HAL
open science

Noise and robustness downstream of a morphogen gradient : Quantitative approach by imaging transcription dynamics in living embryos

Carmina Angelica Perez Romero

▶ **To cite this version:**

Carmina Angelica Perez Romero. Noise and robustness downstream of a morphogen gradient : Quantitative approach by imaging transcription dynamics in living embryos. Cellular Biology. Sorbonne Université; McMaster university (Hamilton, Canada), 2019. English. NNT : 2019SORUS306 . tel-03139877

HAL Id: tel-03139877

https://theses.hal.science/tel-03139877

Submitted on 12 Feb 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Sorbonne Université & McMaster University

Ecole doctorale Complexité du Vivant ED515 & School of Graduate Studies

Nuclear Dynamics & Department of Biochemistry

Noise and Robustness downstream of a morphogen gradient: *Quantitative approach by imaging transcription dynamics in living embryos*

By CARMINA ANGELICA PEREZ ROMERO

Thèse de doctorat de Biologie Cellulaire & Ph.D. Thesis in Biochemistry

Directed by Dr. Nathalie Dostatni & Dr. Cecile Fradin

Presented and publicly defended on the 29th of March 2019

In front of a jury composed of:

- Dr. Paul François (Reviewer, McGill University)
- Dr. Eric Lecuyer (Reviewer, Montreal University)
- Dr. Joshua Milstein (Toronto University)
- Dr. Laure Teysset (Sorbonne University)
- Pr. Ray Truant (McMaster University)
- Pr. Nathalie Dostatni (Sorbonne University)
- Pr. Cécile Fradin (McMaster University)

Except where otherwise noted, this work is licensed under
http://creativecommons.org/licenses/by-nc-nd/3.0/

A mi padre por su inspiración

A mi madre por su apoyo incondicional

A mi hermana por enseñarme a nunca rendirse

A mi hermano por su curiosidad

A mi amor por ser mi soporte

A mis amigos por creer en mí

A mis mentoras por dejarme volver a soñar

A mi sueño por impulsarme

To my father for inspiring me

To my mother for her unconditional support

To my sister for her admiration

To my brother for his curiosity

To my love for holding me

To my friends for believing

To my mentors for letting me dream again

To my dream for pushing me forward

"What's past is prologue" Shakespeare, The Tempest (II.i)

Lay Abstract

Have you ever wondered how a single cell can become a full grown organism? Well it starts when an egg and sperm fuse together. As time passes this single cell divides over and over again until an organism is formed. During this developmental process, somehow the cells know exactly where they are and what they need to become so that they form the organism. However, we don't fully understand this process and this is what we hope to answer with our research: How do the cells know where they are and what they need to become during development?

We study this process in the fruit fly. Although fruit flies might not look a lot like us, during early embryonic development we are quite similar, so we can try to answer these questions in fruit flies and what we find might be relevant to other organisms like us.

During development, the first element that an embryo needs to know is the orientation of its body, where the head and tail, the left and right and the back and front of the body will be. We concentrate on studying how the head to tail axis, which we call the anterior-posterior axis, is formed.

To know where the head is going to be, the embryo releases proteins called morphogens that broadcast instructions to other genes so that cells know where they are and what they should become. We study a morphogen called Bicoid. Its concentration is high in the anterior, the region that will become the head of the embryo, and lower as you move towards the posterior where the tail will form. Bicoid activates a gene called *hunchback*, which ends up dividing the embryo in two large parts, the top and the bottom. However, Bicoid's message fades away during each cell division and needs to be read again at the beginning of each new nuclear cycle. So how is the message read and how long does this process take? This last question is particularly critical during the period of very fast cell division.

My thesis tries to answer this question. We found out that it takes 3 minutes for a nuclei to read the Bicoid concentration, activate *hunchback* and express it correctly. However, in contrast to what was believed before, or namely, that only Bicoid was involved in this process, we found out that other players are involved in helping relay this message. This way *hunchback* can accurately divide the body in two parts exactly in the middle and without mistake in such a short period of time.

Abstract

During development, cell differentiation frequently occurs upon signaling from concentration or activity gradients of molecules called morphogens. These molecules control in a dose-dependent manner the expression of sets of target genes that determine cell identity. A simple paradigm to study morphogens is the Bicoid gradient, which determines antero-posterior patterning in fruit fly embryos. The Bicoid transcription factor allows the rapid step-like expression of its major target gene *hunchback*, expressed only in the anterior half of the embryo. The general goal of my thesis was to understand how the information contained in the Bicoid morphogen gradient is rapidly interpreted to provide the precise expression pattern of its target.

