NORTHWESTERN UNIVERSITY

Analysis of Developmental Gene Regulation via Quantitative Imaging

A DISSERTATION

SUBMITTED TO THE GRADUATE SCHOOL

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS

For the degree

DOCTOR OF PHILOSOPHY

Field of Interdisciplinary Biological Sciences (IBiS)

By

Rachael Bakker

EVANSTON, ILLINOIS

December 2020

Abstract

Proper spatiotemporal expression of genes is essential during development. One method of regulation of signaling-responsive genes is at the level of transcription. In this work, I present the adaptation of single molecule fluorescent *in situ* hybridization for use in *Drosophila* imaginal disc tissues in order to more precisely quantify transcript levels in these tissues. I show the detection of nascent and mature mRNA molecules. I also present the development of robust automated image analysis in order to identify transcripts and localize them to the nearest nuclei in a 3D image volume. Using these methodologies, it is now possible to count individual transcripts in *Drosophila* imaginal disc tissues and to perform spatial analysis of gene expression using this method. Single-cell quantitative studies of transcription have revealed that transcription occurs intermittently, in bursts. I utilized smFISH in the wing imaginal disc in order to quantify mRNA of genes downstream of the evolutionarily conserved Wg and Dpp morphogen gradients. I compared these experimental results with predicted results from *in silico* modeling of transcription in order to predict outcomes when transcriptional burst parameters are varied. My results indicate that the transcription levels of these genes appear to share a common method of control by burst frequency modulation. Additionally, I utilized quantitative analysis of fluorescent proteins in the eye imaginal disc in order to explore the regulation of a key regulator in neuronal fate transition in this tissue, Yan. I show that in the absence of micro-RNA miR-7 regulation, Yan protein levels are mildly derepressed in undifferentiation precursor cells in the eye disc. This is consistent with the known role of miRNAs as weak repressors of gene expression during development. I also present evidence that slowing metabolism makes miR-7 repression of Yan unnecessary, supporting the hypothesis that weak repressors are required for gene expression during rapid growth.

Acknowledgements

I would like to acknowledge my adviser, Dr. Richard Carthew, for his valuable guidance throughout the years. He has been very supportive of my work and had a great deal of confidence in this project. I would also like to thank Dr. Madhav Mani for his advice and encouragement, particularly for the modeling and image analysis portions of my work. Additionally, I appreciate the guidance from my thesis committee: Dr. Robert Holmgren, Dr. Erik Andersen and Dr. Ravi Allada.

Developing smFISH was a very taxing task for a beginning graduate student. I would like to thank Dr. Shawn Little for his assistance in getting the technique working for me. I would also like to acknowledge Dr. Arjun Raj for his advice. Former Carthew lab members Dr. Nicolas Pelaez and Dr. Justin Cassidy both taught me confocal microscopy and provided valuable discussion. Dr. Sebastian Bernasek performed the modeling that inspired my experiments in Chapter V. Without his modeling and analysis, my Yan-YFP and miR-7 data may never have found its way in to any publication, let alone such a wonderful and impactful one.

I would also like to acknowledge all Carthew lab members and especially Dr. Ritika Giri, Dr. Pamela Agbu, Dr. Kevin Nyberg, Joe Nguyen, and Saya Dennis for their helpful advice and companionship in the lab. Undergraduate researcher William Finnegan was great to work with and taught me to be a better programmer. Undergraduate Mayur Popuri helped stress test smFISH, teaching me more about the parameters and limits of my methods. Dr. Sandra Siepka suggested I organize my things in lab in February 2020 because you "never know what might come up during your last trimester of pregnancy." I don't think she meant a global pandemic, but I find myself constantly grateful for this advice in light of what transpired in March.

I would lastly like to thank my wonderful husband, Chris Clow. I would not be able to do anything without his support. And a final acknowledgement to my daughter Ezri Clow, who came into being alongside this thesis.

Table of Contents

List of Figures and Tables

Chapter II: Single molecule fluorescent in situ hybridization: A method for detection of mRNA in Drosophila imaginal discs

Chapter III: Development of Image Analysis for Automated Detection of smFISH Data

Chapter IV: Wg and Dpp Morphogens Regulate Gene Expression by Modulating Frequency of Transcriptional Bursts

Chapter V: Effects of Metabolism on Gene Regulatory Networks

Chapter I: Introduction

Gene regulatory networks in developmental biology

Multicellular animals consist of many genetically identical cells performing specialized functions in order to contribute to the survival and reproduction of the whole. All of these cells descend from a single zygote cell at the moment of conception. Development is the process by which this single cell generates all cells of the body and forms the body plan of an entire organism. The ability to do so is genetically encoded in the nuclear DNA of the organism.

While an animal's complement of genes is hard-coded into their DNA, they do not operate in a vaccum. The function of a gene is entirely dependent on its interactions with other molecules present in the same cell. These interactions were largely determined by years of mutation analysis. Genes were disabled and their effects on expression of other genes were assessed. This is a powerful technique that allows us to draw representational interactions between genes. For example, gene A regulates gene B expression, or vice versa. In actuality, it is the molecular products of gene A that regulate the expression of gene B *in trans*. However, the limitations of this approach are that it cannot give us a systems-level understanding of how an animal is made, beyond a simple list of the gene required and their effects on one another (Alon, 2007; Levine and Davidson, 2005).

As more and more of these interactions were solved, one could see how context-dependent behavior arises. For example, gene interactions can create logical network behaviors, such as 'AND' gates or 'OR' gates. These network interactions can be assembled into complex behaviors and network architectures, much like a computer program (Alon, 2007).

The developmental process can thus be said to be controlled by a collection of gene regulatory networks (GRNs). Each GRN is a group of genes using network logic in order to

produce a functional outcome in the developing animal. During development, GRNs flow from the more general such as dividing the embryo in anterior versus posterior space, to highly specific, such as specifying individual cell types inside a tissue (Levine and Davidson, 2005; Peter and Davidson, 2016).

It has been found that a gene might be able to function in many different GRNs. How can this be? One solution is that its function is spatially and temporally restricted. As such, gene function may change depending on the expression of other genes in that region or time in development. Another solution is expression dynamics. Genes may pulse throughout time, or show other transient behaviors (Chubb et al., 2006; Peter and Davidson, 2016).

Genes are highly regulated in terms of when and where they are expressed in the animal. This regulation occurs via several avenues at each step of the central dogma of biology. The rate of production of mRNA copies from DNA is regulated by transcriptional regulation. Translation rates from mRNA into protein are regulated via ribosome activity and mRNA stability. Protein-protein interactions occur to regulate the stability, location, and catalytic activity of resulting proteins. However, on a global scale across all genes, transcriptional regulation alone can explain much of the variation in protein abundance in the cell (Lee et al., 2011; Liu et al., 2016; Xu et al., 2015). As such, transcriptional regulation is key to a full understanding of how expression of genes are regulated.

In order to understand and interrogate the behaviors of GRNs and transcriptional regulation, we as developmental biologists must develop tools for precise spatial quantification of gene expression. Not only is this important understand spatial restriction, but precise quantification of gene expression can reveal new mechanisms of gene expression.

Methods for the Detection of Transcription

The most basic understanding of transcriptional regulation requires investigation into which genes are transcribed when, where, and how much. One way to assay whether transcription is occurring is the detection of mature RNA transcripts. The earliest methods for detection of RNA involved hybridization of labeled oligonucleotides designed to be complementary to mRNAs of interest. Purified RNA material was run on a denaturing electrophoresis gel and detected with radiolabeled probes, in a northern blot (Alwine et al., 1977). Northern blotting allowed for quantitation of relative transcript levels, using radiolabeled standards as a reference. However, northern blots require homogenization of large quantities of tissue and early hybridization based techniques offered low sensitivity (VanGuilder et al., 2008). Transcripts present at very low numbers simply did not rise above the level of background noise in hybridization-based radiolabeled RNA detection techniques. Another drawback was that these methods were not precisely quantitative – signal did not scale linearly with mRNA abundance (Bartlett, 2002).

In the 1990s, the development of an RNAse protection assay offered improved sensitivity in RNA detection relative to using a northern blot. Similar to a northern blot, RNAs of interest are run on an electrophoresis gel and detected with radiolabeled probes. However, in an RNase protection assay, transcripts are hybridized with probes and digested with a single-strand specific RNase enzyme prior to electrophoresis and detection (Ma et al., 1996). Only hybridized RNAs escape degradation by the enzyme, therefore only RNAs of interest are detectable on the gel. The result was a 10-fold improved sensitivity of detection and linearity of signal to mRNA abundance. Using synthesized double-stranded DNA to create a standard curve allowed for precise quantification of RNAs of interest (Bartlett, 2002; VanGuilder et al., 2008).

In the early 2000s, even more quantitative detection of mRNA was made possible by reverse transcription quantitative polymerase chain reaction (RT-qPCR). RT-qPCR performs the polymerase chain reaction on complementary DNA (cDNA) reverse transcribed from mRNA strands harvested from samples of interest. Using fluorescence, the amount of PCR product can be monitored at the end of each cycle of PCR. It takes many rounds of PCR for a specific PCR product to become detectable in the reaction. The number of cycles it takes for each species to cross the threshold of detectability depends on the number of molecules of cDNA present in the original sample. By comparing small differences in amplification time in the PCR, one could detect small fold-change differences between genes of interest and reference cDNAs in the sample. As PCR theoretically only requires one molecule of cDNA to be present in the original sample for detection to be possible, this method is extremely sensitive. It is also generally less time-consuming and more high-throughput than gel-based detection methods, allowing more genes to be assayed at one time (VanGuilder et al., 2008).

In the 2000s and 2010s, RNA detection technology has been marked by an increase in scale and throughput. Microarrays and high-throughput RNA-sequencing (RNA-seq) have dramatically increased the number of genes that can be assayed for transcription at one time. Microarrays are based on hybridization of cDNA fragments to a chip or array of DNA oligos of known genes. RNA-seq involves sequencing cDNA libraries generated from RNA fragments with adaptors for sequencing. RNA-seq is generally less noisy than microarrays and is quantitative over a greater dynamic range of expression levels. However, RNA-seq presents challenges inherent in high-throughput sequencing approaches, in that it requires mapping the sequenced fragments back to a reference genome. This can be problematic in repetitive sequences of the genome (Lähnemann et al., 2020; Wang et al., 2009).

Both of these approaches massively increase the scale of RNA detection in a single sample, allowing discovery of new transcriptional regulation by measuring the changes in all genes in different treatments or experimental conditions. However, they also require the aggregation of many cells or tissues together, which can obscure spatial or individual heterogeneity in transcript levels (Vera et al., 2016).

Assays for Nascent Transcription

The detection of mRNA transcripts in a tissue is not always sufficient to answer questions about transcription. Though mRNAs have variable half-lives in general, mature transcript levels are a readout of transcriptional activity over relatively long time-scales of tens of minutes to hours (Larson et al., 2011)(Milo et al., 2010). In many contexts, we may seek to understand how much transcription is happening on a much smaller time-scale. To do so, we must turn to techniques that specifically detect nascent transcriptional activity, as opposed to mRNA transcripts in general.

Nuclear run-on assays were developed in the early 1980s to detect active transcription of a gene. Cells of interest were chilled and lysed to stop transcriptional activity, and the nuclei were isolated. Then, transcription was allowed to proceed *in vitro* using radiolabeled UTP as a substrate. No new transcription will be able to initialize, but polymerases already present on transcripts will continue transcription, incorporating radiolabeled nucleotide into the transcript. Transcripts of interest can then be hybridized with DNA probes and detected similar to an RNAse protection assay. The amount of radioactivity detected provided a quantitative readout of how many nascent transcripts were present on a gene (Smale, 2009).

In the 21st century, nuclear run-on assays have largely been replaced by chromatin immunoprecipitation (ChIP) using antibodies against the RNA polymerase II holoenzyme. This

isolates fragments of DNA bound to RNA polymerase II and therefore in the process of active transcription. These fragments can be detected and quantified using either qPCR or high-throughput sequencing. ChiP requires less starting material and shows improved sensitivity over nuclear run-on assays. However, like run-on assays, ChIP requires the homogenization of tissues, potentially losing spatial information about gene transcription (Sandoval et al., 2004).

High Resolution mRNA Assays: Combining Spatial Resolution and Quantitation in a Single Assay

Broadly speaking, the methods I have reviewed thus far either produce sensitive and quantitative estimates of transcript abundance and transcription with poor spatial resolution, or qualitative estimates of transcript abundance with high spatial resolution. If one wants both types of information, several techniques have been developed.

The sensitivity of *in situ* hybridization was improved such that single molecule mRNAs were visible and could be counted. I review the development of the methodology behind single molecule fluorescent *in situ* hybridization in Chapter II. In brief, increased availability of synthesized oligonucleotides enabled researchers to increase signal to background by using many short probes for hybridization instead of a single probe (Kwon, 2013). While, *in situ* hybridization can only be done on a few genes at a time, other techniques allow a transcriptome-widge view of mRNA abundance. Spatial RNA-sequencing involves barcoding mRNAs within a on a hydrogel according to their spatial location. Sequencing of the resulting library preserves the original spatial information of the entire transcriptome (Ståhl et al., 2016).

Additionally, groups have pioneered the use of single cell RNA (scRNA) sequencing. This requires the dissociation of cells prior to RNA sequencing. Each cell is suspended in a droplet and receives a unique barcode for the RNA contained within, prior to sequencing.

Therefore, the sequencing information can be reassembled into which cell it came from. This can produce extremely powerful data that can give us snapshots of entire transcriptional "states" of cells (Wang et al., 2009). However, each cell measurement has a high degree of uncertainty stemming from amplification bias and very low amounts of starting material. The process of dissociation itself may induce significant gene expression changes as well.

The field of reconstructing spatial resolution of cells from scRNA sequencing data has only just begun. Groups have used additional barcoding to finely preserve location information, reconstruction from known gene expression patterns, and the incorporation of spatial-RNA equencing data into cell location reconstruction. While these techniques are promising for producing a high degree of complete transcriptional information, this technique currently serves as a tool of discovery which need to be confirmed via other techniques (Lähnemann et al., 2020).

Imaginal Discs in Drosophila

The fruit fly *Drosophila melanogaster* has long been used as a valuable model organism due to its genetic tractability, relatively short generation time, and large number of progeny. They offer several experimental contexts for study of their tissues, including a set of epithelial structures known as imaginal discs. Imaginal discs are specialized organs set aside from 10-50 epidermal cells during embryogenesis. During the next four days of larval development, these groups of cells rapidly divide into a final size of approximately 10,000 to 50,000 cells, depending on the type of disc. Each disc is organized into a sac-like epithelial bilayer (Aldaz and Escudero, 2010). One layer consists of flat, squamous cells known as the peripodial membrane. The other layer, known as the disc proper, is made up of columnar epithelial cells. These cells differentiate and eventually give rise to the organs and structures of the adult *Drosophila* body. Epithelial cells of the disc proper are analogous in structure to epithelial cells making up human organs.

Each disc corresponds to the set of structures it will eventually produce in the adult body: eye, antenna, wing, leg, haltere, and genitals. During metamorphosis and pupation, discs evert and undergo structural changes necessary to develop the adult body (Aldaz and Escudero, 2010; Beira and Paro, 2016).

Imaginal Discs as a Model System

Imaginal discs are a model for organ formation and morphogenesis, as they undergo significant structural changes during the *Drosophila* life cycle. Imaginal discs have been used as model systems for various biological processes. Physical transplantation and wounding have contributed to studies of growth and size control in imaginal discs. The rapid growth provides an opportunity for the study of organ size control and control of cellular proliferation (Beira and Paro, 2016).

The development of clonal analysis allows for the creation of mitotic clones of mutant cells. This enables the study of mutant phenotypes on a cellular level, and allows observation of interactions between neighboring cells of mutant and wild-type genotypes (Wu and Luo, 2006). This has been used in many ways in the imaginal disc. Mutations may have different effects depending on their location in the disc and their particular signaling milieu, which has helped elucidate the role of genes in specifying compartments, and growth regulation. The partial introduction of tumorigenic mutations can create aggressive, cancer-like tumors in imaginal disc tissue which are being studied as a simple model for cancer (Mundorf and Uhlirova, 2016). Clonal patches of mutant cells can also create scenarios in which competitive cell growth has been studied, which also has application in the invasiveness and metastasis of tumor tissue in humans (Martín et al., 2009).

As all *Drosophila* external adult structures differentiate from the epithelia of the imaginal disc, they are a good model system for cell-cell signaling that gives rise to patterning of tissues. Initially, these structures were transplanted to different regions of the larval body, revealing the role of the *Antennapedia* complex and eventually leading to greater understanding of the function of Hox selector genes in the body plan (Schneuwly and Gehring, 1985). Transplantation experiments have continued to elucidate a finer understanding of cellular identity within the larger context of the organism as whole.

Restricting genetic manipulations to the disc structure itself allows the study of genetic mutations that would be lethal or too severe to study in an embryonic context. The Gal4-UAS system allows for tissue and pattern specific expression of genes of interest (Osterwalder et al., 2001, p. 4).

Phenotypes that manifest in the adult organs of the fly can be used as convenient readouts for defects in patterning that occur at the imaginal disc stage, allowing for rapid genetic screens and efficient epistasis experiments. This has been of particular note in the *Drosophila* eye, where even mild mutations can cause phenotypes that manifest as a disruption of the crystalline lattice of the compound eye by disrupting the recruitment and patterning of many different cell types. Such research led to the elucidation of the major components of EGF signaling and discovery of the function of Notch signaling (Kumar, 2001).

Long range signaling molecules are used extensively in the imaginal disc system. Hedgehog signaling is required to pattern anterior and posterior compartments of the discs, and wingless and decapentaplegic use long-range morphogen activity to control growth and development of the disc. Each of these signaling pathways in conserved in humans, and insights

about function of these pathways in the imaginal disc continues to provide insight into signaling and its role in development today (Kicheva et al., 2012; Wartlick et al., 2011).

Both the genetic tools available for *Drosophila* as a whole, and the ease of experimentation on imaginal discs, make them an intriguing model system for signaling going forward. The imaginal disc has proven to be an analogous signaling environment to human organ tissues and provides a generalizable context for understanding how signaling works during development and disease.

Chapter II: Single molecule fluorescent *in situ* hybridization: A method for detection of mRNA in *Drosophila* imaginal discs

Introduction

Imaging assays of gene expression offer unique advantages for the observation of biological processes. In developing tissues, they allow us to observe gene expression in the context of the morphology of the organism. In disease states, researchers can ascertain heterogeneity of gene expression in tumors. Assays that require the bulk preparation of tissue inherently lose any quantification of cellular, spatial, or individual heterogeneity of gene expression. In terms of observation of mRNA in tissues, *in situ* hybridization has been in use for over fifty years (Gall, 2016).

Initial methods for in situ hybridization employed long single-stranded DNA or RNA probes that were produced either by digestion of a DNA plasmid or *in vitro* transcription. Probes elicited a signal by associating with an enzyme to provide a colorimetric signal or fluor to provide a fluorescent signal (Levsky and Singer, 2003). By nature, these probes produced a background signal because a single nonspecific binding event and a single specific binding event of the probe introduces an equivalent amount of signal into images. While this challenge can be somewhat overcome by blocking with nonspecific nucleic acids, it nonetheless limits the sensitivity and resolution of traditional *in situ* hybridization assays (Gall, 2016). The addition of an enzymatic or fluorescent amplification step to generate signal makes the assays semi-quantitative. Thus, for many years, *in situ* hybridizations were not considered to be strongly quantitative, and researchers relied on other methods such as northern blots and RT-qPCR to obtain complementary quantitative results.

In 1998 the Singer lab invented single molecule fluorescent *in situ* hybridization (smFISH). It detected mRNAs using somewhat smaller probes (50 nucleotide-long oligos) that were each labeled with multiple fluorophores (Femino et al., 1998). However, these probes still

had many drawbacks. Synthesis and purification of multiply labeled probes was costly and difficult, producing partially labeled probes that introduced additional variability in fluorescence. In addition, multiply labeled oligonucleotides appear to be particularly prone to fluorescence quenching (Femino et al., 1998; Levsky and Singer, 2003). For these reasons, the method saw only limited use in the early 21st century.

In 2008, Raj and colleagues used smaller probes, which introduced smFISH to a wider audience (Raj and Tyagi, 2010; Raj and van Oudenaarden, 2008). This advance was facilitated by broader availability of inexpensive oligonucleotide synthesis. Large numbers of 20-nucleotide oligos could be individually labeled with a fluorophore, which was more cost effective. The use of many (at least 24 and up to 96) smaller probes showed an improved signal to background because a single nonspecific binding event contributes only marginally to background fluorescence. Many specific binding events are required to observe a signal. The annealing of the probes along the length of the mRNA molecule mitigates the effects of partial RNA degradation, occlusion by RNA binding proteins, and the possibility of prohibitive secondary structure. The loss of a few potential binding sites does not necessarily prevent detection of that RNA molecule (Raj and Tyagi, 2010).

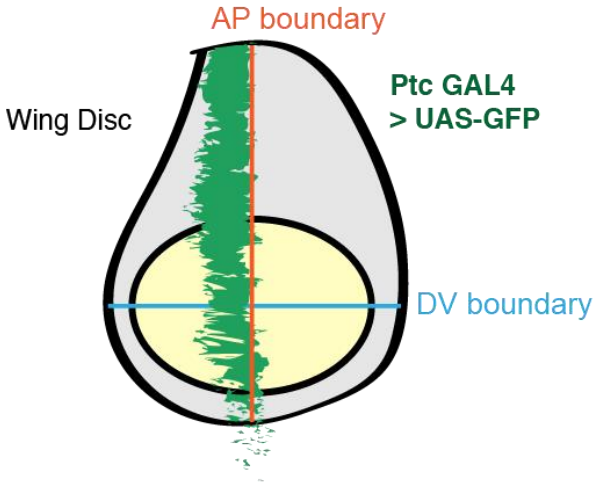
Here, I have adapted smFISH for use in the powerful model system of *Drosophila* imaginal discs.

Results

Initial Detection of GFP Transcripts via smFISH Using an Anti-GFP Probe Set

I performed a 'first pass' experiment in order to assay the viability of a smFISH methodology for *Drosophila* imaginal discs. I designed probes complementary to superfolder-GFP (sfGFP). These oligos were ordered pre-conjugated to a CalFluorRed dye from

A


AP boundary
Wing Disc
Ptc GAL4
> UAS-GFP
DV boundary

B

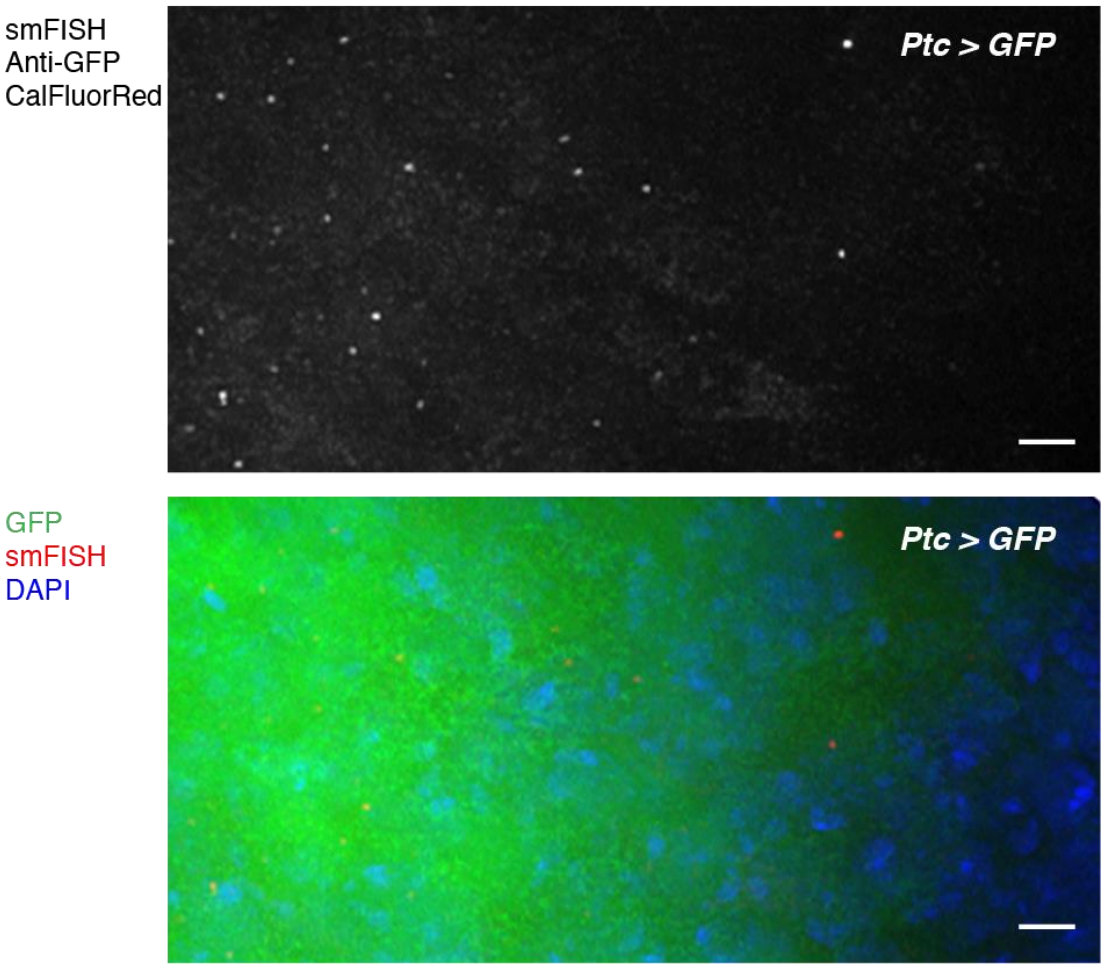
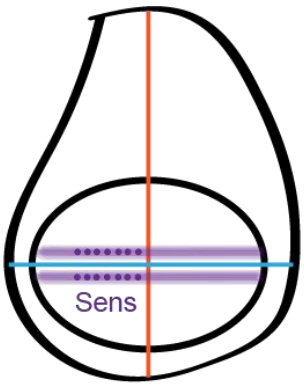

smFISH
Anti-GFP
CalFluorRed
Ptc > GFP
GFP
smFISH
DAPI
Ptc > GFP

Figure 1. Detection of *Patched-GAL4 > UAS-GFP* expression with smFISH. (A) Schematic of expected GFP expression domain in the wing imaginal disc. AP boundary represents the anterior-posterior compartment boundary, DV boundary represent dorsal-ventral wing boundary. **(B-C)** Widefield deconvoluted sections of imaginal discs expressing *Patched-GAL4 > UAS-GFP* . Nuclei are stained with DAPI. Scale bars= 5 µm. **(B)** GFP mRNA visualized with anti-GFP CalFluorRed smFISH probes. **(C)** GFP mRNA (red) visualized along with GFP protein (green) and nuclei stained with DAPI (blue).

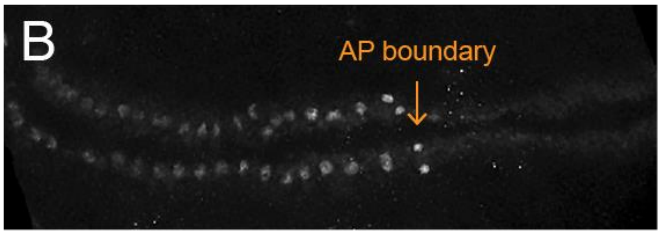
BiosearchTech (See methods for details). smFISH was performed according the protocol described in (Raj and Tyagi, 2010). Imaging was performed using a Deltavision wide-field microscope. Initial experiments used a *Patched-GAL4* driver crossed to a *UAS-GFP* line. This produces flies with wing discs expressing GFP very strongly in the *Patched* pattern, which is a stripe immediately anterior to the anterior-posterior compartment boundary in the wing disc (Figure 1A). I reasoned that such a high amount of GFP expression should produce some smFISH signal if any could be detected. As expected, I saw smFISH signal that colocalized with GFP signal (Figure 1B, C).

Given these results, I decided to proceed to using the sfGFP probes to detect expression from a *sfGFP-senseless* (*sfGFP-sens*) transgene (Giri et al., 2020). I anticipated that this would produce signal that was more likely to give individual spots for the following reasons: first, the probes were designed specifically for this version of GFP, and second, this transgene is expressed at far lower levels than the overexpression caused by the GAL4-UAS system. By confocal microscopy, GFP protein was visible in the expected *sens* expression pattern of two stripes immediately adjacent to the dorsal-ventral boundary of wing discs (Figure 2A-B). After smFISH was performed, the pattern was somewhat visible by widefield microscopy (Figure 2C), showing the approximate pattern. In this region, there were distinct punctae visible in the smFISH channel (Figure 2 D & E, red arrows). It should be noted that these punctae were not very bright and only visible after deconvolution using the Deltavision imaging software. As can be seen for the GFP channel (Fig 2C), my imaging setup, including deconvolution, produced fluorescent signal that tends towards a punctate appearance.

A



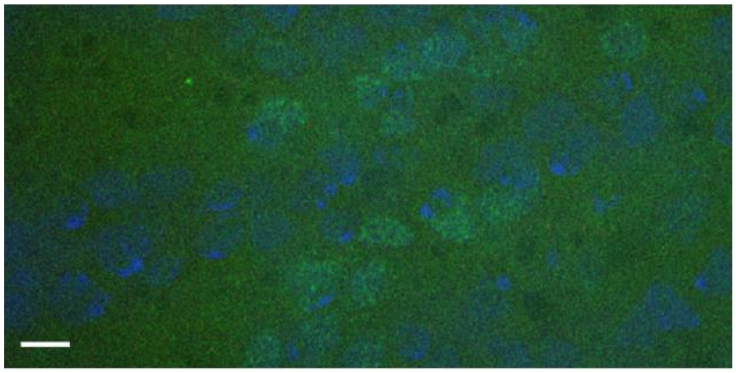
B



Sens:sfGFP Protein

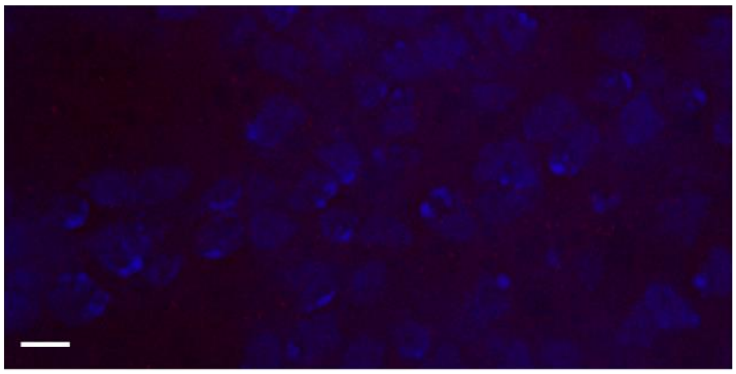
Sens protein

C



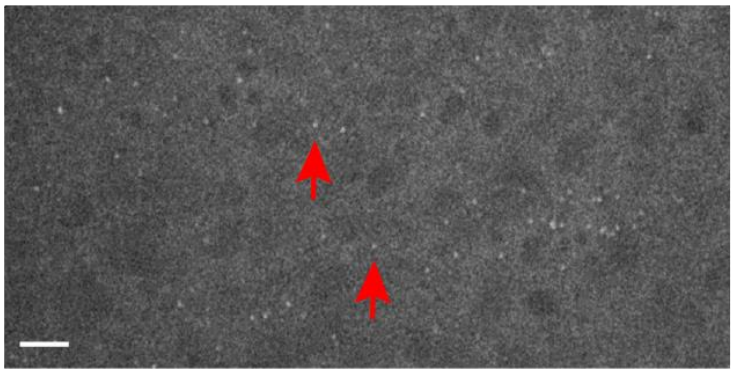
Sens:sfGFP Protein
DAPI

D



Anti-sfGFP
CalFluor Red smFISH
DAPI

E



Anti-sfGFP
CalFluor Red smFISH

Figure 2. Detection of sfGFP:*Sens* with smFISH (A) Schematic of Sens expression relative to the dorsal-ventral boundary of the wing disc. **(B)** Confocal section of sfGFP:Sens GFP expression. **(C-E)** Representative deconvoluted widefield sections of disc expressing sfGFP:Sens. Scale bars= 5 µm. **(C)** SfGFP-Sens protein (green) visualized with nuclei stained with DAPI (blue). Two stripes of Sens expression run roughly parellel to the x axis of the image. **(D)** SfGFP-Sens mRNA visualized with anti-GFP smFISH probes along with nuclei stained with DAPI. (E) SfGFP-Sens mRNA CalFluorRed signal alone. Red arrows indicate presumptive transcription sites.

To determine the nature of these spots, I designed probes complementary to *sens* RNA. These were expected to detect mRNA from the *sfGFP-sens* transgene. The transgene is present in a *sens* protein null background. However, mutant alleles are not mRNA null, and transcription still occurs. These endogenous *sens m*RNAs would also be detectable with anti-*sens* probes. I ordered these probes conjugated to the fluorophore Quasar 670, which fluoresces in the far-red channel, so that both *sens* and GFP probes could be visualized simultaneously in the same tissue. If detection is perfect, I would expect that each GFP spot would have a corresponding spot in the *sens* channel. However, FISH performed using both probe sets clearly shows many more anti-*sens* Quasar 670 spots than anti-GFP CalFluorRed spots (Figure 3B). Interestingly, there were clearly two classes of spots visible in the anti-*sens* Quasar 670 channel: very bright spots, located in the nucleus, and less bright spots. It appeared as if the majority of spots visible in the anti-GFP CalFluorRed channel belong to the class of very bright nuclear spots. I hypothesized that these brighter spots were sites of nascent transcription. In nuclei containing a lot of overall FISH signal, there were clearly two bright spots per nucleus: one that fluoresced in both the anti-GFP and anti-*sens* channel, and one that was only present in the anti-*sens* channel (Figure 3C). This is consistent with sites of transcription from two loci: the *sfGFP-sens* locus on the second chromosome, and the mutant endogenous *sens* locus on the third chromosome. The likely reason that I did not see two separate alleles is that chromosomes undergo spatial pairing in *Drosophila* cells (Metz, 1916). Therefore, each spot likely represents transcription from both alleles of that locus.

It was clear from my images that the anti-*sens* Quasar 670 probes were giving a much stronger signal-to-background than their counterparts in the anti-GFP CalFluorRed channel. I wondered if this was due to the properties of the fluorophore, or if this was a probe-set specific

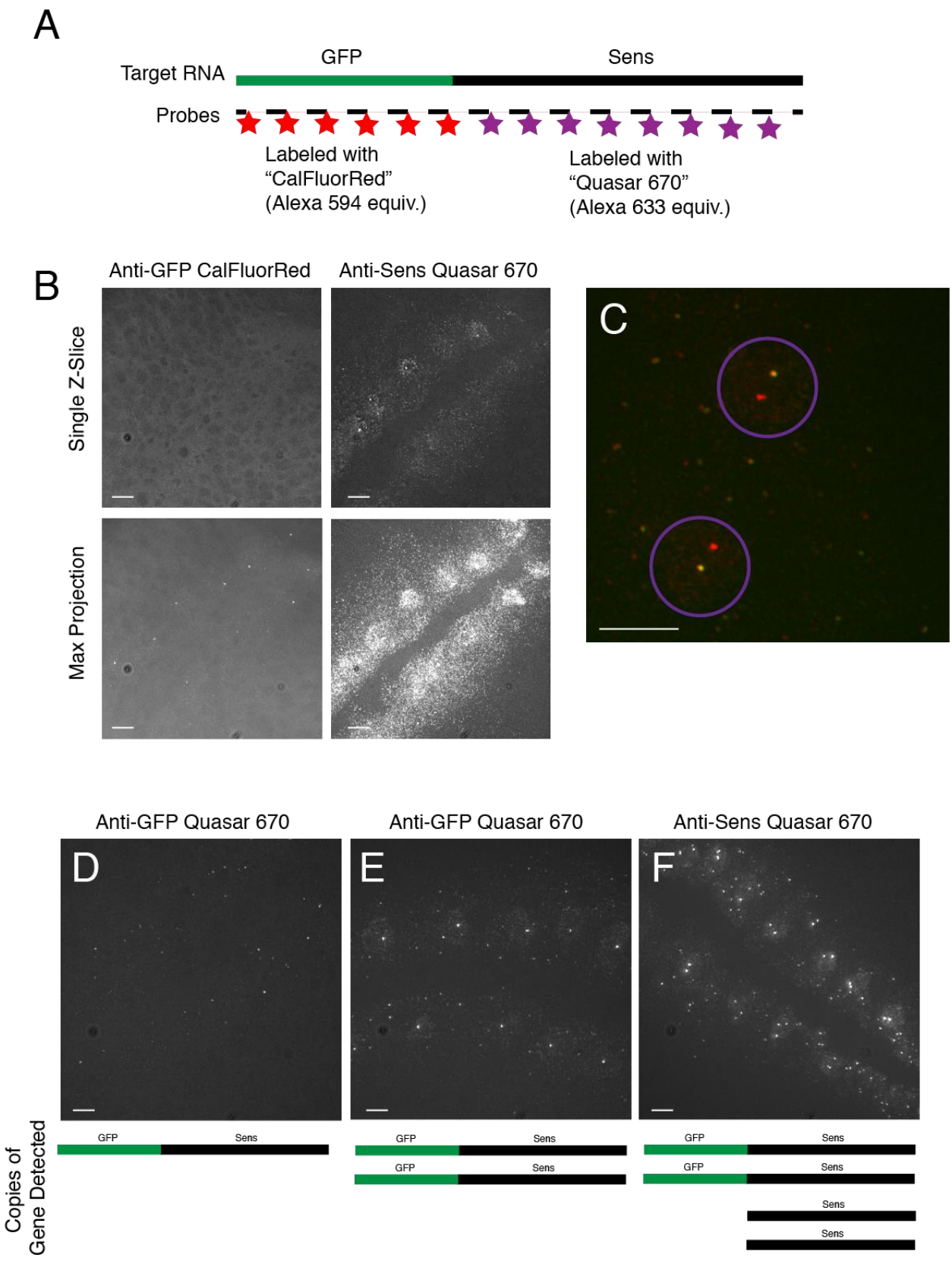

A
GFP
Sens
Target RNA
Probes
Labeled with
“CalFluorRed”
(Alexa 594 equiv.)
Labeled with
“Quasar 670”
(Alexa 633 equiv.)
B
Anti-GFP CalFluorRed
Anti-Sens Quasar 670
Single Z-Slice
Max Projection
C
Anti-GFP Quasar 670
Anti-GFP Quasar 670
Anti-Sens Quasar 670
D
E
F
Copies of
Gene Detected
GFP
Sens

Figure 3. Dual color smFISH detection of sfGFP-sens

(A) Schematic illustrating design method for two independent sets of probes targeting GFP and sens portions of mRNA transcripts. CalFluor Red molecules are indicated by red stars and Quasar 670 fluorescent molecules are indicated by purple stars. **(B)** Single deconvoluted widefield sections and maximum projections of both red and far-red channels for discs expressing sfGFP:sens. Scale bars= 5 µm. **(C)** Representative max projection of red and far-red channels combined. Purple outlines indicate the approximate boundaries of a single nucleus. Scale bar= 5 µm **(D-F)** Widefield maximum projections of discs hybridized with anti-GFP Quasar 670 FISH **(D,E)** or anti-sens Quasar 670 **(F)** probes. Genotypes are schematized below corresponding images to illustrate the DNA copy number producing detectable RNA transcripts in each image. Scale bars= 5 µm. **(D)** Disc with one copy of sfGFP-sens, one copy of mutant $sens^{E1}$, and one $sens^{+}$ probed for GFP mRNA. **(E)** Disc with two copies of sfGFP-sens and two copies of mutant $sens^{E1}$ probed for GFP mRNA. **(F)** Disc with two copies of sfGFP-sens and two copies of mutant $sens^{E1}$ probed for sens mRNA. Although the mutant alleles are mutant for protein, they still make sens mRNA.

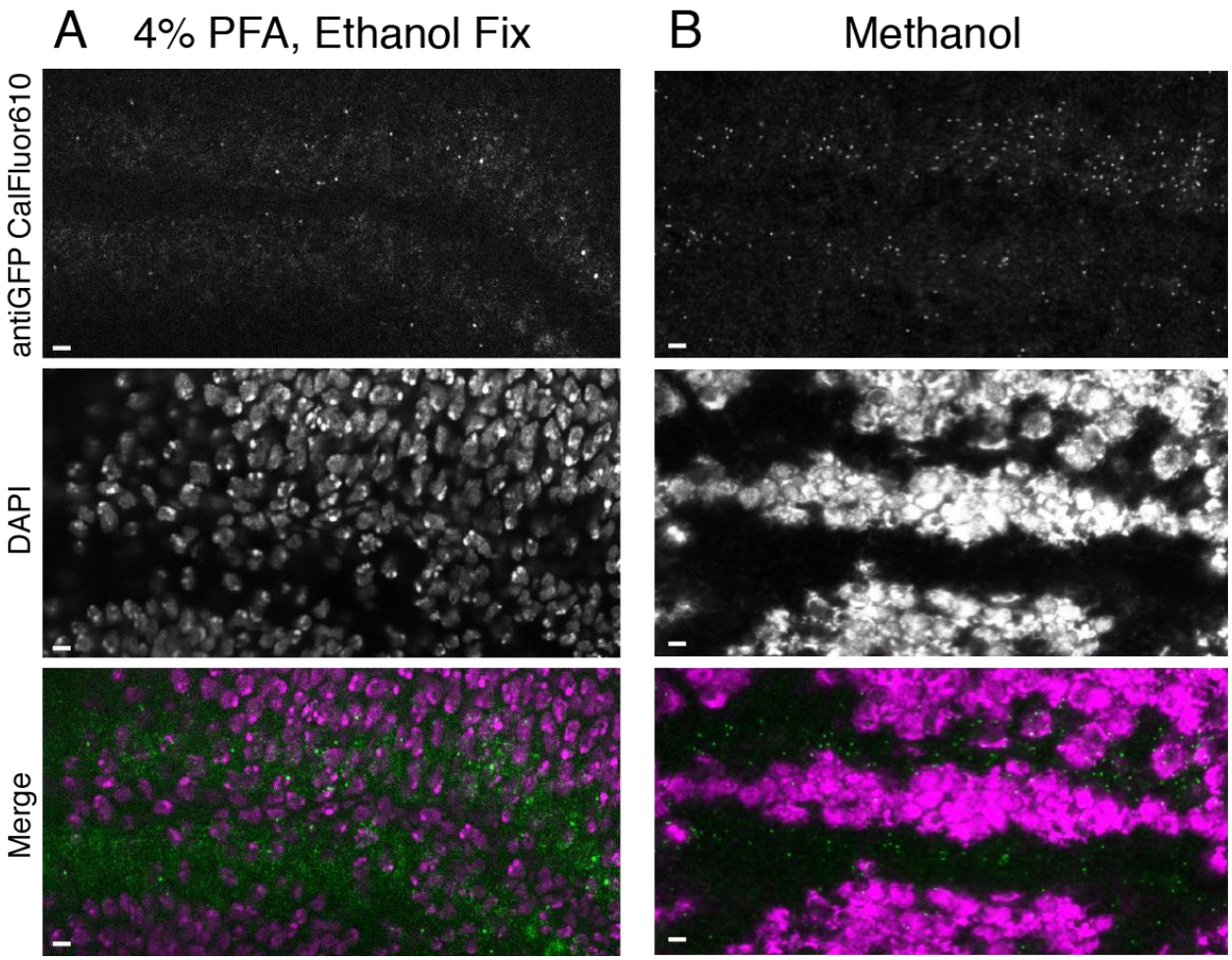


Figure 4. Effect of fixation methods on smFISH signal.

