (2004), 'Reinitiation involving upstream ORFs regulates ATF4 mRNA translation in mammalian cells.', *Proceedings of the National Academy of Sciences of the United States of America* **101**(31), 11269–74.

Wade, J. T. & Struhl, K. (2008), 'The transition from transcriptional initiation to elongation.', *Current Opinion in Genetics & Development* **18**(2), 130–6.

Waite, K. A. & Eng, C. (2003), 'From developmental disorder to heritable cancer: it's all in the BMP/TGF-$\beta$ family', *Nature Reviews Genetics* **4**(10), 763–773.

Walters, M. C., Fiering, S., Eidemiller, J., Magis, W., Groudine, M. & Martin, D. I. (1995), 'Enhancers increase the probability but not the level of gene expression', *Proceedings of the National Academy of Sciences* **92**(15), 7125–7129.

Wang, F., Flanagan, J., Su, N., Wang, L.-C., Bui, S., Nielson, A., Wu, X., Vo, H.-T., Ma, X.-J. & Luo, Y. (2012), 'RNAscope', *The Journal of Molecular Diagnostics* **14**(1), 22–29.

Wang, R. N., Green, J., Wang, Z., Deng, Y., Qiao, M., Peabody, M., Zhang, Q., Ye, J., Yan, Z., Denduluri, S., Idowu, O., Li, M., Shen, C., Hu, A., Haydon, R. C., Kang, R., Mok, J., Lee, M. J., Luu, H. L. & Shi, L. L. (2014), 'Bone Morphogenetic Protein (BMP) signaling in development and human diseases', *Genes & Diseases* **1**(1), 87–105.

Wang, S. N., Lapage, J. & Hirschberg, R. (2001), 'Loss of tubular bone morphogenetic protein-7 in diabetic nephropathy.', *Journal of the American Society of Nephrology* **12**(11), 2392–9.

Wang, X., Harris, R. E., Bayston, L. J. & Ashe, H. L. (2008), 'Type IV collagens regulate BMP signalling in Drosophila', *Nature* **455**(7209), 72–77.

Wang, Y.-C. & Ferguson, E. L. (2005), 'Spatial bistability of DppâĂŞreceptor interactions during Drosophila dorsalâĂŞventral patterning', *Nature* **434**(7030), 229–234.

Wang, Y., Ni, T., Wang, W. & Liu, F. (2019), 'Gene transcription in bursting: a unified mode for realizing accuracy and stochasticity', *Biological Reviews* **94**(1), 248–258.

Wangler, M. F., Yamamoto, S. & Bellen, H. J. (2015), 'Fruit Flies in Biomedical Research', *Genetics* **199**(3), 639.

Weil, T. T., Parton, R. M. & Davis, I. (2010), 'Making the message clear: visualizing mRNA localization', *Trends in Cell Biology* **20**(7), 380–390.

Wharton, K., Ray, R. & Gelbart, W. (1993), 'An activity gradient of decapentaplegic is necessary for the specification of dorsal pattern elements in the Drosophila embryo', *Development* **117**(2).

Whyte, W. A., Orlando, D. A., Hnisz, D., Abraham, B. J., Lin, C. Y., Kagey, M. H., Rahl, P. B., Lee, T. I. & Young, R. A. (2013), 'Master transcription factors and mediator establish super-enhancers at key cell identity genes.', *Cell* **153**(2), 307–19.

Wieschaus, E., Nüsslein-Volhard, C. & Jürgens, G. (1984), 'Mutations affecting the pattern of the larval cuticle in Drosophila melanogaster: III. Zygotic loci on the X-chromosome and fourth chromosome.', *Wilhelm Roux's Archives of Developmental Biology* **193**(5), 296–307.

Wilcockson, S. G. & Ashe, H. L. (2019), 'Drosophila Ovarian Germline Stem Cell Cytocensor Projections Dynamically Receive and Attenuate BMP Signaling', *Developmental Cell* **50**(3), 296–312.

Wilk, R., Reed, B. H., Tepass, U. & Lipshitz, H. D. (2000), 'The hindsight Gene Is Required for Epithelial Maintenance and Differentiation of the Tracheal System in Drosophila', *Developmental Biology* **219**(2), 183–196.

Winstanley, J., Sawala, A., Baldock, C. & Ashe, H. L. (2015), 'Synthetic enzyme-substrate tethering obviates the Tolloid-ECM interaction during Drosophila BMP gradient formation.', *eLife* **4**.

Wittkopp, P. J., True, J. R. & Carroll, S. B. (2002), 'Reciprocal functions of the Drosophila yellow and ebony proteins in the development and evolution of pigment patterns.', *Development (Cambridge, England)* **129**(8), 1849–58.

Wolpert, L. (1969), 'Positional information and the spatial pattern of cellular differentiation', *Journal of Theoretical Biology* **25**(1), 1–47.

Wood, W., Jacinto, A., Grose, R., Woolner, S., Gale, J., Wilson, C. & Martin, P. (2002), 'Wound healing recapitulates morphogenesis in Drosophila embryos', *Nature Cell Biology* **4**(11), 907.

Wozney, J. M., Rosen, V., Celeste, A. J., Mitsock, L. M., Whitters, M. J., Kriz, R. W., Hewick, R. M. & Wang, E. A. (1988), 'Novel regulators of bone formation: molecular clones and activities.', *Science* **242**(4885), 1528–34.

Wu, B., Chao, J. & Singer, R. (2012), 'Fluorescence Fluctuation Spectroscopy Enables Quantitative Imaging of Single mRNAs in Living Cells', *Biophysical Journal* **102**(12), 2936–2944.

Xie, T. & Spradling, A. C. (1998), 'decapentaplegic Is Essential for the Maintenance and Division of Germline Stem Cells in the Drosophila Ovary', *Cell* **94**(2), 251–260.