Using the MS2 system to fluorescently tag specific RNA in living embryos, we were able to show that the ongoing transcription process at the *hunchback* promoter is bursty and likely functions according to a two-state model. At each nuclear interphase, transcription is first observed in the anterior and it rapidly spreads towards the posterior, as expected for a Bicoid dose-dependent activation process. Surprisingly, it takes only 3 minutes from the first hints of transcription at the anterior to reach steady state with the setting of a sharp expression border in the middle of the embryo. Using modeling taking into account this very fast dynamics, we show that the presence of only 6 Bicoid binding sites (known number of sites in the *hunchback* promoter) in the promoter, is not sufficient to explain the establishment of a sharp expression border in such a short time. Thus, either more Bicoid binding sites or inputs from other transcription factors could help reconcile the model to the data. To better understand the role of transcription factors other than Bicoid in this process, I used a two-pronged strategy involving synthetic MS2 reporters combined with the analysis of the *hunchback* MS2 reporter in various mutant backgrounds. I show that the pioneer factor Zelda and the Hunchback protein itself are also critical for *hunchback* expression, maternal Hunchback acting at nuclear cycle 11-12, while zygotic Hunchback is acting later at nuclear cycle 13-14. The synthetic reporter approach indicate that in contrast to Hunchback and Caudal, Bicoid is able to activate transcription on its own when bound to the promoter. However, the presence of 6 Bicoid binding sites only leads to stochastic activation of the target loci. Interestingly, the binding of Hunchback to the Bicoid-dependent promoter reduces this stochasticity while Caudal might act as a posterior repressor gradient. Confronting these experimental data to theoretical models is ongoing and should allow to better understand the role of transcription factors, other than Bicoid, in *hunchback* expression at the mechanistic level.

Résumé

La différenciation des cellules au cours du développement est souvent déclenchée par des gradients de concentration ou d'activité de molécules appelées morphogènes. Ces molécules contrôlent de façon dose-dépendante les ensembles de gènes déterminant l'identité cellulaire. Un paradigme simple pour l'étude des morphogènes est le gradient de Bicoid, qui détermine l'identité cellulaire le long de l'axe antéro-postérieur chez la mouche du vinaigre. Le facteur de transcription Bicoid permet l'expression rapide de son principal gène cible, *hunchback*, dans la moitié antérieure de l'embryon dans un domaine d'expression avec une bordure très franche. L'objectif principal de ma thèse était de comprendre comment l'information graduelle inscrite dans le gradient de concentration de Bicoid est rapidement interprétée pour permettre la formation de la bordure du domaine d'expression de son gène cible.

En utilisant le système MS2 rendant fluorescents des ARN spécifiques dans les embryons vivants, nous avons montré que la transcription au promoteur d'*hunchback* est « bursty », reflétant vraisemblablement le fonctionnement d'un promoteur à deux états. A chaque interphase, la transcription est d'abord détectée à l'antérieur et se propage rapidement vers le postérieur, comme attendu pour un processus dépendant de la dose de Bicoid. De manière surprenante, il suffit de 3 minutes, après la première détection de transcription à l'antérieur, pour que l'expression soit stabilisée et que la bordure franche du domaine d'expression soit précisément positionnée au milieu de l'embryon. Le développement de modèles théoriques prenant en compte la rapidité de cette dynamique, montre que, seuls, 6 sites de liaison à Bicoid (nombre de site connus dans le promoteur d'*hunchback*) ne sont pas capables d'expliquer la formation de cette bordure franche dans un délai aussi court. Ainsi, soit une augmentation du nombre de sites de liaison à Bicoid, soit l'implication d'autres facteurs de transcription doit être prise en compte pour réconcilier données expérimentales et modélisation. Afin de mieux comprendre le rôle des facteurs de transcription autres que Bicoid dans ce processus, j'ai utilisé une double stratégie impliquant des gènes rapporteurs MS2 synthétiques combinés à l'analyse du gène rapporteur *hunchback* MS2 dans des contextes génétiques mutants. J'ai pu montrer que le facteur de transcription « pionnier » Zelda et la protéine Hunchback elle-même sont critiques pour l'expression du gène *hunchback*, la composante maternelle d'Hunchback agissant plutôt aux cycles nucléaires 11-12, alors que la composante zygotique agirait plus tard aux cycles nucléaires 13-14. L'analyse des gènes rapporteurs synthétiques indique que, contrairement à Hunchback et Caudal, Bicoid est capable d'activer la transcription à elle seule en se fixant sur le promoteur. Toutefois, la présence de 6 sites de liaison à Bicoid ne permet qu'une activation stochastique du gène cible. De façon intéressante, la fixation d'Hunchback sur un promoteur régulé par Bicoid réduit cette stochasticité alors que Caudal agirait comme un gradient postérieur répresseur. La confrontation de ces données expérimentales aux modèles théoriques est en cours et devrait permettre de mieux comprendre le rôle de chacun de ces facteurs autres que Bicoid dans l'expression d'*hunchback* d'un point de vue mécanistique.

Acknowledgments

There are so many people to thank during my PhD journey, as many know it was not an easy one and I always had the support of friends, family and amazing strangers reminding me to never give up and strengthen me to keep me going through this whole process. To them all, from the bottom of my heart, I sincerely thank you.

I specially want to thank my family, for though far away at times and with oceans between us; they were always there for me with all their love, and unconditional support. They might not understand the strain of doing a PhD, nor much about what I do other than I have baby flies to care for, and kill; but just knowing they are there has always made those hard days easier, and the triumphs even bigger. Gracias Yesi, Alex, mama y papa por siempre estar ahí para mí y por recordarme que rendirse no es una opción.

I want to thank my husband Mike for believing in me since the start, and believing in us through the complicated journey. For always being there when I needed a shoulder to cry, for all the amazing and complex dinners he cooked when I was exhausted for writing this thesis, on when things in the lab weren't working, or when things were just going great. For understanding my moods, the complexity of science, research and its timings. For withstanding the hardships of this cotutelle and its long-distance relationship. Thanks for always, always believing in me and in us when sometimes the stress and all has make me forget it.