Representative confocal z-sections of sfGFP-sens wing discs hybridized with anti-GFP Quasar 670 FISH probes using **(A)** formaldehyde and ethanol or **(B)** methanol-only fixation methods. Scale bars= 5 µm.

phenomenon. To this end, I performed smFISH with anti-GFP Quasar 670 probes. I saw similar signal-to-background between the anti-GFP and the anti-*sens* probe sets in this channel (compare Figure 3 E,F). I concluded that smFISH performance is enhanced in the far-red channel when compared to the red channel, regardless of probe set.

Next I sought to assay whether intentionally changing the levels of transcript by varying copy number could produce a change in the number of RNA spots that I could detect. To do so, I compared smFISH images from animals with one copy of *sfGFP-sens* and two copies of *sfGFP-sens* probed with anti-GFP Quasar 670 probes. I also compared these to images of animals with two copies of *sfGFP-sens* using anti-sens Quasar 670 (Figure 3 D-F). Qualitatively, there are more spots visible in image stacks with more copy numbers of the transgene present in cells (Figure 3 D-F). This indicates I on the right track to be able to obtain quantitative results with a large perturbation of mRNA levels. However, the signal-to-background quality of the spots remained too poor to precisely quantify spot numbers.

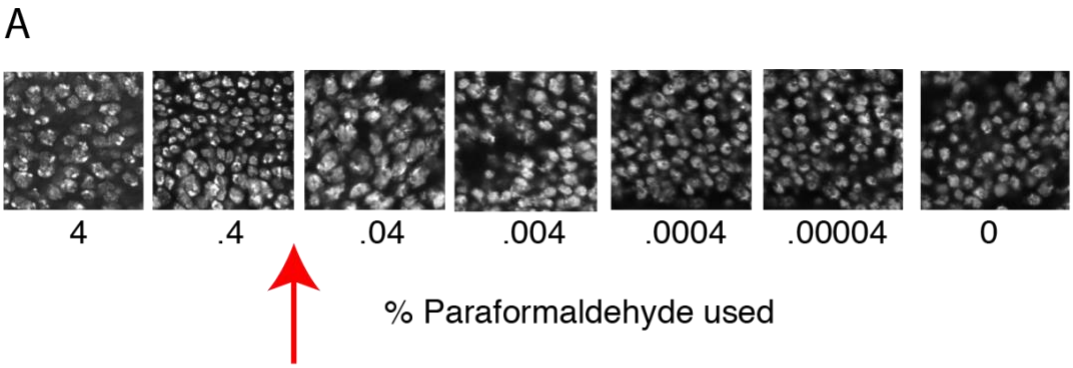
smFISH Signal Quality Improved with Methods Development

I next sought to image smFISH treated discs using confocal microscopy. Protocols for smFISH published at this time largely did not use confocal microscopy. However, these protocols were primarily optimized for cell culture. Imaginal disc tissue is much thicker (approximately 15-20 µm) than cultured cells. While deconvolution can assist with localizing signal to a plane in thick tissue in z-space, confocal imaging can do so via the microscopy setup itself without any need for post-processing. Confocal microscopy also offers a greater breadth of settings for boosting a low signal: laser power, detector gain, and scanner speed can all be manipulated to detect very weak signal. However, using the Leica SP5 confocal to detect smFISH signal resulted in a signal to background very similar to what could be achieved using

the widefield setup on the Deltavision (Figure 4 A). It did however, eliminate some artifacts related to deconvolution.

I further sought to improve signal to background by altering the fixation step of the smFISH protocol. I had been fixing the discs with 4% (w/v) paraformaldehyde (PFA) in PBS. PFA might lower the signal-to-background, possibly because crosslinking decreases access to target mRNAs preserving contact with RNA binding proteins. In contrast, methanol fixes tissue primarily by protein denaturation rather than crosslinking (Hoetelmans et al., 2001). By eliminating PFA in the fixation in favor of methanol, I thought I might increase probe binding efficiency since denatured RNA-binding proteins will not be associated with mRNA. Indeed, methanol fixation greatly increased the signal to background of each fluorescent image (compare Figure 4 A & B). However, methanol fixation caused two significant problems. First, all RNA spots now seemed to be entirely the same intensity. The bright, nuclear RNA signals I had seen in previous imaging were no longer visible. Secondly, the nuclei, as visualized by the DAPI stain, appeared to lack any characteristic structure and appeared as indistinct masses. Taken together, these observations seemed to suggest a severe loss of nuclear structure, possibly due to histone denaturation, which would cause the DNA to lose compaction and the nascent RNAs to no longer be concentrated in one transcription spot. I was concerned that this would make it impossible to detect individual cells. It would also make it impossible to know whether an RNA spot was nascent or mature in origin, potentially impacting RNA measurements.

I thought I could mitigate this loss in structure by pre-fixing the discs with a lower concentration of PFA, thus creating enough crosslinking to maintain nuclear structure without impairing probe binding. I performed a pre-fixation step prior to the methanol fixation and used


A
4
.4
.04
.004
.0004
.00004
0
% Paraformaldehyde used

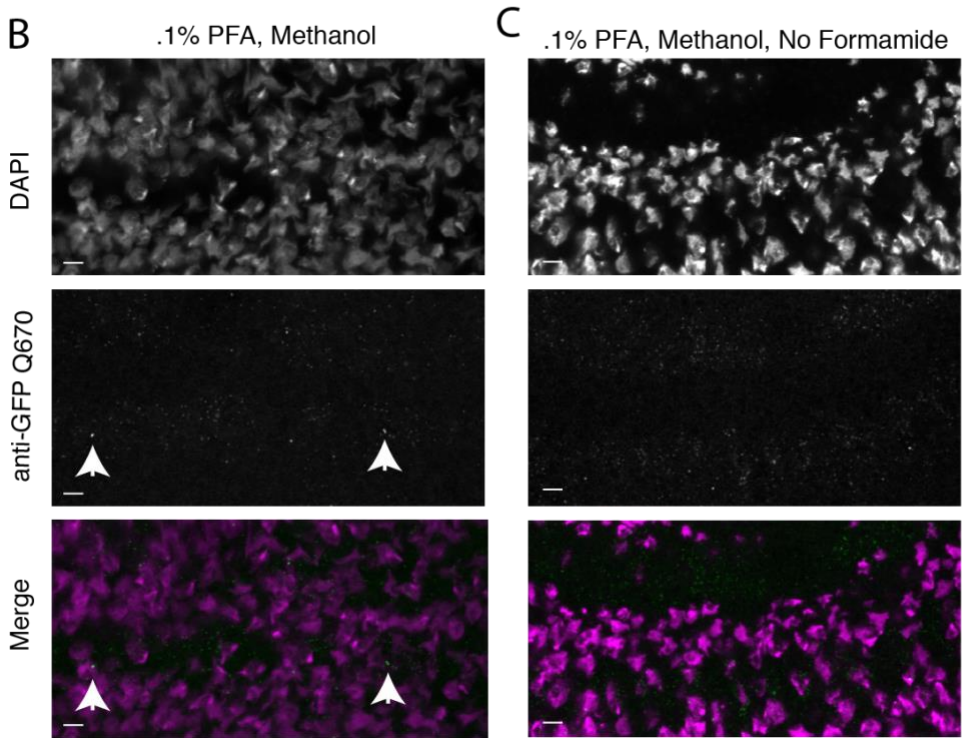

B
.1% PFA, Methanol
C
.1% PFA, Methanol, No Formamide
DAPI
anti-GFP Q670
Merge

Figure 5. Optimal smFISH hybridization conditions.

(A) Confocal z-sections of wing discs treated with a mock hybridization method using a 15-minute fixation in PBS-buffered paraformaldehyde solution of the concentration indicated. Nuclei are stained with DAPI. **(B-C)** Confocal sections of sfGFP-sens wing disc dorsal-ventral margin hybridized with anti-GFP Quasar 670 smFISH probes. Scale bars= 5 µm. **(B)** sfGFP-sens detected in discs hybridized with 30% formamide at 37ºC. **(C)** sfGFP-sens detected in discs hybridized with 0% formamide at 62º C.

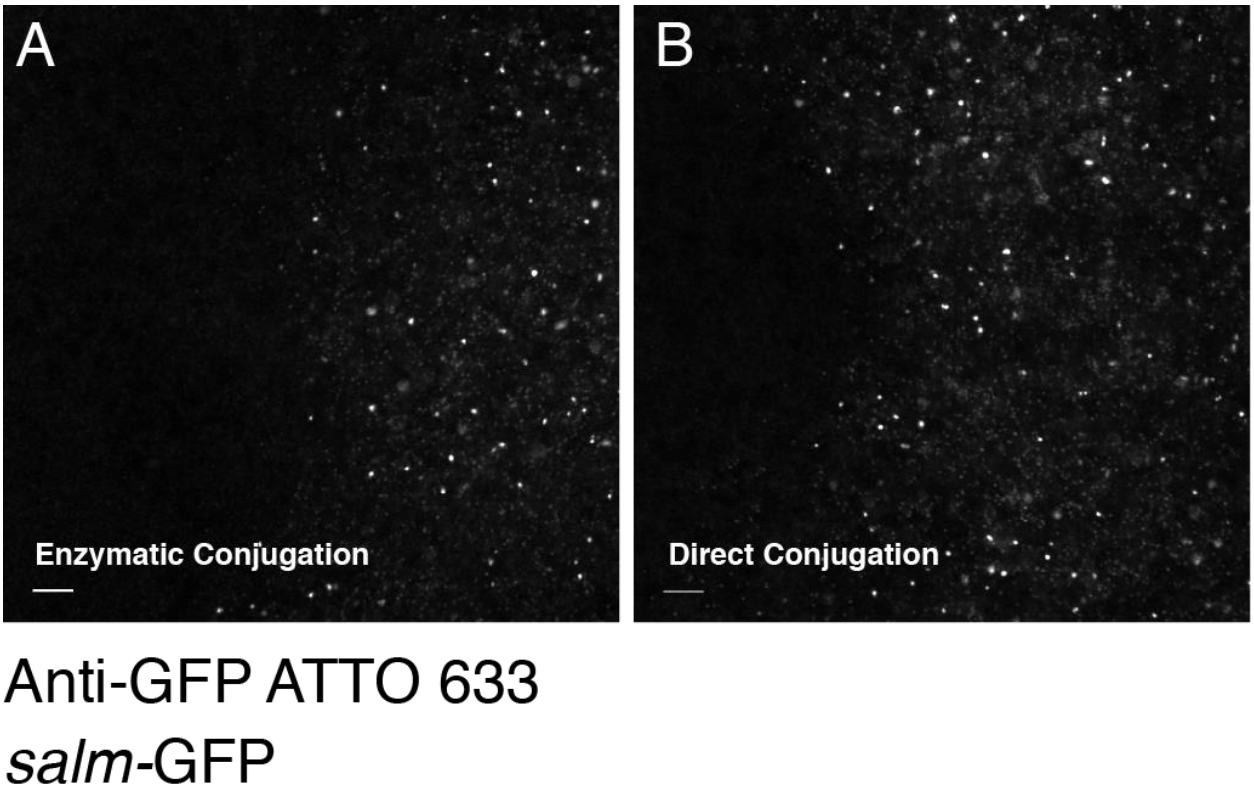
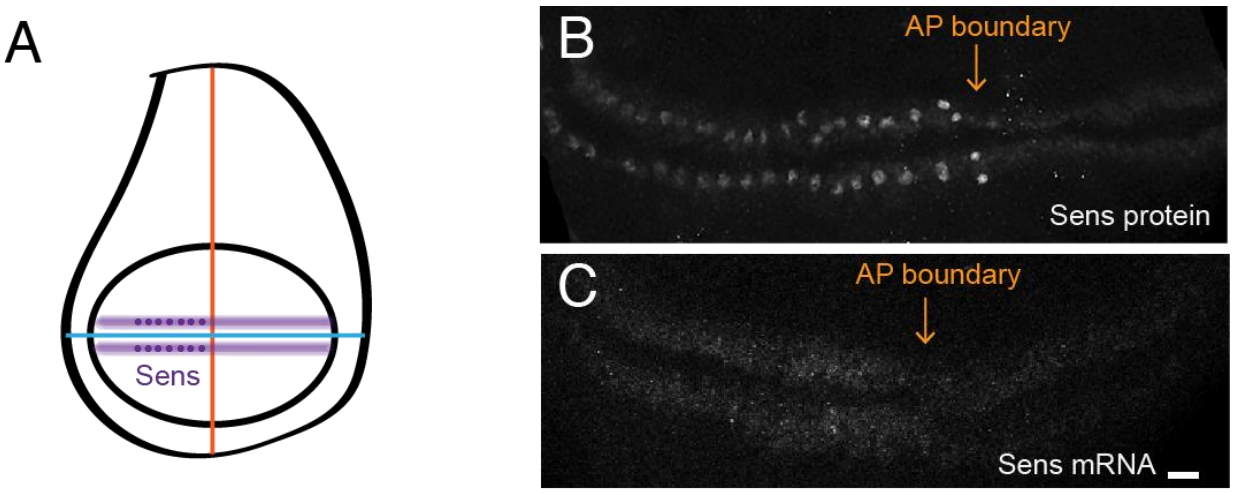


Figure 6. Enzymatic conjugation of FISH probes.

Confocal z-sections of anterior compartment of *salm:*GFP wing discs hybridized with anti-GFP FISH probes prepared by **(A)** enzymatic or **(B)** direct conjugation to the ATTO 633 fluorophore. Scale bars =5µm.

a dilution series of PFA concentrations, followed by DAPI staining and visualization of nuclei. The nuclei of these discs experienced a loss of structure for concentrations of 0.04% PFA and lower (Figure 5A). I concluded that a PFA concentration of 0.1% would be appropriate for retaining structure. When a 0.1% PFA pre-fix was performed along with hybridization, I observed a good signal-to-background and some evidence of bright putative nascent transcription sites (Figure 5B, middle panel). However, the nuclei still appeared structurally diffuse (Figure 5B, top panel). I suspected that this might be due to osmotic stress on the mildly fixed tissue, caused by use of 50% (v/v) formamide in the hybridization step. Formamide decreases the T_M of annealing by approximately 0.7^o C per 1% formamide in hybridization solution for DNA with 50% G-C content (Farrell, 2005). I omitted the formamide and increased the hybridization temperature to compensate for the change in T_M according to this formula. Hybridization using this procedure produced images with superior nuclear signal and nuclei quality, while preserving good signal to background (Figure 5C).

Finally, the most expensive element of the smFISH protocol was the cost of purchasing or manufacturing oligo probes. I intitially used probe sets manufactured by Stellaris/Biosearchtech. These probe sets are conjugated with proprietary fluorophores by Biosearchtech. I noticed significant bleaching over time from these probes, and I hypothesized that a different fluorescent dye would give superior results. I selected ATTO dyes, which purportedly do not bleach as rapidly as other available dyes. For anti-GFP probes, I conjugated these fluorophores directly to oligonucleotides ordered from BioSearchTech (see Methods) using an esterification reaction between NHS-Ester labeled fluorophore and terminal amino labeled oligonucleotides.


A
Sens
B
AP boundary
Sens protein
C
AP boundary
Sens mRNA

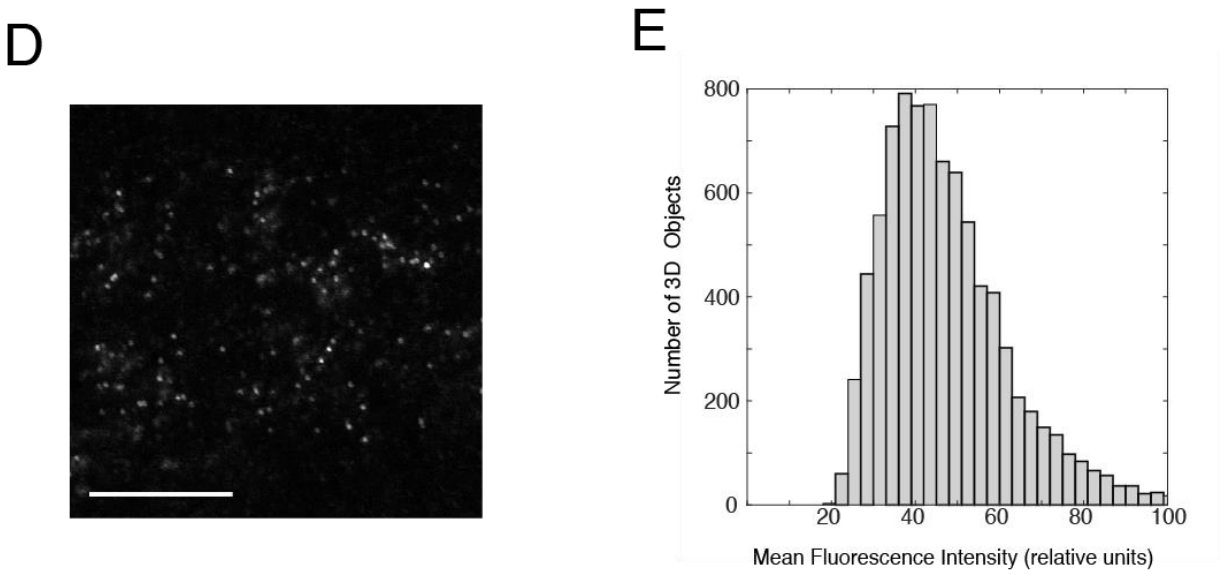

D
E
800
600
400
200
0
Number of 3D Objects
20
40
60
80
100
Mean Fluorescence Intensity (relative units)

Figure 7. Detection of sfGFP-sens using optimized smFISH protocol.

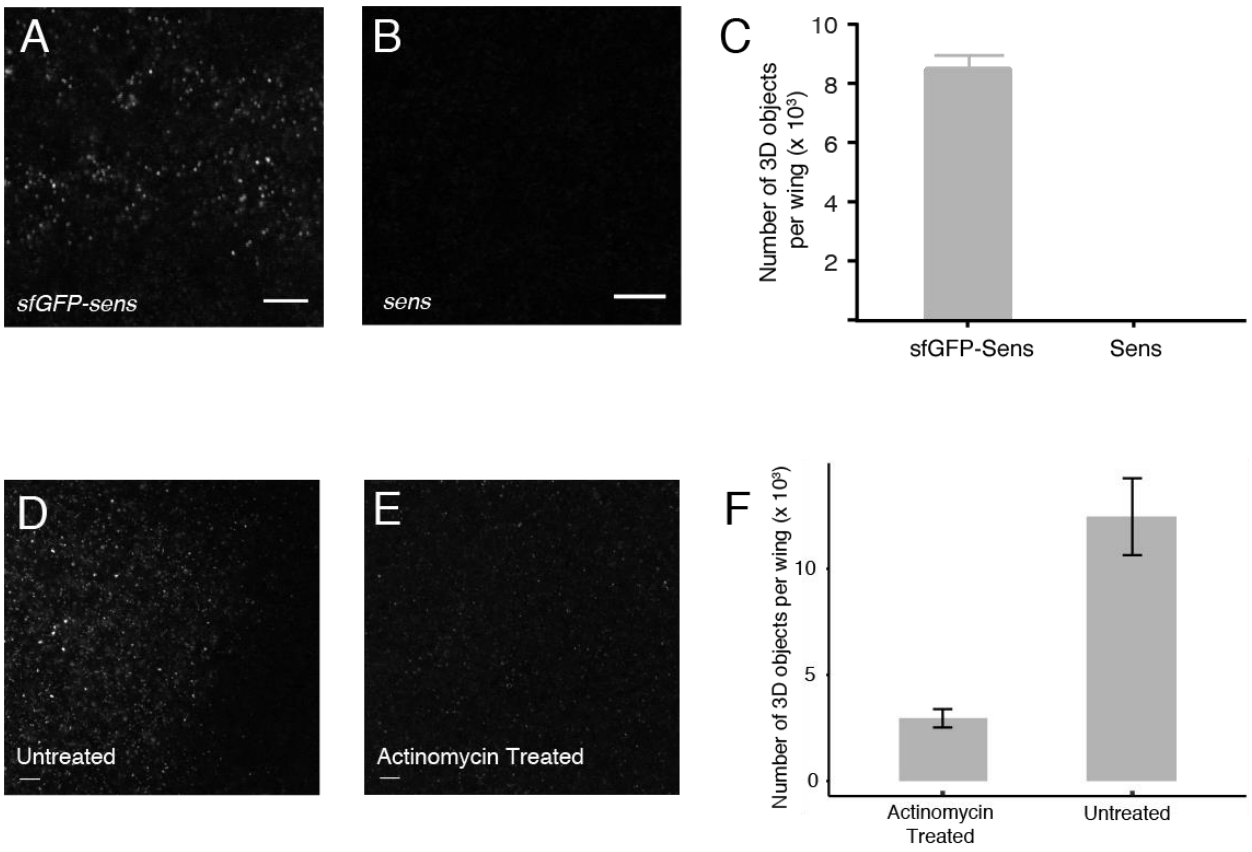
(A) Schematized expression of sens in the wing disc about the D-V margin. sens is also expressed in clusters of cells in the notum, which are not shown. **(B-D)** Confocal sections of wing discs expressing sfGFP:*sens*. Scale bars= 10 µm. **(B)** sfGFP-sens protein fluorescence. **(C)** sfGFP-sens mRNA visualized using smFISH using anti-sfGFP probes. **(D)** Higher magnification of sfGFP-sens mRNAs visualized using anti-sfGFP probes. **(E)** Distribution of mean intensity for all identified fluorescence objects from one wing disc expressing sfGFP-sens.

This procedure was still costly because the 5' nucleotide in each oligo was a chemical derivative that was expensive to add to the oligos. In order to bring down the cost of additional probe sets, I utilized an enzymatic procedure for fluor conjugation to oligos. This involves performing a single NHS-Esterification reaction in order to conjugate fluorophore to dideoxy-UTP (ddUTP). An enzymatic reaction is then used to add the fluorescent ddUTPs to an oligonucleotide that does not require any special chemical modifications and could be purchased from IDT. This reduces the need to repurchase NHS-esterified fluorophores and amino-5'-labeled oligonucleotides, greatly reducing the cost per probe set. Probes prepared in this way with ATTO-633 dyes perform equivalently to probes conjugated chemically (Figure 6).

smFISH in the wing imaginal disc is specific for GFP RNA

Under these optimized conditions, smFISH of *sfGFP-sens* wing discs generates signal specifically in the pattern that is expected for *sens* (Figure 7A-C). This signal at high magnification shows clear, punctate spots that appear well separated from background signal (Fig 7D). Anti-GFP probes give very little signal when used to probe discs from larvae lacking *sfGFP-sens* (compare Figure 8 A-B). A typical *sfGFP-sens* image stack contains approximately 8,500 spots as assigned by my automated spot finding program, whereas an image stack on a disc expressing no GFP construct contains only 40 spots (for details on the spot-finding program development, see Chapter III). These 40 spots were all identified in the first five z-planes, which are always captured near the peripodial membrane of the wing imaginal disc. The source of the spots therefore likely represents non-specific binding of probes to the exterior of the disc.

To further examine the nature of the spots, I incubated wing disc explants in culture media supplemented with actinomycin-D before probing for mRNAs expressed from *spalt-major* (*salm*). This resulted in a significant reduction in the number of spots visible in images (Fig 8


A
sfGFP-sens
B
sens
C
Number of 3D objects per wing (x 10³)
10
8
6
4
2
sfGFP-Sens
Sens
D
Untreated
E
Actinomycin Treated
F
Number of 3D objects per wing (x 10³)
10
5
0
Actinomycin Treated
Untreated

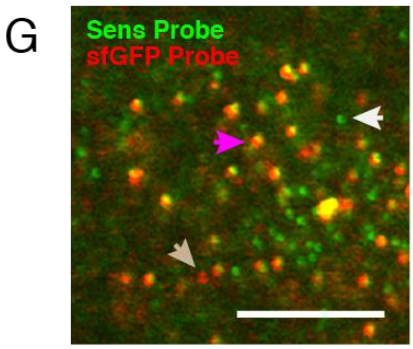

G
Sens Probe
sfGFP Probe

Figure 8. smFISH Positive and Negative Controls.

(A-B) Representative confocal sections of wing discs hybridized with anti-GFP FISH probes. Scale bars= 5 µm. **(A)** A disc from an animal containing two copies of sfGFP-sens and two copies of mutant *sens*^{E1} allele. **(B)** A disc from an animal containing two copies of endogenous *sens*$^{+}$. **(C)** Mean number of fluorescence objects identified in whole wing discs containing either two copies of sfGFP-sens or two copies of endogenous sens hybridized with anti-sfGFP probes. Error bars are SEM. **(D-E)** Representative confocal sections of wing discs hybridized with anti-Salm RNA probes. Error bars=5 µm. **(D)** A disc cultured in growth medium for 30 minutes. **(E)** A disc cultured in growth medium + 5ng/µL actinomycin-D for 30 minutes. **(F)** Mean number of fluorescence objects idnetified in entire wing discs hybridized with anti-Salm RNA probes cultured in growth medium with or without actinomycin-D. Error bars are SEM. **(G)** A representative confocal section from a wing disc expressing sfGFP-sens and the mutant *sens*^{E1} allele. The disc was probes for GFP (red) and sens (green) RNA using independent probe sets. Spots that fluoresce both red and green (purple arrow) are presumptive sfGFP-sens mRNA that have annealed to both probe sets. Spots that only fluoresce with the sens probe set (white arrow) are presumptive sens mRNAs that are generated from the endogenous protein *sens* gene. Although these *sens* alleles are mutant for protein output, they still produce mRNA. The occasional spot (beige arrow) that fluoresces only in the sfGFP set are presumptive sfGFP-sens mRNA that failed to hybridize with the anti-sens probe set. These are false negatives. Scale bar= 5 µm.

D-E). My automated spot detection found on average, approximately a four-fold reduction in the number of RNA spots detected per disc (Figure 8F). This indicates that the spots I detect are likely RNA in origin.

I also visualized *sfGFP-sens* mRNAs detected by both sens and GFP probes with my optimized protocol, in order to estimate the number of mRNAs missed by in situ hybridization with a single probe set. GFP probes conjugated with ATTO 633 detected GFP RNA, whereas anti-sens probes conjugated with ATTO 565 detected sens RNA. This resulted in three sets of spots: colocalized spots (*sfGFP-sens* mRNA), spots detected only in the sens channel, and spots detected only in the GFP channel. Spots detected only in the sens channel may represent endogenous *sens* transcripts, as the *sens* null alleles used in this study are protein null but not RNA null in nature. Theoretically, every GFP spot should have a corresponding *sens* spot in the other channel, so I sought to analyze exactly how many GFP spots had no colocalized ATTO 565 RNA spot. I did this by analyzing the fluorescence intensity of a spot around each identified GFP RNA spot in the 565 channel. I modeled the background fluorescence of the 565 channel as a normal distribution about the mean background pixel intensity. If the intensity of 565 fluorescence in the region associated with the 633 RNA spot was in the 98th percentile or higher, it was considered to be a spot in the 565 channel as well. Using this method, I found that 93% of anti-GFP ATTO 633 spots had a corresponding spot in the anti-sens ATTO 565 channel (Figure 8G). Therefore, I concluded that no more than 7% of mRNAs are being missed with a probe set.

Bright Nuclear Spots are Sites of Nascent Transcription

While optimizing smFISH for *sfGFP-sens*, I was able to visualize and detect bright spots that appeared to be localized to the nuclei in my images. These spots appeared to have a maximum of one spot per nucleus using anti-GFP probes. For anti-sens probes in *sfGFP-sens*

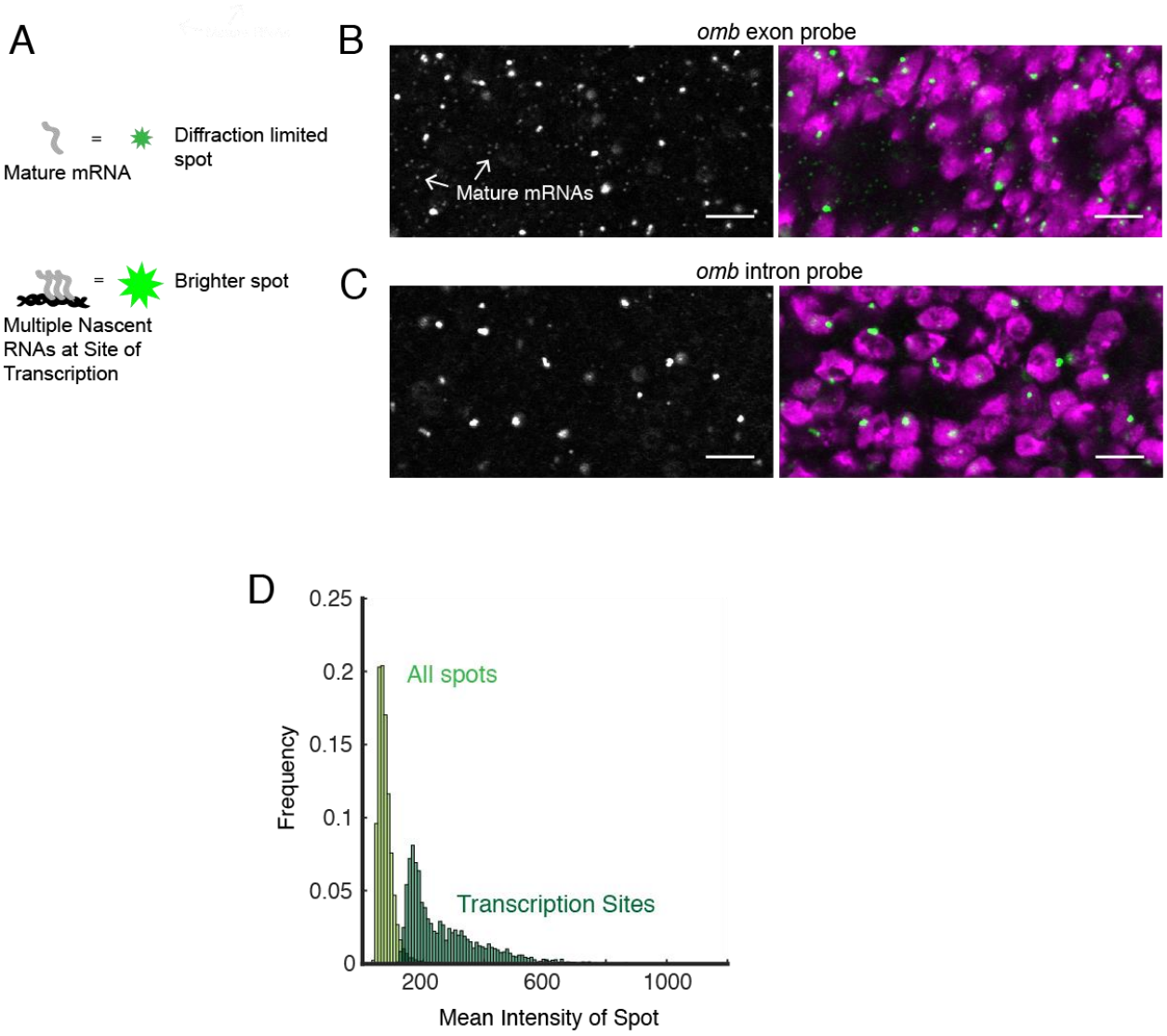

A
=
Diffraction limited spot
Mature mRNA
=
Brighter spot
Multiple Nascent RNAs at Site of Transcription
B
omb exon probe
Mature mRNAs
C
omb intron probe
D
0.25
0.2
0.15
0.1
0.05
0
Frequency
All spots
Transcription Sites
200
600
1000
Mean Intensity of Spot

Figure 9. Sites of nascent transcription are detected by smFISH.

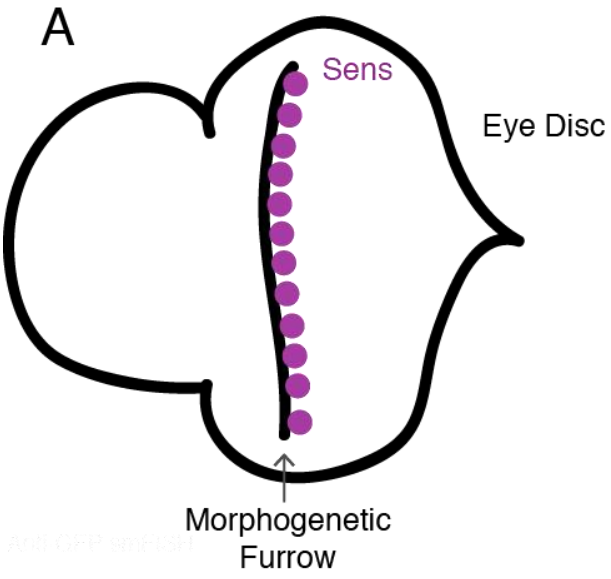
(A) Sites of nascent transcription can fluoresce more brightly due to multiple nascent transcripts localized to one gene locus. **(B)** Probes recognizing an omb exon many small dim spots and a few bright large spots. Right image shows merge of DAPI and probe fluorescence. **(C)** Probes recognizing an omb intron only generate only large bright spots that are associated with nuclei. Scale bars= 5 μm. **(D)** Frequency distribution for all spots in a wing disc probes with *sens* mRNA. Using a threshold of 2.5 times the median spot intensity, single mRNA spots were filtered out, leaving only spots associated with transcription sites. The frequency distribtuion for this class of spot is shown.

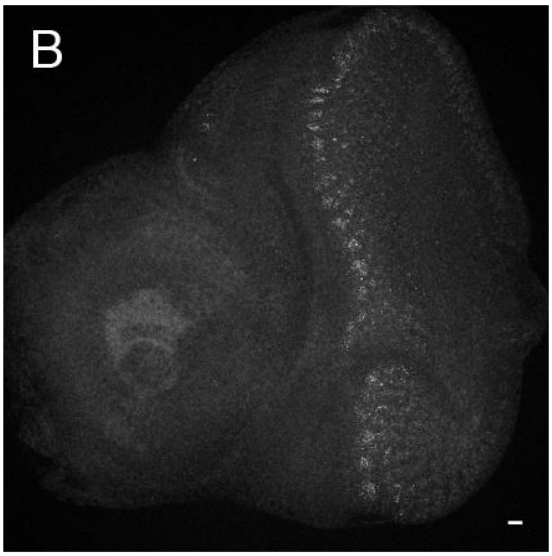
animals, a maximum of two spots per nucleus could be detected (Figure 3C). These seemed likely to be sites of nascent transcription, as have been described elsewhere in smFISH literature (Raj et al., 2006; Raj and Tyagi, 2010) (Figure 9A).

To confirm this, I designed two sets of probes for mRNA transcribed from *optomotor blind* (*omb*), one for exonic RNA and one for intronic RNA. As expected, the intronic RNA probes did not generate any cytoplasmic spots, and only showed bright, nuclear localized spots, whereas the exonic probes detected both types of spots (Figure 9 B & C). This supports the idea that these bright spots represent sites of nascent transcription. My automated spot detection program finds that the bright nuclear spots are 2-25 fold brighter than the weak cytoplasmic spots (Figure 9D for sample quantification).

smFISH methodology also functions in the eye imaginal disc

To test whether the protocol optimized in the wing imaginal disc system was broadly applicable to other imaginal discs, I performed smFISH using eye-antennal discs from *sfGFP-sens* animals and anti-GFP ATTO 633 probes. *sens* is expressed in a stripe of clustered proneuronal cells near the morphogenetic furrow of the eye disc (Figure 10A). I observed smFISH spots of comparable intensity and quality to those in wing discs expressed in this pattern (Figure 10B-C). I therefore conclude that smFISH is able to function similarly in the eye disc and can be used there. It is likely that this methodology is broadly applicable to any imaginal disc, as they all have similar structural properties and physical sizes.


A
Sens
Eye Disc
Morphogenetic
Furrow


B

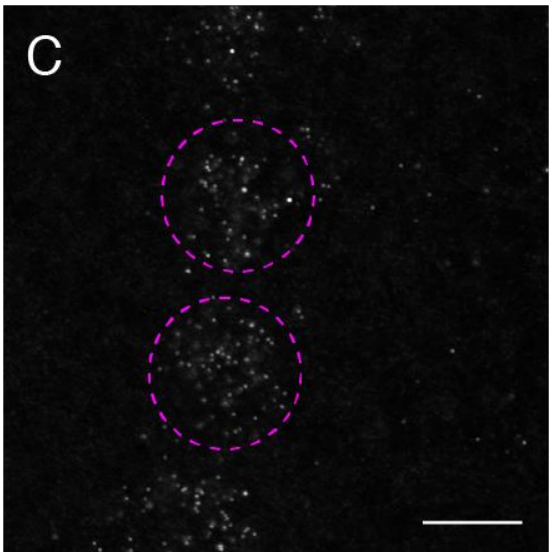
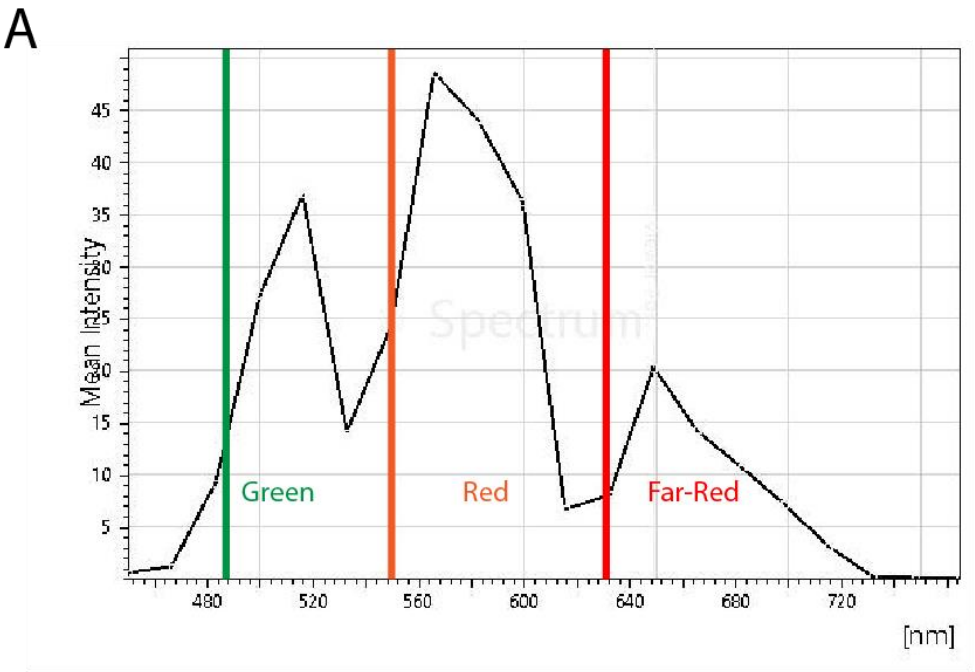
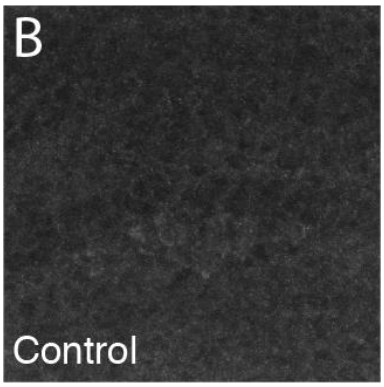
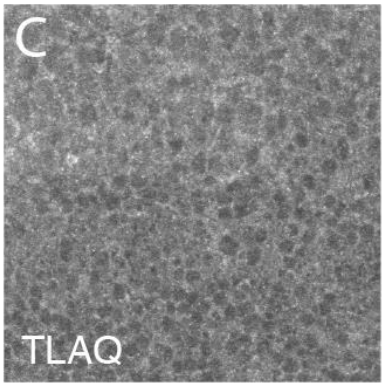

C

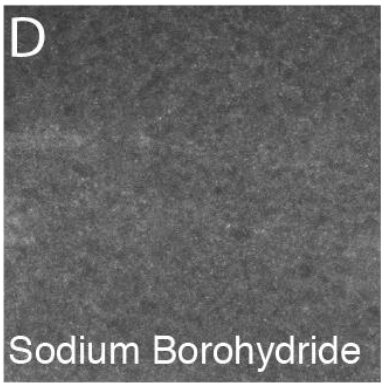
Figure 10. smFISH imaging of the eye antennal disc.

(A) Schematic of the eye antennal disc showing the approximate location of cells expressing *sens*. Anterior is to the left. (**B, C)** Representative confocal sections of discs expressing sfGFP-sens probed for GFP mRNA. Scale bars= 5 μm. **(B)** Low magnification shows a vertical stripe of positive fluorescence that oscillates between high abundance and low abundance of mRNA. This is the pattern that has been reported for cells in the morphogenetic furrow (Nolo et al 2000). **(C)** Higher magnification showing two complete clusters of sens-positive cells (ashed purple lines).









Anti-GFP ATTO 565
sfGFP-mCh-sens

Figure 11. Autofluorescence in smFISH

(A) Mean pixel intensity of images of mock-hybridized single disc imaged every 20 nm along the light spectrum from 440-780nm. Green (488nm), red (550nm) and far-red (633nm) lasers were activated and their locations in the spectrum are indicated. **(B-D)** Representative optical sections of wing discs expressing sfGFP-mCh-sens probed for sfGFP RNA using the ATTO 565 fluorescent dye. **(B)** Disc was hybridized using unaltered optimized smFISH protocol. **(C)** Disc treated with TruBlack Lipofusion Autofluorescence Quencher prior to hybridization. **(D)** Disc treated with sodium borohydride prior to hybridization.

Discussion

smFISH Optimization: General Principles and Strategies for the Future

In my work, the critical points of optimization for smFISH in this context were fixation method and formamide concentration. Fixation chemically prepares the tissue for optimal probe binding while still retaining structural integrity for microscopy. Formaldehyde crosslinks macromolecules with amine groups to one another, preserving secondary and tertiary structure. I found that formaldehyde fixation at concentrations and times used for other smFISH protocols did not produce optimal results in imaginal discs. High degree of cross-linking may decrease probe access to RNAs of interest by preserving the structures of RNA-associated proteins, and the secondary structures of RNA itself (Hobro and Smith, 2017). Methanol fixation is more appropriate for smFISH in that it denatures proteins, exposing RNAs for annealing (Srinivasan et al., 2002). However, it was clear from my results that methanol fixation profoundly altered the structure of the nucleus. It appears likely that methanol fixation alone with no crosslinking was denaturing histones and other DNA associated proteins. It is also possible that this was causing the nuclear membrane to become excessively permeable, resulting in an obscure and unstructured nucleus visible with a DAPI stain (Hoetelmans et al., 2001).