Xu, M., Kirov, N. & Rushlow, C. (2005), 'Peak levels of BMP in the Drosophila embryo control target genes by a feed-forward mechanism', *Development* **132**(7), 1637–1647.

Yamaguchi, Y., Inukai, N., Narita, T., Wada, T. & Handa, H. (2002), 'Evidence that negative elongation factor represses transcription elongation through binding to a DRB sensitivity-inducing factor/RNA polymerase II complex and RNA.', *Molecular and Cellular Biology* **22**(9), 2918–27.

Yao, J., Ardehali, M. B., Fecko, C. J., Webb, W. W. & Lis, J. T. (2007), 'Intranuclear Distribution and Local Dynamics of RNA Polymerase II during Transcription Activation', *Molecular Cell* **28**(6), 978–990.

Yip, M. L., Lamka, M. L. & Lipshitz, H. D. (1997), 'Control of germ-band retraction in Drosophila by the zinc-finger protein Hindsight.', *Development* **124**(11), 2129–41.

Yokoshi, M. & Fukaya, T. (2019), 'Dynamics of transcriptional enhancers and chromosome topology in gene regulation', *Development, Growth & Differentiation* **61**(5), 343–352.

Yu, K., Sturtevant, M., Biehs, B., Francois, V., Padgett, R., Blackman, R. & Bier, E. (1996), 'The Drosophila decapentaplegic and short gastrulation genes function antagonistically during adult wing vein development', *Development* **122**(12).

Zalokar, M. & Erk, I. (1976), 'Division and migration of nuclei during early embryogenesis of Drosophila melanogaster.', *J. Microsc. Biol. Cell.* **23**, 97–106.

Zeitlinger, J., Zinzen, R. P., Stark, A., Kellis, M., Zhang, H., Young, R. A. & Levine, M. (2007), 'Whole-genome ChIP-chip analysis of Dorsal, Twist, and Snail suggests integration of diverse patterning processes in the Drosophila embryo.', *Genes & development* **21**(4), 385–90.

Zenklusen, D., Larson, D. R. & Singer, R. H. (2008), 'Single-RNA counting reveals alternative modes of gene expression in yeast', *Nature Structural & Molecular Biology* **15**(12), 1263–1271.

Zhang, H., Levine, M. & Ashe, H. L. (2001), 'Brinker is a sequence-specific transcriptional repressor in the Drosophila embryo.', *Genes & Development* **15**(3), 261–6.

Zhao, M., Yang, X., Fu, Y., Wang, H., Ning, Y., Yan, J., Chen, Y.-G. & Wang, G. (2013), 'Mediator MED15 modulates transforming growth factor beta (TGF$\beta$)/Smad signaling and breast cancer cell metastasis', *Journal of Molecular Cell Biology* **5**(1), 57–60.

Zoller, B., Little, S. C. & Gregor, T. (2018), 'Diverse Spatial Expression Patterns Emerge from Unified Kinetics of Transcriptional Bursting', *Cell* **175**(3), 835–847.e25.

# Appendices

# A | Extended Materials and Methods

**Table A.1: Primer sequences and restriction sites used to generate *tup* enhancer transgenic fly lines.**

| Name | Primer sequence 5'-3' | Restriction site |
|---|---|---|
| Backbone fwd | gcGCATGCggtccacgggactggcgtcgtgatg | SphI |
| Backbone rev | gcGCATGCtaataaccgggcaggccatgtctgc | SphI |
| *yellow* fwd | ccagtcccgtggaccGCATGCtgttccaggacaaagggtgg | SphI |
| *yellow* rev | cctgcccggttattAGCATGCgcatacttacatttttccg | SphI |
| Proximal enh fwd | cggctcgagggtaccTCTAGAaaccaactccactcaatgtcaagtgg | XbaI |
| Proximal enh rev | tgaattcggcgcgccTCTAGAaatgcctctctttccgtctggccgt | XbaI |
| *tup* sh-enh 1 fwd | ggtgaattcgttaacAGATCTccctctgctccgccgatcttcgcct | BglII |
| *tup* sh-enh 1 rev | tcgagccgcggccgcAGATCTttatacgcttttaattagcgctaat | BglII |
| *tup* sh-enh 2 fwd | ttcgcaggcctcgcACCGGTtagaaattttcccagatatctgttt | AgeI |
| *tup* sh-enh 1 rev | tctgttaacgaattcACCGGTagaggggtaagagatgatgcact | AgeI |

**Table A.2: Primer sequences and restriction sites used to generate *ush* enhancer transgenic fly lines.**

| Name | Primer sequence 5'-3' | Restriction site |
|---|---|---|
| Proximal enh fwd | cagggtacctctagaGGCGCGCCtggccatgtggatgcgtctggtggc | AscI |
| Proximal enh rev | cgcactagtgaattcGGCGCGCCaaaccaaccaaccaaccaaccagcc | AscI |
| *ush sh-enh 1* fwd | cggctcgagggtaccTCTAGAttacatgcatgtgtgtaaaatcag | XbaI |
| *ush sh-enh 1* rev | tgaattcggcgcgccTCTAGAtatctgggagtctcgactgtgact | XbaI |
| *ush sh-enh 2* fwd | tatgctagcggatccAAGCTTactagcacgccataccccacaccag | HindIII |
| *ush sh-enh 2* rev | ggtgcgaggcctgcgAAGCTTacgaagcaaagacaaaaacaaaaa | HindIII |