I want to thank the amazing group of friends, or more like chosen family that I had to support me throughout my entire PhD, together we shared so many good times and hard times, but they were there for me all the time. Carla thank for being the best friend I could wish for, an awesome roommate, and a great role model for any PhD, but also for the latin society; estoy segura que la vida te dapara muchas cosas y estoy honrada de ser tu amiga, muchas gracias por todo. Andres, what can I say, thanks for always making my life so much easier, relaxing me when I needed a break and just being there for me; solo una cosa que decir Pamela Chu. Cristina, thank for being like a little sister to me, and for keeping us busy, out of the laboratory with all the fanclub news, it was great times; que onda un tequilita o que. There are so many people left to thank that would merit an entire thesis itself, so just special mention to Darko, and Mai, thanks for being great friends since the start and always helping me through the way. Finalmente gracias a toda la

banda vieja y nueva por siempre estar ahí unida, dispuesta a apoyar a quien sea y cuando lo necesita son bien chingones, sigan así. Special thanks go to that little gang that I met on the coffee where I wrote these theses, one who became very special to me and I owe the corrections of this thesis, and a lot of enriching moments and conversations of the heart.

Now I want to thank two amazing women who changed my life, and gave me hope when I saw darkness. Cécile, I don't know how to thank for allowing me to become your student, you may never know how much that decision helped me in many aspects of my life, so thank you for believing in me, even when I doubted myself. Especially want to thank you for not only being a great scientists, but also a great person and mentor. I think one does not realize the importance of having a mentor at this stage of life until you have the opportunity to have one. I have met many great professors and scientists but have never had the opportunity to have a real mentor. So I want to really thank you, Cécile, for being that person and taking the time to teach me so much more than just science. Thanks for being a role model to me and our community to follow. Nathalie, thanks for receiving me with open arms into your laboratory, and allowing me to just be myself. For seeing the potential in me and letting me run free around Curie, enriching my creativity and allowing me to just do great science, no matter the administrative headaches it brought you. Thanks for always finding ways to make magic happen, this cotuttelle is a clear example of it, and helping me go to congress, courses, and securing financing for me when needed. Thanks for always believing in me, and helping me realize my mistakes. Thank you for being a great mentor and for letting me feel at home while at Curie.

I want to also thank our two great collaborators Aleksandra and Mathieu for their insightful conversations every week, and their patient reviewing of the results, helping us make sense of the complex data. Thank you for being patient with me and my naïve questions, and for your interdisciplinary knowledge. This collaboration has allowed me to expand my field of view of science and the importance of interdisciplinarity in trying to understand complex biological questions, thanks for letting me be part of the team. Thank you also for your great critique and assisting me during the manuscript writing process, which has led to successful publications that have ensured this PhD is a great one, and a much more complete one.

I want to thank all the past and current members from both the Fradin's and Dostatni's laboratories for making the long days in the lab a good time, aiding with experiments,

troubleshooting, and crisis moments, so science can happens. From the Dostatni's lab, I especially want to thank Huy and Marie, with whom countless hours were shared in the laboratory and without them I wouldn't have been able to complete my PhD: Huy, thanks for making the magic happen with the code in your hands, this allowed me the time to explore the questions we needing answers that hopefully put us closer to where we wanted. Thank you my friends for all the great conversations on our lunch times, and late dinners, I am so glad to say this was all a great team effort, let it not be our last collaboration together. Marie, thanks for teaching with all the patience in the world how fly genetics worked, and for making sure my crosses were correct. Thank you for being my great friend in Paris, for our conversations during countless hours in the fly room, at lunch, and outside of the lab. From the Fradin's lab, I want to thank Mike for the endless help on setting up the microscope, showing me how to calibrate it a thousand times and always having a solution for any problem in the lab. Markus, thank you for showing me around the lab that first day when I was unsure I wanted to continue in science. Thanks for being a friend, and for all the support on our hard journey, we did it.

I want to thank the secretaries, their assistance made my administrative life an easier one. For always finding the right person to contact and connect with, and having a good ear to hear my complaints, and being great confidants, thanks Lisa, Caroline and Tina.

I also want to thank the professor with whom I started my PhD originally, for opening my eyes to problems in science, as well as the vital importance of policy and academic integrity. Albaeit through a tough lesson, for confirming what I intrinsically knew deep within myself that I do belong in science, and encouraging me to never give up, regardless of who or what is in the way of my dreams.

Finally, I want to thank all the past and present executives and staff of CUPE 3906, my union. Here I found a place full of support, to discuss new ideas, to fight for our rights, to help other members, and realize that we can always ask for a better workplace, and thought that change our community for better. They empower me to become the woman I am today, and giving me all the support and training needed for me to become a leader, and an activist. I am forever thankful for my union for showing me the great change we can achieve if we all work together for better societies and a better world.