I found that a balance between first fixing with a low concentration formaldehyde and followed by methanol fixation was optimal for imaginal disc tissues. However, optimal fixation methods may vary between tissue types. For example, a methanol fixation alone may not be problematic in a tissue with few nuclei, where a loss of structure would not obscure the position of said nuclei. Other contexts may also be more amenable to formaldehyde fixation, particularly if the mRNAs of interest are not heavily bound with RNA-binding proteins.

Selection of Fluorophores for FISH

Single molecule detection works most robustly when the signal-to-background ratio is maximized. This can be achieved by maximizing signal and reducing background. Choosing a fluorophore with a high quantum yield and low bleaching helps for maximizing signal. For ATTO 565, incomplete enzymatic conjugation diluted the functional probe concentration by having to compete with many unconjugated oligonucleotides for binding. For further use of single molecule FISH, I would strongly advise probe concentration (and purity) as a starting point for optimization.

Background fluorescence has two sources: unbound probe and autofluorescence from the tissue itself. Care should be taken to select fluorophores in the lowest autofluorescence wavelengths for the tissue being imaged. Typically, the greatest levels of autofluorescence are detected in the 488 nm region of the visible light spectrum. For imaginal disc tissue, I found that the mid-red 552 laser generated the greatest autoflourescence (Figure 11). This was not relieved by any autofluorescence quenching methods that I tried. For this tissue, far red 633 was optimal for low autofluorescence. If performing smFISH in a new tissue or context, I would recommend initially performing a spectral scan on fixed and mounted (unhybridized) tissue before deciding on the correct fluor label to use.

Additionally, residual fluorescence from the fluorescent proteins expressed in the tissue themselves must be avoided. The methanol fixation denatures a significant portion of the expressed fluorescent proteins but does not entirely kill fluorescence. Therefore, fluorophores for smFISH will work best in wavelengths far away from any fluorescently tagged proteins expressed in the tissue.

Hybridization conditions

Hybridization can be favored or disfavored by temperature, salt concentration, and formamide concentration. The hybridization conditions I chose were based on previous smFISH protocols and are somewhat strictly based on the average T_m of the oligonucleotide probes. In theory, the binding of multiple oligos to an RNA is a cooperative process (Kwon, 2013). The initial binding events serve to weaken the secondary structure of the RNA and to lower the binding energy for additional oligonucleotides. Thus, strict hybridization conditions help to prevent nonspecific binding.

I used the Stellaris probe designer to design most of my probe sets. This program allows for a relatively wide variation in T_m for probes. If a sufficient number of oligos are designed, a handful of poor oligos may not affect results. However, one can imagine a scenario where one wants to detect a short transcript, and probe number is limited by the available length of the RNA. When working with a smaller number of different oligonucleotide probes, it is helpful to design such probes to have highly similar melting temperatures, and to tailor the hybridization temperature (or other conditions) to that melting temperature. Keep in mind that melting temperatures calculated by companies that manufacture oligonucleotides are calculated under certain conditions and are for DNA-DNA binding rather than RNA-DNA duplex binding (Freier et al., 1986). Therefore, T_M may be a starting or reference point for calculating hybridization temperature, but independent calculations using proper salt concentrations and empirical testing need to be done.

While hybridization time was never systematically explored by me, I serendipitously found that doubling the hybridization time and the subsequent wash did not significantly alter the FISH signal-to-background ratio in my hands. While I never used the resulting images for

analysis out of an abundance of caution, I speculate that one hour is long enough for probe binding to approach an equilibrium state in this tissue under these conditions.

Immunostaining and smFISH

I never utilized antibody co-staining for any analysis in my data. I made several attempts to use an anti-lamin antibody to detect nuclei, performing the immunostaining both before and after hybridization with FISH probes. These were not successful using the 0.4% PFA and methanol fixation method, and resulted in no antibody signal. I speculate that antibodies for which their corresponding epitopes are intact after methanol fixation (or denaturation in general) could be used for detection.

Why was smFISH difficult in imaginal discs?

I speculate that imaginal discs are challenging with regards to smFISH because of their tissue geometry. The columnar cells and pseudo-stratified epithelial structure make for a densely packed and irregular cell architecture. Compared to *Drosophila* embryos, there is a high density of cell membranes, causing a high lipid content in the tissue that possibly causes additional autofluorescence. Other tissues, (such as zebrafish tissue) have been found to present similar difficulties (Trinh and Fraser, 2013). Thus, additional methodologies for RNA detection may be useful for *Drosophila* tissues in the future.

Future Directions

Several recently developed technologies may offer superior imaging-based detection of single molecule RNAs in thick tissues. Hybridization-based amplification employ use of unlabeled oligo probe sets that are hybridized to target RNAs. The probes themselves can then be detected with secondary fluorescently labeled probes, which can be highly specific and

heavily labeled (Sylwestrak et al., 2016). Additionally, new fluorophore products may be developed that give higher quantum yield, or many new colors. Several exciting technologies allow FISH to be multiplexed in a manner that will allow the development of image-based transcriptomics. Seq-FISH involves repeated hybridization, imaging, and stripping of a fixed sample with multiple sets of FISH probes (Eng et al., 2019). Alternatively, multiplexed error-robust-FISH (mer-FISH) uses combinations of fluorophores to produce more colors with non-overlapping emission spectra than the traditional four-color fluorescence system (Xia et al., 2019). Both of these approaches allow many genes to be assayed in the same sample in parallel, although they each have their own sets of limitations. Even an expanded number of genes assayed in each FISH cannot approach the broad transcriptomic approach offered by a sequencing based approach. Some mRNAs are too short or structurally inaccessible to probes, and multiplexing does not resolve or improve such a shortcoming. Additionally, the cost of probe sets and imaging time can become exorbitant (Chen et al., 2019; Lubeck et al., 2014). Nonetheless, multiplexed RNA detection expands the ease of asking certain types of scientific questions. How are sets of genes co-regulated?

It would allow us to characterize tissues with an unprecedented level of detail, giving us information about cell-to-cell heterogeneity for both transcript types and transcript levels. Combined with single cell RNA sequencing, this could be particularly powerful tool for discovery. Rare sets of cells with a particular set of expressed genes could be discovered via sequencing and then their location and distribution within a tissue characterized, leading to insights into tissue function.

An additional avenue for future research would be to explore how transcript levels change over time, using live imaging approaches. This would require adaptation of the MS2

system for *Drosophila* imaginal discs. While the technology to bring the MS2 system into imaginal discs exists, actually implementing it will require building the proper reagents (Garcia and Gregor, 2018). Culture conditions must be worked out for discs that allow transcription to proceed the same as it does *in vivo*. Additionally, imaging can be much more challenging in live tissues, which have even more autofluorescence than live tissues (Wang et al., 2008). However, the observation of transcription in real time would undoubtedly help answer many questions and lead to new discoveries.

Materials and Methods

Drosophila genetics

All *Drosophila* were raised at room temperature and grown on standard molasses-cornmeal food. The *sfGFP-sens* transgenic line was used as described in (Cassidy et al., 2013).11/24/2020 5:02:00 PM

smFISH Probe Design and Preparation

smFISH oligonucleotide probes were designed using Stellaris Probe Designer (Biosearch Technologies). Probes sets contain between 45 and 48 non-overlapping 20-nucleotide oligos. A full list of all probe sets is provided in Appendix 1. Anti-GFP probes were prepared by conjugating NHS-ester ATTO 633 dye (Sigma 01464) to the 3' end of each oligonucleotide. Anti-sens probes were prepared by conjugating NHS-ester ATTO 565 dye (Sigma 72464) to the 3' end of each oligonucleotide. These oligos bear a mdC(TEG-Amino) 3′ modification to allow conjugation, and were obtained from Biosearch Technologies. Conjugation and purification was performed as described (S. C. Little & Gregor, 2018). All other probe sets were prepared using the enzymatic conjugation protocol as described (Gaspar, Wippich, & Ephrussi, 2017). Briefly,

amino-11-ddUTP (Lumiprobe) was conjugated to NHS-ester ATTO 633. Terminal deoxynucleotidyl transferase (New England Biolabs) was then used to conjugate ATTO 633-ddUTP to the 3' ends of oligonucleotides that had been purchased from IDT. After enzymatic conjugation, oligos were purified from free ATTO 633-ddUTP using G-25 spin columns (GE Illustra) according to manufacturer's instructions. Final concentration of oligonucleotide was 33 µM in water. Probes were stored at -20°C, protected from light, until use.

smFISH: Initial Protocols

Wing discs were dissected from wandering 3rd instar larva in cold phosphate buffered saline (PBS) and immediately fixed in 4% (w/v) paraformaldehyde / PBS for 45 minutes at room temperature. Discs were then fixed for 24 hours in 70% reagent grade ethanol at -20° C. Discs were transferred to hybridization buffer (35% formamide, 10% w/v dextran sulfate, 4X SSC, 0.01% w/v salmon sperm ssDNA (Invitrogen 15632), 1% v/v vanadyl ribonucleoside (NEB S14025), 0.2mg/mL BSA, 0.1% v/v Tween-20). Oligo probes were added to a 1.5 µM final concentration in the hybridization buffer, and hybridization was performed overnight (15 hours) at 37° C. After hybridization, discs were washed once for 30 minutes at 37 ° C in wash buffer (35% formamide, 4X SSC, 0.1% v/v Tween-20, 5 µg/mL 4′,6-diamidino-2-phenylindole (DAPI) (Invitrogen)). Discs were washed with PBS + 0.1% Tween-20 and transferred to Vectashield (Vector Labs) for mounting. Discs were mounted in 15 µl of Vectashield on glass microscope slides using an 18 X 18 mm No. 1 coverslip (Zeiss).

smFISH: Final Protocol

Wing discs were dissected from wandering 3rd instar larva in cold phosphate buffered saline (PBS) and immediately fixed in 0.1% (w/v) paraformaldehyde / PBS for 15 minutes at room temperature. Discs were then fixed for 30 minutes in methanol at room temperature. Discs were

transferred to hybridization buffer (10% w/v dextran sulfate, 4X SSC, 0.01% w/v salmon sperm ssDNA (Invitrogen 15632), 1% v/v vanadyl ribonucleoside (NEB S14025), 0.2 mg/mL BSA, 0.1% v/v Tween-20). Oligo probes were added to a 1.5 μM final concentration in the hybridization buffer, and hybridization was performed for 1 hour at 62° C. After hybridization, discs were washed once for 5 minutes at 62 ° C in wash buffer (4X SSC, 0.1% v/v Tween-20). Discs were then incubated with 2.5 μg/mL 4′,6-diamidino-2-phenylindole (DAPI) (Invitrogen) in PBS + 0.1% Tween-20 for 5 minutes at room temperature. Discs were washed with PBS + 0.1% Tween-20 and transferred to Vectashield (Vector Labs) for mounting. Discs were mounted in 15 μl of Vectashield on glass microscope slides using an 18 X 18 mm No. 1 coverslip (Zeiss). For eye imaginal discs, discs were dissected from late 3rd instar larva in cold PBS with brain and mouth hooks attached, then smFISH was performed as described. Immediately prior to mounting, brain and mouth hooks were removed from eye discs and discarded.

Actinomycin D Treatment

Wing discs were dissected in room temperature Graces' Insect Medium (Sigma 69771) supplemented with 1X Pen-Strep (Gibco 15140-122) and 5 mM Bis-Tris (Sigma B4429). Half of the total dissected discs were transferred to 24-well tissue culture dishes containing this prepared media + 5 μg/mL Actinomycin D, and half were transferred to untreated controls containing culture media + 1:1000 (v/v) DMSO. Discs were incubated with gentle shaking for 30 minutes at room temperature, protected from light, before being washed with fresh culture media, and 1X PBS. SmFISH was then performed as described.

Microscopy

Widefield imaging: 16-bit 3D image stacks were collected on a Deltavision Olympus IX-71 inverted microscope using a EMCCD camera and a 100X oil immersion (NA 1.44) objective. Deconvolution was automatically performed according to default settings.

Confocal imaging: 12-bit 3D image stacks were collected with x-y pixel size of 76 nm and z-intervals of 340 nm on a Leica SP8 scanning confocal microscope, using a pinhole size of 1 Airy unit and a 63X oil immersion (NA 1.4) objective. DAPI, ATTO 565, and ATTO 633 were excited by the 405, 555, and 630 nm lasers, respectively. ATTO dye fluorescence was collected using a HyD detector on photon counting mode and a scanning speed of 200 Hz, with 16X line accumulation. DAPI fluorescence was collected using PMT detector using 8X line averaging.

Chapter III: Development of Image Analysis for Automated Detection of smFISH Data

Introduction

Image analysis seeks to automate the organization of visual data in a way that makes sense for completing certain tasks. Such tasks include: segmentation of image features into discrete objects, classification of different features of an image into categories, the reconstruction of 2D images into 3D, and tracking objects across time and space (Grande, 2012; Pavlidis, 1988).

In order to do so, imaging data can be understood as large data sets that take the form of 3D arrays of numbers. Therefore, they can be transformed, analyzed and classified in the same manner of any data array in order to perform image analysis.

Most image analysis tasks are performed adeptly and unconsciously by the human visual system. Therefore, our own perceptions are often a good metric for how well automated image analysis has completed a task. We compare the output of image analysis to a "ground truth," or unambiguous manually labeled image (Collins, 2007). It is important to keep in mind that if we cannot establish a ground truth, we will never have a strong metric for "how well" image analysis has worked. Sometimes there is a mistaken perception that image processing or analysis can "fix" poor imaging data. However, if you as a human researcher cannot provide the correct answers using your own eyes, you will never be able to train an image processing pipeline or deep learning network to do so for you, and you will need to focus on obtaining higher quality data or adjusting your preprocessing steps until you can do so.

Finding Objects and Feature Extraction

A critical step in any image analysis workflow is feature extraction and selection. This entails identifying the critical data features for performing the image analysis task. Often this is done via the use of various preprocessing filters (Collins, 2007).

For smFISH data, the goal is to perform varying forms of segmentation, first for RNA spots, and then for nuclei. This task requires the classification of pixels in an image into "object" and "background" pixels, followed by clustering them into discrete objects. In other words, the value of grey pixel values is changed into only one of two categories: 0 or 1. The most simple way to achieve this is to call a threshold value, where all pixels with values above the threshold become 1 and all with values below the threshold are transformed to zero. This simple mechanism often does not provide robust segmentation due to insufficient contrast between signal and background, uneven image textures within objects themselves, or uneven background (Grande, 2012).

There are several methods for preprocessing images that can deal with these problems. If background is uneven in a predictable and consistent way, a control background image can be subtracted from the test image to increase contrast. Similarly, images can be added or averaged together. The value of each individual pixel can also be changed according to a nonlinear function in order to artificially increase the contrast of an image (Russ et al., 2018).

Alternatively, there is a class of neighborhood processing functions, that perform a calculation using a moving kernel of pixels. For example, a median filter changes each pixel to the median of the pixel values in a kernel surrounding it. This can be used to blur or reduce the texture of objects that may be interfering in segmentation. Edge detection filters tend to have the opposite effect, enhancing edge features in an image (Collins, 2007; Russ et al., 2018).

Once object and background pixels are satisfactorily discriminated, the pixels must be organized into discrete objects. This is sometimes straightforward if an object's pixels are all connected and objects are far enough apart. However, objects in the image may need to be

dilated, eroded or filled in in order to achieve this. Nuclei tend to follow this pattern as they are often touching each other significantly and have uneven internal structure.

Deep Learning for Image Analysis

A form of artificial intelligence known as deep learning has been extremely helpful to the image analysis community in that it allows automation of the feature extraction process. Some deep learning uses a neural network, or interconnected layers of functions. Data is input into the network and passes through a series of functions or hidden layers before making a prediction, which is then compared to the true output. This is performed iteratively until the network has been trained to perform the image analysis task to some satisfaction. This training is then tested with one or more separate validation sets of data. Different deep learning engines use different types of network architectures (Moen et al., 2019). Network architectures known as convolutional neural networks (CNNs) have met with the most success in the image processing field. CNN networks contain convolution layers and pooling layers. A convolution layer convolutes input images and obtains image features. The pooling layer compares multiple images and selects relevant features (Moen et al., 2019). The result is automated feature extraction before moving onto the classification step. Additionally, some networks have utilized transfer learning to broaden the utility of a trained network. This entails the use of an existing trained and validated deep learning network, applied to a new task. Provided the new task is similar enough, the network will only need a small amount of training for it to achieve the new task. This has been particularly helpful for cell segmentation tasks as the network can be trained on “easier” data and then adapted to more challenging image tasks and styles. Prior to deep learning, cell segmentation for each data set required a unique, user-defined set of preprocessing tasks that can achieve robust segmentation of that data set but are mostly useless for any other

data (Minaee et al., 2020). With deep learning networks and transfer learning, researchers have been able to produce networks that are trained to identify a wider variety of cell shapes and image styles.

For my image data, once I could clearly delineate a ground truth for what was an RNA spot and what was not, I was able to do my data analysis robustly using a simple threshold for segmentation of RNA spots. Due to the clustered nature of nuclei in the imaginal disc, and the uneven features of DAPI staining, I chose to use a CNN network trained for segmentation of imaged nuclei. While still producing imperfect results, these image analysis modalities still allow me to make robust conclusions about my data (see Chapter IV).

Results

Accurate smFISH RNA Spot Detection

My first step in analysis of the smFISH data was to accurately and reproducibly segment RNA spots. I faced several challenges when doing so. First, my data contains a high dynamic range of pixel intensities. Many transcription sites contain pixels with values an order of magnitude higher than pixels in mature RNA spots (see Figure 1A). This is very confusing for many pre-existing segmentation tools, which appear to perform best when each class of objects have similar properties in terms of size, shape, and pixel intensity. From a qualitative assessment, I observed that filters that optimized for RNA spots tend to capture a lot of artifactual light from transcription sites, whereas filters optimized for transcription sites disregard and misidentify many RNA spots. From the outset, it seemed apparent that identifying nascent and mature RNA spots might require somewhat different segmentation processes.

Secondly, mature RNA spots are very small and potentially close together. There is a high propensity for segmentation tools to segment “dumbbell”-like objects when two RNAs are

in close proximity in space, even when these objects appear spatially resolved to a human observer (as in Figure 1B). This undercounts the number of objects and interferes with the quantification of properties of those objects. Thirdly, there is some amount of variation in pixel intensities from image stack to image stack. This could be due to disc-to-disc variability, as well as variability in the laser intensity of the confocal microscope, which is somewhat dependent on how long the lasers have been on.

Initial attempts to segment RNA spots were performed with a machine learning tool called Ilastik, an open-source algorithm that specializes in Random Forest classification tasks. Ilastik was attractive in that it relies on a "point-and-click" GUI wherein the user classifies pixels and clicks on example pixels from their data. This can be a fast, user-friendly, and reliable method for segmentation. Ilastik then produces a segmentation based on this user input, algorithmically determining the types of filtering and classification that will best give the output the user has requested. In an ideal scenario, the user trains a filter in Ilastik that can then be automatically run on each image without any additional input from the user. This filter can be saved and provided to future users of the image processing program for reproducibility.

I sought an alternate approach for several reasons. Due to the variability in image properties between data sets, Ilastik trained on one image rarely performed very well when used to segment another image. Most filters were overfit, or too specific for general use. I reasoned this would be a particular problem for comparing different data sets. Sets of images in which smFISH was performed on the same day and in which microscopy was performed close together were more alike than separate experiments. However, I still wanted to be able to reliably compare them and reduce the impact of experimental variability on image processing results. The other major issue was that when different users trained Ilastik, we got somewhat different

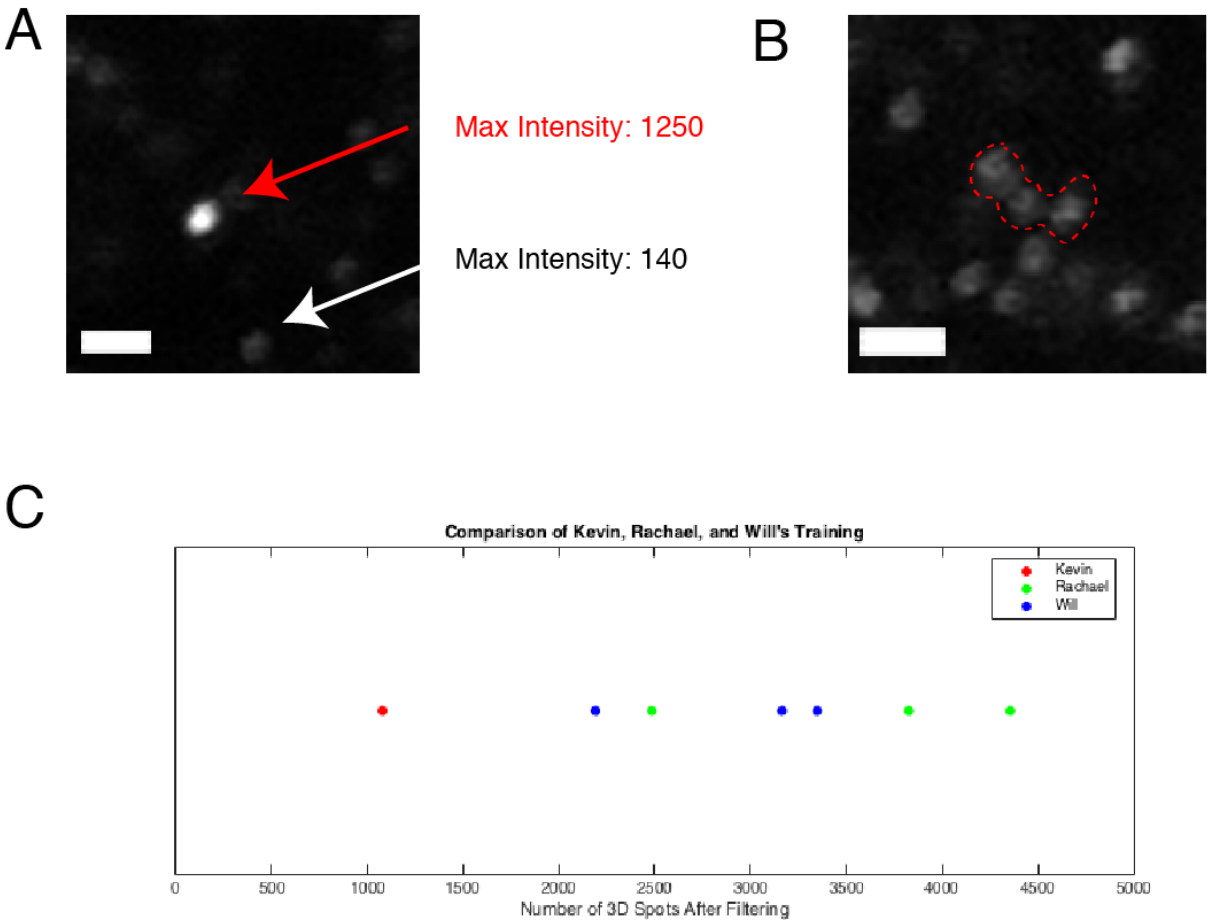


Figure 1. Challenging aspects of smFISH image analysis. (A) Confocal section of smFISH illustrating a transcription site spot (red arrow) relative to a single mRNA spot (white arrow). Scale bar= 1µm. **(B)** Confocal section of smFISH illustrating several mRNA foci within close proximity to one another (dashed red outline). Scale bar=1µm. **(C)** Comparison of number of objects identified from a single smFISH confocal image stack after filtering via an Ilastik filter generated by manual training. Each point represents a separate Ilastik training. Points are colored according to the individual who performed the training. After training and filtering via Ilastik, images were all segmented using an identical spot detection function in Matlab.

results. Even though we agreed on what "should" be a real spot (ground truth), we all had different criteria for when an Ilastik image was adequately trained for obtaining this ground truth (see Figure 1C for example). This was a concern in that I felt I could not reliably describe what criteria I was using to train Ilastik in a reproducible way. Ergo, since a single filter could generate reproducible and reliable data across images, and I could not train separate filters for images in a reproducible manner, I chose to pursue a different route for segmentation. Instead, I applied a simple threshold to my images, wherein objects above the threshold are considered spots and objects below the threshold are considered background. Each group of connected pixels are counted as one object. For each individual z-slice or optical section, I sweep across a range of thresholds and count how many objects are identified at each threshold. At low thresholds, the number of objects increases rapidly as many background pixels are above the threshold value and counted as objects. However, as the threshold goes higher, there is a point of inflection where the number of objects becomes relatively constant over a range of threshold values (Fig 2A). This range of threshold values where the number of objects is constant, identifies a number of objects similar to the number of objects that were manually identified (the ground truth) (Figure 2C). Furthermore, the centroids of the computationally identified objects were an average of 2 pixels distance from the centroids of the manually labeled objects. I concluded that this inflection point corresponds to a break between the background pixel intensities and signal pixel intensities, and that it can be used as a reliable indicator of an appropriate threshold for segmenting spots. Furthermore, I can independently determine an optimal threshold for each individual image stack, minimizing the effects of experimental variation in pixel intensity. In practice however, the necessary threshold for analyzing images within a single experimental condition or group tended to be quite close in value. Additional

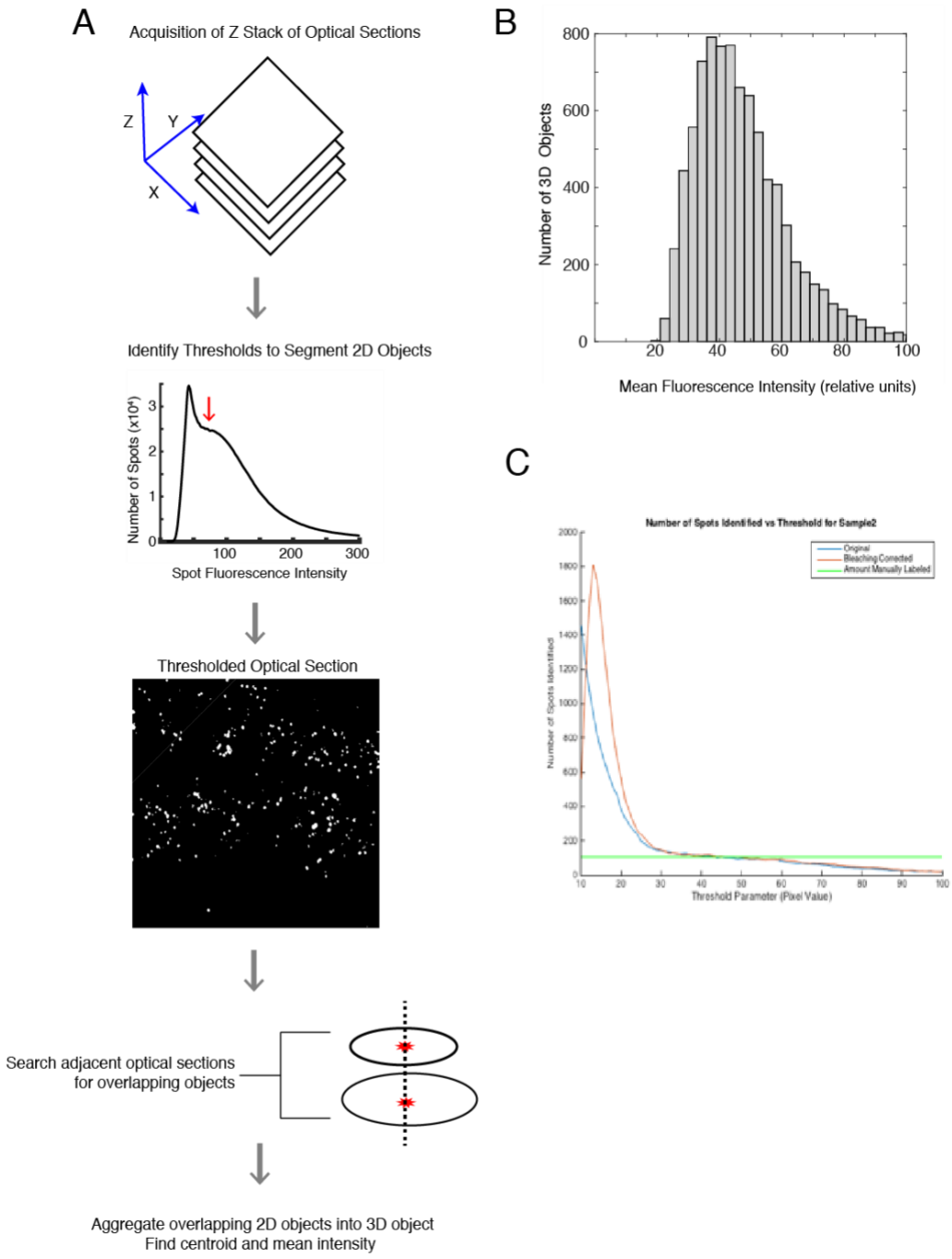

A
Acquisition of Z Stack of Optical Sections
Z
Y
X
Identify Thresholds to Segment 2D Objects
Number of Spots (x10⁴)
Spot Fluorescence Intensity
Thresholded Optical Section
Search adjacent optical sections
for overlapping objects
Aggregate overlapping 2D objects into 3D object
Find centroid and mean intensity
B
Number of 3D Objects
Mean Fluorescence Intensity (relative units)
C
Number of Spots Identified vs Threshold for Sample2
Original
Bleaching Corrected
Amount Manually Labeled
Number of Spots Identified
Threshold Parameter (Pixel Value)

Figure 2. Development of smFISH imaging and analysis. (A) Imaging and analysis pipeline to quantify mRNAs as 3D fluorescent objects. **(B)** Distribution of mean fluorescence intensity for all identified fluorescent objects from one wing disc expressing sfGFP-Sens mRNAs. **(C)** Representative distribution of the number of fluorescent objects identified at a range of threshold values for a subsection of a confocal z-stack. Fluorescence objects were also manually identified for this test set. The number of objects identified manually is denoted by a green horizontal line.

filtering was applied to the identified spots in order to address spots appearing in multiple z-planes and other potential problems (see Methods for details).

Transcription Site Identification

A histogram of the intensity of identified spots highlights a unimodal distribution of intensities (Figure 3A). This suggests that most spots are comprised of a single molecule of mRNA. However, the distribution has a long tail. This highlights the challenge of identification and quantification of spots of nascent RNA. Potential nascent RNA spots will occupy a wide range of intensities, from being nearly equivalent to mature mRNA spots, to being many times brighter than them. I decided to independently classify spots that were transcription sites for analysis. I did so by selecting a new threshold at which to segment spots. An appropriate threshold would be one that excludes single molecule mRNA spots but includes spots with greater than twice the intensity of single molecule spots. Simply doubling the segmentation threshold did not achieve this outcome. The value would often include many single RNA spots because they contain a small population of brighter pixels. Instead, I empirically determined that 2.5X the median value of a single RNA spot was an appropriate threshold to segment spots of nascent RNA for most datasets (see Figure 3B for example).

I then normalized nascent RNA spot intensity by dividing the intensity of each nascent transcription site by the median single RNA spot intensity (Figure 3C). This provides an estimate of the number of nascent RNAs each transcription site contains, although this estimate is likely inaccurate. It also allows me to compare the intensity of transcription sites across samples within a dataset and across datasets, since the value is not dependent on the raw pixel values in the image.

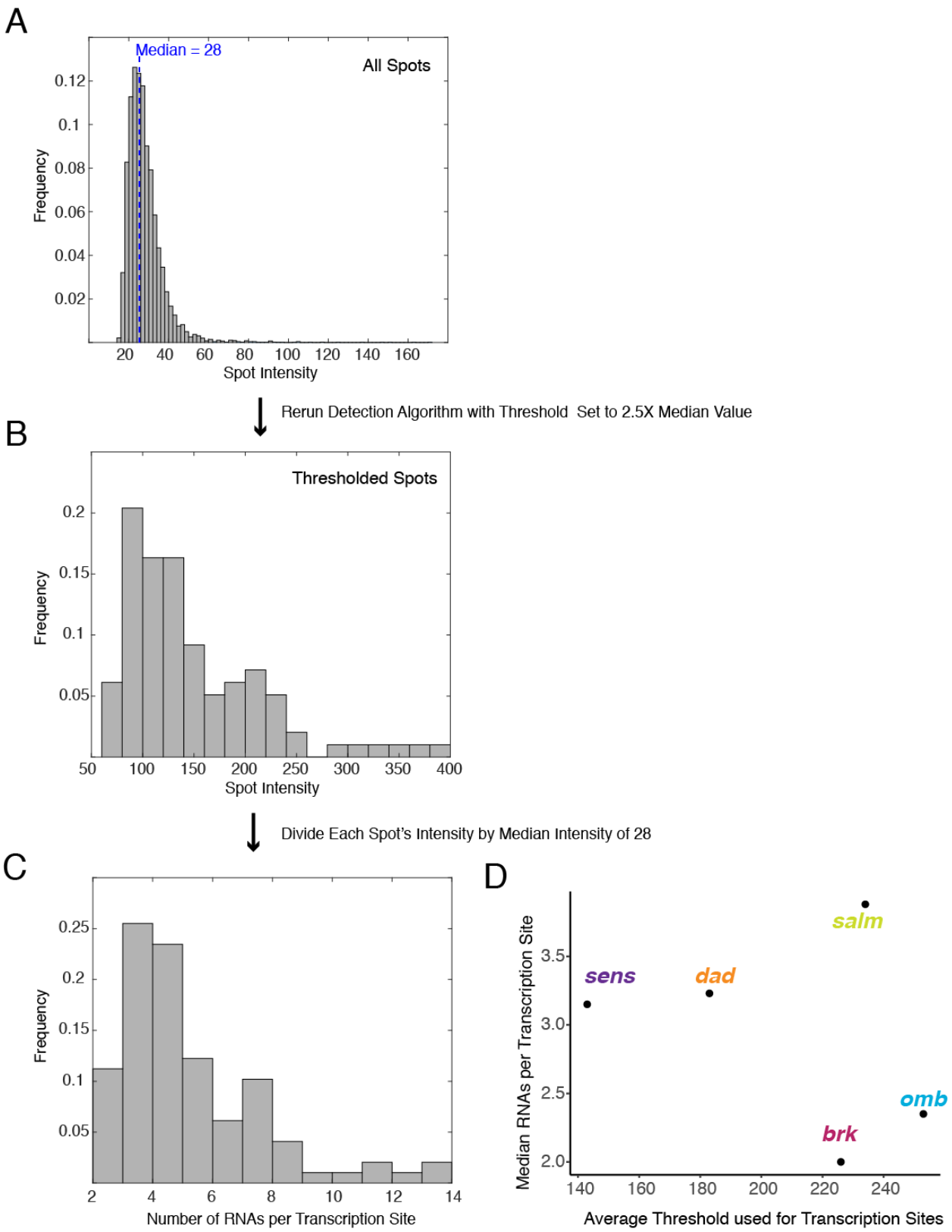

A
Median = 28
All Spots
Frequency
Spot Intensity
Rerun Detection Algorithm with Threshold Set to 2.5X Median Value
B
Thresholded Spots
Frequency
Spot Intensity
Divide Each Spot's Intensity by Median Intensity of 28
C
Frequency
Number of RNAs per Transcription Site
D
salm
sens
dad
omb
brk
Median RNAs per Transcription Site
Average Threshold used for Transcription Sites

Figure 3. Detection of transcription sites and their quantification. (A) A representative frequency distribution of fluorescence intensity for 3D spots identified in one wing disc expressing *sfGFP-sens*. The median intensity is 28 units. **(B)** The same wing disc was reanalyzed for 3D spots but using a threshold of 70 units as a cutoff, below which spots are not counted. **(C)** The fluorescence intensity of each 3D spot in B is divided by the median intensity of 28 units to provide a normalized number of RNAs that are localized to that 3D spot. This is not an actual number of RNA molecules but the output from partially transcribed RNAs annealing to a variable number of probes depending on the composition of binding sites in the RNA composite. **(D)** The mean threshold used for transcription site identification for each data set plotted against the median normalized RNA molecules per transcription site for all transcription sites in that data set.

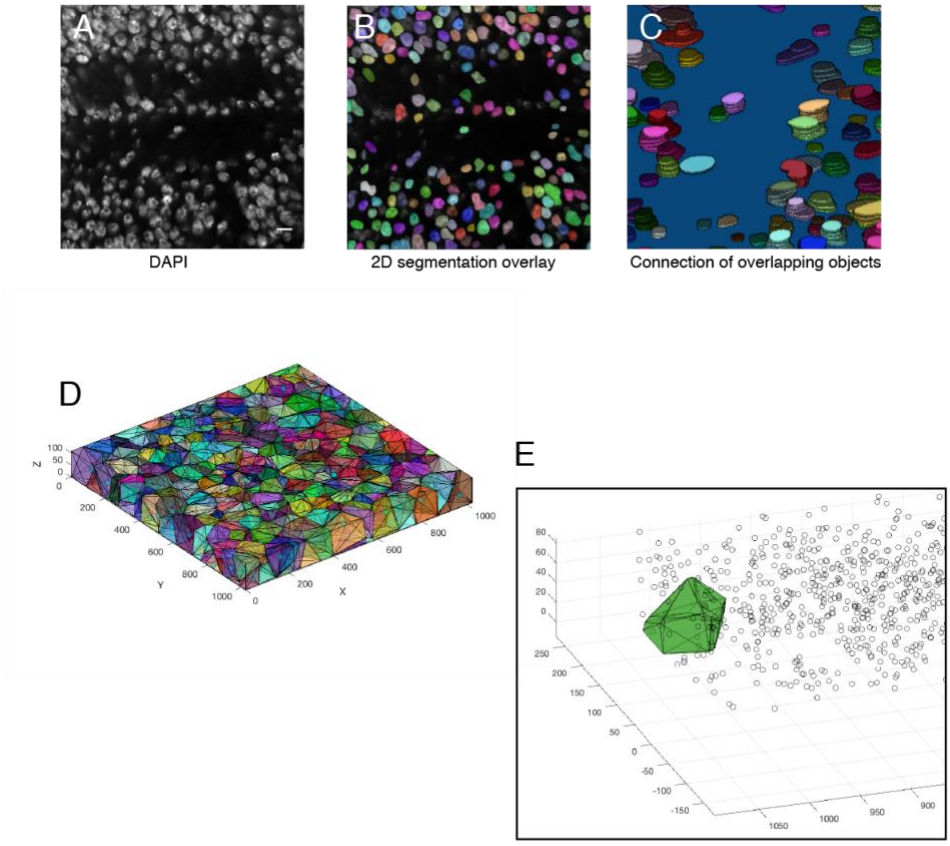


Figure 4. Pipeline for 3D segmentation of cell nuclei. (A) An optical section showing DAPI fluorescence. **(B)** 2D segmentation of this image. **(C)** 3D segmentation by connecting 2D objects in neighboring sections that overlap with one another in the x-y plane. **(D)** 3D Voronoi tessellation of an image stack. The centroids of segmented nuclei (shown as circles) were used to tessellate the image stack, creating virtual cells. Cells are represented with different colors. Numbers in the x-y plane refer to pixel positions in the 1024 x 1024 sections. **(E)** An image stack showing the centroid positions of 3D mRNA objects as circles. One tessellated cell (green) is superimposed to show the mRNA objects that reside in space occupied by the tessellated cell. These mRNAs would be assigned to that particular cell.

Nuclear Segmentation

In order to assign RNAs detected by smFISH to particular cells in the disc tissue, I needed a way to define cell boundaries. Due to the methanol fixation, plasma-membrane associated proteins were not detectable by innate fluorescence or immunofluorescence. Dye-based membrane stains that rely on interactions with lipids in the plasma membrane have not shown a great deal of success in *Drosophila* imaginal discs. The dense, columnar nature of the epithelial cells in the discs makes such staining difficult to interpret and analyze. However, I could segment nuclei, and so I explored several options. Attempts were made to stain for nuclear lamin but were unsuccessful under the fixation conditions in which smFISH was performed. Given these results, I decided to use DAPI as it provided the most robust delineation of nuclei. I segmented each optical section of DAPI fluorescence using NucleAIzer, a masked R-CNN deep learning network (Hollandi et al., 2020).

The network has been trained to do nuclei segmentation on a variety of styles of microscopy images. I used the trained network provided for small fluorescent nuclei, so I had to do no further training or validation, and the results were reproducible (Figure 4 A&B). This offered an improvement over previous methods used in the lab, which use a meanshift-root algorithm (Qi et al., 2013). Segmentation using a mean-shift root identified approximately 75% of nuclei segmented by hand, whereas NucleAIzer identified approximately 85% of nuclei segmented by hand. Regardless of method, segmentation identified fewer nuclei than a manual counting, indicating that these segmentation methods are likely combining or missing nuclei at a greater rate than they are inappropriately splitting nuclei.

Each optical section only contains a slice of one nucleus, and other sections contain other slices of the same nucleus. I did not want to over-count nuclei and so I needed to reconstruct

each nucleus as a 3D object. In order to connect the 2D nuclear objects in each z-section to homologous objects in other sections, I adapted a cell-tracking script to use in z-stacks. The computer identifies the maximum overlapping object in the next z-slice for each segmented object and links these together into one object (see Figure 4C for visualization of this).

Assignment of RNAs to Cells

The next step in analysis was to decide how to assign RNAs to respective cells when I only had identified the nuclei of cells and not their membrane boundaries. I wondered if there was any effect of assigning RNA spots to their nearest nuclei compared to dividing the tissue into random "cell" volumes. I divided a test stack into a 16 X 32 grid of equally sized volumes and summed the RNAs in each cell volume (Figure 5A). I compared this to a two-dimensional Voronoi diagram wherein each cell corresponded to a segmented DAPI nuclear centroid (Figure 5C). RNA spots located within each Voronoi cell were summed. The notable difference between these two methods is that the random grid of cell volumes is an impartial analysis of RNA density agnostic to the number or position of cells in a region. For the Voronoi diagram, cells in areas of the tissue that are more sparse have a greater volume than areas that contain many nuclei.

Both methods produce distributions of RNA per cell with similar median values. The distributions are both skewed towards 0 RNAs per cell with a tail of larger values, although the Voronoi cells produce distributions with a greater variation (Figure 5 B, D and E). This difference is grounded in a reality in the imaging data: the regions of the tissue with the most RNA spots also have fewer nuclei present. This is largely due to the distribution of Senseless in two morphologically distinct stripes, and differentiation of the sensory organ precursor (SOP) cells. Ultimately, despite its flaws, the Voronoi tessellation provides a valuable calculation:

normalization for the number of cells in different regions of the tissue. Areas with fewer cells should reasonably have more RNAs per cell.