**Table A.3: Primer sequences used to generate promoter swap lines.**

| Gene name | Sense primer 5'-3' | Antisense primer 5'-3' | Product size (bps) |
|---|---|---|---|
| *hnt* promoter | tcgcatcgccgctcggtctttt catttcaaattttacgctc | cttttgagttggtttcttgaagc | 202 |
| *ush* upstream region | ttctccggagcggccgcctccg gcacaatatcttatttcag | accgagcggcgatgcgatga | 190 |
| *ush* downstream region | agaaaccaactcaaaagcgtt cgtcataacgagaatttg | gtaggatctgcggccgccgtg ggaattttctttgactc | 67 |
| *ush* promoter | tcttctctcttttctctttggttcgctc ggtatctg | tgcggtatcgtattgtagctcgaagc | 200 |
| *hnt* upstream region | ttctccggagcggccgctgtaa tcctatttgcgcca | agagaaaagagagaagagagaagc | 145 |
| *hnt* downstream region | tacaatacgataccgcacaga caaatcaaataactaacttg | gtaggatctgcggccgccgac ggcattccaactcc | 360 |

**Table A.4: smFISH probes targeting *MS2* RNA sequences.** 4 sequence specific probes conjugated to Quazar 670 fluorophores.

| Name | Probe sequence | Name | Probe sequence |
|---|---|---|---|
| MS2 1 | tcgtgctttcttggcaataa | MS2 3 | aatactggagcgacgcgtga |
| MS2 2 | cgtttgaagattcgacctgg | MS2 4 | accgtaggatctgatgaacc |

**Table A.5: smFISH probes targeting *hnt* RNA sequences.** 48 gene specific probes targeting exonic regions and conjugated to Quazar 570 fluorophores.

| Name | Probe sequence | Name | Probe sequence |
|---|---|---|---|
| hnt 1 | aatggcgaattttgcgcttg | hnt 25 | cggaatagctgctgcatata |
| hnt 2 | tagtccatcacaatggatgc | hnt 26 | gtctgggactggaacatgag |
| hnt 3 | gatagcaccttggagcaaat | hnt 27 | ggtggtgccataaagggaaa |
| hnt 4 | gcaacttgatgctgtttgtg | hnt 28 | gattgatacggtttggtgga |
| hnt 5 | agattgttgttagccttgtg | hnt 29 | aggactccattcttgatgac |
| hnt 6 | atgttgttgttgatggtggt | hnt 30 | agcagtgttcacaggcaaag |
| hnt 7 | ccaaaagagtggcgacatcg | hnt 31 | cgaacgcagcgtaaatctcg |
| hnt 8 | gacaacttaggagcaaggca | hnt 32 | ttgtgcatagaactgcggat |
| hnt 9 | tggtatgtctgagtgcattg | hnt 33 | cgaaatgggagcatgaccat |
| hnt 10 | gacgatgggtggcaaagatc | hnt 34 | caacagatcggtgttcttgt |
| hnt 11 | gtaagcgacaacagtgcgac | hnt 35 | ccaaagtgcaggagaggatt |
| hnt 12 | tgtcttgtgcttcagataca | hnt 36 | ctcatgatcatgctcgtttt |
| hnt 13 | ttcaggatgctctggatatc | hnt 37 | ggctgctcatctagaatcaa |
| hnt 14 | aaactggaactggagctggt | hnt 38 | cttgagcgatcatcgttttc |
| hnt 15 | ttcacggactgctcaaagtt | hnt 39 | acttctttgagggattttcc |
| hnt 16 | tgtactgactcgaattgggc | hnt 40 | atccttgaacggtgaactgg |
| hnt 17 | ggcacgaaggttcgggaaaa | hnt 41 | tggcaatacggacaggagac |
| hnt 18 | aaatgtacccgattgtgacc | hnt 42 | tcttggtggtgaacaacagc |
| hnt 19 | aaacggcgtagggacacatg | hnt 43 | cagcgactcagtttcaatgg |
| hnt 20 | tgaaggcatagttgcacacg | hnt 44 | gcaagcagaaggcacagata |
| hnt 21 | ctcagatgacgttcacagtt | hnt 45 | tgtgcttcagtgtgaacttc |
| hnt 22 | gtgggactgatcgaacggaa | hnt 46 | tcactgatgcccaatagatt |
| hnt 23 | ctgggatttgctctcattta | hnt 47 | tactccagtatttatggctc |
| hnt 24 | gctgagatccaaaacatcca | hnt 48 | ctttggtatacgtaaggcgg |

**Table A.6: smFISH probes targeting *ush* transcript regions.** 48 gene specific probes probes targeting intronic and exonic regions and conjugated to Quazar 570 fluorophores.

| Name | Probe sequence | Name | Probe sequence |
|---|---|---|---|
| ush 1 | tctgcggtatcggaacaatc | ush 25 | gtccagggaagtgtacttaa |
| ush 2 | atctttggaatctctgctgt | ush 26 | ggggcaatagtagttctgat |
| ush 3 | ggaatcgttcaactgatcct | ush 27 | ggacaaggcagaccaatttc |
| ush 4 | cgaactcagcatcttcatca | ush 28 | acaggacacttgttcagact |
| ush 5 | tgaagatcgtgttcctgttc | ush 29 | aatgggttcgaatgtgggtg |
| ush 6 | acagggcaggcacataaatc | ush 30 | agatgcaggtcatgtatagc |
| ush 7 | ttgagggcgaactgaatgca | ush 31 | agggatgatccatggaatcg |
| ush 8 | tccgtgtccttaattctatg | ush 32 | aggtggagacatagttgcag |
| ush 9 | catgtggcgatttagggata | ush 33 | agtcacatcttgatctctgg |
| ush 10 | ctattccgttctggagaaga | ush 34 | accgttggatacatcggaat |
| ush 11 | cggtttgttgactagagcta | ush 35 | gagcagtacttcttcatcac |
| ush 12 | cttgatgttgttgaaccgga | ush 36 | caggtaggtcttcacgtaat |
| ush 13 | cgaactgcagtagtgttgtt | ush 37 | tgcagtagaactgcttgtga |
| ush 14 | aaaggtggatgcgaatgcgg | ush 38 | cgtgtaataccactcaagct |
| ush 15 | acaattatgatcgggtgggt | ush 39 | ggaggctaggattcgattag |
| ush 16 | tgggcctggaataaagctag | ush 40 | ttgcggtttggatagtgtac |
| ush 17 | cagattctctagttaaccct | ush 41 | ggagattttccgggaatact |
| ush 18 | aagctcagtgaattcagggt | ush 42 | tcggtaagtcgaaggagtca |
| ush 19 | cttccacatctaagagagca | ush 43 | tgtgttattcctttaggtgt |
| ush 20 | gagggagcgcaaacgatttg | ush 44 | gtaagctttggcatgcatta |
| ush 21 | actactagtttgggcaggaa | ush 45 | gccctcaatttaatttctgt |
| ush 22 | ccctttttcacatagatctg | ush 46 | cgcagaccattgcaaacttg |
| ush 23 | acaatattgcactccatgca | ush 47 | agaattgctcgctttatggg |
| ush 24 | tgccaagtagttctcatact | ush 48 | cttttattgtcgcacacact |