TABLE OF CONTENTS

Lay Abstract iii

Abstract iv

Résumé v

Acknowledgments vi

Table of Contents ix

List of Figures xi

List of Abbreviations and Symbols xii

Preface 1

Chapter 1 INTRODUCTION 2

1.1 *Drosophila melanogaster* life cycle 4

1.1.1. Oogenesis 4

1.1.2. Early embryonic development 6

1.2. Anterior-Posterior cell fate specification 7

1.2.1 Genes involved in Anterior-Posterior patterning 9

1.2.2. Maternal to Zygotic transition 12

1.3. Noise and robustness during embryonic development 13

1.3.1. The Bicoid gradient 13

1.3.2. The *hunchback* transcriptional response 15

1.3.3. Precision of the transcriptinal response 16

1.3.4. Live imaging of transcription 18

1.4. Main project hypothesis and objectives 20

Chapter 2 MATERIALS AND METHODS 21

2.1 Sample preparation and live embryo imaging 22

Article 1: Live imaging of mRNA transcription in *Drosophila* embryos 23

2.2 Image and data analysis 42

Chapter 3 HOW LONG DOES IT TAKE FOR THE *hunchback* GENE TO PRECISELY MEASURE Bicoid CONCENTRATION? 47

3.1 Visualizing *hb* transcription 48

Article 2: 3 minutes to precisely measure morphogen concentration.....49

3.2 *hb* promoter state.....75

3.3 Modeling *hb* transcription 79

Chapter 4 CONTRIBUTION OF OTHER TRANSCRIPTION FACTORS TO THE DYNAMICS OF *hunchback* TRANSCRIPTION 83

Article 3: A synthetic biology approach to understand the rapid and precise dynamics of *hunchback* transcription 84

Chapter 5 DISCUSSION AND PERSPECTIVES.....133

Afterword.....1410

References.....1421

Copyright Acknowledgments ……147

List of Figures

Figure 1. *Drosophila* embryonic development at a glance …….3

Figure 2. *Drosphila* Life cycle and oogenesis…….5

Figure 3. Early fruit fly embryo development…….7

Figure 4. Early models proposed to explain cell fate determination during development……..9

Figure 5. Anterior-Posterior axis specification pathways……..10

Figure 6. Anterior-Posterior axis development in *Drosophila*……..12

Figure 7. Overview of the maternal to zygotic transition in *Drosophila*……...13

Figure 8. The Bicoid-*hunchback* system to study precision in *D.* embryonic patterning……..18

Figure 9. The MS2 system to study transcription in living *Drosophila* embryos……..20

Figure 10. Overview of *LiveFly*…….46

Figure 11. Summary of results from autocorrelation analysis……...79

Figure 12. Summary of results from modeling analysis……84

Figure 13. The memory hypothesis…….144

List of Abbreviations and Symbols

RNAs – ribonucleic acids

Bcd – Bicoid protein

Hb – Hunchback protein

Nos – Nanos protein

Cad – Caudal protein

Zld – Zelda protein

bcd – bicoid RNA/gene

hb – hunchback RNA/gene

nos – nanos RNA/gene

cad – caudal RNA/gene

zld – zelda RNA/gene

nc – nuclear cycle

AP – Anterior-Posterior axis

GFP – Green Fluorescent Protein

RFP – Red Fluorescent Protein

FRAP – Fluorescence Recovery After Photobleaching

FCS – Fluorescence Correlation Spectroscopy

FISH – Fluorescence In-Situ Hybridization

SDD – Synthesis Diffusion Degradation

MCP – MS2 Core Protein

Nup – Nucleoporin 107

His – Histone2B

UTR – Untranslated region

S phase – Synthesis phase

EL – Egg Length

TFs – Transcription Factors

NLS – Nuclear Localization Signal

ECP – Embryo Collection Plates

NA – Numerical Aperture

ROI – Region of Interest

GUI – Graphical User Interface

PWMs – Position Weights Matrixes

MZT – Maternal to Zygotic Transition

ZGA – Zygotic Genome Activation

Preface

Five years ago, when I started my Ph.D. I would never have thought where this journey would take me, which gave me the opportunity to end up working with two amazing women role models, nor would I have thought this will end up giving me an opportunity to visit a new country, travel a bit, make new friends and have fun doing amazing science along the way.

I started my thesis in another laboratory which I decided to leave, I was broken, had lost trust and love in science, and wasn't even sure if I was good enough to continue pursuing a Ph.D degree or what to do next.

Thankfully, I was able to meet my supervisor Dr. Fradin 3½ years ago, who reminded me of the many reasons I love science and the values I cared of a researcher. She reminded me of why I love science and made me believe in it again. I am forever in grateful for helping me rise when I fallen, taking me under her wing and mentoring me, giving me her very best helping me to spread my wings and fly. Literally, to France where I met her collaborator and my second supervisor Dr. Dostatni. Who welcomed me in her team with open arms and gave me all the tools and liberty to make my time there and Ph.D. a great one, thanks to her idea I was able to establish a co-toutelle between McMaster University and Sorbonne Université which allowed me to further develop as a student and a researcher.

I owe this great two mentors my success in completing this thesis, I am grateful to all the people I meet on my way that reminded me each day to never give up, and continue giving science a chance. Today I am glad to say that it was totally worth continuing and finishing my Ph.D. So without further or due, I hope you enjoy reading this thesis as much as I enjoyed the ride.

So now let me take you through a journey of mystery, enigmas and questions waiting to be solved… The journey begins in your kitchen with that annoying fruit-fly in your bananas, you might have little love for this little pest, but have you ever asked yourself how they arrive there in the first place? How are they formed? How a single cell can form an organism so complex as them? How do the cells know where they are and what they need to become? Well let me welcome you into this fascinating world of developmental biology.