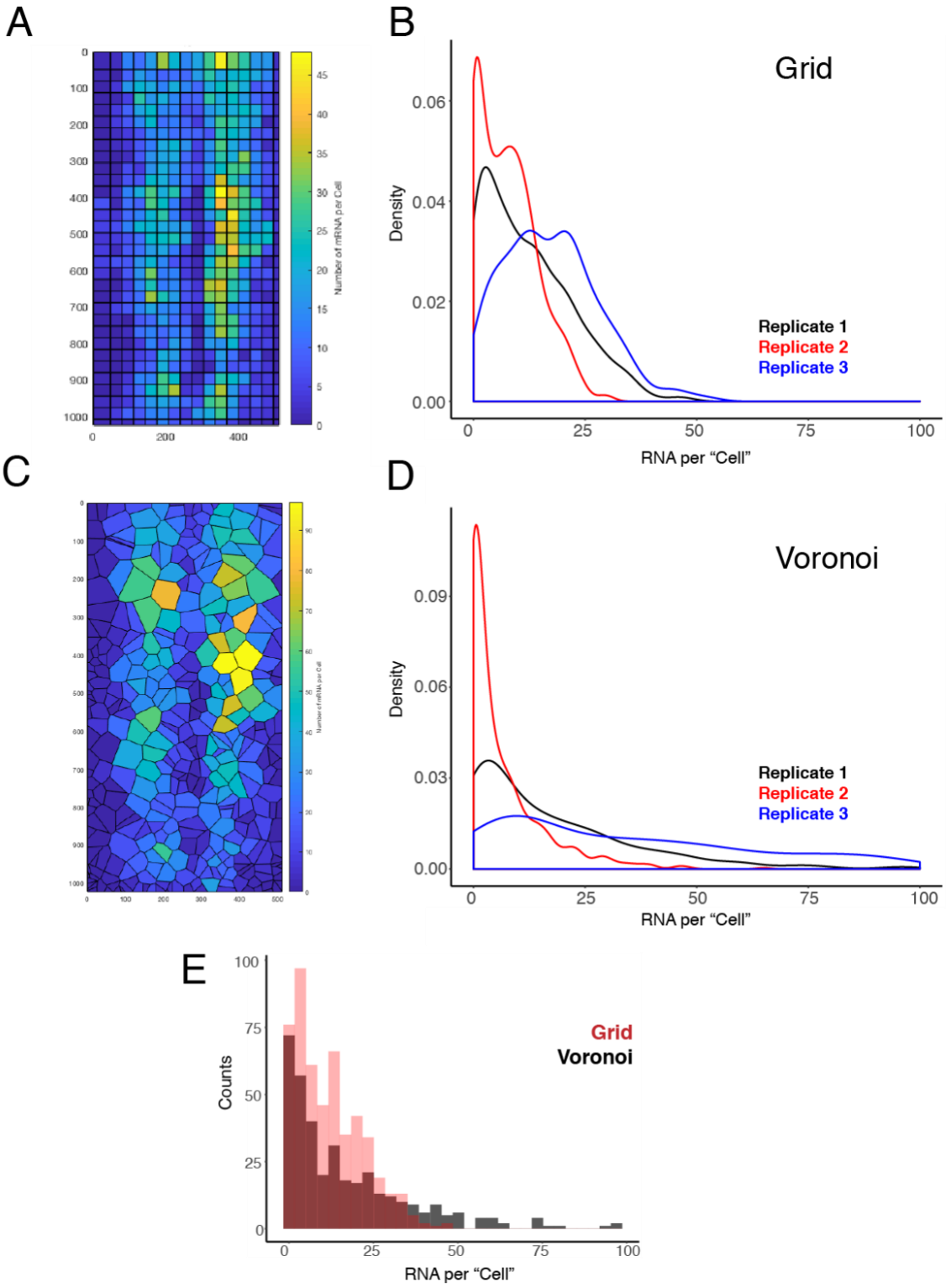

A
B
Grid
Density
RNA per "Cell"
Replicate 1
Replicate 2
Replicate 3
C
D
Voronoi
Density
RNA per "Cell"
Replicate 1
Replicate 2
Replicate 3
E
Grid
Voronoi
Counts
RNA per "Cell"

Figure 5. Comparison of 2D Grid and Voronoi methods (A) Representative heat map of a sample disc partitioned into equally spaced cubes. RNA fluorescence objects were assigned to the nearest cube region. Regions with warmer colors were assigned more RNA fluorescence objects. **(B)** Smoothed distributions of RNA per spatial region for three test z-stacks of images. Each distribution represents one disc. **(C)** Same imaginal disc partitioned using a Voronoi diagram based on the centroids of segmented nuclei. Regions with warmer colors were assigned more RNA fluorescence objects. **(D)** Smoothed distributions of RNA per spatial region for three test z-stacks of images. Each distribution represents one disc. These are the same test image stacks analyzed in (B). **(E)** Combined distributions for all three replicates, comparing a grid (red) and Voronoi (grey) methods of partitioning the tissue.

Discussion

Limitations of Transcription Site Detection

In the data, spots of nascent RNA are treated in the same manner as mature RNA spots: the largest object in the 3D stack of objects is counted, and a 2D circle is drawn about it in order to extract an intensity measurement. This underestimates the intensity of nascent RNA spots for several reasons. First, I measure intensity within the diffraction limit, whereas many nascent RNA spots appear to be somewhat larger objects than the diffraction limit. Second, due to their small size, mature mRNA spots only occupy 2-3 z-planes, and are usually only present as a measurable spot in one z-plane. Spots of nascent RNA can occupy more z-planes than this because they are larger. The result of these limitations is that subtle changes in nascent RNA intensity may not be detectable through this analysis. This analysis has been sufficient for the conclusions I have drawn discussed in Chapter IV.

However, it has limitations. The *sfGFP-sens* transgene landed in two different loci produce statistically identical data using this analysis (Ch. IV Figure 11). However, analysis of protein data of these two constructs predicts that they would show a small difference in number of nascent RNAs present at the transcription site (Giri et al., 2020). However, I see no statistical difference in the intensity of nascent RNA spots. Perhaps this difference does not exist as predicted. Alternatively, the experiment is not precise enough in order to produce data sufficient for asking the question. It is possible that differences between the two landing sites exist, yet simply are too small to be detected by the analysis presented here.

Thus, when trying to ask questions that deal with very subtle differences between sets of data, one should draw intensity measurements from entire 3D objects or from all included pixels below threshold. I did not implement this because including intensity data from a variable

number of pixels (depending on intensity and threshold) makes the intensity measurements very sensitive to the selection of threshold. This amplifies small differences between replicates, diminishing reproducibility. If one wanted to implement this data analysis, great care would need to be taken to standardize the threshold used for all replicates and conditions. One would expect there to be greater variation between replicates using this method as well.

Limitations of Cell Assignment

As discussed above, there are significant challenges in assigning RNAs to cells in the imaginal disc, particularly with respect to smFISH. Nuclear markers are the most robust way to mark cells in this tissue but given the complex columnar nature of the epithelial cells, assigning cells to their nearest nuclei as I have done is not a particularly accurate method of determining exact RNA counts per cell. The method I have chosen normalizes RNA numbers in a region to the number of nuclei in that region and estimates the variability of RNA numbers from region to region. Although errors will occur on a cell-by-cell basis, the errors will be systematic and therefore, large-scale trends will not be affected by using this method. This method may not be appropriate to draw conclusions about data that are highly dependent on accurate cell assignment (like calculating RNA noise per cell).

The best way to improve cell assignment, in my view, would be to develop a reliable method of membrane staining and 3D cell reconstruction in the imaginal disc. This presents many challenges. Finding a robust membrane stain that is compatible with methanol fixation is a significant challenge. Alternatively, dyes that interact with lipids could be used. Also 3D reconstruction of cells is a non-trivial task. Alternatively, experiments could be performed only on RNAs that are localized to one part of the cell. For example, *wingless* mRNAs are apically located, making individual cells fairly easy to pick out and segment using the RNA channel

alone. Other gene constructs could be produced that also apically localize RNA for analysis. Finally, I suggest that we keep in mind alternative systems in which to ask similar questions. *Drosophila* embryos and ovaries may be more simple to manipulate and image. Other systems should be considered on a case-by-case basis, depending on what manner of question you want to ask.

Materials and Methods

Image Processing

Raw smFISH images were processed using a custom MATLAB script with no prior preprocessing. My pipeline is available at github.com/bakkerra/smfish_pipeline.
The pipeline consists of several modules:

Selection of mRNA Segmentation Threshold: Spot segmentation is performed by applying a threshold value to an smFISH image and transforming all pixels above the threshold ('objects') to white and pixels below this threshold ('background') to black. To robustly identify RNAs, it is therefore important to select a threshold where real RNA fluorescent spots are above the threshold, and background fluorescence is below the threshold. Using this threshold method, I classify an object in each 2D image when white components have a connectivity of 8 pixels or more. When the number of objects in an image stack is counted across a range of segmentation thresholds, the number of objects reaches an inflection point and plateaus at a threshold approximately equal to the level of fluorescence that separates real RNA spots from background (Figure 2).

I manually identified and labeled 347 RNA spots from sub-regions of four independent image stacks and found that when a threshold is selected within the plateau after the inflection point, the number of objects identified is no more than +/- 5% different than the ground truth

manual curation. Furthermore, the centroids of identified objects have an average displacement of only 2 pixels from the manually identified centroids. Therefore, this plateau is an appropriate threshold for accurate segmentation of RNA spots.

I reasoned that the position of inflection might vary from sample to sample. Therefore, for each image stack, a range of thresholds is tested, and a threshold is selected within the plateau to collect the segmented data. As a result, each image stack has the potential for a unique threshold, allowing robust segmentation of spots despite variation in raw fluorescence between image stacks. In practice, replicates from the same experiment captured in the same imaging session did not require thresholds for segmentation more than 15 fluorescent units apart. If image stacks did not show an identifiable inflection point and plateau, the signal-to-background of that sample was determined to be insufficient and it was not used for analysis. The smFISH protocol and imaging is robust enough that in my hands, this occurs in less than 10% of image stacks collected. Once a threshold is selected, the following properties of each object are recorded: x-y centroid position, z-plane, and a list of the connected pixels.

Connecting Segmented Objects into mRNA Spots: Diffraction-limited fluorescent spots captured with the 63X objective at 633 nm wavelength are estimated to be approximately 600 nm in diameter. This corresponds to a diameter of 8 x-y pixels in my images (Lipson et al., 1995). As each z-plane is 340 nm in depth, it is assumed that genuine diffraction-limited RNA spots will appear in 2 or 3 consecutive z-planes, depending on the spot's position along the z-axis. Therefore, candidate RNA spots must satisfy two criteria in order to be counted:

1. Candidate must have a corresponding object centroid at least one neighboring z-plane within a diffraction limited radius of 4 pixels.

2. Candidate must be larger (contain more pixels) than corresponding objects in neighboring z-planes. This criterion prevents RNA fluorescence spots from being counted in multiple z-planes.

A candidate that satisfies these criteria is recorded as an mRNA spot, and only the largest 2-D object is recorded.

The analysis program shows the FISH images overlaid with markers indicating recorded spots so that each image stack can be manually inspected for any significant errors or inconsistencies. The most common problem detected at this stage resulted from images taken of discs that were "drifting," or moving significantly between z-slices, which can cause a large number of identified spots to be filtered out during processing for not meeting criterion 1. Excessive bleaching across the z-stack can also cause clear inconsistencies. In this study, such problems were rare enough that any sample experiencing these problems was considered to have failed quality control and was simply not included for further analysis.

Intensity measurements are recorded from a circle of pixels of radius 4 about the centroid of each recorded RNA spot. By keeping the area of each intensity measurement fixed, I uncouple user-generated variation in selection of segmentation thresholds from spot intensity measurements. A 2D circle was used instead of a 3D sphere to extract intensity measurements because the spots only appear in 2 or 3 z-planes. This makes their 3D geometry variable from spot to spot, and they cannot be consistently described using a sphere or ellipse.

***Segmentation of Transcription Sites*:** In my images, nascent RNA spots tend to contain pixels that are many times brighter than mature RNA spots. As a result, the brightest transcription sites are frequently misidentified during segmentation of mature RNA spots because the second criterion for spot identification only records the largest object within the diffraction limit in z.

For transcription sites, this is not always the brightest plane. Therefore, I segment transcription sites independent of mature RNAs using a higher threshold.

The objective in threshold selection for transcription sites is to select one that includes objects with a total fluorescence intensity of twice the average mature RNA and excludes mature RNA spots. I define the "average" intensity for a spot containing a single mRNA to be the median of the distribution of all identified mature RNA spot objects. I empirically determined that merely doubling the threshold for segmentation does not achieve this, because mature mRNAs may contain a few pixels above the threshold, enough to still be identified as objects and included in analysis. Therefore, I use a threshold calculated by multiplying the median mature RNA intensity by a factor of 2.5 (Figure 3).

To test the accuracy of this segmentation procedure, I manually inspected three particularly RNA-dense regions in independent images where automated segmentation found a total of 103 transcription sites and 4,066 mature RNAs. I determined that only 7 of 4,066 mature RNAs were misidentified as transcription sites, and found no examples of transcription sites that had been missed by automated segmentation.

After identification, object intensity measurements are recorded from a circle of pixels of radius 4 (the diffraction limit) about the centroid of each identified transcription site (Figure 3B). The average transcription site threshold selected for replicates in a dataset show no correlation with the average intensity of transcription sites in that dataset (Figure 3D). Therefore, the differences in transcription site intensity between genes cannot be explained merely by differences in threshold selection or variability in image fluorescence between datasets.

Estimation of Nascent RNA Number per Transcription Site: The intensity measurement of each identified transcription site in an image stack is divided by the median intensity of

identified mature RNAs in that sample (Figure 3). This serves two purposes. First, it serves to normalize these measurements within each sample so transcription site intensity measurements can be pooled across replicates without the effects of image-to-image variability in fluorescence. Secondly, each transcription site object is presumed to be the sum of intensities of multiple nascent RNA molecules elongating at the transcription site. By dividing each transcription site intensity by the average intensity of a single RNA, I obtain an estimate of the number of nascent RNAs present at the transcription site. Because some transcripts are partially elongated, this number cannot be completely accurate, and I attempt to compensate for this in my computational model when interpreting results.

Nuclei Segmentation: DAPI fluorescence images are output as labeled 16-bit images, where each nuclear object corresponds to a 'level' in the 16-bit image. These images are input to a nuclei segmentation pipeline, which flattens the images to white nuclei objects and black background. Nuclei images are segmented in 2D using the NucleAIzer platform maskRCNN Network, trained as described in (Hollandi et al. 2020) This is available online at nucleaizer.org. It requires the user to define an expected nuclear radius, which I set at 32 pixels (Figure 4B). To ascertain the accuracy of segmentation, I compared results to manually labeled nuclei in four randomly selected disc images. The automated method identified at least 85% of nuclei objects identified manually for each image.

The segmented black and white images are then processed using a custom Matlab script in order to join overlapping 2D objects into 3D. Each nucleus object in each z-slice is assigned an identity index. For each object in the first z-slice, the object with the highest number of overlapping pixels in the next z-slice is identified, and this object's identity index is altered to be identical to its overlap partner. This proceeds through the entire z-stack of images, creating

objects that resemble ‘pancake stacks’ of linked 2D objects in 3D (Figure 4C). The 3D-centroid and list of included pixels of these new objects is then recorded. Objects not incorporated into a 3-D object are disregarded.

Generation of Voronoi Diagrams: A 3D Voronoi tessellation divides a geometric volume into spatial regions with boundaries equidistant from a set of points (Voronoi, 1908). I use this method to assign RNA objects to the nearest nuclear centroid. The set of segmented 3D nuclear centroids are used to divide the z-stack of images into a 3-D Voronoi tesselation using a polytope bounded Voronoi diagram available for Matlab, which uses the DeLaunay triangulation to calculate the Voronoi diagram (Park, 2020) . The result of this tesselation is a list of 3-D vertices of each Voronoi ‘cell’ in space, which is recorded along with the associated nuclear centroid (Figure 4D).

Assignment of RNA to nearest nuclei: Mature mRNAs and transcription spots located within a Voronoi spatial cell are assigned to that particular cell. To assign spot objects to cells, a 3D convex hull of the each Voronoi cell is constructed from the vertices data for that cell. An entire set of image points, either the mRNA or transcription spot centroids, are tested to determine whether they fall inside or outside of each hull (Figure 4E). This is performed using a Matlab function called inhull, which uses dot products to shorten calculation times (D’Errico, 2012) . Spots that fall inside a given cell’s Voronoi hull are assigned to that cell’s nuclear centroid, and the number of assigned spots, as well as their centroid and z-plane information are recorded. This is then repeated for every Voronoi cell in the image stack. The final result is a list of cells, their nuclear centroids, the total number of RNA spots assigned, and a list of each assigned spot’s centroids.

Chapter IV: Wg and Dpp Morphogens Regulate Gene Expression by Modulating Frequency of Transcriptional Bursts

The majority of work in this chapter is published at the online journal eLife (Bakker et al., 2020). I developed the methodology and performed all experiments and analyses. Mathematical modeling was conceptualized by me and Dr. Madhav Mani. The manuscript was written by me and Dr. Richard Carthew.

Introduction

Paracrine signaling is a highly conserved means for cells within a tissue to communicate with one another to regulate diverse activities including proliferation, differentiation, apoptosis, and movement. Many of these activities are mediated by changes in gene transcription that are brought about by reception of the signals. Paracrine factors acting as morphogens are a particularly important class of gene regulators. Morphogens form spatially-extended gradients from the source of their synthesis, and elicit different transcription outputs from target genes, depending on local concentration of the morphogen (Tabata and Takei, 2004).

Many paracrine signals regulate gene transcription via control of the availability or activity of sequence-specific transcription factors. Some transcription factors regulate assembly of the preinitiation complex (PIC) composed of Pol II and general factors at the transcription start site (Esnault et al., 2008). Other factors recruit coregulators that modify nucleosomes or remodel the chromatin architecture of the gene (Bannister and Kouzarides, 2011). However, transcription is a dynamic process, and thus, molecular models of regulation via PIC assembly or chromatin structure, do not adequately capture what kinetic steps in transcription initiation are being regulated. Recently developed methods have uncovered greater complexity in the transcription initiation process than previously imagined. Genes that are constitutively expressed rarely show uniform and continuous mRNA synthesis. Rather, mRNA synthesis occurs in bursts that are interrupted by periods of dormant output. This phenomenon is known as transcriptional bursting (Chen et al., 2019; Chubb et al., 2006; Dey et al., 2015; Raj et al., 2006; Suter et al., 2011)

Various studies have explored how mechanisms of gene regulation affect the size and frequency of transcriptional bursts, and thereby affect transcription output. The availability of transcription factors has been shown to affect burst frequency (Ezer et al., 2016; Larson et al., 2011; Senecal et al., 2014). For example, the *Drosophila* transcription factors Bicoid and Dorsal have been studied in great detail with respect to their effects on transcription burst frequency in the embryo (Garcia and Gregor, 2018; Holloway and Spirov, 2017; Little et al., 2013; Xu et al., 2015). Enhancer strength and enhancer-promoter contact correlate with burst frequency of genes (Bartman et al., 2016; Bothma et al., 2014; Chen et al., 2019; Fukaya et al., 2016; Larsson et al., 2019) . These studies altogether suggest that bursting frequency is potentiated by enhancer-promoter contact and is mediated by transcription factors binding to DNA.

In this study, I have explored how the Wnt protein Wingless (Wg) and BMP protein Decapentaplegic (Dpp) regulate transcription dynamics of genes in the *Drosophila* wing imaginal disc. The Wnt and BMP families of proteins are two highly conserved paracrine factors that can act as morphogens. In canonical Wnt signaling, the binding of extracellular Wnt protein to its transmembrane receptor Frizzled causes β-catenin to be stabilized and free to enter the nucleus, where it relieves repression of Wnt-responsive genes by binding to the sequence-specific transcription factor TCF (Clevers and Nusse, 2012; Swarup and Verheyen, 2012). In canonical BMP signaling, ligand binding to receptor triggers phosphorylation of SMAD proteins, which translocate to the nucleus along with co-SMADs, bind to responsive genes, and activate their transcription (Hamaratoglu et al., 2014; Shi and Massagué, 2003) .

To explore the effects of Dpp and Wg signaling on transcription dynamics, I have adapted single molecule fluorescent in situ hybridization (smFISH) for use in imaginal disc tissues. I use smFISH to quantify nascent and mature mRNAs for several genes expressed in

highly diverse patterns within the wing disc. Despite having different expression patterns, all of the genes are regulated by modulation of transcription burst frequency by Dpp and Wg.

Results

In this study, I have explored how the Wg and Dpp morphogens regulate transcription dynamics in the wing disc. Each morphogen is synthesized in a narrow stripe of cells within the disc. Wg is produced in cells at the boundary between Dorsal and Ventral (DV) compartments of the wing pouch, while Dpp is produced in cells at the boundary between Anterior and Posterior (AP) compartments (Figure 1A). These factors diffuse from their sources forming concentration gradients across the disc and regulate gene expression in a concentration-dependent manner.

smFISH detection of *Sens* mRNA

I first probed for expression of the *senseless* (*sens*) gene in the wing disc. *Sens* is required for cells to adopt a sensory organ fate, and the gene is expressed in two stripes of cells adjacent to and on either side of the DV boundary in the wing pouch (Figure 1B,C). *Sens* expression in the wing pouch is induced by Wg, which is expressed by cells located at the DV boundary (Nolo et al., 2000; Zecca et al., 1996) I probed for *sens* mRNAs expressed from a transgenic version of the *sens* gene. I did so for a number of reasons. First, the genomic transgene rescues the endogenous gene based on function and expression (Cassidy et al., 2013). Second, the transgene is tagged such that the amino-terminal coding sequence corresponds to super-fold GFP (sfGFP). By using oligo probes directed against sfGFP, I could easily determine the specificity of detection.

Samples were probed and imaged, revealing the expected pattern of fluorescence localized to two stripes adjacent to the DV midline in the wing pouch (Figure 1D). The abundance of *sens* mRNAs within the DV stripes varied from one to fifty molecules per cell

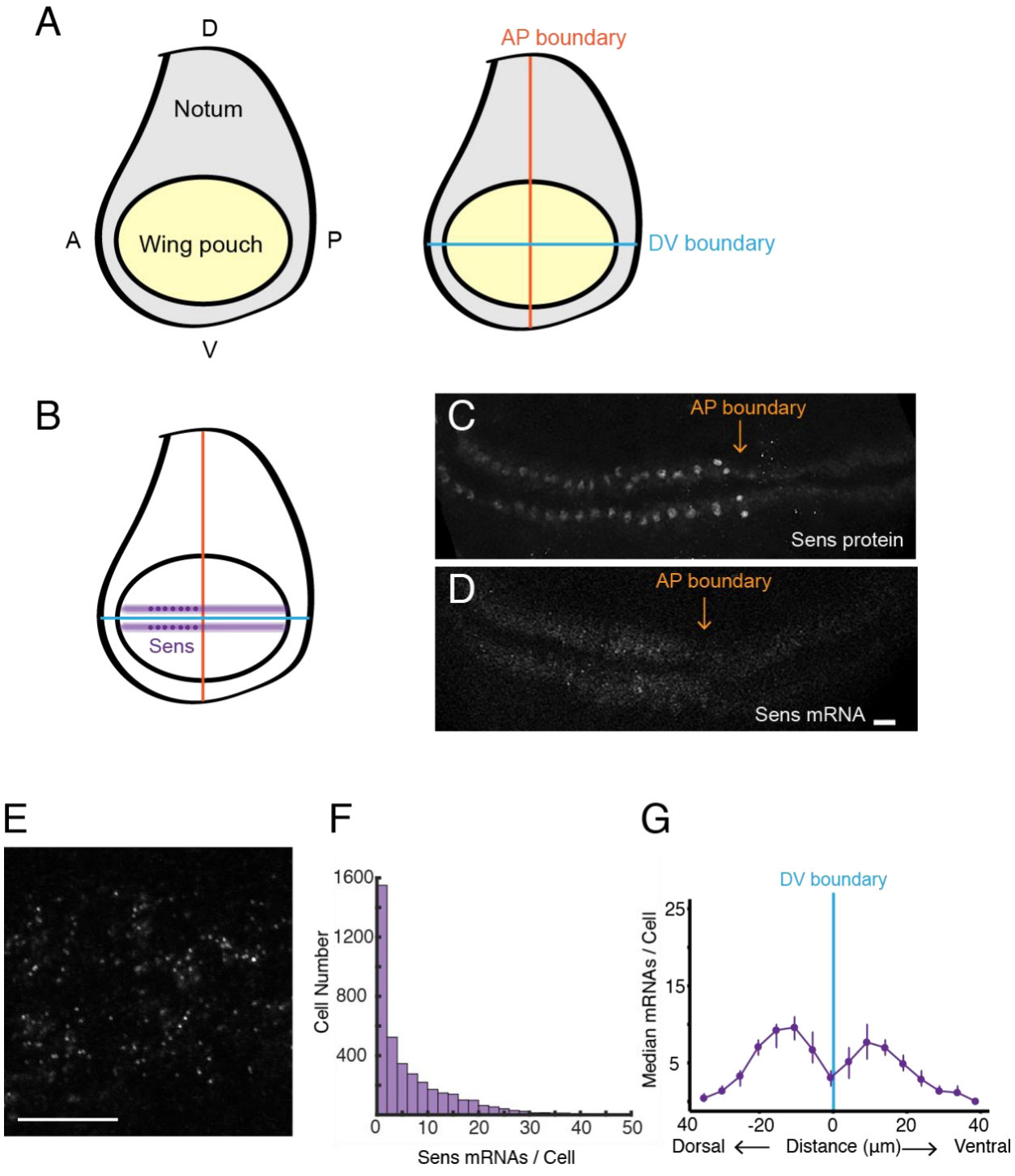

A
D
Notum
A
Wing pouch
P
V
AP boundary
DV boundary
B
Sens
C
AP boundary
Sens protein
D
AP boundary
Sens mRNA
E
F
1600
1200
800
400
Cell Number
0
10
20
30
40
50
Sens mRNAs / Cell
G
DV boundary
25
15
5
Median mRNAs / Cell
40
-20
0
20
40
Dorsal
Distance (μm)
Ventral

Figure 1. smFISH analysis of sfGFP-sens mRNA levels in wing imaginal discs. (A) Schematic of a wing disc outlining different regional domains, and the positions of boundaries betIen Dorsal (D) - Ventral (V) and Anterior (A) - Posterior (P) compartments of the disc. Each wing disc is composed of roughly 50,000 cells organized in a pseudostratified epithelium. **(B)** Schematized expression pattern for Sens inside the wing pouch centered around the DV boundary. Sens is also expressed in clusters of cells in the notum, which are not shown. **(C-E)** Confocal sections of wing discs expressing sfGFP-Sens. **(C)** sfGFP-Sens protein fluorescence. **(D)** sfGFP-Sens mRNAs as visualized by smFISH using sfGFP probes. Scale bar = 10 μm. **(E)** Higher magnification of sfGFP-Sens mRNAs as visualized by smFISH using sfGFP probes. Scale bar = 10 μm. **(F)** Distribution of wing disc cells as a function of the number of Sens mRNA molecules per cell. **(G)** Sens mRNA number as a function of cell distance from the DV boundary. The shortest path length from each cell centroid to the DV boundary was calculated. Cells were then binned according to this path length and whether they were dorsal or ventral cells. The median mRNA number/cell for each bin is plotted. Error bars represent bootstrapped 95% confidence intervals. A bimodal distribution captures the expression pattern of Sens.

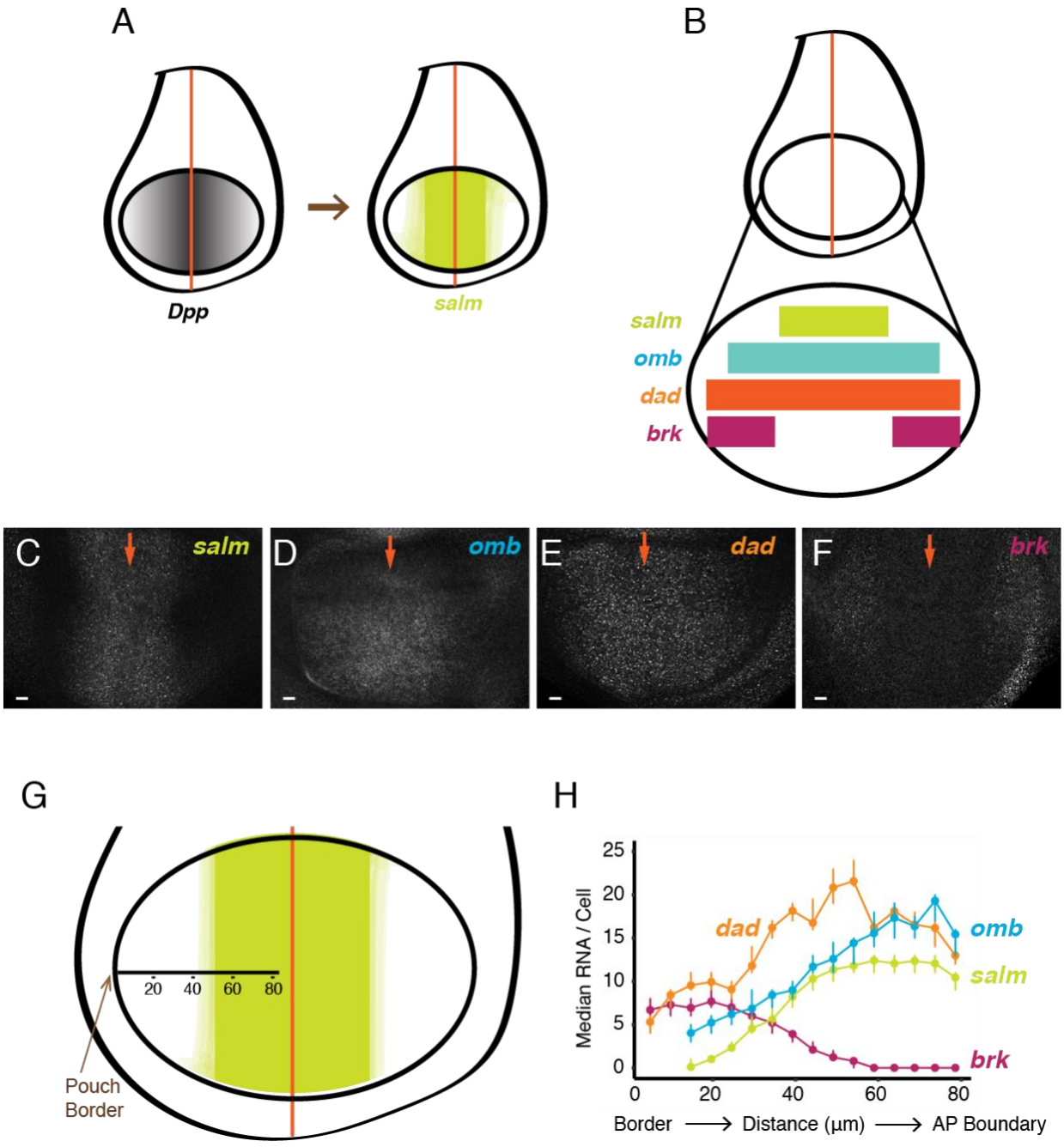

A
Dpp
salm
B
salm
omb
dad
brk
C
salm
D
omb
E
dad
F
brk
G
20
40
60
80
Pouch
Border
H
Median RNA / Cell
25
20
15
10
5
0
dad
omb
salm
brk
0
20
40
60
80
Border
Distance (μm)
AP Boundary

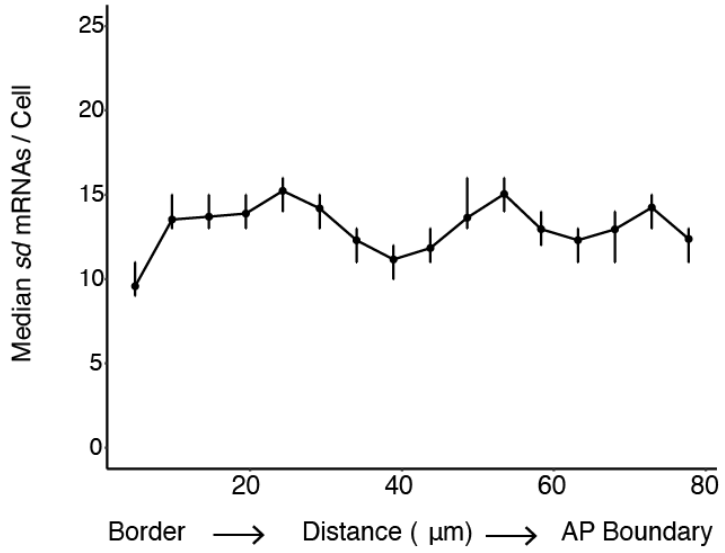
Figure 2. smFISH analysis of mRNA levels from Dpp-responsive genes. (A) Schematic of wing discs highlighting the graded distribution of Dpp protein in the wing pouch, centered around the AP boundary, and the expression domain for *salm*, one of the targets of Dpp regulation. Not shown is Dpp localization in the notum domain of the disc. **(B)** Expression domains of four target genes of Dpp signaling. **(C-F)** Confocal sections of wing pouches probed for mRNAs synthesized from the *salm* **(C)**, *omb* **(D)**, *dad* **(E)**, and *brk* **(F)** genes. Orange arrows mark the position of the AP boundary in each image. **(G, H)** mRNA number as a function of cell distance from the anterior-most border of the wing pouch. **(G)** An axis tangential to the AP boundary is used to map cell position. Numbers refer to distance in μm from the wing pouch border. **(H)** Cells were binned according to position along the axis. The median mRNA number/cell for each bin is plotted. Error bars represent bootstrapped 95% confidence intervals.

(Figure 1F). This was because the Wg morphogen induces a graded expression pattern of Sens protein across the width of each stripe (Nolo et al., 2000; Zecca et al., 1996). Therefore, I binned cells according to their distance from the DV boundary, and I observed peaks in mRNA number per cell as a function of distance from the boundary (Figure 1G).

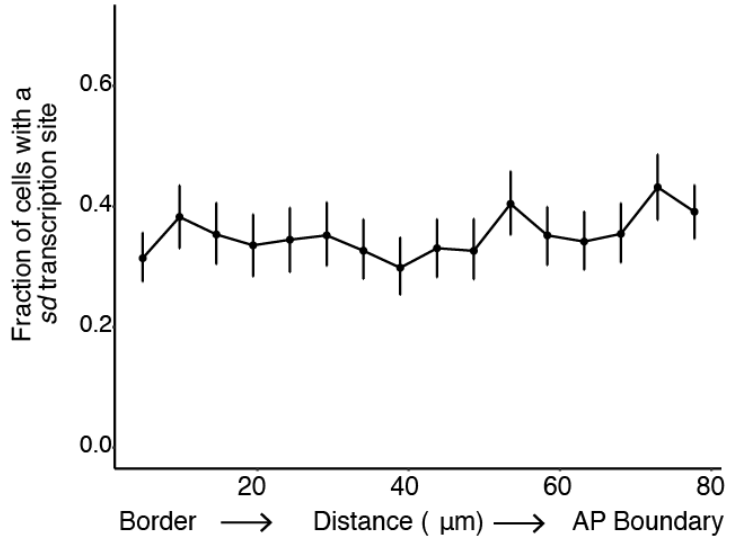
smFISH detection of gene expression regulated by Dpp

I extended the analysis to genes downstream of the BMP family protein Dpp. Dpp is expressed in a stripe of cells located at the AP boundary of the wing disc, orthogonal to the Wg stripe (Figure 2A). Dpp protein is transported bidirectionally to form gradients across the disc, and several genes are regulated by Dpp in a concentration-dependent manner. S*palt-major* (*salm), optomoter blind* (*omb*), *daughters against dpp* (*dad),* and *brinker* (*brk*) are expressed in symmetric domains within the anterior and posterior compartments of the wing pouch (Figure 2A,B). *Salm* is symmetrically expressed in a domain somewhat broader than the Dpp stripe, whereas *omb* and *dad* are expressed more broadly, and *brk* is expressed only near the wing pouch border (Celis et al., 1996; Grimm and Pflugfelder, 1996; Tabata and Takei, 2004). When smFISH was used to detect mRNAs of these genes, it qualitatively recapitulated their known expression patterns (Figure 2C-F). I quantified the number of mRNAs per cell and attempted to map the distribution to cell position within the wing pouch. Since the only landmark I could reliably use was the border between the wing pouch and the rest of the disc, I measured cell position as a function of distance from the border (Figure 2G). When I did so, the distributions in mRNA number per cell gave profiles that were somewhat consistent with previous qualitative descriptions of their expression patterns (Figure 2H). To ensure that these distributions are not an artifact of landmarking the border, I probed for mRNAs produced from the *scalloped* (*sd*) gene. The *sd* gene is expressed uniformly throughout the wing pouch

A


Median sd mRNAs / Cell
25
20
15
10
5
0
20
40
60
80
Border ⟶ Distance (μm) ⟶ AP Boundary

B


Fraction of cells with a sd transcription site
0.6
0.4
0.2
0.0
20
40
60
80
Border ⟶ Distance (μm) ⟶ AP Boundary

C

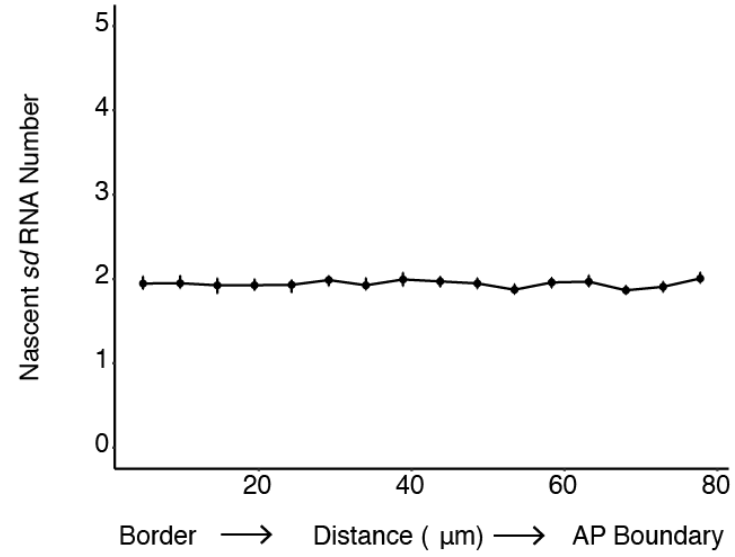

Nascent sd RNA Number
5
4
3
2
1
0
20
40
60
80
Border ⟶ Distance (μm) ⟶ AP Boundary

Figure 3. SmFISH analysis of mRNA levels of the *sd* gene. (A-C) An axis was tangential to the AP boundary was used to map cell position. Numbers refer to distance in in microns from the wing pouch border. Cells were binned according to position along the axis. **(A)** mRNA number as a function of cell distance from the anterior-most border of the wing pouch. Median mRNA number/cell for each bin is plotted. Error bars represent bootstrapped 95% confidence intervals. **(B)** Fraction of cells with a transcription site in each bin is plotted. Error bars represent bootstrapped 95% confidence intervals. **(C)** Median nascent RNA per transcription site for each bin is plotted. Error bars represent bootstrapped 95% confidence intervals.

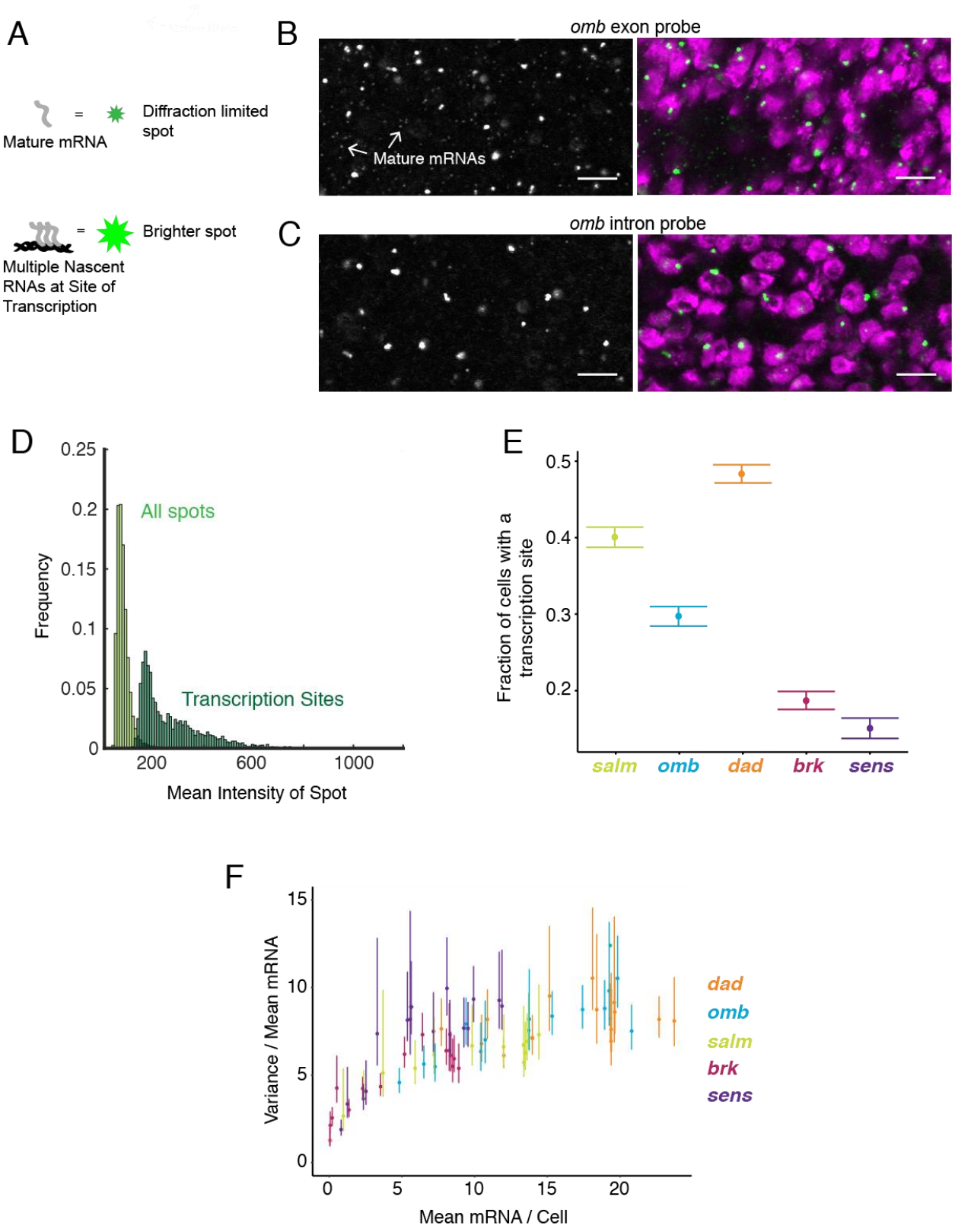

A
Mature mRNA
Diffraction limited spot
Multiple Nascent RNAs at Site of Transcription
Brighter spot
B
omb exon probe
Mature mRNAs
C
omb intron probe
D
All spots
Transcription Sites
Frequency
Mean Intensity of Spot
E
Fraction of cells with a transcription site
salm
omb
dad
brk
sens
F
Variance / Mean mRNA
Mean mRNA / Cell
dad
omb
salm
brk
sens

Figure 4. Sites of nascent transcription are detected by smFISH. (A) Sites of nascent transcription can fluoresce more brightly than single mRNA molecules due to multiple nascent transcripts localized to one gene locus. **(B)** Probes recognizing an *omb* exon generate many small dim spots and a few large bright spots. Right image shows the merge of probe and DAPI fluorescence. The bright spots are associated with nuclei whereas most dim spots are not. **(C)** Probes recognizing an *omb* intron only generate large bright spots that are associated with nuclei. Scale bars = 5 μm. (D) Frequency distribution of intensity for all spots identified in a wing disc probed for *sens* RNAs. Using a threshold of twice the median spot intensity, all single mRNA spots were filtered out, leaving only spots that are associated with transcription sites. The frequency distribution for this class of spots is shown. **(E)** Transcription sites are assigned to cells. For each cell that contains one or more mRNA molecules, it is scored for whether it also has one or more transcription sites. The average fraction of all such cells with a transcription site is shown for each gene. Error bars represent 95% confidence intervals. **(F)** The variance of mRNAs/cell is ratiometrically compared to the mean mRNAs per cells for all genes. This ratio is much larger than one, irrespective of the median mRNA number for binned sub-populations of cells and the gene type. Error bars represent 95% confidence intervals.