**Table A.7: smFISH probes targeting *tup* RNA sequences.** 46 gene specific probes targeting exonic regions and conjugated to Quazar 570 fluorophores.

| Name | Probe sequence | Name | Probe sequence |
|---|---|---|---|
| tup 1 | caatctctgccattaccat | tup 24 | gggtcgcttgtccctaata |
| tup 2 | tagctggtgagccaaatgg | tup 25 | cattgagcacggtccgaac |
| tup 3 | ttgtggttgtgattgtgca | tup 26 | tgagcgtatgcagctgttt |
| tup 4 | cgatggctgtaatccagtt | tup 27 | gcggattagcattgtagca |
| tup 5 | ccaaatccaggtgatggtt | tup 28 | tccttcatgagcgcatcag |
| tup 6 | tatcacatcgtgtccatgg | tup 29 | gatgactcgcggcgacagg |
| tup 7 | catccgacgcagtgagata | tup 30 | cgcttgttctggaaccata |
| tup 8 | tactgatcgtggatctggc | tup 31 | aatggtcttcttcttgtcc |
| tup 9 | actccagatcgggggcaac | tup 32 | ctgcatttgcagcttcatc |
| tup 10 | tcctggcacttcagacacg | tup 33 | aactggcgatcatggggat |
| tup 11 | tacagctttcgtccaggaa | tup 34 | cctgaagattcagtgggga |
| tup 12 | tgccatcgcgcacaaaaca | tup 35 | cggttgatatgtctgcaca |
| tup 13 | aatcacgcttgcagtaggt | tup 36 | aagtcgcttaaggctttcc |
| tup 14 | ccgcatttatcacattttg | tup 37 | gtcgagatcggcgtgaagg |
| tup 15 | tcatttttgctgaaggagt | tup 38 | ggtattgatcgctccattg |
| tup 16 | aagattttcgttttggccc | tup 39 | aactgctgaaatgcgggcg |
| tup 17 | ggagcatcgaaagcactcg | tup 40 | catcccgttcaggtcgtag |
| tup 18 | cagcaattgtcgcgcacac | tup 41 | cggatgcggcggcagaatg |
| tup 19 | gcatcgcgtaacgcgaatt | tup 42 | ctgctggttctgatggttg |
| tup 20 | tcctccttgcagtacaaag | tup 43 | gagtccaaactgctgcctc |
| tup 21 | acgatttctccagcacatc | tup 44 | atggtggtgcgaggtgatc |
| tup 22 | ccgacgaagaagtgaggct | tup 45 | gacgtaggagtcggtgctg |
| tup 23 | tttgtgtgagcccgattcg | tup 46 | gtcatcgctctccaggtag |

**Table A.8: smFISH probes targeting *yellow* RNA sequences.** 48 gene specific probes targeting exonic regions and conjugated to Quazar 670 fluorophores.

| Name | Probe sequence | Name | Probe sequence |
|---|---|---|---|
| yellow 1 | tccaactatatcgctcctga | yellow 25 | agtacagggtacgataacca |
| yellow 2 | gggaaagcaaagtccagctg | yellow 26 | cgatgacttgctaacggact |
| yellow 3 | ttggtcctttagtcgggtat | yellow 27 | aaaatcctcgtggatacggc |
| yellow 4 | tataatctccactagccaga | yellow 28 | catgatagctatcttccgtc |
| yellow 5 | aacaggtagagcattttgcg | yellow 29 | ccgttcatctaaggcaacaa |
| yellow 6 | cgattgccaaagtgttcgac | yellow 30 | cacgtgaagtggtatgggag |
| yellow 7 | aaactgcggtccatgtttat | yellow 31 | acagctcaattccatcatcg |
| yellow 8 | gccaatctggatacggaatt | yellow 32 | gagtacggcattgatgagtg |
| yellow 9 | caatctccagctgtatttga | yellow 33 | ccacaatgccatgaaattgc |
| yellow 10 | gtaggcagtggtaatactgt | yellow 34 | ttttcacatcggccggaaaa |
| yellow 11 | ccacactcatccactttaat | yellow 35 | acccaaacgtttttgttctc |
| yellow 12 | cacggattagtggtggtatt | yellow 36 | gcaagaaaacgggcatccta |
| yellow 13 | gtatccgtggtcaagtcaaa | yellow 37 | aagggagccgtgtaaattcg |
| yellow 14 | tagctcgtatctccgaattc | yellow 38 | aggcgttattcctcaaatca |
| yellow 15 | gtatttggatttgtgtccac | yellow 39 | acggcttgttttggtattga |
| yellow 16 | cacggcaatgttagctatga | yellow 40 | atataacggtggacccattg |
| yellow 17 | atcatcgcaatttttgccta | yellow 41 | tttctgtggcaagacaggac |
| yellow 18 | tatcccaattcatcggcaaa | yellow 42 | cgggcaaataagtgcgactt |
| yellow 19 | cccaggagtaagcaatcaag | yellow 43 | tggagactacattgcctgaa |
| yellow 20 | agaatctccaggacttgttc | yellow 44 | ggacccacagaatttgtaga |
| yellow 21 | cctcaatggatcggggaaaa | yellow 45 | ccgttgtgctggttgaaaat |
| yellow 22 | cccattggaagttaatacca | yellow 46 | gaccacttgtctcgtaattt |
| yellow 23 | ataccaaatataccctcctc | yellow 47 | gggttgatgggtgggaaata |
| yellow 24 | cgatcgaatgggcgaaaggg | yellow 48 | aaccttgatgctgatgatgc |