Except where otherwise noted, this work is licensed under
http://creativecommons.org/licenses/by-nc-nd/3.0/

Chapter 1
INTRODUCTION

Some of the most intriguing questions in developmental biology have been: How can one complex organism arise from the meeting of two simple cells, an oocyte and a spermatozoid? How can the information contained in these cells, in the form of nucleic acids and proteins, give a precise and robust blueprint for the formation of a body? How can an individual cell know its place in the blueprint and decide what types of cell it needs to become? What happens if these processes go wrong and how can these failures be prevented? These are the type of questions developmental biologists seek to answer. However, the complex dynamics found in embryos has awoken the interests not only of biologists, but also of physicists, all aiming to explain how a single cell can robustly give rise to a living individual like you and me.

One of the first features that any embryo needs to establish is the axes of its body and their respective orientation, in other words, where its head/tail (anterior/posterior), back/front (dorsal/ventral), and left/right parts of its body will be. While most animals (vertebrates, annelids, and some arthropods) specify cellular location in the embryo by sequentially adding segments to their body during growth, some recently evolved insects form segments while simultaneously subdividing the embryo (Jaeger, 2011).

For example, in *Drosophila melanogaster,* also known as the fruit fly, the future head/tail (antero-posterior axis) is determined by maternal ribonucleic acids (RNAs) that are laid and anchored at each extremity of the mother's egg (oocyte). After fertilization, these RNAs are translated into gradients of proteins called morphogens, which diffuse from the source of synthesis and broadcast information to the cells about their spatial location in the form of their concentration. In response to that information, the cells can then activate expression of different target genes (the gap genes) which will help divide the body in large domains along the anterior-posterior (AP) axis, and further activate other sets of genes (pair-rule) into smaller domains that will themselves further activate other sets of (segment polarity) genes, dividing segments into smaller regions. Finally, the cells in each region will get distinct signals and express specific sets

of genes (homeotic genes) which will reflect its position along the AP axis and determine its identity and function (Fig. 1). In this way, when morphogenesis happens, each cell knows where it is and what it should become. Cells can then move to form the right organ in that specific place, and start robustly shaping the living organism (Martinez-Arias & Stewart, 2002).

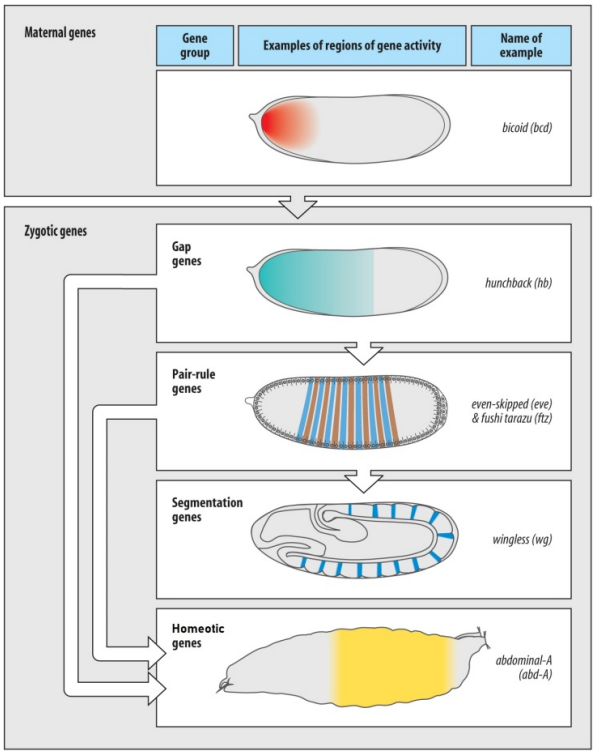


Figure 1. *Drosophila* embryonic development at a glance: By convention the anterior pole is on the left, the posterior on the right, the dorsal side is up and the ventral side is down. The early embryonic patterns are established by maternal genes that form morphogen protein gradients. One such protein is Bicoid, a transcription factor that activates gap genes like *hunchback*, which in turn define broad territories in the embryo. This then enables the activation of the pair-rule genes (*eve, ftz*), which further divide the body into smaller regions. The segment polarity genes then divide the embryo into segmented units (*wg*), which define the spatial domain of the homeotic genes (*abd-A*), with each segment conferring a unique identity. Reproduced with permission from (Martinez-Arias & Stewart, 2002).

My research focuses on the very first 2 hours of the anterior axis development in the fruit fly. It aims to understand how the signal contained in the smooth concentration gradient formed by Bicoid (Bcd), a protein morphogen involved in anterior formation, activates one of its target gap-genes, *hunchback* (*hb*), whose RNAs are expressed in a domain harboring a step-like expression border dividing the body in half after only a very short time.

1.1 *Drosophila melanogaster* life cycle

Drosophila melanogaster, commonly known as the fruit fly, has been used as a powerful model organism to study development. This insect undergoes a complex metamorphosis through its life cycle. Emerging from the fusion of an oocyte and a spermatozoid, an embryo develops in 24 hrs and hatches into a crawling larva which will spend around 24 hours in three successive larval stages. At the end of the last larval stage, the larva stops moving and becomes a pupa in which metamorphosis will occur and from which the adult fly will emerge. During each part of the life cycle different developmental processes can be studied (Fig. 2a). For example, during the embryonic stage, scientists aim to understand how spatial patterns emerge, how segmentation arises and how cells get to know their location and function. In the pupal stage, on the other hand, they aim to understand how more complex organs and systems are formed like the nervous system or the imaginal discs which will give rise to the adult appendages. In the adult fly, behavior or immunity can be studied, among other things (He, et al., 2011). In all the stages of development, the function of specific proteins can be studied via the effect of mutations on the individual and its progeny.