(Campbell et al., 1992; Williams et al., 1991), and thus I anticipated a uniform distribution of mRNAs/cell if our method was accurate. Indeed, there was a fairly constant level of mRNAs/cell across the wing pouch as determined by my smFISH pipeline (Figure 3A).

Transcription occurs in bursts

Transcription sites were counted by applying a threshold that only included spots with at least twice the intensity of a mature mRNA spot (Figure 4D). There was a broad distribution of transcription site intensities, suggesting a large range of nascent RNA numbers that were present on a gene at a given time.

Strikingly, many cells did not have a detectable transcription site even though the cells contained mature mRNAs (Figure 4E). From 50 - 80% of cells had this feature, and it was observed for all genes. This observation is not an artifact of segmentation erroneously assigning mature mRNAs to a cell, since the presence of transcription sites was highly variable before segmentation (Figure 5).

I wanted to know why cells with mature mRNAs lacked detectable transcription sites. One explanation is that each gene's promoter is always open, but since transcription is fundamentally stochastic, there would be times when zero or just a few Pol II molecules are transcribing the gene. In this scenario, the birth and death of mRNAs can be described as a Poisson process. For simple Poisson processes, the ratio of the variance to the mean number is one. In our case, Poisson-like birth-death of mRNAs would yield a ratio of variance in mRNA number to mean mRNA number to be one (Munsky et al., 2012; Raj and van Oudenaarden, 2008). Since mRNA number per cell varied systematically across the wing disc because of Wg and Dpp signaling, I binned cells according to their position in the disc, as had been described earlier (Figure 1G, 2H). Strikingly, the ratio of variance to mean mature mRNA number per cell

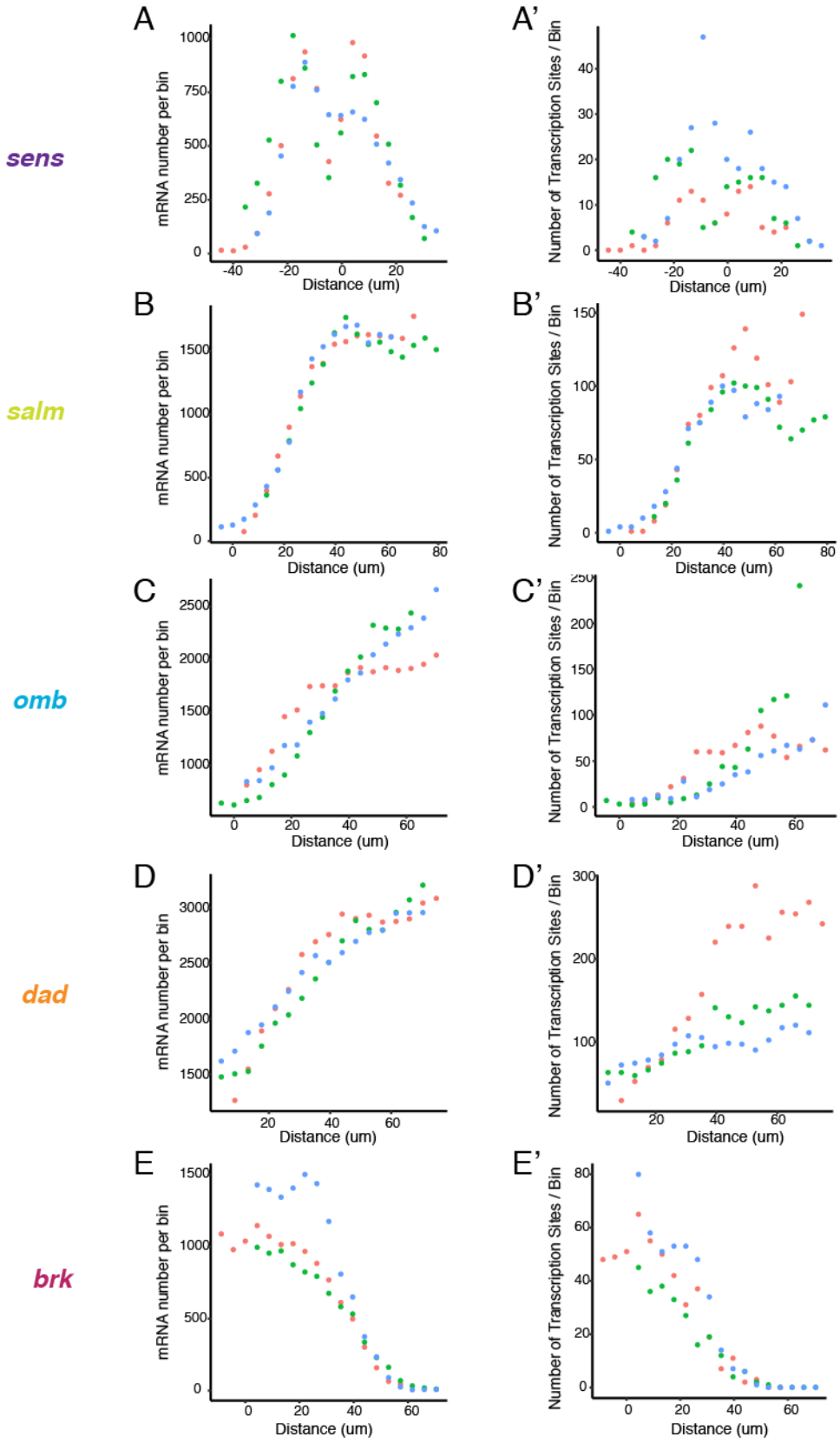

sens
salm
omb
dad
brk
A
A'
B
B'
C
C'
D
D'
E
E'
mRNA number per bin
Number of Transcription Sites / Bin
Distance (um)

Figure 5. Transcription sites and mRNA patterns in unsegmented images. (A,B,C,D,E) Three discs were analyzed independently (green, blue, orange dots) for spots that corresponded to the mRNAs from *sens* **(A)**, *salm* **(B)**, *omb* **(C)**, *dad* **(D)** and *brk* **(E)**. Spots were binned according to their positions along the AP or DV axes, and total mRNAs per bin were plotted. Note the strong concordance of independent discs for all genes. **(A',B',C',D',E')** The same three discs were analyzed independently (green, blue, orange dots) for spots that corresponded to transcription sites from *sens* **(A)**, *salm* **(B)**, *omb* **(C)**, *dad* **(D)** and *brk* **(E)**. Spots were binned according to their positions along the AP or DV axes, and total transcription sites per bin were plotted. Note the strong concordance of independent discs for all genes.

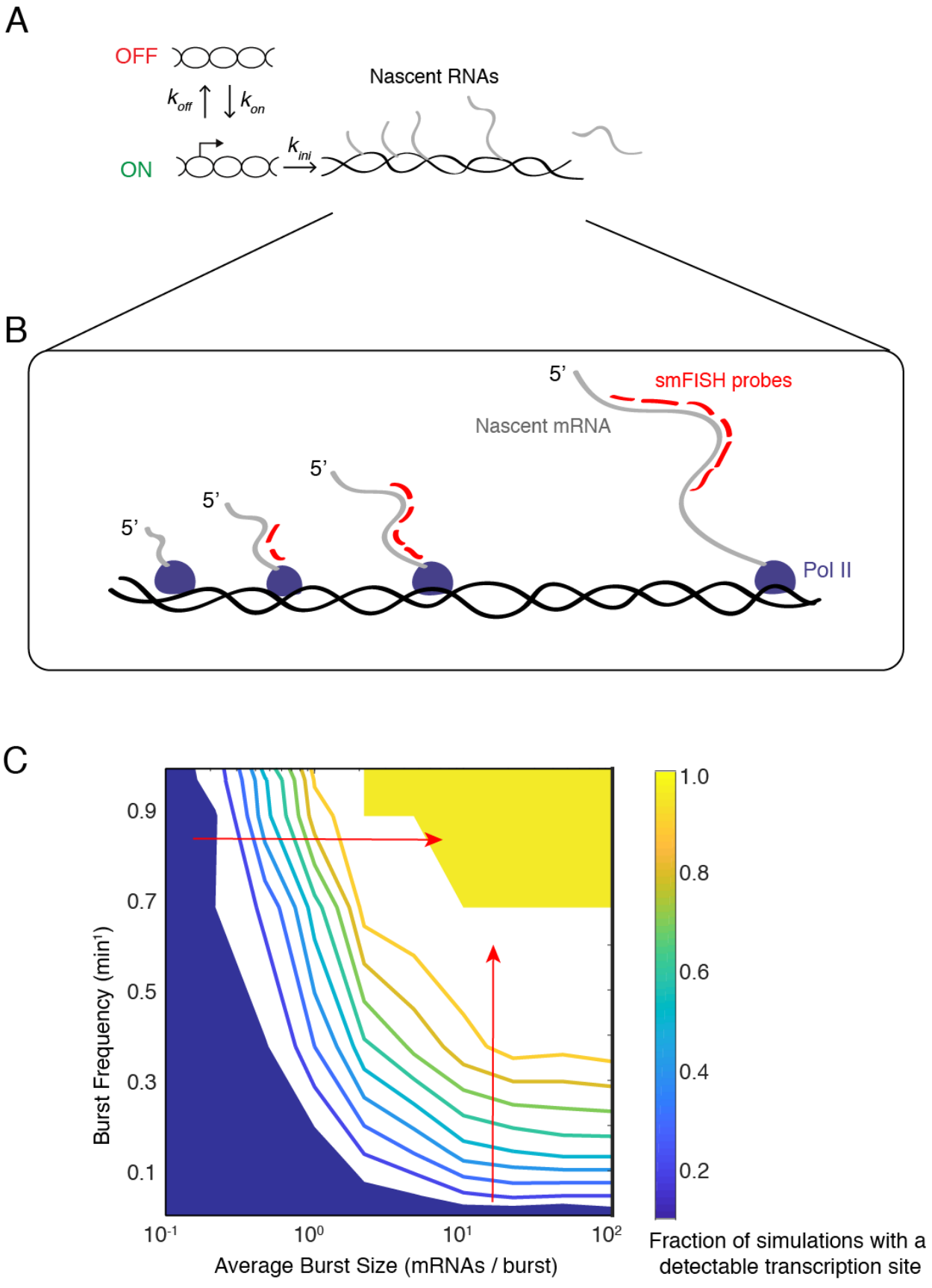

A
OFF
k_{off}
k_{on}
ON
k_{ini}
Nascent RNAs
B
5'
smFISH probes
Nascent mRNA
5'
5'
5'
Pol II
C
Burst Frequency (min^{-1})
0.9
0.7
0.5
0.3
0.1
10^{-1}
10^{0}
10^{1}
10^{2}
Average Burst Size (mRNAs / burst)
1.0
0.8
0.6
0.4
0.2
Fraction of simulations with a detectable transcription site

Figure 6. Modeling transcription sites using bursting dynamics. (A) Model framework showing the three rate parameters affecting transcription initiation. Two parameters affect the promoter state, while the third parameter only affects how many initiation events occur when the promoter is ON. **(B)** Pol II molecules in elongation mode are distributed along the transcription unit. If Pol II is upstream of the probe binding sites, the nascent transcript will not be detected. If Pol II is downstream, the nascent transcript will be detected as 100% signal. If Pol II is transcribing within the binding sites, the nascent transcript will be detected as a partial signal. These three different scenarios are all found in model simulations. For example, in the simulation result shown here, four Pol II's are situated such that a total of 12 virtual probe binding sites are present. Since each mRNA has 6 binding sites, it means that this simulated transcription site has 12/6 or 2 units of normalized signal. Applying our filter cutoff for identifying a transcription site as 2 or more units, this simulated site would be scored as a positive. **(C)** The distribution of normalized signal intensity for 1,000 transcription site simulations. Shown are two distributions from simulations with different initiation rate parameters. Those simulations that result in signal intensities of 2 or more units are classified as detectable transcription sites. **(D)** The phase diagram of transcription site detection in the model. When burst size increases at low burst frequency, the likelihood of detecting a transcription site remains fairly constant. When burst size increases at high burst frequency (horizontal red arrow), the likelihood of detecting a transcription site is ultrasensitive to burst size. Likewise, when burst frequency increases at low burst size, the likelihood of detecting a transcription site remains fairly constant. When burst frequency increases at high burst size (vertical red arrow), the likelihood of detecting a transcription site is ultrasensitive to burst size.

was between 5 and 10 for all genes and was also fairly independent of mRNA output (Figure 4F). This indicated that a simple Poisson process could not explain why I failed to detect transcription sites in every cell expressing mRNA.

To determine if my observations were possibly caused by transcription bursting, I invoked a classical two-state model of transcription (Figure 6A). A promoter exists in one of two possible states - ON and OFF. The promoter switches between states at particular rates k_{on} and k_{off}. When the promoter is in the ON state, Pol II is permitted to initiate transcription that is subject to a rate constant k_{ini}. When the promoter is in the OFF state, Pol II is unable to initiate transcription. The model also includes a transcription elongation step, which is assumed to be 100% processive, and whose timescale depends on the gene length and the rate of elongation. The latter is assumed to be 1,100 nucleotides/min, which is a value that has been experimentally determined in *Drosophila* (Ardehali et al., 2009).

In the model, transcriptional bursts have a characteristic size (number of transcripts per burst) and frequency (rate at which bursts occur). The average burst size is defined as k_{ini} / k_{off}, whereas the average burst frequency is defined as $(k_{on}^{-1} + k_{off}^{-1})^{-1}$. I systematically and independently varied the parameters k_{on}, k_{off}, and k_{ini} to tune the frequency and size of virtual bursts. For each parameter set, I ran 1,000 simulations of the master equation. To capture the stochastic nature of gene expression, most reactions in the model were treated as probabilistic events, with the exception of transcript elongation time. Therefore, simulations with identical parameter values nevertheless gave variable output.

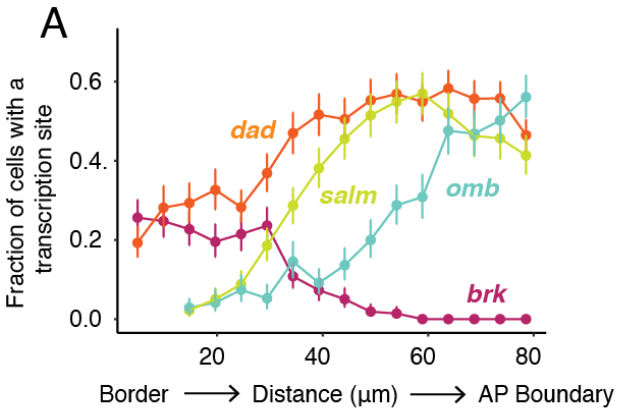
To better relate the results of model simulations to experimental data, I performed the following treatments. First, I randomly paired two independent simulations to mimic the total transcription site activity of paired alleles within a nucleus. Second, I transformed the two

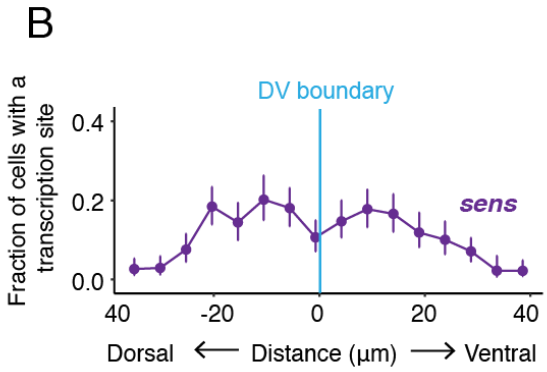
simulations' output to capture the heterogeneity in fluorescence signal intensity at a transcription site. The intensity of each site depends on how many binding sites for probe are present in all nascent transcripts at the site (Figure 6B). This varies with the number of elongating Pol II molecules on the gene and the fraction of molecules that are elongating within or downstream of the region complementary to the probe set. Since this variable is highly dependent upon the position of the complementary region relative to the transcription start and stop sites, I adjusted model conditions to match each particular gene and its region of probe set complementarity. I used these constraints to estimate transcription site intensities from 1,000 pairs of simulations for each parameter set.

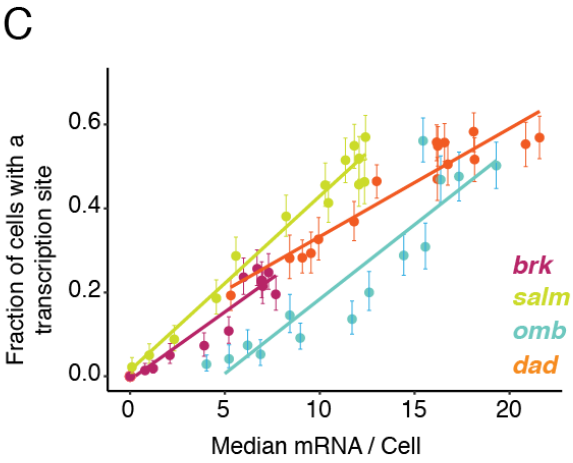
When a simulated transcription site intensity fell below the threshold of twice the number of probe binding sites per mRNA, I counted that simulation as having no "detectable" transcription site. This mimicked the threshold that was applied to the experimental data for identifying a transcription site. I then asked what combination of burst size and frequency could theoretically account for the observed frequency of finding cells with a transcription site (this ranged from 20 to 50% of cells). A phase diagram revealed that a broad range of burst sizes and frequencies would explain our observations (Figure 6C). Therefore, according to my model results, tuning burst frequency and/or size can produce a variable likelihood of detecting a transcription site by smFISH.

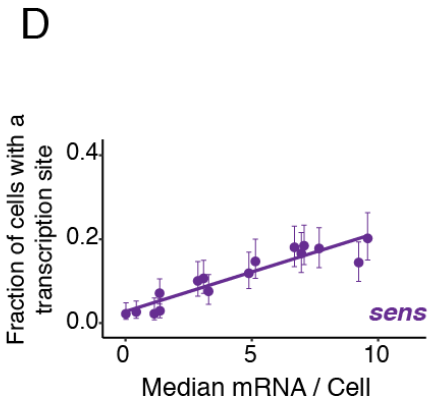
Burst frequency is regulated by Dpp and Wg

I quantified the frequency of detecting a transcription site as a function of cell position within the wing pouch (Figure 7A,B). This frequency varied across the disc in a manner that was gene-specific. Strikingly, for all genes, the spatial distributions of transcription site frequency strongly paralleled the mRNA number per cell (compare Figure 7A,B and Figures 1G,


A
Fraction of cells with a transcription site
0.6
0.4
0.2
0.0
dad
salm
omb
brk
20
40
60
80
Border ⟶ Distance (µm) ⟶ AP Boundary


B
DV boundary
Fraction of cells with a transcription site
0.4
0.2
0.0
sens
40
-20
0
20
40
Dorsal ⟵ Distance (µm) ⟶ Ventral


C
Fraction of cells with a transcription site
0.6
0.4
0.2
0.0
brk
salm
omb
dad
0
5
10
15
20
Median mRNA / Cell


D
Fraction of cells with a transcription site
0.4
0.2
0.0
sens
0
5
10
Median mRNA / Cell

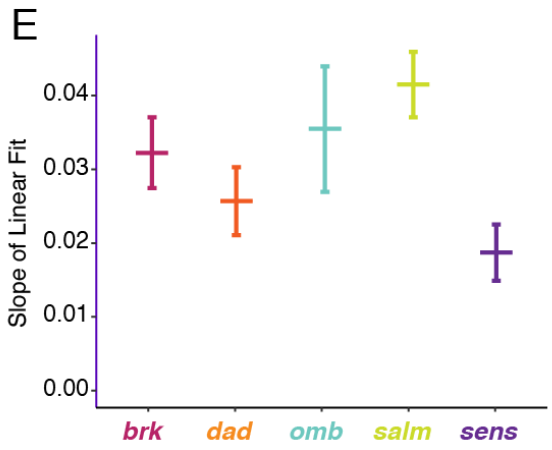

E
Slope of Linear Fit
0.04
0.03
0.02
0.01
0.00
brk
dad
omb
salm
sens

Figure 7. Transcription site detection correlates with mRNA number. (A,B) The probability of detecting a cell with a transcription site varies with the cell's location relative to the source of morphogen. Error bars are 95% confidence intervals. **(A)** Cells are binned according to their distance from the pouch border, and the fraction of cells in each bin with a transcription site are shown for each Dpp-responsive gene. **(B)** Cells are binned according to their distance from the DV boundary, and the fraction of cells in each bin with a transcription site is shown for the *sens* gene. **(C,D)** The probability of detecting a cell with a transcription site varies linearly with the number of mRNA molecules in the cell. Fitted lines are from linear regression. Error bars are 95% confidence intervals. **(C)** Cells are binned according to the number of mRNAs they contain, and the fraction of cells in each bin with a transcription site are shown for each Dpp-responsive gene. **(D)** Cells are binned according to the number of mRNAs they contain, and the fraction of cells in each bin with a transcription site is shown for the *sens* gene. **(E)** Linear regression analysis was performed on samples from C and D, shown is the slope with a parametric 95% confidence interval.

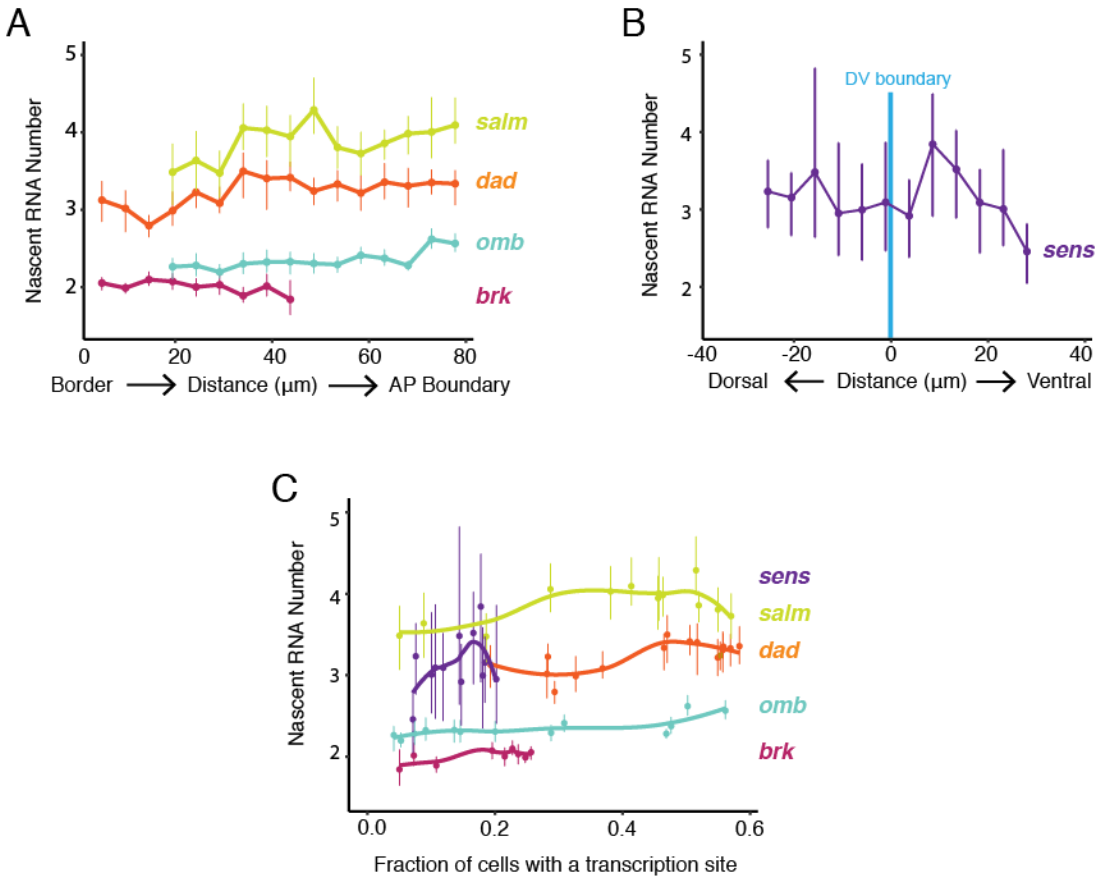

A
Nascent RNA Number
salm
dad
omb
brk
Border → Distance (μm) → AP Boundary
B
DV boundary
sens
Dorsal ← Distance (μm) → Ventral
C
sens
salm
dad
omb
brk
Fraction of cells with a transcription site
D
Burst frequency (min-1)
0.04
0.2
0.4
Fraction of cells with a transcription site
kini
E
Burst size (RNAs / burst)
10
4
1
Fraction of cells with a transcription site
kon

Figure 8. Burst frequency is regulated by Dpp and Wg. (A,B) The average number of nascent RNAs in a transcription site does not vary with the cell's location relative to the source of morphogen. Error bars are bootstrapped 95% confidence intervals. (A) Cells are binned according to their distance from the pouch border, and the average number of nascent RNAs per site in each bin are shown for each Dpp-responsive gene. (B) Cells are binned according to their distance from the DV boundary, and the average number of nascent RNAs per site in each bin is shown for the *sens* gene. (C) The average number of nascent RNAs in a transcription site does not vary with the probability of detecting a cell with a transcription site. Error bars are 95% confidence intervals. (D,E) Modeling the relationship between average number of nascent RNAs in a transcription site and the probability of detecting a site for the *dad* gene. (D) Simulations are performed where the rate parameter k_{ini} has been systematically varied so that burst size is variable. Resulting values for nascent RNA number and fraction of cells with a site are shown. Each datapoint is the average of 1,000 simulations. Simulations are repeated for three different values of k_{on} to specifically set the burst frequency to 0.04, 0.2 and 0.4 min^{-1}. (E) Simulations are performed where the rate parameter k_{on} has been systematically varied so that burst frequency is variable. Resulting values for nascent RNA number and fraction of cells with a site are shown. Each datapoint is the average of 1,000 simulations. Simulations are repeated for three different values of k_{ini} to specifically set the burst size to 1, 4 and 20.

2H). I further examined the relationship between mRNA number per cell and transcription site frequency (Figure 7C,D). Indeed, average mRNA number per cell and the probability of detecting transcription sites in cells were strongly correlated with one another. Remarkably, the slopes of linear fits for three Dpp-responsive genes, *brk*, *omb*, and *salm*, were not significantly different from one another (Figure 7E). The slope for *dad* was similar to *brk* and *omb* but smaller than for *salm*. The slope for *sens* was smaller still. The linear correlation between frequency of transcription site detection and mRNA number confirms that gene regulation by Dpp and Wg is primarily determined through control of transcription initiation.

The likelihood of detecting a transcription site increases because either the promoter is spending more total time in the ON state or more RNAs are being transcribed while in the ON state. These properties are affected by burst size and burst frequency in different ways. I sought to determine whether burst size or frequency was being regulated. I did so by estimating the number of nascent RNAs at each transcription site, which was quantified as a multiple of the median pixel intensity of mature RNA spots (Chapter III Figure 2). The average number of nascent RNAs per transcription site did not significantly vary between cells that were receiving different levels of Dpp and Wg signal (Figure 8A,B). This was observed for all genes. Moreover, the average number of nascent RNAs per transcription site was also independent of the likelihood that transcription was occurring in a cell (Figure 8C). Therefore, the propensity for a cell to generate nascent transcripts does not correlate with the number of nascent transcripts.

To understand the relationship between these observed features, I turned to the modeling framework. I first considered whether modulation of transcription burst size by Wg and Dpp could explain our observations. I modulated burst size by systematically varying the k_{ini} parameter, and from simulations, then calculated the number of nascent RNAs per transcription

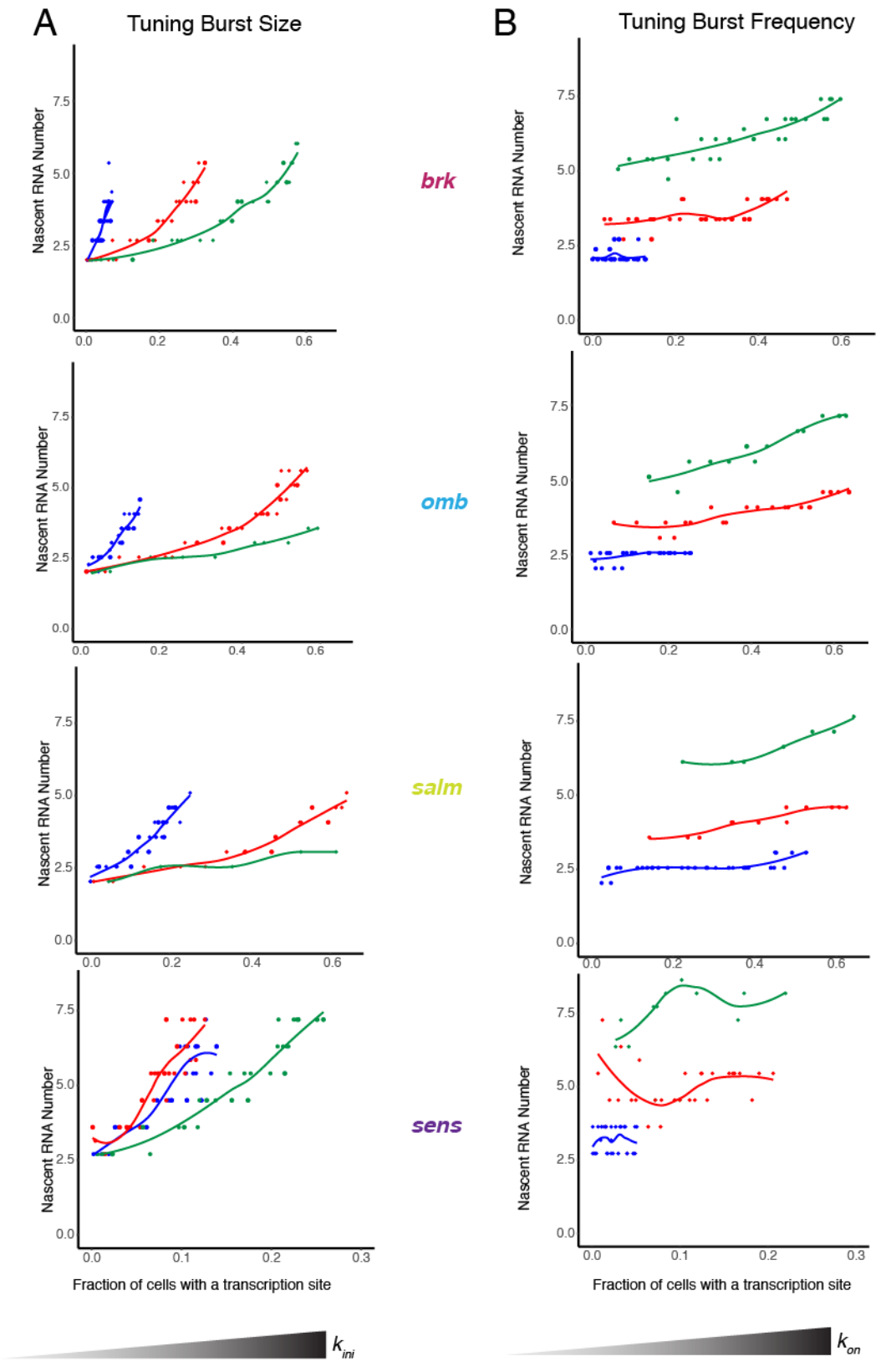

A
Tuning Burst Size
B
Tuning Burst Frequency
brk
omb
salm
sens
Nascent RNA Number
Fraction of cells with a transcription site
k_{ini}
k_{on}

Figure 9. Modeling the relationship between average number of nascent RNAs in a transcription site and the probability of detecting a site for the *brk, omb, salm,* and *sens* genes. (A) Simulations are performed where the rate parameter k_{ini} has been systematically varied so that burst size is variable. Resulting values for nascent RNA number and fraction of cells with a site are shown. Each datapoint is the average of 1,000 simulations. Simulations are repeated for three different values of k_{on} to specifically set the burst frequency to 0.04, 0.2 and 0.4 min^{-1}. **(B)** Simulations are performed where the rate parameter k_{on} has been systematically varied so that burst frequency is variable. Resulting values for nascent RNA number and fraction of cells with a site are shown. Each datapoint is the average of 1,000 simulations. Simulations are repeated for three different values of k_{ini} to specifically set the burst size to 1, 4 and 10.

site and the transcription site detection frequency. There was a positive correlation between nascent RNA number in a transcription site and the probability of detecting a transcription site (Figure 8D and Figure 9A). This was observed across a wide range of fixed burst frequencies. When nascent RNA number was 3 or higher, the correlation with transcription site frequency was strongest. Moreover, when the probability of a transcription site was very low, nascent RNA number converged to a common value irrespective of burst frequency. None of these model predictions were observed in the experimental results with the target genes (Figure 8C). It suggests that transcription burst size is not strongly regulated by Dpp and Wg.

I then modulated burst frequency in the model by systematically varying k_{on} and calculated the number of nascent RNAs per transcription site and the transcription site frequency. There was little change in nascent RNA number as transcription site frequency changed, even across a wide range of fixed burst sizes (Figure 8E and Figure 9B). Interestingly, the burst size appeared to determine what nascent RNA number value was held at a constant. Moreover, there was no convergence of nascent RNA number when the probability of a transcription site was very low, irrespective of burst size. All of these model predictions agree well with the experimental results (Figure 8C). This suggests that Dpp and Wg regulation of genes in the wing disc primarily occurs by modulation of transcriptional burst frequency.

A reduced Wg gradient affects Sens expression

I next sought to investigate what would happen to downstream gene expression if the upstream morphogen gradient was manipulated. To do so, I utilized a membrane tethered variation of Wg that replaces endogenous Wg with a construct bearing a fusion with the type 2 transmembrane protein Neurotactin. This eliminates the ability of Wg to form a gradient but allows juxtacrine Wnt signaling. Animals homozygous for this variation of Wg

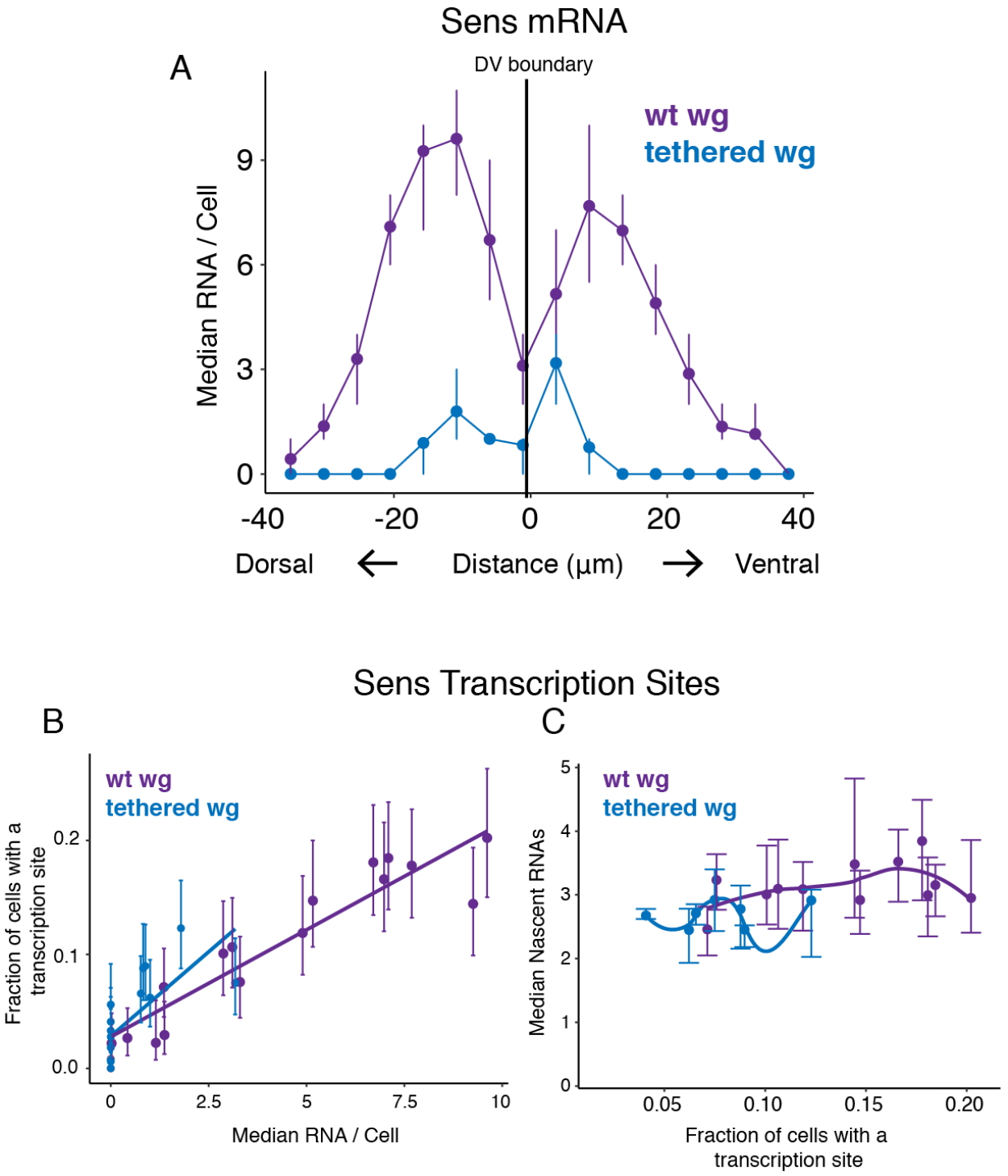

Sens mRNA
A
DV boundary
wt wg
tethered wg
Median RNA / Cell
9
6
3
0
-40
-20
0
20
40
Dorsal
Distance (µm)
Ventral
Sens Transcription Sites
B
C
wt wg
tethered wg
Fraction of cells with a transcription site
0.2
0.1
0.0
0
2.5
5
7.5
10
Median RNA / Cell
Median Nascent RNAs
5
4
3
2
1
0
0.05
0.10
0.15
0.20
Fraction of cells with a transcription site

Figure 10. smFISH analysis of Sens in tethered Wg wing discs. (A) Cells are binned according to their distance from the DV boundary, and the median *sens* mRNA per cell is plotted for each bin. Error bars are 95% confidence intervals. **(B)** Fraction of cells with a *sens* transcription site in each bin is plotted against the median mRNA/cell in that bin. Error bars are 95% confidence intervals. Linear regression line is shown for each data set. **(C)** Median nascent RNA per transcription site in each bin is plotted against the fraction of transcribing cells in that bin. Error bars are 95% confidence intervals. Loess fit for each data set is shown.

have a growth delay. Expression domains of Wg-dependent genes in the wing disc are reduced, but produce normally patterned wings that are slightly reduced in size (Alexandre et al., 2014)).

I analyzed *sfGFP-sens* expressions in animals bearing both *sfGFP-sens* and tethered Wg. I found that the expression domain of *sfGFP-sens* was limited to 15 µm (diameter of approximately three nuclei) beyond the D-V midline, as opposed to 40 µm for wild type animals. Additionally, Sens expression as measured by the median RNA per cell was reduced by approximately 65% in tethered Wg discs (Figure 10A).

This suggests that *sfGFP-sens* expression is not being activated by tethered Wg as strongly as in wild type animals. Next, I analyzed transcription sites for *sfGFP-sens* in tethered Wg discs. If Wg signaling is controlling *sfGFP-sens* expression via burst frequency, I would expect to see fewer overall transcription sites in the tethered Wg discs. Since I predict that burst size is unchanged by Wg activity, I would expect these transcription sites to contain approximately the same nascent RNAs per transcription site.

The proportion of cells with transcription sites in animals with tethered Wg decreased proportionally to the decrease in mature RNA expression. The slope of the linear relationship between median RNA level and fraction of cells with a transcription site was not significantly altered in the tethered Wg discs (Figure 10B). Additionally, the median intensity of these transcription sites was statistically identical to those in wild type discs (Figure 10C). Therefore, the characteristics of transcription sites are exactly what would be predicted for these levels of mature RNA expression in discs with wildtype Wg. There are fewer transcribing cells, but their transcription sites have approximately the same number of nascent RNAs per site. This evidence suggests that these cells receive lower overall levels of Wg signaling from the tethered Wg than from wildtype Wg, and *sfGFP-sens* transcription responds concordantly by decreasing burst

frequency of transcription. While eliminating the Wg morphogen gradient alters the levels and expression pattern of its downstream target *sens*, my data suggests this does alter the underlying transcription dynamics in terms of burst size and frequency change in response to Wg signaling.

Notably, these decreased Sens levels do not lead to a strong wing phenotype in adult animals. However, it has been observed that tethered Wg causes growth delays and smaller wings (Alexandre et al., 2014). All animals for this study were dissected at the white prepupal stage and should therefore be at the same developmental stage, but it is possible that the wing discs have a stronger growth delay than the overall animal. Therefore, the changes in Sens expression domain that I observe may be more due to the earlier developmental age of these wing discs relative to controls. This does not alter my conclusions about the effect of decreased Wg signaling on *sens* transcription dynamics, only the cause of the decreased Wg signaling.