**Table A.9: smiFISH probes targeting *hnt* RNA sequences.** 48 gene specific probes targeting intronic regions. All probes are fused to the Z-flap primer sequence: CCAGCTTCTAGCATCCATGCCCTATAAG at the 5' end.

| Name | Probe sequence | Name | Probe sequence |
|---|---|---|---|
| hnt 1 | gttcacgcacaaatcacaga | hnt 25 | ttcattttggttggctgatt |
| hnt 2 | gcgctcgattatttatccaa | hnt 26 | tttgttgtccattggctttc |
| hnt 3 | acttttccaatttcatcctt | hnt 27 | atttctaccgataacgagcc |
| hnt 4 | tgttagcttagctgaagacc | hnt 28 | tgaaacctcagaggctgatc |
| hnt 5 | acactgctgcacttcaattg | hnt 29 | cccccgaaaggcaacaaaaa |
| hnt 6 | ttccttttgatttaatcgga | hnt 30 | gtggtagtggatcgatttga |
| hnt 7 | gggttttgggaatggacttg | hnt 31 | ctatgactatgagcggtgtc |
| hnt 8 | ttgacccaaaacaccaaggg | hnt 32 | cactcattagcatttaccgt |
| hnt 9 | atttttgacccggctaatt | hnt 33 | gtgtgggtgaaaaatgtggg |
| hnt 10 | gttttagcatcggtcatggg | hnt 34 | gctgctcttatcaagaattt |
| hnt 11 | cgtggtgtgtcaaagttcaa | hnt 35 | aattcaatacacatctggca |
| hnt 12 | ttttcggactggactgtctg | hnt 36 | ccgcagaaaggtgcgaagaa |
| hnt 13 | cctgaccacggaaatttctg | hnt 37 | aattggcgactgttgcaagt |
| hnt 14 | ctataaacgtcgcctcatca | hnt 38 | agttcctgtgaagagcgaat |
| hnt 15 | cccctgcgatattaaacaga | hnt 39 | aacagctgattcaggcaagt |
| hnt 16 | aaccgtttgacatggagcac | hnt 40 | gtcttcgattgtcgcgtaat |
| hnt 17 | cggaatcaagccgagtgaag | hnt 41 | aaaatcgatagtggggcacc |
| hnt 18 | caaatttctaatcgccatct | hnt 42 | aactgggtgtagggttcaat |
| hnt 19 | gaacacaacttggtcagtga | hnt 43 | atgatatccacactgacagc |
| hnt 20 | atgacccttgggaattatca | hnt 44 | gcttctttcgatattcct |
| hnt 21 | aattgttgggcggttttcac | hnt 45 | agtcctagcacatctgtc |
| hnt 22 | atatatacttttttggggcgg | hnt 46 | aacctacggctagctctt |
| hnt 23 | tgagataaagattccctccc | hnt 47 | aatcaggcgatcaggggt |
| hnt 24 | cgttgtccatgagtcatgaa | hnt 48 | ggatggtgggagaaacga |

**Table A.10: smiFISH probes targeting *ush* RNA sequences.** 48 gene specific probes targeting intronic regions. All probes are fused to the Z-flap primer sequence: CCAGCTTCTAGCATCCATGCCCTATAAG at the 5' end.

| Name | Probe sequence | Name | Probe sequence |
|---|---|---|---|
| ush 1 | attgctagtttgcttttctt | ush 25 | ccatttcgtttaagctgcaa |
| ush 2 | ttgtaattggcatcatcgcg | ush 26 | cagataaattcccatccact |
| ush 3 | agcaccactcataactttgc | ush 27 | aggagggcatgccaaaagaa |
| ush 4 | atcggctaaacgtggcttaa | ush 28 | taacccaactagccgattag |
| ush 5 | aggctctacgaaattgagcg | ush 29 | acaaacatcaaggggcggac |
| ush 6 | acatgacgtgcaggttgatt | ush 30 | tctgcaagataagccgaagc |
| ush 7 | attttcagtttttgcagtcc | ush 31 | atgcttcccgcataaacaag |
| ush 8 | actacagtgtcggacaagtc | ush 32 | gtcgatgtttttgttgttgc |
| ush 9 | cgaaatgcggacttggactt | ush 33 | tgaaatcccgacaccacaat |
| ush 10 | cgcagataggcgactgataa | ush 34 | ccaatatttcatcgcaaccg |
| ush 11 | aactgaggccgtggaatgaa | ush 35 | gctattttagctatccacac |
| ush 12 | atgcagatacatatgagccg | ush 36 | gactttccttcgacttgatg |
| ush 13 | gcacacacagatactcgtac | ush 37 | gggcatagaattcatttcca |
| ush 14 | tcaccgcagatcgtagatag | ush 38 | ggcaacttggagatgatgga |
| ush 15 | tccttcgttttaattcagct | ush 39 | acacgaaggcagaggatgtg |
| ush 16 | gctttgtgttgtttttctca | ush 40 | ggagccgaaaggaggcaaaa |
| ush 17 | aactacagtacccagagtac | ush 41 | taggccccgaaaaaagtgtg |
| ush 18 | gtttactacgcacagtagtt | ush 42 | agacacatcattacctacca |
| ush 19 | ccagtatctttgcatctttt | ush 43 | ggcttgcactaagttgaagt |
| ush 20 | cattgtttgcttttgtccac | ush 44 | aaataaaggaccccacagca |
| ush 21 | ttccagccaaatggttttac | ush 45 | attcccaaaaagaggggagc |
| ush 22 | tttatgcatagtccggacta | ush 46 | taatctgccatcgactgttg |
| ush 23 | gtagtttgaagcgaataccc | ush 47 | atcaaggaaaagaaggcccc |
| ush 24 | gaggtttgcttgttcatttc | ush 48 | gaactactgtctccgaaacc |