1.1.1 Oogenesis

Female flies have a pair of ovaries, each composed of ~15 ovarioles which contain a linear sequence of egg chambers at different developmental stages. Each egg chamber produces a single egg. It is composed of 16 germline cells (15 nurse cells and one oocyte) which are surrounded by ~600 somatic follicle cells (Bastock & St Johnston, 2008). The egg chamber arises from the germarium, hosting at its tip the germline stem cells, which divide asymmetrically to give rise to a germline stem cell and a cystoblast. After 4 uncomplete divisions, this cystoblast will give rise to the 15 nurse cells and one oocyte connected by cytoplasmic bridges. Once differentiated, the new egg chamber buds off from the adjacent egg chambers staying linked by stalk cells (He et al., 2011). In the next stages, the egg chamber rotates (Fig. 2b), polarizing its actin cytoskeleton, which leads to its adherence to the extracellular matrix and allows the egg chamber to elongate (Bastock & St Johnston, 2008). As the egg chamber elongates, the border cells containing the polar cells start migrating towards the oocyte. This leads to periodic basal actomyosin contractions which activate polarization of

maternal mRNAs (He et al., 2011). Motor proteins travelling along the microtubules help transport these polarized maternal mRNAs to a specific pole of the oocyte. As an example, *bcd* mRNAs are transported by Dyneins towards the anterior pole of the embryo, while *nanos* (*nos*) is tethered to its posterior end (Gilbert & Barresi, 2016) (Fig. 2b). Finally, to form the mature egg, follicle cells synthesize yolk polypeptides, transport them to the oocyte, and produce the egg shell. Meanwhile, other maternal RNAs such as *hb* or *caudal* (*cad*) mRNAs are produced in the nurse cells and dumped into the oocyte cytoplasm through the cytoplasmic bridges. These maternal mRNA distribute evenly in the oocyte where they remain translationally quiescent until fertilization and egg laying (Gilbert & Barresi, 2016; He et al., 2011).

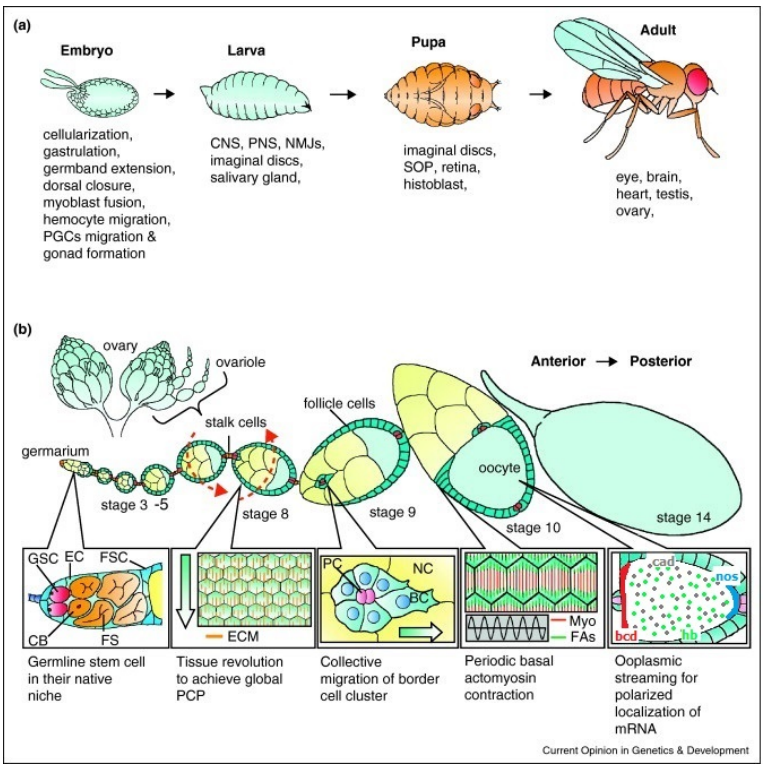


Figure 2. *Drosophila* Life cycle and oogenesis: A) Major stages of the *Drosophila* life cycle showing commonly studied developmental events. PGC: primordial germ cell; CNS: central nervous system; PNS: peripheral nervous system; NMJs: neuromuscular junctions; SOP: sensory organ precursor. **B)** Anatomy of fruit fly ovary and expanded view of egg chambers in a single ovariole. Germline stem cell self-renewal, follicle rotation, border cell migration, periodic actomyosin contraction, and polarized mRNA localization are further illustrated. Arrows in the bottom panels indicate the direction of movement (GSC: germline stem cell; CB: cystoblast; EC: escort cell; FSC: follicle stem cell; FS: fusome; ECM: extracellular matrix; PC: polar cell; BC: border cell; NC: nurse cell; Myo: myosin; FAs: focal adhesion; MT: microtubule; Nos: nanos; Grk: gurken; and PCP: planar cell polarity; *bicoid*: bcd, red: *nanos*: nos, light blue; *hunchback:* hb, green; *caudal:* cad, gray. Modified with permission from (He et al., 2011).