Salm-GFP shows altered dynamics from endogenous Salm

Initial experiments to determine the Dpp morphogen gradient on *salm* were performed using a transgenic version of *salm-GFP*. This construct contains the entire known enhancer construct for *salm*, the *salm* gene, and a GFP tag ("FlyBase Reference Report: model organism Encyclopedia of Regulatory Network (modERN) Project, 2015-, Genomic BAC constructs containing epitope tagged proteins, generated by the modERN Project.," n.d.). Probing this construct with anti-GFP FISH probes revealed an apparent increase in *salm-GFP* fluorescence intensity in the anterior compartment of the disc relative to the posterior (Figure 11A). Such a discrepancy between compartments had not been reported in *salm* literature, where *in situ* hybridization and antibody stains show a more homogeneous stripe about the anterior-posterior midline. This was in contrast to the three other GFP constructs utilized in this study (*brk, dad,*

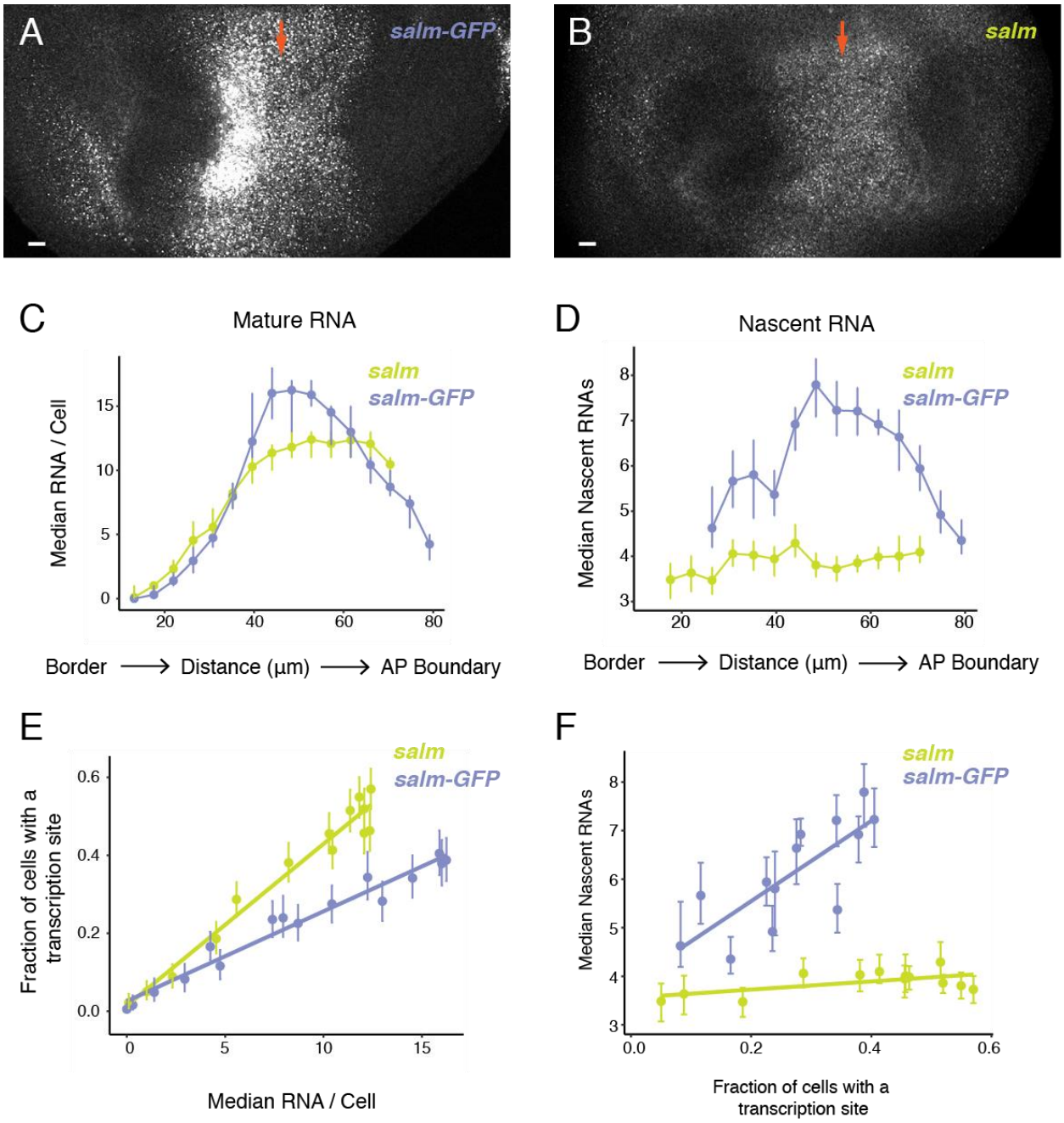

A
salm-GFP
B
salm
C
Mature RNA
Median RNA / Cell
salm
salm-GFP
Border ⟶ Distance (µm) ⟶ AP Boundary
D
Nascent RNA
Median Nascent RNAs
salm
salm-GFP
Border ⟶ Distance (µm) ⟶ AP Boundary
E
Fraction of cells with a transcription site
salm
salm-GFP
Median RNA / Cell
F
Median Nascent RNAs
salm
salm-GFP
Fraction of cells with a transcription site

Figure 11. SmFISH analysis of Salm-GFP (A-B) Optical sections of smFISH. Scale bars=5μm. Orange arrows denote approximate location of the AP compartment boundary. **(A)** Salm-GFP animals probed for GFP mRNA. **(B)** w^{1118} animals probed for *salm* mRNA. **(C-D)** An axis tangential to the AP boundary was drawn and cells were divided into bins based on their position along this axis. Numbers refer to the distance in μm from the anterior most border of the wing pouch. **(C)** Median mature RNA per cell in each bin is plotted. Error bars are bootstrapped 95% confidence intervals. **(D)** Median nascent RNA per transcription site in each bin is plotted. Error bars are bootstrapped 95% confidence intervals. **(E)** Fractions of cells with a transcription site in each bin is plotted as a function of median mature RNA per cell in that bin. Linear regression lines are plotted. Error bars are bootstrapped 95% confidence intervals. **(F)** Median nascent RNA per transcription in each bin is plotted as a function of fraction of transcribing cells in that bin. Linear regression lines are plotted. Error bars are bootstrapped 95% confidence intervals.

and *sens*), all of which showed expression domains that were consistent with previously reported *in situ* hybridization and antibody stains.

When I probed for endogenous *salm* mRNA in non-transgenic discs, the pattern of fluorescence I observed was roughly equivalent in the anterior and posterior compartments (Figure 11B). This pattern agrees with the literature with regards to the *salm* expression pattern. When I compared mRNA numbers between the endogenous *salm* and *salm-GFP*, I see that the median RNA per cell is elevated in *salm-GFP* relative to endogenous *salm* in the anterior compartment (Figure 11C). This suggests that *salm-GFP* might be responding to Dpp signaling in an altered manner to the endogenous gene. The Dpp signaling gradient is not symmetrical in the anterior and posterior compartments of the discs and therefore alterations in sensitivity to this gradient might generate asymmetry in the *salm* stripe.

To investigate this further, I compared *salm-GFP* and endogenous *salm* transcription sites. Interestingly, while endogenous *salm* shows no change in nascent RNA per transcription sites across space, *salm-GFP* shows a marked increase in transcription site intensity across space, which correlates with the increase in mature RNA expression (Figure 11D,E). This indicates more nascent RNAs per transcription site for *salm-GFP*. Additionally, a smaller fraction of cells contained a transcription site for any given mature RNA expression level in *salm-GFP* animals than endogenous *salm* (Figure 11F). *Salm-GFP* cells are less likely to contain a transcription site, but they contain more nascent RNAs per transcription site.

There are two explanations for these results. First, Dpp signaling is altering *salm-GFP* burst size along the expression gradient. According to my simulations, this would cause the number of nascent RNA per transcription site to be more labile in general (Figure 8D, 9B). Alternatively, *salm-GFP* could have an overall higher burst size. As burst frequency increases,

this causes multiple bursts to be in close proximity to one another at a transcription site and increases the overall median nascent RNA per transcription site (Figure 9A). It is not possible to strongly distinguish between these two possibilities. However, given that all the other assayed genes appear to be regulated by burst frequency, it seems more likely that the *salm-GFP* is also modulated by burst frequency, and that is has a higher baseline burst size. This might also explain the difference between the posterior and anterior compartments, as small asymmetries in the Dpp gradient may be amplified somewhat by a larger baseline burst size.

sfGFP-sens transcription sites are primarily monoallelic

I investigated how copy number impacts the number of detected transcripts and transcription sites by comparing animals bearing one copy of *sfGFP-sens* and one copy of wild-type *sens* to animals bearing two copies of *sfGFP-sens*. I probed these discs with anti-GFP FISH probes. The number of mature mRNA spots per cell detected was not straightforwardly halved in cells with one copy of *sfGFP-sens*. At the DV midline, the two had approximately the same median RNA per cell. At the peak of expression, 15 µm from the D-V midline, one copy *sfGFP-sens* had a median RNA per cell approximately 65% of the level with two copies. In regions further from the DV midline, one copy *sfGFP-sens* showed approximately 50% the numbers of RNA per cell (Figure 12A). With regards to numbers of transcription sites, the single copy *sfGFP-sens* showed approximately half the number of transcribing cells as two copies of *sfGFP-sens* across the entire disc (Figure 12B). Both groups had statistically identical numbers of nascent RNAs present at these sites (Figure 12C). If each observed transcription site was the product of both alleles together, I would expect a similar number of transcription sites and a reduction in nascent RNAs at each site in discs with one copy of *sfGFP-sens*. Taken together, the reduction in number but not intensity of transcription sites indicates that most of the

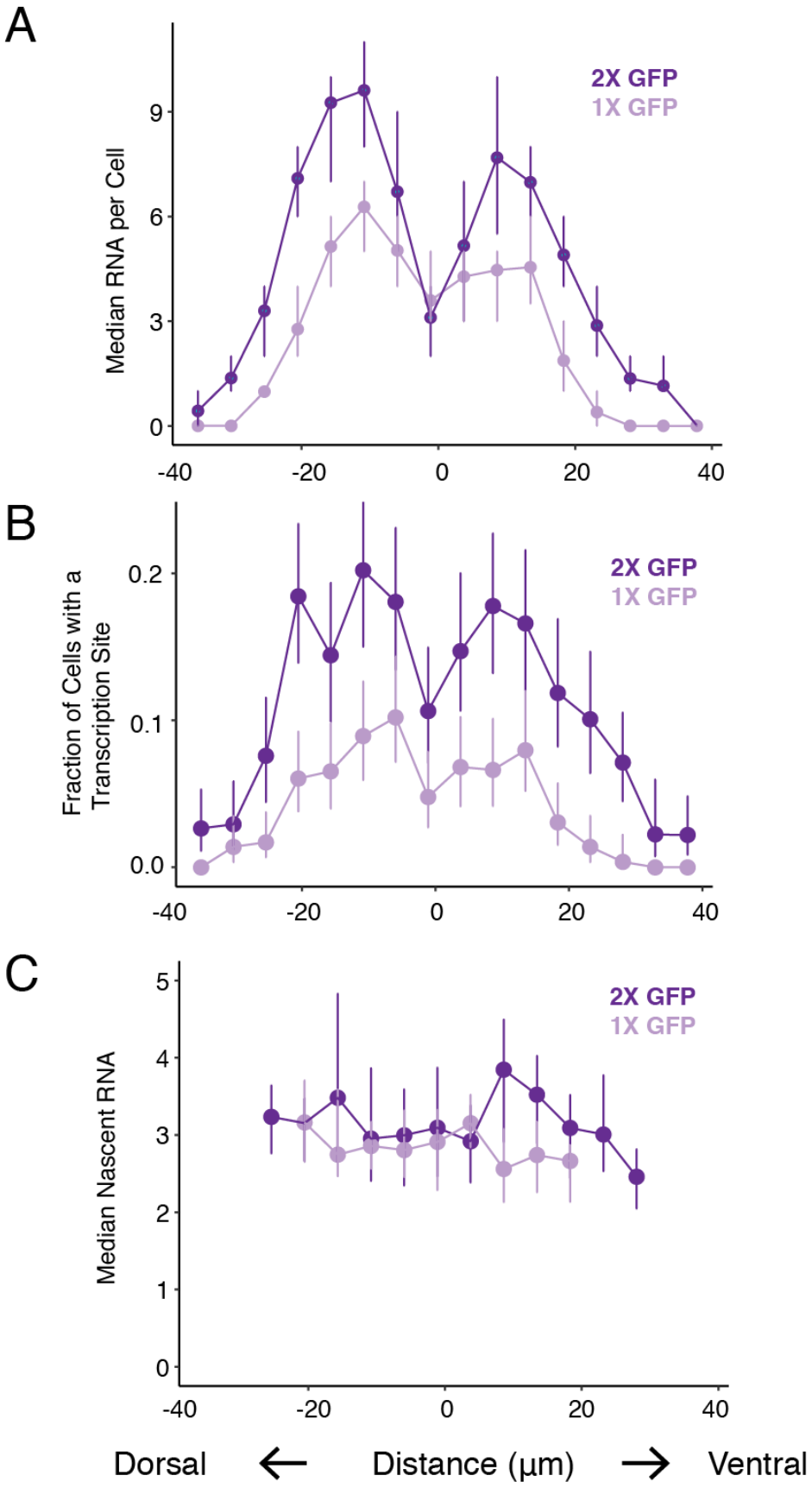

A
Median RNA per Cell
2X GFP
1X GFP
B
Fraction of Cells with a Transcription Site
2X GFP
1X GFP
C
Median Nascent RNA
2X GFP
1X GFP
Dorsal
Distance (µm)
Ventral

Figure 12. SmFISH analysis of one and two copies of *sfGFP-sens*. (A-C) Cells are binned according to their distance from the DV boundary. **(A)** Median mature RNA per cell in each bin is plotted. Error bars are bootstrapped 95% confidence intervals. **(B)** Fraction of transcribing cells in each bin is plotted. Error bars are bootstrapped 95% confidence intervals. **(C)** Median nascent RNA per transcription site in each bin is plotted. Error bars are bootstrapped 95% confidence intervals.

transcription sites observed for *sfGFP-sens* comprise a single allele. This suggests that burst frequency is low enough in these cells that two alleles bursting at the same time is relatively rare. Each allele spends most of its time in an "Off" state.

Estimating burst frequency

The *omb* gene comprises of two exons separated by a 63 kb intron (Grimm and Pflugfelder, 1996). I created two sets of probes for the *omb* gene, one for each exon. RNA polymerase II is predicted to take approximately 53 minutes to transcribe this intron. Therefore, the 5' exon is transcribed approximately 53 minutes before the 3' exon (Ardehali et al., 2009). Probes designed to detect the 5' exon will detect all nascent RNAs initiated in the previous 53 minutes. Therefore, even at relatively low burst frequencies, most transcribing cells should show a transcription site. I observed that the fraction of cells with a transcription site increased in a non-linear manner, rapidly approaching a maximal value of 40-60%, even in areas of the disc with a median level of 1 or 2 transcripts per cell (Figure 13A). This may represent a ceiling of detection in analysis, as no other gene had measured fractions of transcribing cells above 60%.

I observed that in regions of higher *omb* expression in the disc, the average number of nascent transcripts at that site increased linearly (Figure 13B). Averaged over such a large timescale, this is not informative about whether transcription is controlled by burst size or frequency. With such a long transcription time, smFISH spots of the 5' exon are likely to contain transcription products that represent multiple transcription bursts. Therefore, I expected to see an increase in the number of nascent RNA per transcription site, as these bursts would accumulate in regions of higher transcription regardless of transcription dynamics. However, if I were to assume that transcription is controlled primarily by burst frequency as my other data suggests, I can estimate the number of bursts occurring in 53 minutes of transcription time. At

the highest levels of *omb* transcription in my dataset, each 5’ exonic spot contains a median of 4.5-fold more nascent RNAs than a 3’ exonic spot (Figure 13B). If 3’ spots represent a single transcription burst, I can conclude that in 53 minutes, there was an average of 4.5 transcription bursts, or one burst every 11.7 minutes. However, this estimation has many caveats. It assumes that the average burst size is entirely captured by 3’ exon spots, which may not be the case. It also is likely an underestimation, because my analysis pipeline must underestimate the transcription intensity of very bright transcription sites.

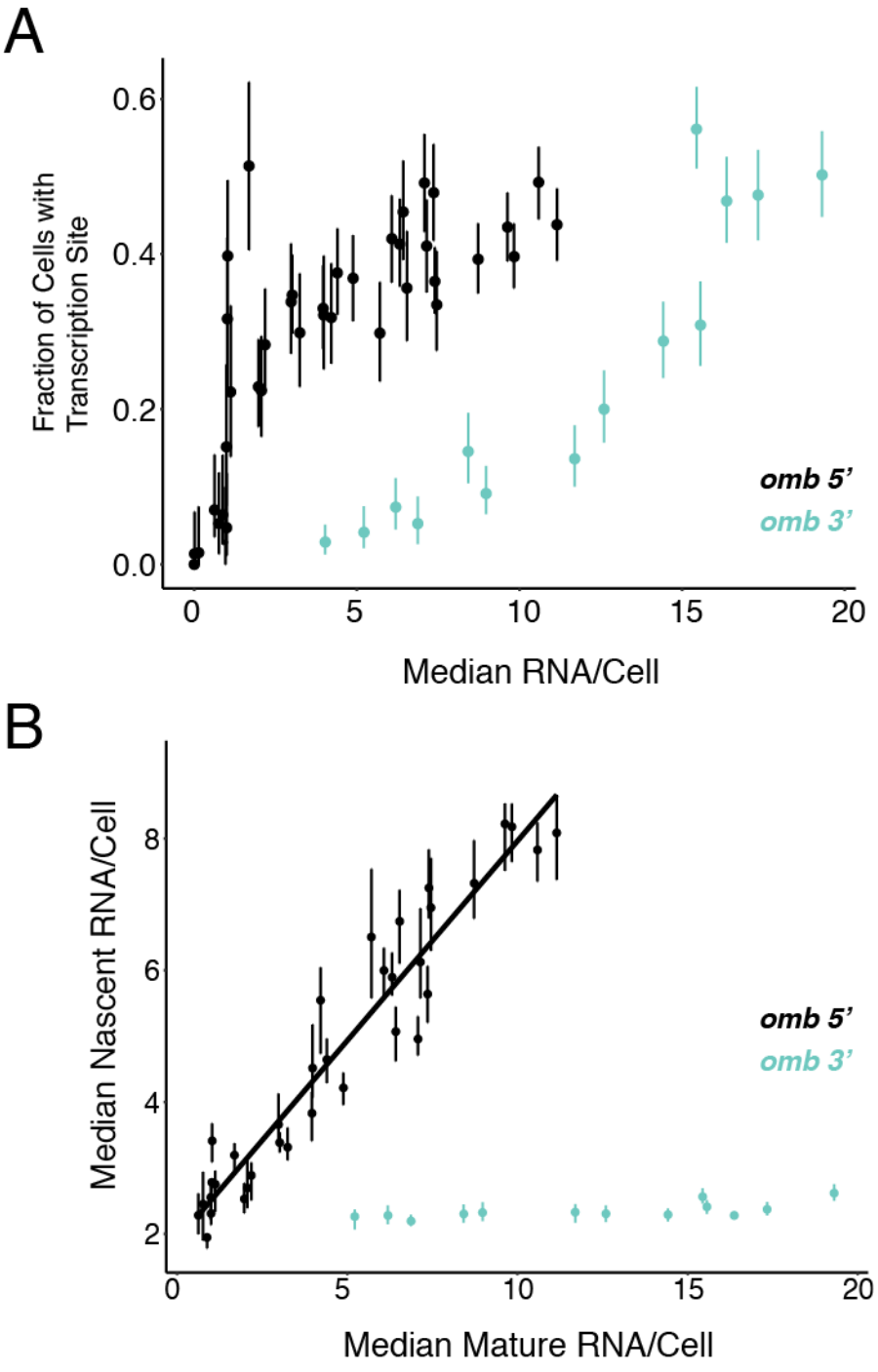

A
Fraction of Cells with Transcription Site
0.6
0.4
0.2
0.0
0
5
10
15
20
Median RNA/Cell
omb 5'
omb 3'
B
Median Nascent RNA/Cell
8
6
4
2
0
5
10
15
20
Median Mature RNA/Cell
omb 5'
omb 3'

Figure 13. smFISH analysis of 5' and 3' exons of the *omb* gene. (A) Fraction of cells with a transcription site in each bin is plotted as a function of median mature RNA per cell in that bin. Error bars are bootstrapped 95% confidence intervals. **(B)** Median nascent RNAs per transcription site in each bin is plotted as a function of median mature RNA per cell in that bin. Error bars are bootstrapped 95% confidence intervals.

Discussion

Morphogens elicit different transcriptional outputs from target genes, depending on local concentration of the morphogen. The targets of Dpp signaling in the wing offer a well-studied example of this concept. Transcription of the gene *brk* is directly regulated by Mothers-against-dpp (Mad), the effector of Dpp (Minami et al., 1999; Moser and Campbell, 2005). In complex with Medea and Schnurri, Mad represses *brk* transcription (Cai and Laughon, 2009). This generates a gradient of Brk protein expression that is inverted to the Dpp gradient. In turn, the level of Brk protein is instrumental in repressing the expression of genes that are induced by Dpp, including *omb* and *salm* (Campbell and Tomlinson, 1999). Thus, opposing gradients of activation and repression define the expression domains of *omb* and *salm*. Since *omb* is less sensitive to Brk repression than *salm,* its expression domain is broader than that of *salm*. Transcription of *salm* is directly activated by Dpp, but in this case, Mad and Medea without Schnurri activate *salm* transcription (Moser & Campbell, 2005). Curiously, *omb* transcription does not directly depend on Dpp signaling, and its transcriptional activation is brought about by unknown factors (Sivasankaran et al., 2000) .

Given the diverse mechanisms by which genes such as *omb*, *brk*, and *salm* are regulated, it is illuminating that the frequency of transcription bursting is the regulated step for all of these genes. Burst size appears to be independent of Dpp signaling for these genes. If our two-state model for initiation is accurate, then k_{on} is the most likely step that is being regulated directly and indirectly by Dpp. This is because k_{on} specifically affects burst frequency and not size whereas k_{off} affects both frequency and size. If k_{on} is the kinetic step under regulation for all of these genes, how is it rendered rate-limiting given such diverse enhancer architectures and transcription factor inputs? It has been found that burst frequency correlates with enhancer

strength and enhancer-promoter contact, suggesting that k_{on} is potentiated by enhancer-promoter contact and is mediated by transcription factor binding to DNA (Bartman et al., 2016; Bothma et al., 2014; Chen et al., 2019; Fukaya et al., 2016; Larsson et al., 2019)(Bartman et al., 2016; Bothma et al., 2014; Chen et al., 2019; Fukaya et al., 2016; Larsson et al., 2019).

In spite of this universal regulation of burst frequency by Dpp, there are other factors that also help determine the expression domains of these genes. For example, *salm* and *omb* are expressed in nested domains; cells close to the source of Dpp contain mRNAs for both genes while lateral cells more distant from the source contain predominantly *omb* mRNAs (Figure 2H). Although burst frequency control might solely dictate these differences, it appears not to be the case. The likelihood a cell is transcribing either gene does not strictly correlate with the breadth of their expression domains. Lateral cells that predominantly contain *omb* mRNAs nevertheless are more likely to be transcribing *salm* than *omb* (Figure 5A). Looking at the relationship between transcription likelihood and mRNA number (Figure 5C), the relationship for both genes is linear with similar slopes. But for *omb*, the curve is shifted such that more mRNAs are found in cells that have a lower likelihood to be undergoing transcription. This shift is not due to a greater burst size of *omb* transcription (Figure 6C). Rather, the simplest interpretation is that the half-life of *omb* mRNA is greater than the half-life for *salm* mRNA, so *omb* mRNAs are more readily detected in cells between transcription bursts. Thus, the broader domain of *omb* expression might be accounted for by mRNA stability.

This conclusion is at odds with previous studies of *omb* regulation that used an enhancer trap reporter for *omb* expression. There, the reporter was expressed in a broad domain within the wing. However, this reporter expressed lacZ in response to *cis*-regulatory elements near the most distal promoter driving a minor species of *omb* transcript (Mayer et al., 2013). There is a more

proximal promoter 13 kb that appears to be the major site of transcription initiation for *omb* (Flybase).

My results also challenge the view that *salm* and *omb* expression domains have sharp boundaries due to transcription thresholds set by Brk and Dpp. I find that *omb* and *salm* mRNA numbers per cell drop gradually with distance from the source of Dpp (Figure 2H). As well, their gradients in mRNA number are inversely correlated with the gradient in *brk* mRNA number. *Salm* has relatively constant mRNA number in cells near the AP boundary, and those numbers gradually diminish in cells located more laterally. A similar pattern is seen with *omb*, except the domain with constant *omb* mRNA number is smaller than for *salm*. However, the *salm* and *omb* enhancer trap reporters as well as anti-Salm immunohistochemistry have reported expression domains with sharp boundaries. Possibly, the discrepancy hints at some threshold of mRNA expression below which protein output drops sharply. It is also possible that the previously characterized expression domains for *salm* and *omb* were distorted by non-linear detection of antibodies that recognize Salm and the protein product of lacZ, β-galactosidase.

Materials and Methods

Drosophila genetics

All *Drosophila* were raised at room temperature and grown on standard molasses-cornmeal food. The *sfGFP-sens* transgenic line was used as described in (Cassidy et al., 2013). The Nrt-Wg stock was used as described in (Alexandre et al., 2014). Experiments were performed on *dad-GFP, salm-GFP,* and *brk-GFP* transgenes obtained from Bloomington Drosophila Stock Center (stocks 81273 and 38629, respectively). For all transgenic experiments, smFISH was performed on homozygous individuals. Experiments were performed on endogenous *omb* and *salm* in w^{1118} individuals.

Data Analysis

Binning of data: Each disc is imaged with the DV boundary located at the y-coordinate midline of the image. Therefore, the x-coordinate of the image corresponds to position along the disc's AP axis, and the y-coordinate corresponds to position along the DV axis. In order to analyze data across developmental axes, each image is divided into spatial bins of 64 pixels each, approximately equal to the diameter of one cell nucleus. RNA spots are assigned to a bin according to the position of their associated nuclear centroid.

Sample size and replicates: I analyzed image stacks from three independent discs for each experiment. Each image stack contains approximately 1,700 identified nuclei. Therefore, the total sample size is approximately 5,000 cells per experiment. Similar trends in RNA and transcription spots feature are observed in each disc individually, and hence, the analysis is not distorted by artifacts in pooling and cell segmentation (Figure 5).

Alignment of replicates along developmental axes: While each disc is imaged roughly in the same region, there is not an unambiguous landmark that precisely registers different disc images with one another. To pool data across space as accurately as possible, I register discs to each other based on their mRNA spot distributions over space. For each image data set, the number of RNAs per spatial bin is summed, and the distributions across bins are compared. Bins are then manually registered such that the distribution profiles of the 3 datasets line up with one another (Figure 5 A-E). The overlapping bins from the three datasets are then assigned to a pooled bin. Pooling includes the nuclei centroids as well as the transcription and RNA spots. This is repeated for all bins.

Calculations**:** *Median mature mRNAs per cell* is calculated from total number of mature mRNA spots for each cell within a spatial bin of pooled data. As the distribution of mRNAs per cell is

not normally distributed and has a long tail, I ascertained that the median was a more robust descriptor of the "center" of the distribution than mean.

Median nascent RNAs per cell is calculated from normalized intensity measurements for each transcription spot within a spatial bin of pooled data. All nascent RNA spots are included. As the distribution of RNA per cell is not normally distributed and has a long tail, I ascertained that the median was a more robust descriptor of the "center" of the distribution than mean. Because the number of transcription sites varies over space, sample sizes vary for calculating median nascent RNAs per cell. For bins where fewer than 5% of cells contain a transcription site, median nascent RNAs per cell was not calculated, as the sample size was determined to be too small (<15).

Fraction of cells with a transcription site is calculated by dividing the number of cells in a pooled spatial bin with at least one transcription site assigned to them by the total number of cells in that spatial bin.

Fano factor is calculated for each spatial bin by dividing the variance in the mRNA per cell distribution by the mean mRNA per cell for all cells assigned to that pooled spatial bin.

Curve and line fitting: Linear models are produced by unweighted least squares linear regression. LOESS fits are performed using the loess fitter in R, with an optimized span to minimize residuals.

Statistics are calculated by bootstrap resampling analysis using the bias-corrected and accelerated method. I resample data within each bin of pooled data and calculate the statistic of interest 10,000 times. The mean value of the statistic and a 95% confidence interval are calculated from these resampled values.

Stochastic Simulation Model

I model the various steps of gene expression, based on central dogma, as linear first order reactions. To simulate the stochastic nature of reactions, I implement the model as a Markov process using Gillespie's Stochastic Simulation Algorithm (Gillespie, 1977). Simple Markov processes can be analyzed using a chemical master equation to provide a full probability distribution of states as they evolve through time. The master equation defining our gene expression Markov process is as follows:

$$\begin{aligned}\frac{\partial P(N_m, N_g, t)}{\partial t} &= K_{ini}\left[(N_m - 1)P(N_m - 1, N_g, t) - P(N_m, N_g, t)\right] \\ &+ K_{deg}\left[(N_m + 1)P(N_m + 1, N_g, t) - N_m P(N_m, N_g, t)\right] \\ &+ K_{on}\left[\left((N_g - N_{gtot})\right)P(N_m, N_g - 1, t) - (N_g - N_{gtot})P(N_m, N_g, t)\right] \\ &+ K_{off}\left[(N_g + 1)P(N_m, N_g + 1, t) - N_g P(N_m, N_g, t)\right]\end{aligned}$$

where N_m ,N_g, and t are defined as the number of RNA molecules present, as the number of transcriptionally active gene copies, and simulation time, respectively. N_{gtot} is defined as the total number of gene copies present, and thus is the maximum number of active gene copies that can exist in the simulation. K_{ini} , k_{deg} , k_{on} , and k_{off} are rate constants defining the rates of transcription initiation, RNA degradation, promoter state switching from off to on, and promoter state switching from on to off, respectively.

As the Markov process gets more complex, the master equation can become too complicated to solve. Gillespie's Algorithm is a statistically exact method that generates a probability distribution identical to the solution of the corresponding master equation given that a large number of simulations are realized.

A brief description of how the Gillespie simulation produces each probability distribution is as follows:

1. I initialize all simulations to start with no mRNA molecules and promoter state is set to OFF.
2. For each event *i* in the simulation, a total rate r_{tot} is calculated by summing all r_i reaction rate constants in the model, given the current promoter state and the total number of mRNA molecules present.
3. A time-step τ is generated from an exponential probability distribution with mean $1/r_{tot}$. This τ is the time interval between the current event and the next event.
4. Each event *i* is selected from the list of reaction steps in the model available at that time (promoter switching, transcription initiation, mRNA decay). The probability a reaction step is selected is equal to r_i / r_{tot}. An event is selected at random given these probabilities. For each event, the following actions are taken:
    - Promoter switches to ON: Promoter is now in ON state, transcription initiation is now included in r_{tot},
    - Promoter switches to OFF: Promoter is now in OFF state, transcription initiation is no longer included in r_{tot}.
    - Transcription Initiation: Number of mature mRNA molecules is increased by 1.
    - RNA degradation: Number of mature mRNA molecules is decreased by 1.
5. Simulation time is updated as $t + \tau$ where *t* is the total time elapsed in the simulation.

Each simulation is run for 10,000 iterative events to approximate steady-state conditions, at the end of which the number of mRNA molecules present in the simulation is recorded. Independent simulations are then randomly paired to mimic the two alleles within a cell, and the

sum of mRNA numbers is recorded as the mRNA output per cell. A minimum of 1,000 simulation pairs are generated for each set of rate parameter values.

The RNA decay parameter k_{deg} is fixed at 0.04/min for all simulations, as this rate had been experimentally determined for *sens* mRNA (Giri et al., 2020). The transcriptional rate parameters are varied in accordance with the specific hypothesis being tested. I constrain them loosely to be within an order of magnitude of reported values for these rates from the literature (Milo, Jorgensen, Moran, Iber, & Springer, 2010). I also constrain these rates so as to produce steady state mRNA numbers similar to experimental data.

- k_{ini} is varied from 0.2 to 60 /min
- k_{on} is varied from 0.008 to 38/min
- k_{off} is varied from 0.016 to 20/min

To perform a parameter sweep, I vary the relevant parameter across the defined range. Each rate parameter value in the sweep is used to make 1,000 paired simulations as described above.

***Nascent Transcripts*:** Thus far I have described how model simulations generate *in silico* data for mature mRNA numbers. I also use the same simulations to approximate the number of nascent RNAs per gene. After 10,000 iterative events are completed in a simulation, the number of nascent RNAs is counted. A single nascent RNA is counted if a single transcription initiation event has occurred within an interval of time (τ_{elong}) equal to the time it is estimated that RNA polymerase takes to elongate from the binding site for the 5'-most oligo probe to the 3' end of the RNA. To calculate τ_{elong} for each gene, I divide the number of nucleotides from 5' probe-binding site to 3' end by the transcription elongation rate. This rate is assumed to be 1,100 nucleotides/min, as experimentally determined (Ardehali et al., 2009).

Gene	τ_{elong} (min)
brk	1.35
dad	2.05
sens	5.15
salm	5.30
omb	3.05

I weight the count of nascent RNAs in a simulation to mimic the fluorescence output from these nascent RNAs if they are hybridized to probes. I define τ_{probe} to be the time interval for RNA polymerase to elongate from the 5'-most probe-binding site to the 3'-most probe-binding site. If a nascent RNA had been initiated in a time less than τ_{probe}, then I weight the counting of that nascent RNA as 0.5 rather than 1. I do this because the probe-binding region of the nascent RNA is partially transcribed at this point. For simplicity, the exact locations of probes and RNA polymerase are not taken into account to calculate the weighting, and instead I assign the overall probability of fluorescence for an ensemble of such partially transcribed RNAs. If a nascent RNA had been initiated in a time greater than or equal to τ_{probe} and less than τ_{elong}, then I weight the counting of that nascent RNA as 1. These RNAs are assumed to produce 100% of the fluorescence of a mature RNA spot, since all probe-binding sites are transcribed at this point.

I randomly pair two simulations and sum the number of weighted nascent transcripts. This mimics the experimental conditions where the two gene alleles are physically paired and

thus their nascent RNAs are co-localized in space. I collate 1,000 paired simulations for each parameter set and calculate the following statistics:

Fraction of virtual cells with a transcription site is calculated by counting how many paired simulations have a total number of weighted nascent RNAs of 2.0 or more. This is done in order to be consistent with the limitations of the experimental data; only nuclear spots with fluorescence greater or equal to 2 mature mRNA spots were called as transcription sites. When this number of paired simulations is divided by the total of 1,000 paired simulations, it is the fraction of virtual cells with a transcription site.

Median number of nascent RNAs per virtual cell is calculated from those paired simulations with a total number of weighted nascent RNAs of 2.0 or more.

Chapter V: Effects of Metabolism on Gene Regulatory Networks

Introduction

Growth and development are interlinked processes in the formation of an animal. In order for development to proceed appropriately, events must be coordinated in the body through time. The timing of induction of morphogenic events relative to tissue growth can cause changes in resulting organ size relative to shape. Changes in developmental timing of events relative to one another are a mechanism of evolutionary change in biology (Gould, 1977). Therefore, if growth rate of an animal is altered during its development, the processes governing development must be controlled in a manner that can accommodate such changes. As gene expression dynamics are a critical component of developmental processes, it is reasonable to suggest that gene expression must contain some mechanisms conferring robustness to perturbations in growth rate.

Environmental factors influence growth and development via metabolism

Several variables have been shown to influence growth and developmental timescales. Fasting and other means of nutrient limitation slows growth and development in many animals, meaning that nutritional status affects animal's developmental rate. Nutrient limitation also extends overall lifespan (Arendt, 1997; Pontzer et al., 2016). Ambient temperature can also influence the growth rate of ectothermic animals, including *Drosophila* (Kuntz and Eisen, 2014; Zuo et al., 2012). These environmental factors have complex effects. Each of these directly effects biochemical reactions: temperature alters the speed at which they occur, and nutrient deprivation alters the availability of reactants. We might then hypothesize that the mechanism by

which environmental factors affect growth rates is via metabolism. A model has been invoked wherein developmental timing can be explained by body size and temperature as they impact the rate of biochemical reactions in the cell (Gillooly et al., 2002). However, both temperature and nutrient limitation also activate respective signaling pathways which direct cells to coordinate a variety of responses across the animal. Therefore, more precise tools for altering metabolism are needed. Indeed, when cellular metabolism is more directly manipulated by mutations in metabolic respiration, the growth and developmental rate of resulting animals is affected (Nakai et al., 2001). In *Drosophila*, direct genetic alterations in glucose metabolism also causes developmental delays (Rulifson et al., 2002).

An observation in animals with perturbed developmental rates is altered penetrance of developmental phenotypes. Lowering temperature and fasting appear to suppress developmental phenotypes of many mutations (Child, 1939, 1935). Likewise, increasing growth rate of broiler chickens in an agricultural setting by breeding results in increased musculoskeletal defects (Julian, 2005). It seems intuitive that given more time to complete development, animals are more able to correctly develop even with genetic perturbations. However, it is more difficult to precisely conceive of how this occurs on a cellular level. One method may be to explore the effects of metabolism on expression of genes that influence development. Many developmental processes are influenced by gene regulatory networks (GRNs), that control the specification of cell types from pluripotent cells. GRN's coordinate inputs from many genes to act on a cell's decision if and when to differentiate (Levine and Davidson, 2005).

The Yan Network

One example of a GRN critical for differentiation is a network with the *Drosophila* protein Yan as a central component. The Yan protein is a transcriptional repressor with an ETS

domain (Lai and Rubin, 1992). Yan is involved in patterning multiple *Drosophila* tissues where receptor tyrosine kinases (RTKs) are used for patterning, including the embryonic ventral midline, the mesoderm, the trachea and the head (Price and Lai, 1999; Rebay and Rubin, 1995). Yan is thought to be general repressor of differentiation in contexts where RTK activity is an inductive signal. When RTKs activate the Mitogen Activated Protein Kinase (MAPK) signal transduction pathway, it down-regulates Yan, allowing differentiation to proceed. In the larval and pupal eye imaginal disc, Yan is expressed in all undifferentiated cells in the morphogenetic furrow and is lost as the nuclei rise apically and differentiate (Rebay and Rubin, 1995).

Yan down-regulation is controlled via multiple pathways. MAPK activity leads directly to phosphorylation and decay of Yan protein, but also to the activation of three factors that contribute to Yan's deactivation. MAPK activates the transcriptional activator Pointed-P2 (Pnt-P2), which competes with Yan for DNA binding sites in target gene enhancers. Pnt-P2 also activates transcription of Mae and miR-7. Mae inactivates Yan and facilitates its phosphorylation by MAPK (Vivekendand et al. 2004). miR-7 post transcriptionally represses *yan* expression (Li and Carthew, 2005). Thus, an entire network of redundant interactions exists to ensure the downregulation of Yan in response to RTK signaling.

Results

miR-7 regulates Yan Levels in Eye Imaginal Disc Precursor Cells

In order to test the hypothesis that redundant repression is critical for control of gene expression in development, I utilized methodology and tools for gene quantification in the eye imaginal disc in order to quantify the developmental gene Yan.

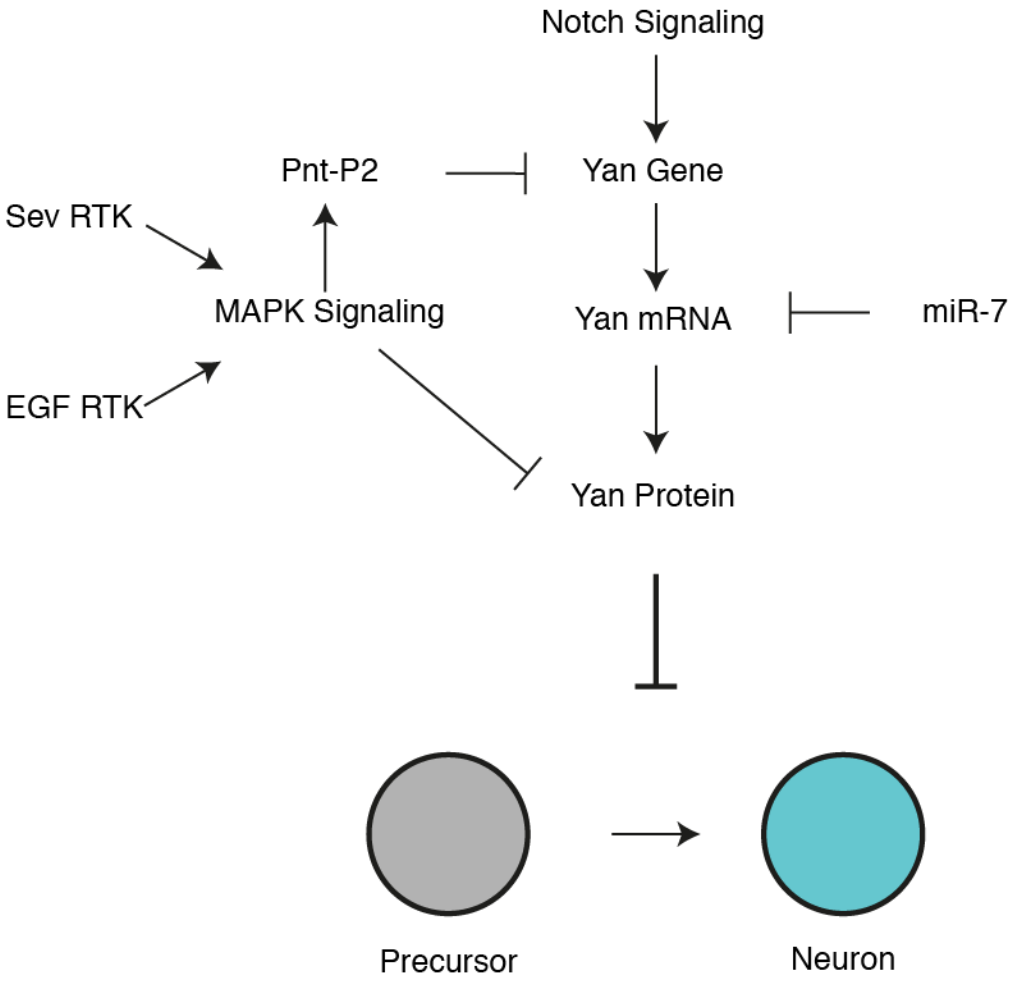


Figure 1. Redundant repression in the Yan network. Schematic of the network inputs controlling Yan protein expression in the eye imaginal disc. Yan's function in the eye disc is to repress differentiation of precursor cells (grey) into neuron cells (blue).

A transgenic form of YFP-tagged Yan was used as described in (Peláez et al., 2015)

Confocal z-stacks are captured of fixed eye-antennal imaginal discs from animals bearing Yan-YFP and fluorescently tagged histone H2 protein. The FlyEye Silhouette software package segments these images using a meanshift-root algorithm. Nuclei of differentiating neurons rise apically in the tissue and change shape, allowing unambiguous identification by a trained observer. The FlyEye Silhouette graphical user interface allows the user to manually define and classify cell types in the image based on nucleus position. The fluorescence data from these cells is extracted and plotted as ratio of Yan-YFP fluorescence to histone-RFP fluorescence in order to control for nuclear size and other variations in fluorescence across the disc. Since a column of ommatidia is generated every two hours, each column in a single disc can be analyzed successively to give an approximation of the time course of gene expression through development (Peláez et al., 2015).

Pelaez et al. (2015) found that Yan-YFP expression is induced strongly in all cells just posterior to the morphogenetic furrow of the eye imaginal disc. In differentiating neurons, Yan levels rapidly decline to basal levels. In precursor cells, Yan-YFP levels slowly decrease over time. There is a great deal of heterogeneity in the levels of Yan-YFP in this precursor cell population.

Several lines of evidence strongly suggested that miR-7 regulates Yan. Overexpression of miR-7 in eye cells phenocopies Yan loss-of-function, producing ectopic R7 cells. The Yan 3' UTR contains four binding sites for miR-7. A GFP reporter fusion to the Yan 3' UTR lacking these binding sites is upregulated compared to a reporter with the wild-type Yan UTR (Li and Carthew, 2005). Furthermore, overexpression of the MAPK-resistant Yan^{ACT} allele is partially

rescued by miR-7 overexpression. In miR-7 mutant animals, Yan expression as visualized with immunohistochemistry appears to be qualitatively higher and more variable, but only in animals exposed to fluctuating temperature conditions (Li et al., 2009).