**Table A.11: smiFISH probes targeting *tup* RNA sequences.** 48 gene specific probes targeting intronic regions. All probes are fused to the X-flap primer sequence: CCTCCTAAGTTTCGAGCTGGACTCAGTG at the 5' end.

| Name | Probe sequence | Name | Probe sequence |
|---|---|---|---|
| tup 1 | gcatttgttaacacggcttt | tup 25 | cgaattgagatgcggccg |
| tup 2 | gttagacaactgtcagccaa | tup 26 | gggcaatcatcaggggttta |
| tup 3 | cgactacttgagaagttgcc | tup 27 | ggaaattagataatgcccccc |
| tup 4 | agcgatctcggacttacata | tup 28 | tttcgaatgagagttgggcc |
| tup 5 | agcattcggtgaatttgctg | tup 29 | tgcctagggctgagcaaata |
| tup 6 | tttcactgacgtttcagcag | tup 30 | caagcgcatgacaccgagtg |
| tup 7 | acaaacctagcgctatgact | tup 31 | tttgccgaacgctacaattg |
| tup 8 | accactgtcgtattgttttt | tup 32 | ctaacagaacagcggcctaa |
| tup 9 | gtttttatgtggcgaacgga | tup 33 | aattggctgggctaacaagc |
| tup 10 | attttcaagcacccactatt | tup 34 | ctttttgtattggcatcagc |
| tup 11 | atatttttctacagccgagc | tup 35 | ggaccagctgttgattagaa |
| tup 12 | gaaatgcaaaacgctcgcca | tup 36 | tgcacgtcatacatttagcg |
| tup 13 | aacgacgccattaaagtggc | tup 37 | cgatgcaaacacaccgattt |
| tup 14 | actcataaagccagcgacta | tup 38 | tgcatggcaaatagttgcga |
| tup 15 | agtatttgaagtcctcatcc | tup 39 | agctttgcacgtgttaatca |
| tup 16 | atttttggtcggattgtgc | tup 40 | catcaatcatccgagcgaac |
| tup 17 | agatgtttgagctggcttag | tup 41 | atcgatgagtgagccgagtg |
| tup 18 | ttaggcagctagaatgcttc | tup 42 | aaattcttttgtcagtgccg |
| tup 19 | gatctgcaaactcctcaact | tup 43 | cagagcttgtttatggctat |
| tup 20 | cgaaaaaaggcgggcagtcg | tup 44 | acacaatttcaaggggggtg |
| tup 21 | caaagagccgaagagctctg | tup 45 | acagccttgaatgagttcac |
| tup 22 | acagtttctcagtttctgtc | tup 46 | gaagatcccatgattggagc |
| tup 23 | gtacctgattcagatccaac | tup 47 | ggtagtatttcccaaatcga |
| tup 24 | acatttttgcattgctgtcg | tup 48 | accagaatcctctttgtttt |

**Table A.12: smiFISH probes targeting *yellow* RNA sequences.** 48 gene specific probes targeting intronic regions. All probes are fused to the X-flap primer sequence: CCTCCTAAGTTTCGAGCTGGACTCAGTG at the 5' end.

| Name | Probe sequence | Name | Probe sequence |
|---|---|---|---|
| yellow 1 | gggcttcatatttaacttcc | yellow 25 | aacagtaatcttgtcccgta |
| yellow 2 | gaatgtcccatttacgatct | yellow 26 | gaatcggatacaacttcccg |
| yellow 3 | gatatctctgatgccctaag | yellow 27 | gagctgagtcgatatcggaa |
| yellow 4 | catatttattcctctgccaa | yellow 28 | ggtggcctaatgaatctgta |
| yellow 5 | ttacatgagggtgctctcaa | yellow 29 | gctcaattgactggggtata |
| yellow 6 | aggtaccaaccggatttatg | yellow 30 | tttatgcgttcactcgttga |
| yellow 7 | catgctattggcttcgattt | yellow 31 | ttggaatgggccgcaaatat |
| yellow 8 | ggatttccatttctatgtgg | yellow 32 | ttttcgactgtcttgaggtt |
| yellow 9 | tgacagctaatcgtcggtta | yellow 33 | cctcgcaactgacagattta |
| yellow 10 | atttcgccactcttacattt | yellow 34 | ttgacatgtgctctttcagt |
| yellow 11 | gtgagtcacagaacaatgcc | yellow 35 | tgagcataatgcacctttca |
| yellow 12 | gcattcctagacgtactgaa | yellow 36 | gctttacaaagtaaccgggg |
| yellow 13 | ccaattgtaagccatgcata | yellow 37 | accgacccatcttcttaaaa |
| yellow 14 | ttaacatcgccgtgtttttc | yellow 38 | tgtttttatagcaggcgagt |
| yellow 15 | gtcagaagtttgggtgtttc | yellow 39 | tgcgtctgacatcattatca |
| yellow 16 | ccttgaagagcgtgtttttt | yellow 40 | atgtcattagccgtctgaag |
| yellow 17 | gaggtagtcagtcgatatct | yellow 41 | ttgcaagttgcgtgcaattt |
| yellow 18 | gggtttatttttgccgatt | yellow 42 | aaaagatccccggctatatg |
| yellow 19 | atgttttgtttgccgtaact | yellow 43 | taatggctctttttctgct |
| yellow 20 | ccattgaagcatgcctttaa | yellow 44 | taagcaatggtaactgccgt |
| yellow 21 | ctaccgatttgctttgacta | yellow 45 | atcgcttatgtttggcattt |
| yellow 22 | tattctgtagcctgttgttg | yellow 46 | ttatcgctgtctgttttctg |
| yellow 23 | aatcataccccataactgtc | yellow 47 | ctgcctaaattccttctttg |
| yellow 24 | cagcattggtggtcaatgat | yellow 48 | agaattagcagggccaaact |