1.1.2 Early embryonic development

Drosophila melanogaster early embryonic development is under the control of maternally supplied RNAs (maternal RNAs), which are translated upon egg laying (Lefebvre et al., 2018). During this period, the nuclei are located in the center of the embryo and divide rapidly going though succesive S phases and mitoses (no G2 phase). These rapid divisions occur synchronously and without cytokinesis, forming a syncytial pre-blastoderm (Fig. 3)(Farrell & O'farrell, 2014). After 8 extremely fast nuclear cycles (nc) of ~8 min/each, nuclei start to migrate from the center to the cortical region of the embryo. The first hints of zygotic transcription are detected at nc 8: this marks the onset of the zygotic genome activation and it is accompanied by a gradual lengthening of the nuclear cycles (Lefebvre et al., 2018). At nc 10 (after 9 nuclear divisions), the (2^9 ~ 500) nuclei have reached the periphery of the cell forming a single layer and this distribution facilitates their imaging in living embryos (Farrell & O'farrell, 2014; Perez-Romero et al., 2018). While nc 10 lasts 10 minutes, nc 13 lasts 25 minutes and is the last synchronous cycle of the syncytial blastoderm (Farrell & O'farrell, 2014; Gilbert, 2000). During these early stages, the embryo undergoes a transition during which the control of its further development is transferred to the zygotic genome (Lefebvre et al., 2018). At nc 14, which lasts more than one hour, the cellularization process begins through invagination of the cell membrane around each nucleus. Once this process is completed, the embryo reaches the cellular blastoderm stage and starts to undergo gastrulation (Farrell & O'farrell, 2014; Gilbert & Barresi, 2016).

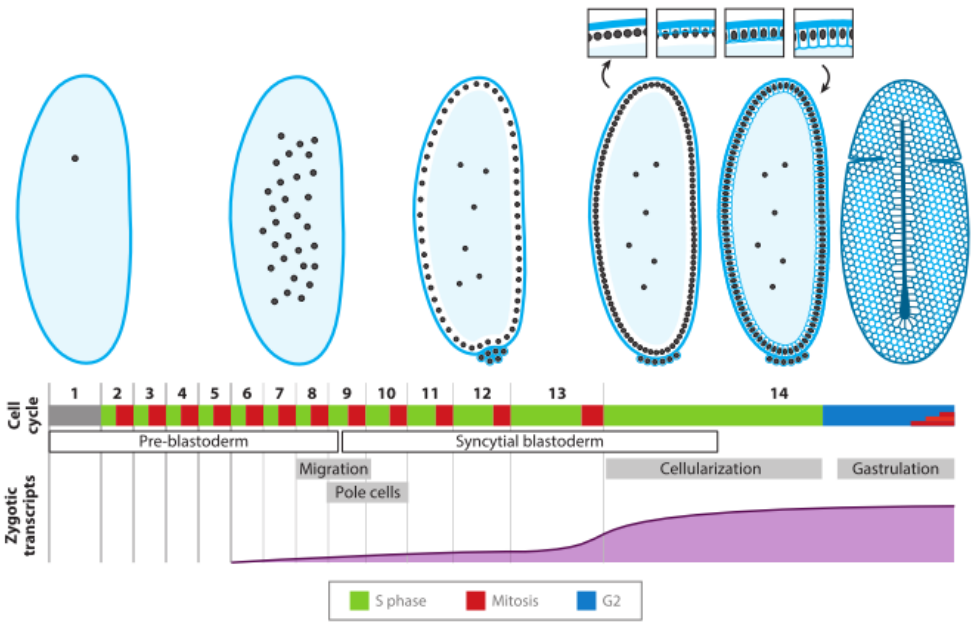


Figure 3. Early fruit fly embryo development: The top panel illustrates the morphological changes taking place during the first 200 minutes of fruit fly embryogenesis (anterior is up and posterior down). The horizontal bar shows the duration of each nc: S phase (green), mitosis (red), and G2 (blue). Nuclear cycle 14 is the first to present a G2 phase and asynchronous division time. The preblastoderm and blastoderm stages are indicated with white boxes. The timing of notable morphological events is demarcated in gray boxes: migration of the nuclei to the periphery, germline emergence by pole cell individualization, cellularization and gastrulation (marked by the formation of the ventral and cephalic furrows). Finally the approximate number of genes for which zygotic transcripts have been detected is represented by the purple curve at the bottom. Reproduces with permission from (Farrell & O'farrell, 2014).

1.2 Anterior-Posterior cell fate specification

The determination of polarity axes (Anterior-Posterior, Dorso-Ventral, Left-Right) is essential for development. As mentioned above, it is generally under the control of gradients of molecules called morphogens. The existence of these substances was first proposed by Thomas Morgan in the early 20th century (Morgan, 1927). The term was used by Alan Turing before their actual discovery, although he thought of them as "a system of chemical substances, reacting together and diffusing through a tissue" or embryo (Fig. 4A) (Turing, 1952). The definition of a morphogen was later revisited by Lewis Wolpert who proposed them to be molecules/substances with two specific properties: i) a morphogen is distributed as a concentration or an activity gradient; ii) a morphogen is able to induce different cell fates at different concentrations. A morphogen provides positional information to each nucleus/cell, which "measures" the

morphogen concentration (or activity) supposedly through a threshold mechanism, allowing the expression of specific sets of target genes responsible for nucleus/cell identity. This model is known as the French Flag model (Kerszberg & Wolpert, 2007; Wolpert, 1969) (Fig. 4B). One important prediction that arises from this model is that a change in the morphogen concentration should induce a shift in cell fate determination along the axis. The first identified substance shown to have the two essential morphogen properties and fulfilling Wolpert's prediction was Bicoid, which defines polarity along the anterior axis of the fruit fly embryo (Driever & Nüsslein-Volhard, 1988b, 1988a).