Initial experiments sought to describe the effects of the deletion of *yan*'s miR-7 binding sites on Yan-YFP fluorescence. To do so, a version of the *Yan-YFP* gene was created with all four miR-7 binding sites scrambled. This represents a more specific and targeted perturbation than previous work from the lab, which had perturbed miR-7 itself. *Yan-YFP* and *Yan$^{\Delta miR-7}$-YFP* animals were generated with their endogenous *yan* gene mutated, so that the entire active pool of Yan protein would be fluorescently labeled. A very mild eye roughening was observed in both *Yan-YFP* and *Yan$^{\Delta miR-7}$-YFP* animals, with mutant animals showing a slightly increased prevalence of the rough eye relative to the wild-type version (Figure 3B-D). Rearing the animals at a higher temperature resulted in increased penetrance of this phenotype for both variants of *Yan-YFP*. The fold-change of the effect was not significantly different between miR-7 insensitive and wild-type *Yan-YFP* alleles (Figure 3D). Therefore, while there seems to be a miR-7 dependent effect on this phenotype and a temperature related effect on this phenotype, it does not appear that these effects interact in a synergistic way. While this observation could argue that miR-7 function itself is not dependent on temperature, it is also possible that the observed phenotype is influenced by the different genetic backgrounds of the two strains.

Eye imaginal discs from animals bearing either the *Yan-YFP* or *Yan$^{\Delta miR-7}$-YFP* were dissected and imaged using confocal microscopy in order to assay YFP fluorescence across the disc. Animals also contained an RFP-labeled Histone H2A in order to normalize fluorescence to nucleus size, shape and other factors. Cells were segmented using FlyEye Silhouette software

package, where relevant nuclei of neurons and precursors were manually selected. The ratio of YFP to RFP was calculated for each cell. I assume that each column of ommatidia represents on

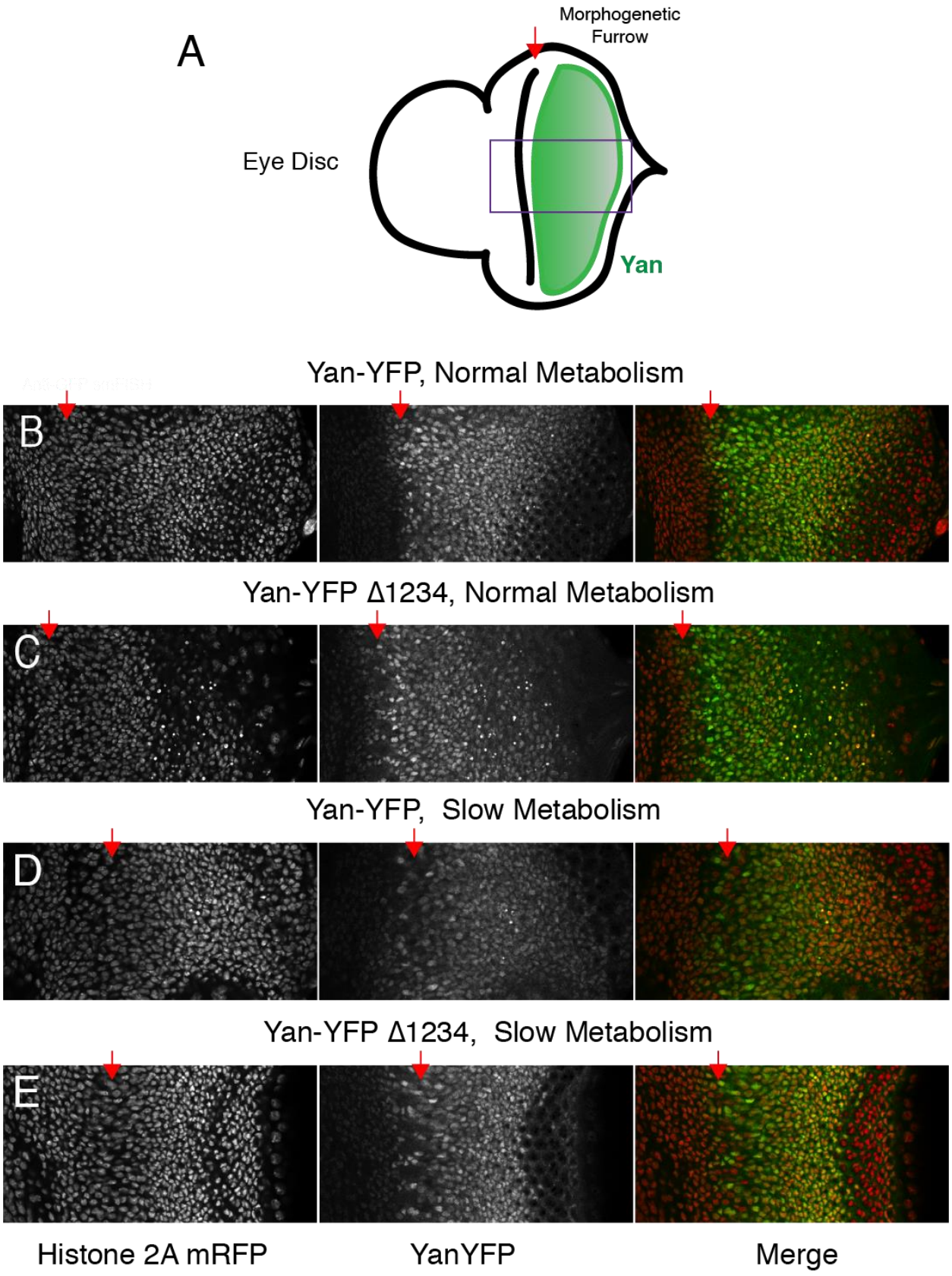
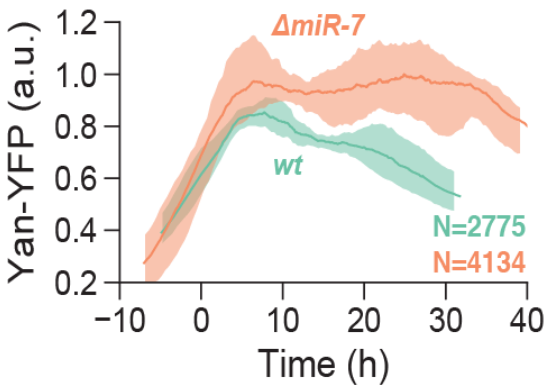
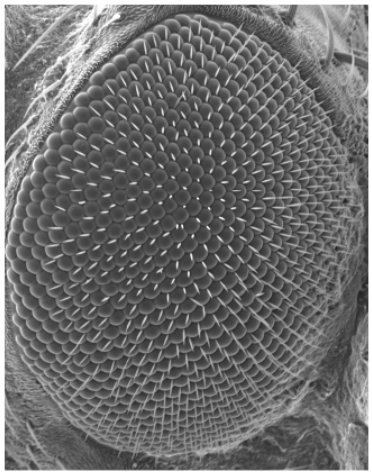

A
Morphogenetic
Furrow
Eye Disc
Yan
Yan-YFP, Normal Metabolism
B
Yan-YFP Δ1234, Normal Metabolism
C
Yan-YFP, Slow Metabolism
D
Yan-YFP Δ1234, Slow Metabolism
E
Histone 2A mRFP
YanYFP
Merge

Figure 2. Imaging Yan-YFP fluorescence. (A) Schematic of eye antennal imaginal disc complex, with approximate late third-instar expression pattern of Yan expression colored in green. Red arrow denotes morphogenetic furrow. Purple box indicates approximate region imaged for analysis. Anterior is to the left. **(B-E)** Representative confocal sections of eye imaginal discs expressing Histone-RFP (left) and *Yan*-YFP (center). Approximate location of morphogenetic furrow is indicated by red arrow. All sections are imaged about the equator of the disc, anterior is to the left. **(B)** *Yan*-YFP and Histone-RFP visualized in animals with normal metabolism, (dILP2-GAL4 only). **(C)** *Yan* $^{\Delta miR\text{-}7}$-YFP and Histone-RFP visualized in animals with normal metabolism, (dILP2-GAL4 only). **(D)** *Yan*-YFP and Histone-RFP visualized in animals with slow metabolism caused by genetic ablation of insulin-producing cells (dILP2-GAL4> UAS-*rpr*). **(E)** *Yan* $^{\Delta miR\text{-}7}$-YFP and Histone-RFP visualized in animals with slow metabolism caused by genetic ablation of insulin-producing cells (dILP2-GAL4> UAS-*rpr*).

A

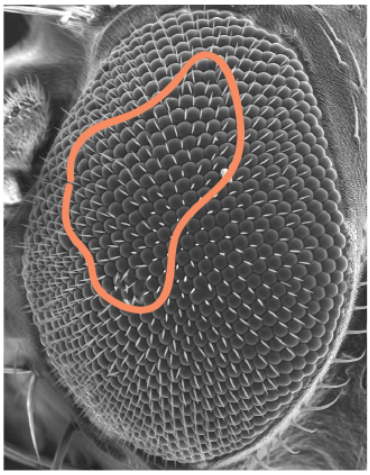


B



Wild-Type Eye

C



Mild Rough

D

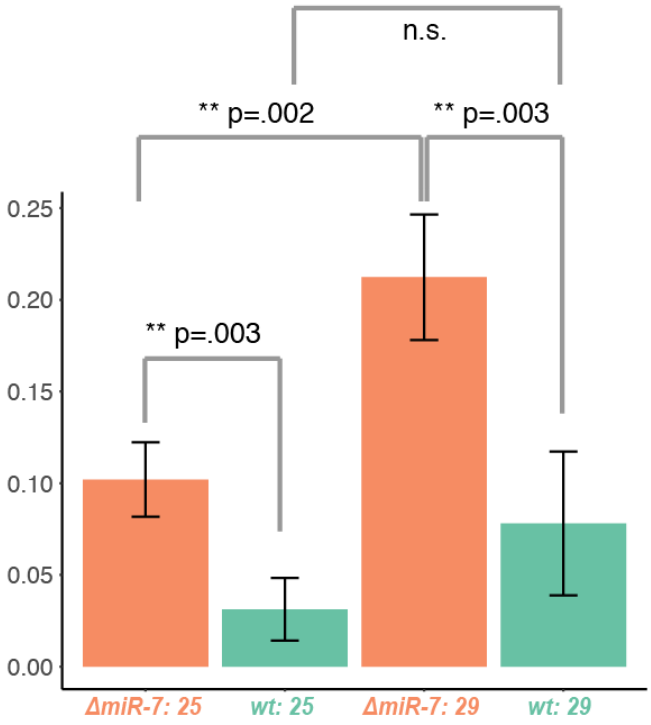


Figure 3. Loss of miR-7 repression on Yan effects in the eye. **(A)** Quantification of Yan-YFP protein dynamics from eye imaginal discs bearing Yan-YFP. Solid lines represent moving averages. Shaded regions denote 95% confidence intervals. Zero marks induction of Yan-YFP expression. **(B)** Representative SEM image of a wild type *Drosophila* eye. **(C)** Representative SEM image of a *Drosophila* eye scored as mildly rough. **(D)** Quantification of penetrance of mild eye roughening phenotype among Yan-YFP lines. Each bar represents the mean penetrance of the mild rough eye phenotype of 3 independent populations (derived from the same fly line) of 80-250 individuals. Error bars are SEM. P-values are result of Fisher's Exact Test.

average two hours of developmental time, allowing me to use column number as a proxy of developmental time by measuring each cell's distance relative to the nearest column.

As expected, both *Yan-YFP* and *Yan$^{\Delta miR\text{-}7}$-YFP* levels increased rapidly in precursor cells just posterior to the morphogenetic furrow. Upon specification to a neuronal fate, YFP levels decreased rapidly to basal levels as Yan-YFP responds to RTK signaling and is repressed. In precursors, Yan-YFP levels decrease over time at a more modest rate, remaining elevated relative to basal levels. Therefore, each single cell experiences a "pulse" of Yan expression followed by a decay over time. The decay is rapid in cells fated to become neurons and long in cells that remain uncommitted. The maximum level of Yan$^{\Delta miR\text{-}7}$-YFP seems to be elevated in precursor cells relative to Yan-YFP maximum level. Yan$^{\Delta miR\text{-}7}$-YFP decay appears also to be delayed somewhat in these precursor cells relative to Yan-YFP. Therefore, I conclude that miR-7 is acting as a weak repressor of Yan in precursor cells. This is consistent with known information about microRNA function on a molecular scale.

Metabolic Effects on miR-7 Insensitive Yan

A mathematical model was constructed by Sebastian Bernasek in the Amaral group at Northwestern to predict how metabolic rate would affect the pulse of Yan. Modeling was performed using a general engineering framework known as control theory. In this framework, an input stimulates a pulse of protein in a cell, which is then attenuated by the simulated inputs of many repressive control mechanisms. This process is iterated many times to create many simulated cells. Noise is introduced to simulate gene expression stochasticity, making each cell have a slightly different pulse trajectory through time. The ensemble of simulated cells created generated a range of likely outcomes. When a weak repressor such as a microRNA is removed from the model, the maximum output levels of the protein pulse increase, and attenuation is

delayed (Figure 4A). This behavior recapitulates what is observed in the dynamics of Yan$^{\Delta miR-7}$-YFP relative to Yan-YFP (Figure 3A and 4C). Slow metabolism was simulated in the model by reducing the rate parameters of all ATP-dependent reactions in the model. This results in a decreased protein output overall. Crucially, in the slow metabolism simulation, the removal of a weak repressor such as a microRNA does not alter the dynamics of the protein pulse (Figure 4B). I quantified Yan-YFP and Yan$^{\Delta miR-7}$-YFP dynamics in eye discs from larvae with genetic ablation of insulin producing cells (IPCs). In normal flies, IPCs secrete insulin-like peptides which regulate glucose uptake by body cells. IPC ablation results in increased circulating glucose, indicating that cells are taking up less glucose for metabolism. Overall energy metabolism is reduced 30%, as determined by whole-body calorimetry. As a result, IPC-ablated animals experience a decreased rate of growth and development. As predicted by the mathematical model, Yan-YFP dynamics were identical in these (Broughton et al., 2005; Rulifson et al., 2002). Slow metabolism flies with or without miR-7 regulation (Figure 4D). This result supports the hypothesis that redundant repressors such as miR-7 in gene regulatory networks are unnecessary under conditions of reduced metabolism and slowed growth.

Caloric restriction affects *salm* RNA levels

I wondered how gene expression at the RNA level was affected by slowed growth. To test this, I raised flies on food containing a decreased concentration of yeast, the major source of calories for *Drosophila* larvae (Ferreira and Milán, 2015). Calorically restricted larvae took nearly twice as long to reach pupariation. I performed smFISH on these animals for the imaginal disc developmental gene *salm.* Salm mRNA levels per cell were quantified per cell for both mature and nascent RNAs. Salm is not known to directly interact with pathways governing metabolism in the cell.

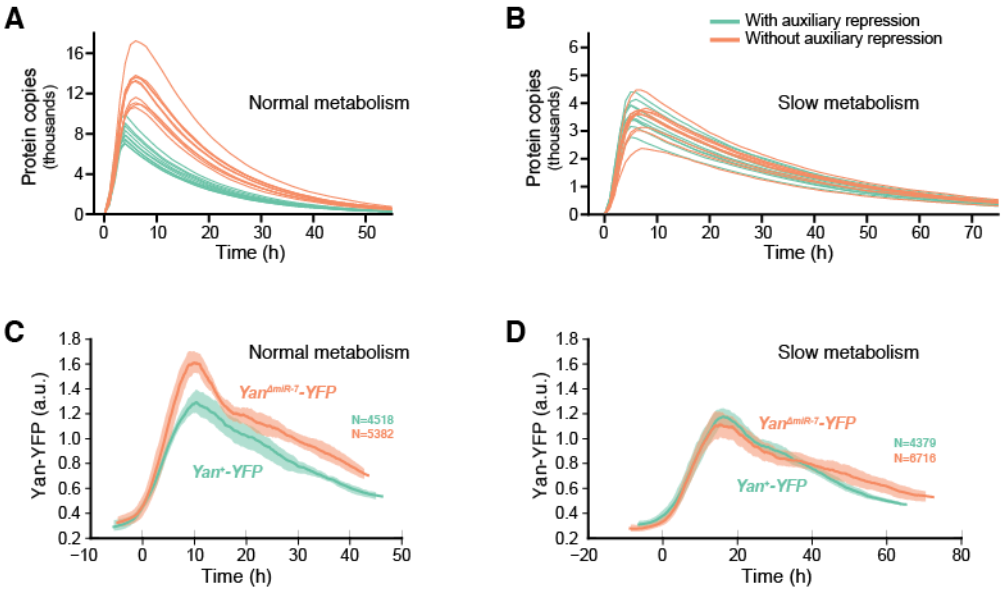
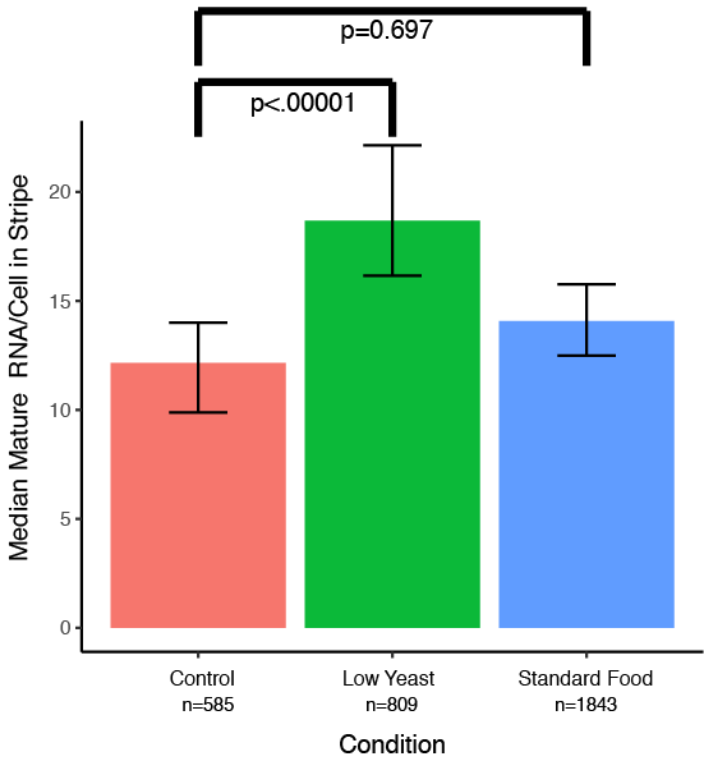


Figure 4. Gene expression dynamics with slow metabolism. (A-B) Simulated protein output under the control of an auxiliary post-transcriptional regulator (green) and when the repressor is removed (orange). Shown are ten randomly chosen samples from a total population of 5,000 simulated trajectories for each condition. Sebastian Bernasek performed these simulations. **(A)** Simulations performed with normal reaction rates. **(B)** Simulations performed following 50% reduction in the rate of ATP-dependent reactions. **(C-D)** Quantification of Yan-YFP and Yan$^{\Delta miR\text{-}7}$-YFP in precursor cells. Solid lines represent moving averages. Shaded regions denote 95% confidence intervals. Zero marks induction of Yan-YFP expression. **(C)** Yan-YFP dynamics under normal metabolic conditions, in which the animals contain the *dILP2-GAL4* gene only. **(D)** Yan-YFP dynamics in animals with genetic IPC ablation via the combination of *dILP2-GAL4* and UAS-*Reaper*.

A


p=0.697
p<.00001
Median Mature RNA/Cell in Stripe
20
15
10
5
0
Control
n=585
Low Yeast
n=809
Standard Food
n=1843
Condition

B

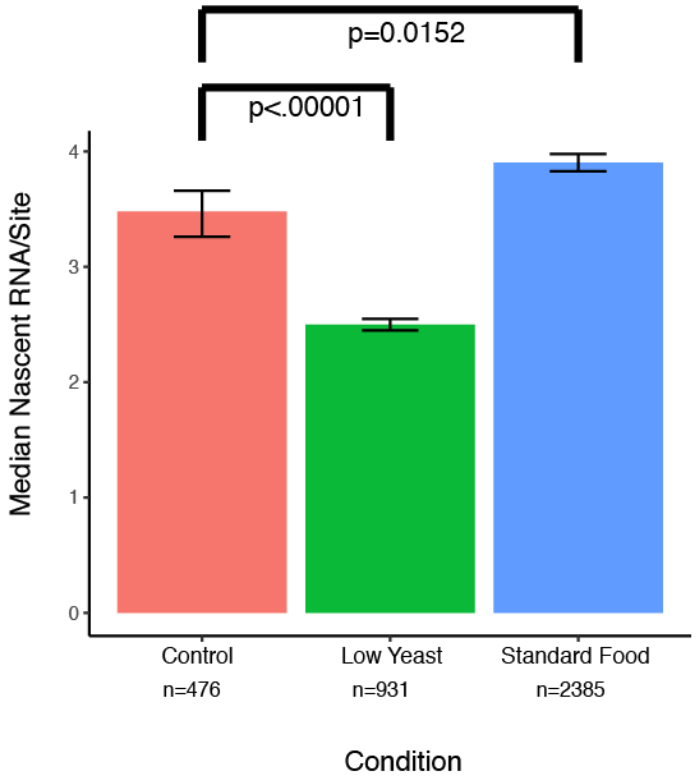

p=0.0152
p<.00001
Median Nascent RNA/Site
4
3
2
1
0
Control
n=476
Low Yeast
n=931
Standard Food
n=2385
Condition

Figure 5. Caloric restriction alters mature and nascent *salm* RNA levels. (A) Median mature RNA per cell within the *salm* stripe for glucose-cornmeal food (red), food with low concentration of yeast (green) and molasses-cornmeal food (blue). Error bars are 95% bootstrapped confidence intervals. P-values are the results of a permutation test for significance. **(B)** Median nascent RNAs per transcription site for *salm*. Error bars are 95% bootstrapped confidence intervals. P-values are the results of a permutation test for significance.

One might expect overall gene expression levels to decrease due to anabolic processes requiring more energy and calories to complete than catabolic processes. However, calorically restricted flies showed a significant increase in median *salm* mRNA levels per cell (Figure 5A). Conversely, flies under starvation conditions showed a significant decrease in the number of nascent *salm* RNAs per cell (Figure 5B). This suggests that the rate of *salm* transcription was reduced in some respect. Reduced transcription of *salm* coupled with an increase in mature RNAs seems to suggest that mRNA degradation was lower in these animals, or that some other means of post-transcriptional regulation had been affected. Overall, the results of these experiments serve as example of how such overarching processes as nutrient restriction can have varied effects on gene regulation. It is difficult to make broad claims about metabolism and gene regulation.

Discussion

My results show that miR-7 mediated repression of Yan in the eye imaginal disc progenitor cells serves to control Yan levels. Without it, progenitor cells show increased levels of Yan throughout eye development. When cellular metabolism is slowed by genetic ablation of IPCs, Yan levels are identical with or without miR-7 regulation. This suggests that under slowed metabolic conditions, miR-7 is not required to regulate Yan expression levels. MiR-7 regulation is one of many forms of auxilliary repression that ensure Yan levels remain within acceptable limits in the progenitor cells of eye disc during development. The requirement for these repressive inputs is relaxed when metabolism is slowed.

Many mutations in repressors cause phenotypes that are suppressed in slow metabolic conditions. My results help explain this behavior by suggesting that gene expression levels do not need to be as tightly regulated in slow metabolic conditions. However, the $Yan^{\Delta miR\text{-}7}$-YFP

gene has not been unambiguously linked with a phenotype in the adult eye. While the subtle effect of miR-7 on Yan expression doesn't appear to have noticeable consequence for final form of the eye, the existence of many redundant mechanisms for downregulating Yan in developing cells indicates that excessive cellular Yan is detrimental to development on some level. I speculate that miR-7 may be even less dispensable for proper Yan regulation outside of standard laboratory conditions.

It was previously observed that animals lacking any miR-7 regulation had elevated Yan immunofluorescence only in conditions of fluctuating temperature. At the time, this was conceptualized as miR-7 providing robustness to Yan gene expression in the eye under variable environmental conditions. The data presented here adds to that understanding. Presumably, the changes in Yan levels that are mediated by miR-7 are subtle enough that they are difficult to quantify without the genetic and image analysis tools that I used in this study. Temperature fluctuation has a wide range of organismal effects on animals. Temperature will presumably affect the reaction rates of all biochemistry in cells. Therefore, elevated temperatures put some stress on the animal and may result in an increased requirement for regulation of gene expression to coordinate growth and development.

It is also possible that suboptimal metabolic conditions result in a stress-induced reorganization of gene regulatory networks that suppresses mutant phenotypes. In the case of Yan, the introduction of genetic ablation of IPCs resulted in qualitative reduction in fertility and population size, indicating a that these animals may have been generally less healthy than their normal metabolism counterparts. Temperature fluctuations could also induce a stress response. Overall, its stands to reason that we require more precise tools for understanding the results of

changing conditions with wide-range effects, such as temperature manipulation or nutrient starvation.

Changes in glucose uptake will alter every cellular function that requires ATP. This includes fundamental mechanisms of gene expression such as transcription and translation. I found that in the case of *salm*, caloric restriction resulted in an increase of mature mRNA levels but not nascent mRNA levels. The simplest explanation for this is that mRNA decay rates have been altered by caloric restriction, resulting in accumulation of *salm* mRNAs in cells. The process of mRNA catabolism in not strictly ATP-dependent, therefore this effect would have to be achieved via regulation of ribonuclease activity. Very little is known the regulation of ribonuclease activity in any context (Houseley and Tollervey, 2009). Therefore, I can only speculate that caloric restriction might cause some cascading effect that slows mRNA decay.

Caloric restriction activates a wide array of gene expression responses that may interact with *salm* in unknown ways. Additionally, I do not have strong developmental markers in the wing imaginal disc. In these studies, I used entry into the white pre-pupa stage as the marker for developmental time, but imaginal disc development may be slightly out of step with whole entry into the pupal stage. To date, there are no longitudinal studies of *salm* expression through developmental time in the wing imaginal disc. It is entirely possible that *salm* transcript and expression levels decrease over developmental time, and that is what I observed.

Materials and Methods

Genetics

Unless otherwise indicated, flies were raised at 25° C on standard cornmeal-molasses food. The recombineered *Yan-YFP* transgene was described in (Webber et al., 2013). *Yan*$^{\Delta miR\text{-}7}$-

YFP was generated as described in (Cassidy et al., 2019). *Yan-YFP* and *Yan$^{\Delta miR-7}$-YFP* were recombined with *His2Av*-mRFP and placed in a *yan^{ER443}/yan^{E884}* mutant background.

Insulin producing cells were ablated by constructing flies bearing a *dILP2-GAL4* gene on the third chromosome, and a UAS-*Reaper (Rpr)* construct on the first chromosome. *dILP2-GAL4* consists of the promoter for the *insulin-like peptide* 2 gene fused to *GAL4*. This drives GAL4 expression specifically in brain IPCs (Rulifson et al., 2002). The *Rpr* gene creates a pro-apoptotic protein that causes cell death where expressed (Lohmann et al., 2002). The combination of *dILP2-GAL4* UAS-*Rpr* completely ablates insulin-producing cells in the larval brains of animals. These animals show elevated blood glucose and slow growth, taking nearly twice as long to reach maturity.

For slow-growth experiments with variants of *Yan-YFP*, flies bearing only the *dILP2-GAL4* gene were used as a control. Flies with *Yan-YFP* also contained 2 copies of endogenous *yan*, because introduction of the *yan* mutant background drastically reduced viability when combined with *dILP2-GAL4* UAS-*Rpr*.

Caloric Restriction

All flies were raised at 25° C. ~30 *w^{1118}* flies were allowed to lay eggs on molasses egg-laying plates for ~12 hours. Approximately 150 eggs were then collected and transferred to sugar-yeast-cornmeal food containing either a standard concentration of yeast, or food containing 10% of the standard amount of yeast. Wing discs from white prepupa were dissected, and fixation and smFISH was performed as described in Chapter IV.

Scanning Electron Microscopy

Adult flies were placed in small petri dishes and frozen at -20 degrees C for at least 20 minutes with dessicant to absorb excess moisture. Whole flies were mounted on stub specimen

mounts with carbon paint resin (ElectroDag 502). Scanning electron microscopy was performed using a Hitachi S-3400N-II cold-source field emission scanning electron microscope in low vacuum mode (30 pascals). Images were obtained using the 5-segment solid state backscatter detector. All imaging was performed at 190X magnification.

Yan-YFP Imaging

White pre-pupal eye discs were fixed in 4% paraformaldehyde/PBS, mounted in Vectashield, and stored in the dark at 4^{o} C until imaging a maximum of 24 hours later. 1024 X 512 images were captured using a Leica SP5 confocal microscope and 40X, NA 1.25 oil-immersion objective. Yan-YFP was visualized using an Argon 488 nm laser and Histone-RFP was visualized with a green HeNe 561 nm laser. Both channels were detected using HyD detectors. Images were oriented with the equator approximately parallel to the x-axis of the image, centered along the y-axis. Optical slices were set at 0.8um thickness. Zoom was set at 1.2-1.4X to minimize blank space in images, with 6-8 rows of ommatidia visible on either side of the equator. All discs for each condition were fixed, mounted, and imaged in parallel. Imaging was performed first thing in the morning to reduce variability in length of time the laser had been active.

Summary of Data Analysis

Segmentation. Cell segmentation and labeling were performed using FlyEye Silhouette, a custom GUI for cell labeling and analysis (Pelaez et al 2015, link to download). Automated cell segmentation was performed on each optical section individually, using H2Av-mRFP as the reference for nuclei cell boundaries. A single 2D contour was manually selected and labeled for each cell. Cell type identities were assigned based on nuclear position and morphology (Wolff and Ready 1993). Yan-YFP measurements were obtained by normalizing YFP fluorescence to

H2Av-mRFP fluorescence in each nucleus in order to normalize for differences in nucleus morphology and imaging conditions.

Distance to time calculations. Cell positions were mapped to developmental time as described previously in Pelaez et al. (2015). This method assumes each column of ommatidia is generated at a constant rate of 2 hours per R8 cell, as determined experimentally (Tomlinson and Ready, 1987). For each image z-stack, median distance between R8s was determined using a Delaunay triangulation and a time-to-distance conversion factor was calculated.

Calculation of Moving Averages. First, all replicates for each condition were aligned with one another in time. One disc was randomly chosen to be a reference disc. Each subsequent disc was adjusted in time to this reference disc by a factor calculated to maximize cross-correlation of moving averages (using a 10-cell window) of normalized Yan-YFP fluorescence. Once aligned in time, all cells were pooled and a 500-cell window moving average was calculated using hierarchical bootstrapping within each group.

Data analysis was done in collaboration with Dr. Sebastian Bernasek. Modeling was performed by Dr. Bernasek.

Chapter VI: Discussion

“Reports that say that something hasn't happened are always interesting to me, because as we know, there are known knowns; there are things we know we know. We also know there are known unknowns; that is to say we know there are some things we do not know. But there are also unknown unknowns—the ones we don't know we don't know. And if one looks throughout the history of our country and other free countries, it is the latter category that tend to be the difficult ones.”

-Donald Rumsfeld, United States Secretary of Defense 2001-2006

New tools for quantification yield new insights

The data presented in this thesis serve as an example of how quantification can yield new insights into gene expression. Less quantitative methods of fluorescent in situ hybridization could not detect the key observation that nascent RNA per transcription site does not seem to change in concert with overall mRNA levels across many different genes in the wing disc. Quantifying these developmental genes across the wing disc has also yield intriguing observations for further study. The existence of sharper stripe boundaries in protein expression relative to mRNA for *salm* and *omb* suggest some level of post-transcriptional regulation may be occurring for those genes. The asymmetric expression of *salm*-GFP relative to endogenous *salm* strongly suggests the existence of a distal control element for *salm* is missing in the *salm*-GFP transgene. These observations were not expected when I was developing the smFISH technology for the imaginal disc, but they offer further avenues for study.

Quantification of gene expression allowed me to see subtle changes in gene expression that are difficult to describe otherwise. The difference between Yan-YFP and Yan$^{\Delta miR\text{-}7}$-YFP is difficult for untrained eyes to detect from imaging alone but becomes clearer when nuclear YFP is quantified. Without the years of effort placed into the FlyEye Silhouette software, I would not have been able to detect this change, nor use it to test hypotheses about metabolic effects on gene expression.

These are examples of how new tools and technologies can lead to greater understanding of gene expression. The development and application of new tools for old problems needs to be a continuing research priority. New technologies allow us to further penetrate into the “unknown

unknowns," illuminating phenomena that we didn't even know existed before they could be observed.

What controls transcriptional bursting?

Transcriptional bursting is a phenomenon observed across all life, from bacteria to animals. It is reasonable to posit that it is an emergent property from the fundamental process of transcription initiation by RNA polymerase. What can bursting tell us about gene regulation on a molecular scale? Transcription is a step-wise process that consists of many biochemical reactions each characterized by kinetic rates. The existence of transcriptional bursts seems to suggest the existence of at least one rate-limiting step in this process that requires a longer period of time to overcome relative to the other steps in the process (Nicolas et al., 2017).

Experiments in bacteria suggest that transcriptional bursting occurs even in the presence of only the minimal components necessary for transcription initiation. Here, bursting might be induced by molecular torsion of DNA caused by RNA polymerase. Torsion induces upstream supercoiling of DNA that halts further polymerase elongation. Transcription is arrested until supercoiling can be relieved by DNA gyrase, a process which is dependent on availability and binding kinetics of these enzymes (Chong et al., 2014; Fujita et al., 2016). However, this hypothesis does not adequately explain the large differences in bursting kinetics seen between genes in eukaryotes.

Another source of delay in transcription could be the necessary formation of large protein complexes prior to transcription initiation. The formation of a pre-initation complex at the promoter has been described in detail. In eukaryotes, it consists of greater than 85 polypeptides which are present in diverse combinations at different genes. These are transcription factors, large complexes such as Mediator, and RNA polymerase subunits. Together, they prime the

promoter and surrounding chromatin for transcription initiation and elongation. Multi-step or progressive complex formation can result in the early protein-DNA interactions lowering the activation energy for additional protein-protein or protein-DNA interactions (Esnault et al., 2008; Larson et al., 2011). Thus, the initial formation events occur on longer timescales than later events, causing a complex to coalesce. Such a mechanism could also explain some gene specific properties of burst dynamics. Variation in promoter sequence and structure influences the binding and unbinding kinetics of components of the transcription initiation machinery, potentially influencing the behavior of the complex as a whole (Spitz and Furlong, 2012).

Additionally, I think of RNA polymerase binding kinetics as possibly having some "search time" across DNA where the molecule must diffuse to an appropriate promoter site before it can actually act. Once the elements of the pre-initiation complex are in place, this "search time" is greatly reduced, as they are likely to undergo cycles of DNA binding and unbinding before diffusing away. In such a manner, the formation of this complex could have switch-like dynamics.

Alternatively, the recruitment of RNA polymerase to the initiation complex may not be a limiting step. It has been suggested that what I might think of as initiation rate, or how often a polymerase initiates transcription, is not governed by polymerase recruitment, but rather by the degree of pause-release of RNA polymerase (Bartman et al., 2019). That is to say that initiation rate depends on how favorable conditions are for initiating transcription in a regulated manner. Such a system would certainly be mechanistically more viable in an evolutionary context: a method for the regulation of polymerase firing seems easier to regulate than binding energy of the polymerase itself, which is a more immutable property of the transcription start site. It seems that once the initiation complex is formed, RNA polymerase can initiate transcription at a given

rate. This concept of transcription certainly mirrors what I think of as a "burst": the slower entry into a state competent for transcription (the formation of the complex) and then a faster initiation rate (polymerase recruitment and firing).

Another cause of the dynamics may be the function of distal enhancers. Enhancers function as sites for binding of transcription factors. When contacting the promoter, they help foment chromatin changes and the formation of the preinitiation complex. In metazoans, enhancers are often located far away on the chromosome from the transcription start site. Therefore, there exists some DNA bending or architectural changes to the chromosomes in order for these distal enhancers to contact the promoter and affect transcriptions(Spitz and Furlong, 2012). This process of DNA bending and physical motion could be a source of burst-like dynamics. Bartman et al. 2016 showed that two separate genes entered states of active transcription only when in contact with a shared enhancer. Their transcription at any single moment of time was mutually exclusive. When the enhancer location moved away from the promoter, transcription did not occur (Bartman et al., 2016).

Recently, it has been suggested that the behavior of so called "super-enhancers" can be described as a process in which many components of the transcription machinery coalesce and form a liquid-phase body in the nucleus. This liquid phase body is so dense with proteins that it could resemble a droplet, and thus kinetic parameters that rely on the assumption of a uniform aqueous solution do not apply (Levine et al., 2014). Perhaps all enhancers work in a fundamentally similar manner separated more by degree: one can speculate that a transcriptional burst is caused the formation of a concentrated cluster of transcription proteins that are held together by intra-phase forces. Certainly, such behavior would enhance nonlinear transcription dynamics in these genes.

Lastly, we should consider the timescale for bursting behavior. In eukaryotes, many studies have found OFF times for promoters on the scale of tens of minutes, to hours (Milo et al., 2010; Suter et al., 2011). My data concurred with these findings, as the model presented in Chapter IV fits my smFISH data well with OFF times in this range.

Both DNA bending and complex formation are processes that generally have not been found to take this amount of time to complete. Formation of transcription initiation complexes, as well as diffusion of an enhancer around the nucleus, appear to occur on the order of seconds. Thus, while they may have some input into bursting behavior, they cannot completely explain the long off times observed between transcriptional bursts in many genes. In contrast, the process of chromatin remodeling and nucleosome positioning can operate on the timescale of minutes to hours. Therefore, it is reasonable to consider this process as a control mechanism for burst frequency. This idea is supported by the finding that many (if not most) transcription factors, recruit chromatin remodelers in order to modulate transcription output (Cai and Laughon, 2009; Hill, 2016; Spitz and Furlong, 2012). This idea is further supported by studies using model selection techniques. In our modeling, the simplest bursting model was used, in which the promoter has only an active or inactive state. In model selection, the promoter can have any number n states. The data is fit to these various models, and the best model is selected. Studies vary with respect to criteria for which a model is selected as the "best fit." Doing so, studies such as those by (Zoller et al., 2015) have found that the long OFF times observed in living cells are best described by up to five inactive states of the promoter. If such theoretical states have a physical correspondence, it is most likely to be a form of chromatin accessibility.

Ultimately, no one molecular process can be said to be the sole source of bursting dynamics. The phenomenon of transcriptional bursting is likely influenced by all of these molecular processes to various degrees, depending on the organism and context.

Tuning transcriptional bursting and noise

Individuals within genetically identical populations can show significant differences in gene expression. This is attributed to the stochastic influence on gene expression, or gene expression "noise". Transcriptional bursting influences gene expression noise. For example, cells with large, infrequent bursts will experience large fluctuations in levels of mRNA transcripts over time (Munsky et al., 2012). Fluctuations in mRNA can be propagated to fluctuations in protein levels. Fluctuations in one gene product can then be propagated throughout a biochemical network, causing noisy gene expression even in genes with low intrinsic noise.

Gene expression noise varies from gene to gene. Gene expression noise as a trait might be important to a gene's proper function in the cell (Dar et al., 2016). Therefore, it is possible that noise as a gene expression trait is acted upon by natural selection. Tuning of transcriptional bursting may be part of the mechanism by which natural selection acts on gene expression noise. Understanding if and how such tuning occurs may be important for synthetic biology, as we seek to harness biological circuits for industrial and medical purposes.

Future directions: Live imaging and the importance of time

All of the methods utilized in this thesis used fixed tissue. As such I can only make inferences about the dynamics of gene expression. I have done this either through the use of modeling, or by using position as a proxy for time as in the eye disc. Time courses are difficult to perform in the wing imaginal disc context, because of the difficulty of precise developmental

staging of discs. To make definitive conclusions about gene expression dynamics, I will ultimately need to observe these dynamics in real time using live imaging strategies.

In many ways, studying transcription via the MS2 system, which allows live-imaging of transcript production, is an attractive next step. This would allow us to visualize nascent transcription as it happens and physically measure burst size and frequency *in vivo*. Some of the pieces for such studies are already in place. Imaginal discs can be cultured as explants in growth media for several hours. *Drosophila* genetic constructs needed for the MS2 system exist (Garcia and Gregor, 2018). One can imagine that the combination of a microfluidics device for culture and imaging could allow for the performance of powerful experiments. Imaginal disc tissue could be given drugs that affect transcription via the culture media, and the effects of transcriptional bursting could be measured in real time.

Other non-trivial challenges remain to be solved, such as optimal imaging conditions and image and data analysis. Without fixing, clearing and mounting to help improve image quality, live-imaging can produce data that requires extensive cleanup and analysis. Additionally, the problems from imaging fixed tissue still exist primarily segmentation of single cells in this pseudostratified epithelium.

References

Aldaz, S., Escudero, L.M., 2010. Imaginal discs. Curr. Biol. 20, R429–R431. https://doi.org/10.1016/j.cub.2010.03.010

Alexandre, C., Baena-Lopez, A., Vincent, J.-P., 2014. Patterning and growth control by membrane-tethered Wingless. Nature 505, 180–185. https://doi.org/10.1038/nature12879

Alon, U., 2007. An introduction to systems biology: design principles of biological circuits, Chapman & Hall/CRC mathematical and computational biology series. Chapman & Hall/CRC, Boca Raton, FL.

Alwine, J.C., Kemp, D.J., Stark, G.R., 1977. Method for detection of specific RNAs in agarose gels by transfer to diazobenzyloxymethyl-paper and hybridization with DNA probes. Proc. Natl. Acad. Sci. U. S. A. 74, 5350–5354.