# B | Extended Results

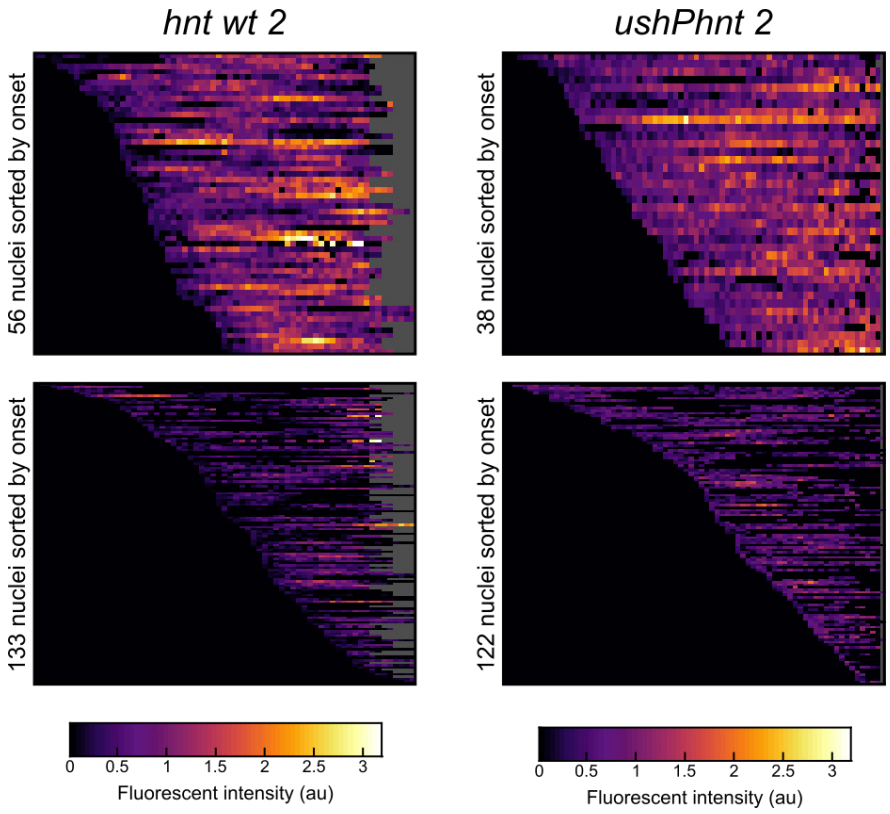



**Figure B.1: K-Means clustering heatmaps of *MS2-hnt* and *MS2-ushPhnt* embryos.** Single cell *MS-hnt* and *MS-ushPhnt* transcription profiles were grouped into two clusters using a K-Means algorithm. Heatmaps visualise transcriptional profiles of single-cell traces for individual replicates. Replicate 1 shown in main text. Traces are sorted according to transcriptional onset (scale as indicated, black = no expression, yellow = high expression; grey indicates periods where nuclei were not tracked due to temperature related lengths of time-lapses).