Ardehali, M.B., Yao, J., Adelman, K., Fuda, N.J., Petesch, S.J., Webb, W.W., Lis, J.T., 2009. Spt6 enhances the elongation rate of RNA polymerase II in vivo. EMBO J. 28, 1067–1077. https://doi.org/10.1038/emboj.2009.56

Arendt, J.D., 1997. Adaptive Intrinsic Growth Rates: An Integration Across Taxa. Q. Rev. Biol. 72, 149–177. https://doi.org/10.1086/419764

Bakker, R., Mani, M., Carthew, R.W., 2020. The Wg and Dpp morphogens regulate gene expression by modulating the frequency of transcriptional bursts. eLife 9, e56076. https://doi.org/10.7554/eLife.56076

Bannister, A.J., Kouzarides, T., 2011. Regulation of chromatin by histone modifications. Cell Res. 21, 381–395. https://doi.org/10.1038/cr.2011.22

Bartlett, J.M.S., 2002. Approaches to the analysis of gene expression using mRNA: a technical overview. Mol. Biotechnol. 21, 149–160. https://doi.org/10.1385/MB:21:2:149

Bartman, C.R., Hamagami, N., Keller, C.A., Giardine, B., Hardison, R.C., Blobel, G.A., Raj, A., 2019. Transcriptional Burst Initiation and Polymerase Pause Release Are Key Control Points of Transcriptional Regulation. Mol. Cell 73, 519-532.e4. https://doi.org/10.1016/j.molcel.2018.11.004

Bartman, C.R., Hsu, S.C., Hsiung, C.C.-S., Raj, A., Blobel, G.A., 2016. Enhancer Regulation of Transcriptional Bursting Parameters Revealed by Forced Chromatin Looping. Mol. Cell 62, 237–247. https://doi.org/10.1016/j.molcel.2016.03.007

Beira, J.V., Paro, R., 2016. The legacy of Drosophila imaginal discs. Chromosoma 125, 573–592. https://doi.org/10.1007/s00412-016-0595-4

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. Proc. Natl. Acad. Sci. 111, 10598–10603. https://doi.org/10.1073/pnas.1410022111

Broughton, S.J., Piper, M.D.W., Ikeya, T., Bass, T.M., Jacobson, J., Driege, Y., Martinez, P., Hafen, E., Withers, D.J., Leevers, S.J., Partridge, L., 2005. Longer lifespan, altered metabolism, and stress resistance in Drosophila from ablation of cells making insulin-like ligands. Proc. Natl. Acad. Sci. U. S. A. 102, 3105–3110. https://doi.org/10.1073/pnas.0405775102

Cai, Y., Laughon, A., 2009. The Drosophila Smad cofactor Schnurri engages in redundant and synergistic interactions with multiple corepressors. Biochim. Biophys. Acta BBA - Gene Regul. Mech. 1789, 232–245. https://doi.org/10.1016/j.bbagrm.2009.01.001

Campbell, G., Tomlinson, A., 1999. Transducing the Dpp Morphogen Gradient in the Wing of Drosophila: Regulation of Dpp Targets by brinker. Cell 96, 553–562. https://doi.org/10.1016/S0092-8674(00)80659-5

Campbell, S., Inamdar, M., Rodrigues, V., Raghavan, V., Palazzolo, M., Chovnick, A., 1992. The scalloped gene encodes a novel, evolutionarily conserved transcription factor required for sensory organ differentiation in Drosophila. Genes Dev. 6, 367–379. https://doi.org/10.1101/gad.6.3.367

Cassidy, J.J., Bernasek, S.M., Bakker, R., Giri, R., Peláez, N., Eder, B., Bobrowska, A., Bagheri, N., Amaral, L.A.N., Carthew, R.W., 2019. Repressive Gene Regulation Synchronizes Development with Cellular Metabolism. Cell 178, 980-992.e17. https://doi.org/10.1016/j.cell.2019.06.023

Cassidy, J.J., Jha, A.R., Posadas, D.M., Giri, R., Venken, K.J.T., Ji, J., Jiang, H., Bellen, H.J., White, K.P., Carthew, R.W., 2013. miR-9a minimizes the phenotypic impact of genomic diversity by buffering a transcription factor. Cell 155, 1556–1567. https://doi.org/10.1016/j.cell.2013.10.057

Celis, J.F. de, Barrio, R., Kafatos, F.C., 1996. A gene complex acting downstream of dpp in Drosophila wing morphogenesis. Nature 381, 421. https://doi.org/10.1038/381421a0

Chen, L.-F., Lin, Y.T., Gallegos, D.A., Hazlett, M.F., Gómez-Schiavon, M., Yang, M.G., Kalmeta, B., Zhou, A.S., Holtzman, L., Gersbach, C.A., Grandl, J., Buchler, N.E., West, A.E., 2019. Enhancer Histone Acetylation Modulates Transcriptional Bursting Dynamics of Neuronal Activity-Inducible Genes. Cell Rep. 26, 1174-1188.e5. https://doi.org/10.1016/j.celrep.2019.01.032

Child, G., 1939. The effect of increasing time of development at constant temperature on the wing size of vestigial of drosophila melanogaster. Biol. Bull. 77, 432–442. https://doi.org/10.2307/1537653

Child, G., 1935. Phenogenetic Studies on Scute-1 of Drosophila Melanogaster. I. the Associations between the Bristles and the Effects of Genetic Modifiers and Temperature. Genetics 20, 109–126.

Chong, S., Chen, C., Ge, H., Xie, X.S., 2014. Mechanism of Transcriptional Bursting in Bacteria. Cell 158, 314–326. https://doi.org/10.1016/j.cell.2014.05.038

Chubb, J.R., Trcek, T., Shenoy, S.M., Singer, R.H., 2006. Transcriptional Pulsing of a Developmental Gene. Curr. Biol. 16, 1018–1025. https://doi.org/10.1016/j.cub.2006.03.092

Clevers, H., Nusse, R., 2012. Wnt/β-Catenin Signaling and Disease. Cell 149, 1192–1205. https://doi.org/10.1016/j.cell.2012.05.012

Collins, T.J., 2007. ImageJ for microscopy. BioTechniques 43, 25–30. https://doi.org/10.2144/000112517

Dar, R.D., Shaffer, S.M., Singh, A., Razooky, B.S., Simpson, M.L., Raj, A., Weinberger, L.S., 2016. Transcriptional Bursting Explains the Noise–Versus–Mean Relationship in mRNA and Protein Levels. PLOS ONE 11, e0158298. https://doi.org/10.1371/journal.pone.0158298

D'Errico, J., 2012. Inhull.

Dey, S.S., Foley, J.E., Limsirichai, P., Schaffer, D.V., Arkin, A.P., 2015. Orthogonal control of expression mean and variance by epigenetic features at different genomic loci. Mol. Syst. Biol. 11. https://doi.org/10.15252/msb.20145704

Eng, C.-H.L., Lawson, M., Zhu, Q., Dries, R., Koulena, N., Takei, Y., Yun, J., Cronin, C., Karp, C., Yuan, G.-C., Cai, L., 2019. Transcriptome-scale super-resolved imaging in tissues by RNA seqFISH. Nature 568, 235–239. https://doi.org/10.1038/s41586-019-1049-y

Esnault, C., Ghavi-Helm, Y., Brun, S., Soutourina, J., Van Berkum, N., Boschiero, C., Holstege, F., Werner, M., 2008. Mediator-Dependent Recruitment of TFIIH Modules in Preinitiation Complex. Mol. Cell 31, 337–346. https://doi.org/10.1016/j.molcel.2008.06.021

Ezer, D., Moignard, V., Göttgens, B., Adryan, B., 2016. Determining Physical Mechanisms of Gene Expression Regulation from Single Cell Gene Expression Data. PLOS Comput. Biol. 12, e1005072. https://doi.org/10.1371/journal.pcbi.1005072

Farrell, R.E., 2005. RNA methodologies: a laboratory guide for isolation and characterization, 3. ed. ed. Elsevier, Amsterdam.

Femino, A.M., Fay, F.S., Fogarty, K., Singer, R.H., 1998. Visualization of Single RNA Transcripts in Situ. Science 280, 585–590. https://doi.org/10.1126/science.280.5363.585

Ferreira, A., Milán, M., 2015. Dally Proteoglycan Mediates the Autonomous and Nonautonomous Effects on Tissue Growth Caused by Activation of the PI3K and TOR Pathways. PLOS Biol. 13, e1002239. https://doi.org/10.1371/journal.pbio.1002239

FlyBase Reference Report: model organism Encyclopedia of Regulatory Network (modERN) Project, 2015-, Genomic BAC constructs containing epitope tagged proteins, generated by the modERN Project. [WWW Document], n.d. URL http://flybase.org/reports/FBrf0228168.html (accessed 9.16.20).

Freier, S.M., Kierzek, R., Jaeger, J.A., Sugimoto, N., Caruthers, M.H., Neilson, T., Turner, D.H., 1986. Improved free-energy parameters for predictions of RNA duplex stability. Proc. Natl. Acad. Sci. U. S. A. 83, 9373–9377.

Fujita, K., Iwaki, M., Yanagida, T., 2016. Transcriptional bursting is intrinsically caused by interplay between RNA polymerases on DNA. Nat. Commun. 7, 13788. https://doi.org/10.1038/ncomms13788

Fukaya, T., Lim, B., Levine, M., 2016. Enhancer Control of Transcriptional Bursting. Cell 166, 358–368. https://doi.org/10.1016/j.cell.2016.05.025

Gall, J.G., 2016. The origin of in situ hybridization – A personal history. Methods, Methods for the Detection of Transcription and mRNA Localization in Cells and Tissues 98, 4–9. https://doi.org/10.1016/j.ymeth.2015.11.026

Garcia, H., Gregor, T., 2018. Live imaging of mRNA synthesis in Drosophila. Methods Mol. Biol. Clifton NJ 1649, 349–357. https://doi.org/10.1007/978-1-4939-7213-5_23

Gillespie, D.T., 1977. Exact stochastic simulation of coupled chemical reactions. J. Phys. Chem. 81, 2340–2361. https://doi.org/10.1021/j100540a008

Gillooly, J.F., Charnov, E.L., West, G.B., Savage, V.M., Brown, J.H., 2002. Effects of size and temperature on developmental time. Nature 417, 70–73. https://doi.org/10.1038/417070a

Giri, R., Papadopoulos, D.K., Posadas, D.M., Potluri, H.K., Tomancak, P., Mani, M., Carthew, R.W., 2020. Ordered patterning of the sensory system is susceptible to stochastic features of gene expression. eLife 9. https://doi.org/10.7554/eLife.53638

Gould, S.J., 1977. Ontogeny and Phylogeny. Harvard University Press.

Grande, J.C., 2012. Principles of Image Analysis. Metallogr. Microstruct. Anal. 1, 227–243. https://doi.org/10.1007/s13632-012-0037-5

Grimm, S., Pflugfelder, G.O., 1996. Control of the Gene optomotor-blind in Drosophila Wing Development by decapentaplegic and wingless. Science 271, 1601–1604. https://doi.org/10.1126/science.271.5255.1601

Hamaratoglu, F., Affolter, M., Pyrowolakis, G., 2014. Dpp/BMP signaling in flies: From molecules to biology. Semin. Cell Dev. Biol., RNA biogenesis & TGFβ signalling in embryonic development 32, 128–136. https://doi.org/10.1016/j.semcdb.2014.04.036

Hill, C.S., 2016. Transcriptional Control by the SMADs. Cold Spring Harb. Perspect. Biol. 8, a022079. https://doi.org/10.1101/cshperspect.a022079

Hobro, A.J., Smith, N.I., 2017. An evaluation of fixation methods: Spatial and compositional cellular changes observed by Raman imaging. Vib. Spectrosc., Prominent Young Vibrational Spectroscopists 91, 31–45. https://doi.org/10.1016/j.vibspec.2016.10.012

Hoetelmans, R.W., Prins, F.A., Cornelese-ten Velde, I., van der Meer, J., van de Velde, C.J., van Dierendonck, J.H., 2001. Effects of acetone, methanol, or paraformaldehyde on cellular structure, visualized by reflection contrast microscopy and transmission and scanning electron microscopy. Appl. Immunohistochem. Mol. Morphol. AIMM 9, 346–351. https://doi.org/10.1097/00129039-200112000-00010

Hollandi, R., Szkalisity, A., Toth, T., Tasnadi, E., Molnar, C., Mathe, B., Grexa, I., Molnar, J., Balind, A., Gorbe, M., Kovacs, M., Migh, E., Goodman, A., Balassa, T., Koos, K., Wang, W., Caicedo, J.C., Bara, N., Kovacs, F., Paavolainen, L., Danka, T., Kriston, A., Carpenter, A.E., Smith, K., Horvath, P., 2020. nucleAIzer: A Parameter-free Deep Learning Framework for Nucleus Segmentation Using Image Style Transfer. Cell Syst. 10, 453-458.e6. https://doi.org/10.1016/j.cels.2020.04.003

Holloway, D.M., Spirov, A.V., 2017. Transcriptional bursting in Drosophila development: Stochastic dynamics of eve stripe 2 expression. PLOS ONE 12, e0176228. https://doi.org/10.1371/journal.pone.0176228

Houseley, J., Tollervey, D., 2009. The Many Pathways of RNA Degradation. Cell 136, 763–776. https://doi.org/10.1016/j.cell.2009.01.019

Julian, R.J., 2005. Production and growth related disorders and other metabolic diseases of poultry – A review. Vet. J. 169, 350–369. https://doi.org/10.1016/j.tvjl.2004.04.015

Kicheva, A., Cohen, M., Briscoe, J., 2012. Developmental pattern formation: insights from physics and biology. Science 338, 210–212. https://doi.org/10.1126/science.1225182

Kumar, J.P., 2001. Signalling pathways in Drosophila and vertebrate retinal development. Nat. Rev. Genet. 2, 846–857. https://doi.org/10.1038/35098564

Kuntz, S.G., Eisen, M.B., 2014. Drosophila Embryogenesis Scales Uniformly across Temperature in Developmentally Diverse Species. PLOS Genet. 10, e1004293. https://doi.org/10.1371/journal.pgen.1004293

Kwon, S., 2013. Single-molecule fluorescence in situ hybridization: Quantitative imaging of single RNA molecules. BMB Rep. 46, 65–72. https://doi.org/10.5483/BMBRep.2013.46.2.016

Lähnemann, D., Köster, J., Szczurek, E., McCarthy, D.J., Hicks, S.C., Robinson, M.D., Vallejos, C.A., Campbell, K.R., Beerenwinkel, N., Mahfouz, A., Pinello, L., Skums, P., Stamatakis, A., Attolini, C.S.-O., Aparicio, S., Baaijens, J., Balvert, M., Barbanson, B. de, Cappuccio, A., Corleone, G., Dutilh, B.E., Florescu, M., Guryev, V., Holmer, R., Jahn, K., Lobo, T.J., Keizer, E.M., Khatri, I., Kielbasa, S.M., Korbel, J.O., Kozlov, A.M., Kuo, T.-H., Lelieveldt, B.P.F., Mandoiu, I.I., Marioni, J.C., Marschall, T., Mölder, F., Niknejad, A., Raczkowski, L., Reinders, M., Ridder, J. de, Saliba, A.-E., Somarakis, A.,

Stegle, O., Theis, F.J., Yang, H., Zelikovsky, A., McHardy, A.C., Raphael, B.J., Shah, S.P., Schönhuth, A., 2020. Eleven grand challenges in single-cell data science. Genome Biol. 21, 31. https://doi.org/10.1186/s13059-020-1926-6

Lai, Z.-C., Rubin, G.M., 1992. Negative control of photoreceptor development in Drosophila by the product of the yan gene, an ETS domain protein. Cell 70, 609–620. https://doi.org/10.1016/0092-8674(92)90430-K

Larson, D.R., Zenklusen, D., Wu, B., Chao, J.A., Singer, R.H., 2011. Real-Time Observation of Transcription Initiation and Elongation on an Endogenous Yeast Gene. Science 332, 475–478. https://doi.org/10.1126/science.1202142

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, 251. https://doi.org/10.1038/s41586-018-0836-1

Lee, M.V., Topper, S.E., Hubler, S.L., Hose, J., Wenger, C.D., Coon, J.J., Gasch, A.P., 2011. A dynamic model of proteome changes reveals new roles for transcript alteration in yeast. Mol. Syst. Biol. 7, 514. https://doi.org/10.1038/msb.2011.48

Levine, M., Cattoglio, C., Tjian, R., 2014. Looping Back to Leap Forward: Transcription Enters a New Era. Cell 157, 13–25. https://doi.org/10.1016/j.cell.2014.02.009

Levine, M., Davidson, E.H., 2005. Gene regulatory networks for development. Proc. Natl. Acad. Sci. U. S. A. 102, 4936–4942. https://doi.org/10.1073/pnas.0408031102

Levsky, J.M., Singer, R.H., 2003. Fluorescence in situ hybridization: past, present and future. J. Cell Sci. 116, 2833–2838. https://doi.org/10.1242/jcs.00633

Li, X., Carthew, R.W., 2005. A microRNA mediates EGF receptor signaling and promotes photoreceptor differentiation in the Drosophila eye. Cell 123, 1267–1277. https://doi.org/10.1016/j.cell.2005.10.040

Li, X., Cassidy, J.J., Reinke, C.A., Fischboeck, S., Carthew, R.W., 2009. A microRNA Imparts Robustness Against Environmental Fluctuation During Development. Cell 137, 273–282. https://doi.org/10.1016/j.cell.2009.01.058

Lipson, S.G., Lipson, H., Tannhauser, D.S., 1995. Optical Physics. Cambridge University Press.

Little, S.C., Tikhonov, M., Gregor, T., 2013. Precise Developmental Gene Expression Arises from Globally Stochastic Transcriptional Activity. Cell 154, 789–800. https://doi.org/10.1016/j.cell.2013.07.025

Liu, Y., Beyer, A., Aebersold, R., 2016. On the Dependency of Cellular Protein Levels on mRNA Abundance. Cell 165, 535–550. https://doi.org/10.1016/j.cell.2016.03.014

Lohmann, I., McGinnis, N., Bodmer, M., McGinnis, W., 2002. The Drosophila Hox Gene Deformed Sculpts Head Morphology via Direct Regulation of the Apoptosis Activator reaper. Cell 110, 457–466. https://doi.org/10.1016/S0092-8674(02)00871-1

Lubeck, E., Coskun, A.F., Zhiyentayev, T., Ahmad, M., Cai, L., 2014. Single-cell in situ RNA profiling by sequential hybridization. Nat. Methods 11, 360–361. https://doi.org/10.1038/nmeth.2892

Ma, Y.J., Dissen, G.A., Rage, F., Ojeda, S.R., 1996. RNase Protection Assay. Methods 10, 273–278. https://doi.org/10.1006/meth.1996.0102

Martín, F.A., Herrera, S.C., Morata, G., 2009. Cell competition, growth and size control in the Drosophila wing imaginal disc. Development 136, 3747–3756. https://doi.org/10.1242/dev.038406

Mayer, L.R., Diegelmann, S., Abassi, Y., Eichinger, F., Pflugfelder, G.O., 2013. Enhancer trap infidelity in Drosophila optomotor-blind. Fly (Austin) 7, 118–128. https://doi.org/10.4161/fly.23657

Metz, C.W., 1916. Chromosome studies on the Diptera. II. The paired association of chromosomes in the Diptera, and its significance. J. Exp. Zool. 21, 213–279. https://doi.org/10.1002/jez.1400210204

Milo, R., Jorgensen, P., Moran, U., Weber, G., Springer, M., 2010. BioNumbers—the database of key numbers in molecular and cell biology. Nucleic Acids Res. 38, D750–D753. https://doi.org/10.1093/nar/gkp889

Minaee, S., Boykov, Y., Porikli, F., Plaza, A., Kehtarnavaz, N., Terzopoulos, D., 2020. Image Segmentation Using Deep Learning: A Survey. ArXiv200105566 Cs.

Minami, M., Kinoshita, N., Kamoshida, Y., Tanimoto, H., Tabata, T., 1999. brinker is a target of Dpp in Drosophila that negatively regulates Dpp-dependent genes. Nature 398, 242. https://doi.org/10.1038/18451

Moen, E., Bannon, D., Kudo, T., Graf, W., Covert, M., Van Valen, D., 2019. Deep learning for cellular image analysis. Nat. Methods 16, 1233–1246. https://doi.org/10.1038/s41592-019-0403-1

Moser, M., Campbell, G., 2005. Generating and interpreting the Brinker gradient in the Drosophila wing. Dev. Biol. 286, 647–658. https://doi.org/10.1016/j.ydbio.2005.08.036

Mundorf, J., Uhlirova, M., 2016. The Drosophila Imaginal Disc Tumor Model: Visualization and Quantification of Gene Expression and Tumor Invasiveness Using Genetic Mosaics. J. Vis. Exp. JoVE. https://doi.org/10.3791/54585

Munsky, B., Neuert, G., Oudenaarden, A. van, 2012. Using Gene Expression Noise to Understand Gene Regulation. Science 336, 183–187. https://doi.org/10.1126/science.1216379

Nakai, D., Yuasa, S., Takahashi, M., Shimizu, T., Asaumi, S., Isono, K., Takao, T., Suzuki, Y., Kuroyanagi, H., Hirokawa, K., Koseki, H., Shirsawa, T., 2001. Mouse homologue of coq7/clk-1, longevity gene in Caenorhabditis elegans, is essential for coenzyme Q synthesis, maintenance of mitochondrial integrity, and neurogenesis. Biochem. Biophys. Res. Commun. 289, 463–471. https://doi.org/10.1006/bbrc.2001.5977

Nicolas, D., Phillips, N.E., Naef, F., 2017. What shapes eukaryotic transcriptional bursting? Mol. Biosyst. 13, 1280–1290. https://doi.org/10.1039/C7MB00154A

Nolo, R., Abbott, L.A., Bellen, H.J., 2000. Senseless, a Zn finger transcription factor, is necessary and sufficient for sensory organ development in Drosophila. Cell 102, 349–362. https://doi.org/10.1016/s0092-8674(00)00040-4

Osterwalder, T., Yoon, K.S., White, B.H., Keshishian, H., 2001. A conditional tissue-specific transgene expression system using inducible GAL4. Proc. Natl. Acad. Sci. 98, 12596–12601. https://doi.org/10.1073/pnas.221303298

Park, H., 2020. Polytope bounded order-2 Voronoi diagram in 2D/3D.

Pavlidis, T., 1988. Image Analysis. Annu. Rev. Comput. Sci. 3, 121–146. https://doi.org/10.1146/annurev.cs.03.060188.001005

Peláez, N., Gavalda-Miralles, A., Wang, B., Navarro, H.T., Gudjonson, H., Rebay, I., Dinner, A.R., Katsaggelos, A.K., Amaral, L.A., Carthew, R.W., 2015. Dynamics and heterogeneity of a fate determinant during transition towards cell differentiation. eLife 4, e08924. https://doi.org/10.7554/eLife.08924

Peter, I.S., Davidson, E.H., 2016. Chapter Thirteen - Implications of Developmental Gene Regulatory Networks Inside and Outside Developmental Biology, in: Wassarman, P.M. (Ed.), Current Topics in Developmental Biology, Essays on Developmental Biology, Part B. Academic Press, pp. 237–251. https://doi.org/10.1016/bs.ctdb.2015.12.014

Pontzer, H., Brown, M.H., Raichlen, D.A., Dunsworth, H., Hare, B., Walker, K., Luke, A., Dugas, L.R., Durazo-Arvizu, R., Schoeller, D., Plange-Rhule, J., Bovet, P., Forrester, T.E., Lambert, E.V., Thompson, M.E., Shumaker, R.W., Ross, S.R., 2016. Metabolic acceleration and the evolution of human brain size and life history. Nature 533, 390–392. https://doi.org/10.1038/nature17654

Price, M.D., Lai, Z.-C., 1999. The yan gene is highly conserved in Drosophila and its expression suggests a complex role throughout development. Dev. Genes Evol. 209, 207–217. https://doi.org/10.1007/s004270050245

Qi, J., Wang, B., Pelaez, N., Rebay, L., Carthew, R.W., Katsaggelos, A.K., Amaral, L.A.N., 2013. Drosophila eye nuclei segmentation based on graph cut and convex shape prior, in: 2013 IEEE International Conference on Image Processing. Presented at the 2013 IEEE International Conference on Image Processing, pp. 670–674. https://doi.org/10.1109/ICIP.2013.6738138

Raj, A., Peskin, C.S., Tranchina, D., Vargas, D.Y., Tyagi, S., 2006. Stochastic mRNA Synthesis in Mammalian Cells. PLOS Biol. 4, e309. https://doi.org/10.1371/journal.pbio.0040309

Raj, A., Tyagi, S., 2010. Chapter 17 - Detection of Individual Endogenous RNA Transcripts In Situ Using Multiple Singly Labeled Probes, in: Walter, N.G. (Ed.), Methods in Enzymology, Single Molecule Tools: Fluorescence Based Approaches, Part A. Academic Press, pp. 365–386. https://doi.org/10.1016/S0076-6879(10)72004-8

Raj, A., van Oudenaarden, A., 2008. Nature, Nurture, or Chance: Stochastic Gene Expression and Its Consequences. Cell 135, 216–226. https://doi.org/10.1016/j.cell.2008.09.050

Rebay, I., Rubin, G.M., 1995. Yan functions as a general inhibitor of differentiation and is negatively regulated by activation of the Ras1/MAPK pathway. Cell 81, 857–866. https://doi.org/10.1016/0092-8674(95)90006-3

Rulifson, E.J., Kim, S.K., Nusse, R., 2002. Ablation of insulin-producing neurons in flies: growth and diabetic phenotypes. Science 296, 1118–1120. https://doi.org/10.1126/science.1070058

Russ, J.C., Neal, F.B., Neal, F.B., 2018. The Image Processing Handbook. CRC Press. https://doi.org/10.1201/b18983

Sandoval, J., Rodríguez, J.L., Tur, G., Serviddio, G., Pereda, J., Boukaba, A., Sastre, J., Torres, L., Franco, L., López-Rodas, G., 2004. RNAPol-ChIP: a novel application of chromatin immunoprecipitation to the analysis of real-time gene transcription. Nucleic Acids Res. 32, e88. https://doi.org/10.1093/nar/gnh091

Schneuwly, S., Gehring, W.J., 1985. Homeotic transformation of thorax into head: Developmental analysis of a new Antennapedia allele in Drosophila melanogaster. Dev. Biol. 108, 377–386. https://doi.org/10.1016/0012-1606(85)90041-7

Senecal, A., Munsky, B., Proux, F., Ly, N., Braye, F.E., Zimmer, C., Mueller, F., Darzacq, X., 2014. Transcription Factors Modulate c-Fos Transcriptional Bursts. Cell Rep. 8, 75–83. https://doi.org/10.1016/j.celrep.2014.05.053

Shi, Y., Massagué, J., 2003. Mechanisms of TGF-β Signaling from Cell Membrane to the Nucleus. Cell 113, 685–700. https://doi.org/10.1016/S0092-8674(03)00432-X

Sivasankaran, R., Vigano, M.A., Müller, B., Affolter, M., Basler, K., 2000. Direct transcriptional control of the Dpp target omb by the DNA binding protein Brinker. EMBO J. 19, 6162–6172. https://doi.org/10.1093/emboj/19.22.6162

Smale, S.T., 2009. Nuclear Run-On Assay. Cold Spring Harb. Protoc. 2009, pdb.prot5329. https://doi.org/10.1101/pdb.prot5329

Spitz, F., Furlong, E.E.M., 2012. Transcription factors: from enhancer binding to developmental control. Nat. Rev. Genet. 13, 613–626. https://doi.org/10.1038/nrg3207

Srinivasan, M., Sedmak, D., Jewell, S., 2002. Effect of Fixatives and Tissue Processing on the Content and Integrity of Nucleic Acids. Am. J. Pathol. 161, 1961–1971. https://doi.org/10.1016/S0002-9440(10)64472-0

Ståhl, P.L., Salmén, F., Vickovic, S., Lundmark, A., Navarro, J.F., Magnusson, J., Giacomello, S., Asp, M., Westholm, J.O., Huss, M., Mollbrink, A., Linnarsson, S., Codeluppi, S., Borg, Å., Pontén, F., Costea, P.I., Sahlén, P., Mulder, J., Bergmann, O., Lundeberg, J., Frisén, J., 2016. Visualization and analysis of gene expression in tissue sections by spatial transcriptomics. Science 353, 78–82. https://doi.org/10.1126/science.aaf2403

Suter, D.M., Molina, N., Gatfield, D., Schneider, K., Schibler, U., Naef, F., 2011. Mammalian Genes Are Transcribed with Widely Different Bursting Kinetics. Science 332, 472–474. https://doi.org/10.1126/science.1198817

Swarup, S., Verheyen, E.M., 2012. Wnt/Wingless Signaling in Drosophila. Cold Spring Harb. Perspect. Biol. 4, a007930. https://doi.org/10.1101/cshperspect.a007930

Sylwestrak, E.L., Rajasethupathy, P., Wright, M.A., Jaffe, A., Deisseroth, K., 2016. Multiplexed Intact-Tissue Transcriptional Analysis at Cellular Resolution. Cell 164, 792–804. https://doi.org/10.1016/j.cell.2016.01.038

Tabata, T., Takei, Y., 2004. Morphogens, their identification and regulation. Dev. Camb. Engl. 131, 703–712. https://doi.org/10.1242/dev.01043

Tomlinson, A., Ready, D.F., 1987. Neuronal differentiation in the Drosophila ommatidium. Dev. Biol. 120, 366–376. https://doi.org/10.1016/0012-1606(87)90239-9

Trinh, L.A., Fraser, S.E., 2013. Enhancer and gene traps for molecular imaging and genetic analysis in zebrafish. Dev. Growth Differ. 55, 434–445. https://doi.org/10.1111/dgd.12055

VanGuilder, H.D., Vrana, K.E., Freeman, W.M., 2008. Twenty-five years of quantitative PCR for gene expression analysis. BioTechniques 44, 619–626. https://doi.org/10.2144/000112776

Vera, M., Biswas, J., Senecal, A., Singer, R.H., Park, H.Y., 2016. Single-Cell and Single-Molecule Analysis of Gene Expression Regulation. Annu. Rev. Genet. 50, 267–291. https://doi.org/10.1146/annurev-genet-120215-034854

Voronoi, G., 1908. Nouvelles applications des paramètres continus à la théorie des formes quadratiques. Deuxième mémoire. Recherches sur les parallélloèdres primitifs. J. Für Reine Angew. Math. 1908, 198–287. https://doi.org/10.1515/crll.1908.134.198

Wang, Y., Shyy, J.Y.-J., Chien, S., 2008. Fluorescence Live-Cell Imaging: Principles and Applications in Mechanobiology, in: Artmann, G.M., Chien, S. (Eds.), Bioengineering in Cell and Tissue Research. Springer, Berlin, Heidelberg, pp. 65–84. https://doi.org/10.1007/978-3-540-75409-1_4

Wang, Z., Gerstein, M., Snyder, M., 2009. RNA-Seq: a revolutionary tool for transcriptomics. Nat. Rev. Genet. 10, 57–63. https://doi.org/10.1038/nrg2484

Wartlick, O., Mumcu, P., Jülicher, F., Gonzalez-Gaitan, M., 2011. Understanding morphogenetic growth control -- lessons from flies. Nat. Rev. Mol. Cell Biol. 12, 594–604. https://doi.org/10.1038/nrm3169

Webber, J.L., Zhang, J., Cote, L., Vivekanand, P., Ni, X., Zhou, J., Nègre, N., Carthew, R.W., White, K.P., Rebay, I., 2013. The Relationship Between Long-Range Chromatin Occupancy and Polymerization of the Drosophila ETS Family Transcriptional Repressor Yan. Genetics 193, 633–649. https://doi.org/10.1534/genetics.112.146647

Williams, J.A., Bell, J.B., Carroll, S.B., 1991. Control of Drosophila wing and haltere development by the nuclear vestigial gene product. Genes Dev. 5, 2481–2495. https://doi.org/10.1101/gad.5.12b.2481

Wu, J.S., Luo, L., 2006. A protocol for mosaic analysis with a repressible cell marker (MARCM) in Drosophila. Nat. Protoc. 1, 2583–2589. https://doi.org/10.1038/nprot.2006.320

Xia, C., Babcock, H.P., Moffitt, J.R., Zhuang, X., 2019. Multiplexed detection of RNA using MERFISH and branched DNA amplification. Sci. Rep. 9, 7721. https://doi.org/10.1038/s41598-019-43943-8

Xu, H., Sepúlveda, L.A., Figard, L., Sokac, A.M., Golding, I., 2015. Combining protein and mRNA quantification to decipher transcriptional regulation. Nat. Methods 12, 739–742. https://doi.org/10.1038/nmeth.3446

Zecca, M., Basler, K., Struhl, G., 1996. Direct and long-range action of a wingless morphogen gradient. Cell 87, 833–844. https://doi.org/10.1016/s0092-8674(00)81991-1

Zoller, B., Nicolas, D., Molina, N., Naef, F., 2015. Structure of silent transcription intervals and noise characteristics of mammalian genes. Mol. Syst. Biol. 11, 823. https://doi.org/10.15252/msb.20156257

Zuo, W., Moses, M.E., West, G.B., Hou, C., Brown, J.H., 2012. A general model for effects of temperature on ectotherm ontogenetic growth and development. Proc. R. Soc. B Biol. Sci. 279, 1840–1846. https://doi.org/10.1098/rspb.2011.2000

Appendix 1: Single Molecule FISH Probe Sequences

EGFP	Senseless	Spalt Major
gattctggatctttacttttt	agacgaacacgtccgtatcg	tgcggcactgagagaaatcg
cgtgatccttgtaatccatt	cggttttatcttgactcact	gcgcgtgtttagctaattta
tcgtggtccttgtagtcgcc	cttcgatgggtcttattcag	cagtgggaaacgcgtagtac
gtcatcgtccttgtaatcaa	gtgatcttggattcgaacgg	tcagcttgatcattgcacta
ggaagtacaagttttccttg	caaatctcactttcttggcg	acacagaacgttggacaggt
tgggggtgggaccagttgct	ttttctcaatactgttcact	tatccttgaagtcactacgc
catgccgctgcccttctcga	ttgctttaacagtctttgga	ttttattgatggtctcctgg
acagctcctcgcccttggac	tcactggatctgatttacgt	ccttcagctgtttgacaatg
aggatgggcaccacgccggt	ctggatctttacttttttgga	aaccattcgcaatggtcttg
cacgtcgccatccagctcca	cggcgataggtgattcattt	tcataggtatctgtttcctg

gcacgctgaacttgtggccg	tatcacagtgttgaaggcgg	tttattagatgctgacgcgc
gcgtcgccctcgccctcgcc	gcgggttgtacaaaaagggc	cttcagagattcagctggag
cagggtcagcttgccgttgg	agtagggcactcgagtagag	tgtgttatccaaactcggtc
tgccggtggtgcagatgaac	cgtggccgaggacaacaaaa	tggttggtgataatgctgga
gtgggccagggcacgggcag	ctgacccaaaaccacagagg	tttctcaccgaaggcaaagt
gtaggtcagggtggtcacca	tcaaggcaaagtcacgatcc	tgaagaacggctcattgcga
agcggctgaagcactgcacg	ttcatttggagttcatgctc	atcttatgtggatctggagc
tgctgcttcatgtgatcggg	cttgctgttctcattgttat	aagtgaaccttaaggttgcc
catggcgctcttgaagaaat	gtgcttagatttagtggcat	tgaaacttgtccatgtgctc
gctcctgcacgtagccctcg	gaatcatccgatgtgatgcg	gggaatgattgggagagcta
tcatccttgaagctgatggt	ctgctgtgatactgatctcg	tagagattctgaaggccagg
ggcgcgggtcttgtaggtgc	cacttcactgctggaactgc	ttaaggatttccataggcgg
tatcgccctcgaacttcacc	taacgttcagcgaggtcatc	tctgatgaaggagagggtgt
agctcgatgcggttcaccag	tttagattcacagcgctcag	gaagaaatcttcgcggggtg
ctccttgaaatcgatgccct	ttgctgtggtgtactcgaac	cggtggtaaagttaaggggg
tgtggcccaggatgttgcca	ggtgaccagatgatgttacc	tcgtgcttgatggggttaaa
ctgttgaagttgtactccag	cgatctctcgcacatcgagg	aggtgttcgaaacctcgatg
ggcggtgatgtacacgttgt	acaataggatccacctggtg	atcacagaccacacactgat
tgatgccgttcttctgctta	gcgctcatacttgaactttc	cactggtggaagtttctcat
tggcggatcttgaagttggc	gcgaggatatggatgcagtt	gcattcgagtacttcttgtg
cacggagccatcctccacat	agcgaggagatcggcgattg	ataaagtgaccgggcatcat
gctggtagtgatcggccagc	tcgaactccagatcctggac	acgacacatcatcgtcgtag
ccgtcgccgatgggggtgtt	atgggcatacaactgctgtt	ctcagcttttgctcgaacaa
gttatcgggcagcagcactg	cgtcaatccggccatgaatg	cgaagagctgcgcttgataa
cgctctgggtgctcaggtag	tgagcagctccagattgttg	ctgagaagcctctgaacgag
tcgttggggtccttggacag	ctcttcagcttcaagtgctg	tcacagatgctgcacttgaa
cagcaccatgtgatcgcgct	ttgatgcgatgctgttgctg	tacttattcgcgatcttcgt
cggcggcggtcacgaactcc	tgccaccaaattcatgagtg	tcttgatagcggattaagcc
tcatccatgcccagggtgat	gtgttgattacgcatgtagc	gcgagaatcgattgcatctt
gccgctgctctccttgtaca	tgatgcaactgttgttgctg	ccccggctagatcataaaaa
ctgggcagcagttcaggaag	tgatgctgatgttgctgcgg	gcatctttggatacgtttca
tgccgcccggttccatgcag	cgtcgagtcaggatgttgct	agattccgttctcagttttg
ctccgcttcaccgctacctc	gacgaggaggacgaacgacg	tcggcgatcgcatctaaatt
aaggcatccccatggtgatg	ttcgttttctccttggtaag	gcaaagcgtttggatttcga

	gtggatgagcagatctggaa	gtggttcttcatctgtgata
	ccacaatattggcagggata	ttaatccttccttccaatga
	atgtccgacttttggtggaa	gtacgatcgctcatagactt
	gtatgtatgtacgtgtgctt	acattggcatccgtgtataa

Omb-5' Exon	Omb-3' Exon	Omb- Intron
ggcaaaagcgtttgacgtga	ttcaacttgtccgaatcgga	agggagagagtctgagtttc
tttggtttatctgacgacgc	tcgtcctgctgattatcaag	caagtgtgagtgtgtctcag
cacacgtaaacacggcgaga	ccaagtgattaaaccaggca	atcttgcactcacttgacac
tgccttcggattgaagcaaa	tttccgagcaactcgaatcg	attgttacgtaagcctttgc
ttcttgattccttatcgtcg	caccgatatattcggtgacg	gctgtgcagagcttagaata
agaatctgcagggtgatgtt	aaggcaacaaatgcggcgac	ttgaatttggccagctatga
tttctttaatcttgtcaggt	gatacaggccatgtggatac	ctttctattggtgttgttgc
cacactttacgcacttttca	ttgggcattgaagagcagac	ccaagtgcgaacaaccaaca
ccttgctgttgtttgatttt	caaacaaagccggatgctgg	tattttgttcaccgacttgc
tttatattctttggacctcc	cttgtaatggtgataccggt	agcatttgccacagaaagga
aaatctcaggcggaaaccag	gcgaaaagcgatggctcttt	ttttgcagggtcatgaactc
ggtcgcattcaaactgcatt	aaactgccgggcaaactgta	gtgtgctaggtacaagtact
ggcgatgacgacacgaacta	caggagtcactgcatcgaag	aaaagggtggcatggtgttt
agggaaacggcgatttggat	ggtgacatcgaaatcggtcg	ttcaagtggcactcacatga
tagttagctggctgctgtaa	tggttgcttcatcagggatg	gccaagtgaatgcaaatcca
tcacttggcagttttaagtt	atcttctccatgctcttaag	aaactatagccacaagcctc
tcgtttcttttgctttgcaa	attgtgttgaacctccagtc	acacgtcattctgaaggagc
gtgtttaaaccttgttctgt	catggtgcgagtgtagatgg	gaagtaggagcaccgtattg
tgttttggcgacggcaaaac	agtaattgtcactgatccgt	tgcgcttagacagcaatttc
catgtttcattgcttttcgt	ccaactctagttccaattcg	ggtgtggaaaacgaggggag
tttctcttcttgttgtttca	tcgacgtccagatcaagatc	ttctcgcactgaaagtgtcg
ggacactcactgacttgtac	ttcagaggctagaggcaatg	ccaataaattgccgctttca
ctctctagtttgacttttga	cgctaaaccacaaagacggt	ctgtcacaaagtccgacatc
gcgatctctccaaaaaaggg	gctacaaatgcctatagtct	tgaactttgacaggcggttg
gtcatgcagaaagggtttcc	tgagttcgcaacgagagagg	catcgataatttttcctgcc
accgtttcaaacaaccaagt	aacactggcgataaacgggc	aaggaggtttttgcccaaat

tggacgtcgtatctcatcac	gctaggctttgcttagaaac	gatatgtgtatctaccgcac
tcctcagcagactgatgaaa	gaatcgccagcgatttctag	attttccatacgctctacac
attggcgaacctggcgaatg	attagacctctccatcactt	aaagtagctaattgcccagt
gcagaaatggatgatatgcc	tttttgctaccttctgtctt	atatgcgtatctttcgttgc
ctgaagtcagtgggtcgttg	ttttcctgtttttttgtttcc	ccgggatctcggaaatatga
cgccgtcaacagcgaggaaa	agttacgacgattctcttgc	gtggattgtagttttgacca
cgctgttgttattgttgcta	cttgaagtctggaaccttca	ttttgctgtctttctttctc
agttgttgtttccgctgttg	ctcattttattccgtctgtt	gcgacggaaacgcttactta
ttggtgttgttgttggaatt	aatttgtctgaggatagccc	catcaattttcggctggttg
gttggtgttattggtgtttg	ataaagaggcgccatctttg	cctccaccttaaatgtgata
ttggtgaaacggccacgaga	tcatccacctaagactaaga	tgtacatttccatcgtttct
tagaccaacaattgtgggcg	gtgtttacggtatcttggta	gtcgtaagattgggtctgta
cgtactgggtgagtgatgtg	gcattttctatcgtctacga	aatattccattccatgcagt
agatgatgggcgtgcggatg	tcaggggttccgaaattaa	tttgtgcatttccaagttgt
tgaccttgggatcatcgacg	cacagaactcttgccctaaa	aaggacgtcatcatcgtcat
ttgtggaacttctcccacag	tgcgaggaacatgcgttgaa	ttgtcgttatttttctggcg
ctcttggtgatgaccatttc	ctttgtcgctctctcttaaa	cgatccgatttttaagcgga
	ttcttgctttttctcctta	cagcaacagccacgatatta
	gaagattgtgtggtgttgca	tgctttttgagagaggctt
	tcataagaatgcctatgcga	ttcttttagtcttcgcagtt
	ggttgtttgcgtttgtttta	gaccgttctcattgtttttg
	cagtgacatgaggtgttgtt	

Scalloped
gagctggtgatgtttttcat
aggttgttctgcaattgcag
caacttccaactcgctgcag
caactgcctgttgttctgtc
gtgtccacggtgatggtatg
atttgtgtctgccgatccaa
acatccaggtttttgctatc
tttcgtcatcgctcatatca

gcatcagcggatgacaagtc
cgatatctggactccataca
aaagcctcttgaaagctctg
tgatttttctacgtccgcac
attttaccctcgtcggataa
tgcgattagctcgttgcgac
cctgtgcgcagttttatata
cacttggatgtgcgaactga
tggatctcgcggagtttacg
gccagaattgcactttgatt
aaatcttgggacgtgcttgg
gaagggcttgatgctgtaat
aaaccgcagtcgacgttttg
caattgtgagggcggaattc
gaatttgtgcgtggcaatgg
gaacgccgtaaactcgagta
tttcatctctctggatttcc
gaacgaatagatgccggtga
cggaaaaggatggcttgccg
tcaacagtctcaagcaatgg
acttgtcgaatatttgccgt
ttaagtcccccagatttctc
tggacccttttcgtagagat
cccagcatttaactaggtaa
gttagatcggtattcaggtc
atagaaatcacccgtttcgc
tttcgtattggctggttacg
cacacgagcacgacattctc
tgcttgccaaaggagcaaac
gattgttctccagtcgagag
gcgttgaatgcgatagacgt
gatcatgtactcgcacatgg
gttcttcagcttctgaataa
tcatcatatagcgttccggt

gtaaagttttccagcacact
ggccctcattacttgcaata
atgcacaacagtgtctcctg
cgccacctcaaacacatagg
gacggtatatgtgatgggtg