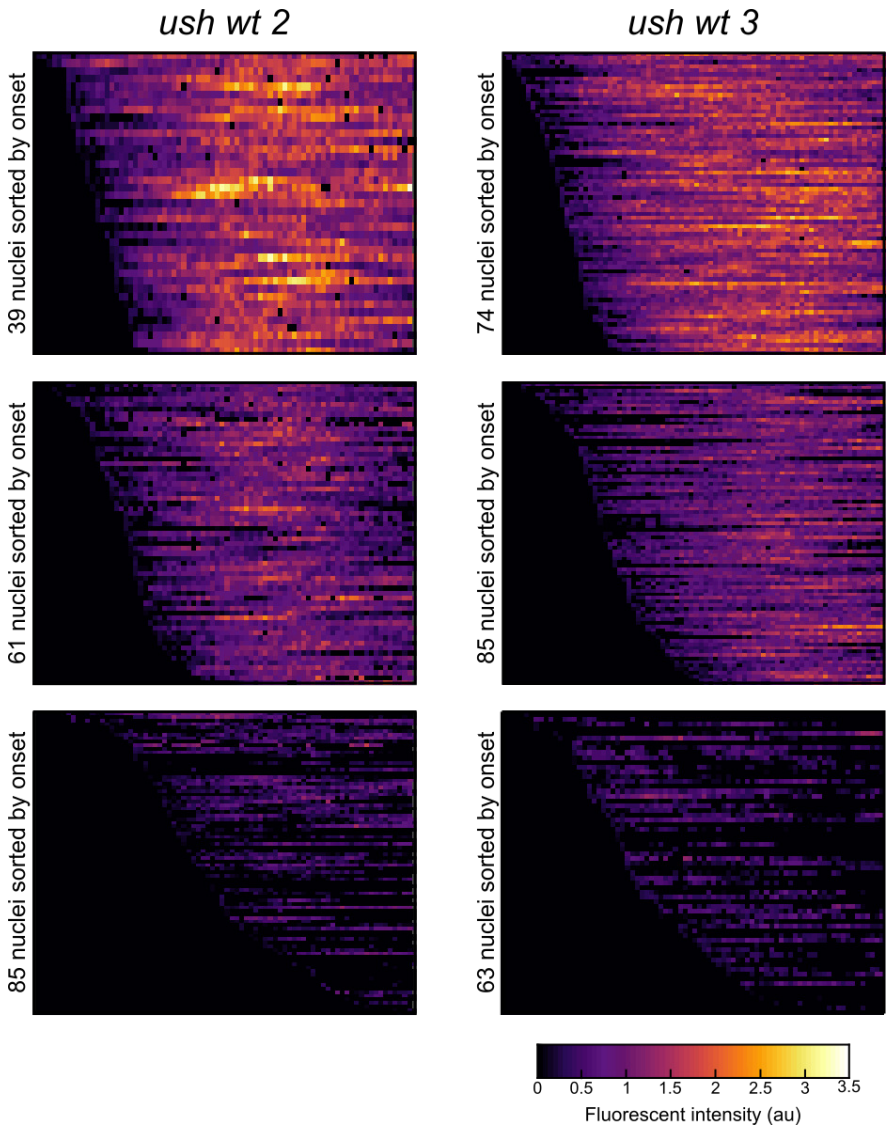



**Figure B.2: K-Means clustering heatmaps of *MS2-ush* embryos.** Single cell *ush* transcription profiles were grouped into three clusters using a K-Means algorithm. Heatmaps visualise transcriptional profiles of single-cell traces for individual replicates. Replicate 1 shown in main text. Traces are sorted according to transcriptional onset (scale as indicated, black = no expression, yellow = high expression; grey indicates periods where nuclei were not tracked due to temperature related lengths of time-lapses).

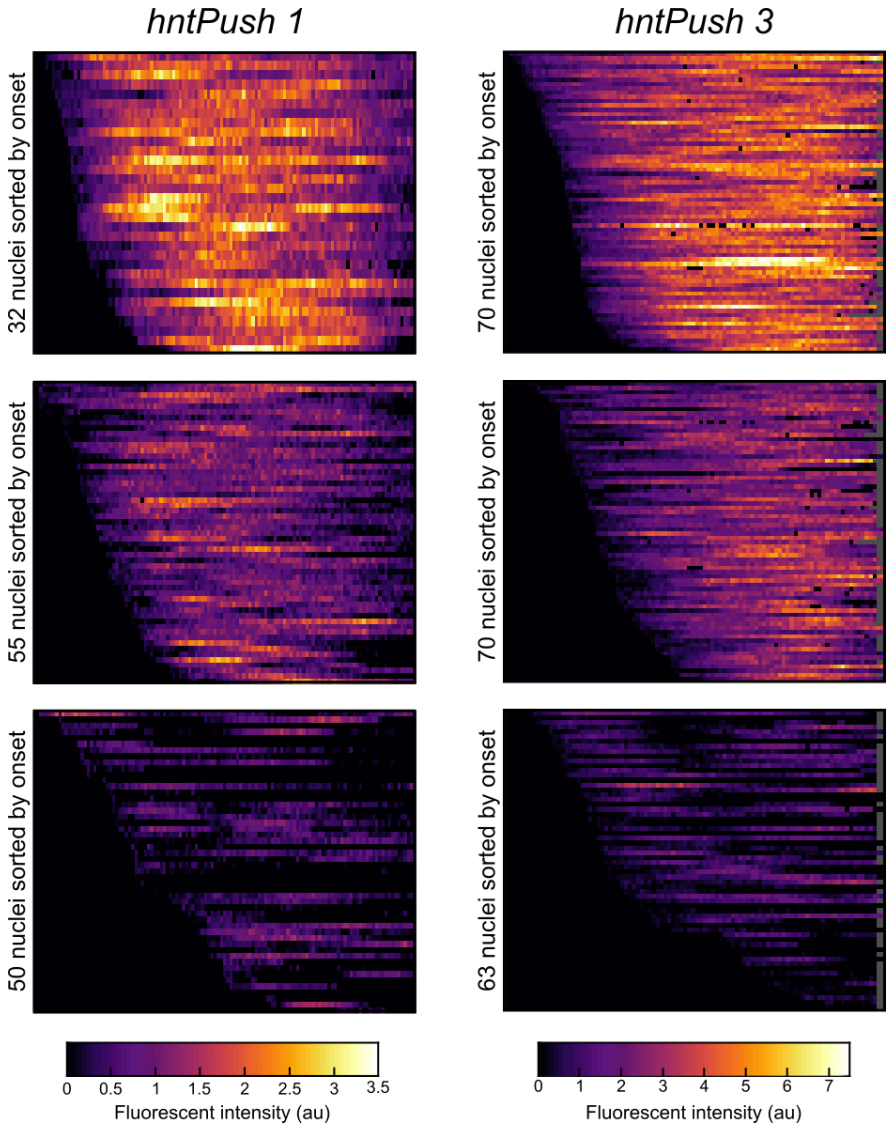



**Figure B.3: K-Means clustering heatmaps of *MS2-hntPush* embryos.** Single cell *ush* transcription profiles were grouped into three clusters using a K-Means algorithm. Heatmaps visualise transcriptional profiles of single-cell traces for individual replicates. Replicate 1 shown in main text. Traces are sorted according to transcriptional onset (scale as indicated, black = no expression, yellow = high expression; grey indicates periods where nuclei were not tracked due to temperature related lengths of time-lapses).