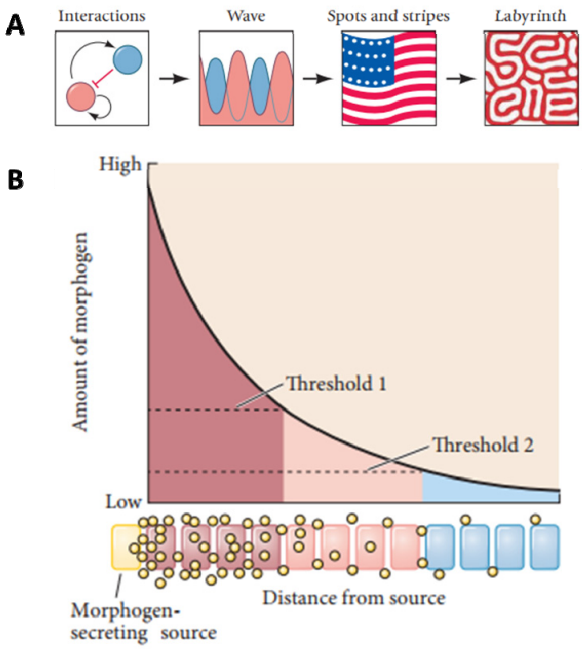


Figure 4. Early models proposed to explain cell fate determination during development: A) The Turing model is based on the interaction between two factors, one that is both auto-activating and able to activate its own inhibitor. It leads to a self-generating pattern of alternative cell fates that can resemble different patterns like: stripes, flags, or labyrinth. This model relies on short range interactions between cells. **B)** The French Flag Model proposed by Lewis Wolpert to explain the specification of different cell types by a morphogen gradient. The morphogen (yellow dots) is secreted from its source nuclei/cell (yellow) and forms a concentration gradient within the responsive nuclei/cell/tissue. Nuclei/cells exposed to morphogen concentrations above threshold 1 activate certain genes (red), while nuclei/cells exposed to intermediate concentrations (between thresholds 1 and 2) activate a different set of genes (pink) and also inhibit the genes induced at the higher concentrations. Those nuclei/cells encountering low concentrations of morphogen (below threshold 2) activate a third set of genes (blue). This model relies on long range interactions between cells. Reproduced with permission from (Gilbert & Barresi, 2016).

1.2.1 Genes involved in Anterior-Posterior patterning

The determination of the fruit fly embryo anterior-posterior (AP) axis is mainly defined by four maternal RNAs: *bcd*, *nos*, *hb* and *cad*. In a mature oocyte, the *bcd* RNAs are anchored at the anterior pole of the embryo while *nos* RNAs are anchored at the posterior pole. In contrast, *hb* and *cad* RNA are initially evenly distributed along the AP axis (Fig. 5A). Upon egg laying, these RNAs are translated. Unlike their RNAs, the Bcd and Nos proteins are free to diffuse in the syncytial embryo from their respective places of synthesis (where their RNAs are anchored). They form opposing concentration gradients (Fig. 5B). Both these proteins are translation regulators: Bcd represses the translation of *cad* mRNA in the anterior while Nos represses the translation of *hb* mRNA in the posterior (Fig. 5C) (Johnston & Nüsslein-Volhard, 1992). The whole process leads to the localized expression of Bcd and maternal Hb in the anterior, and of Nos and maternal Cad in the posterior.

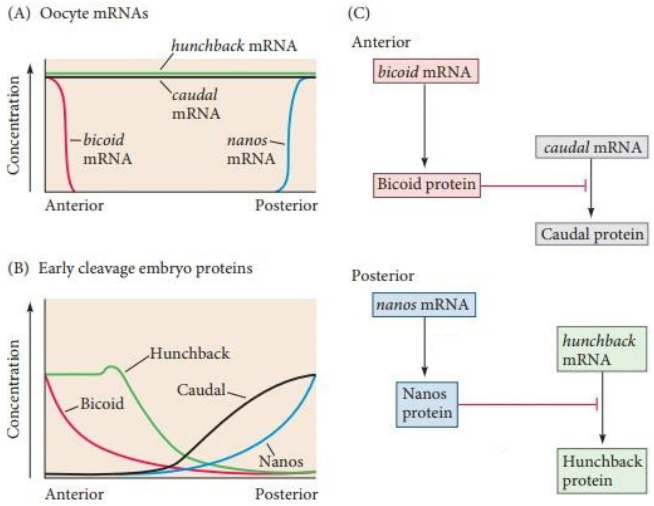


Figure 5. Anterior-Posterior axis specification pathways: A) *bicoid* (anterior) and *nanos* (posterior) maternal RNAs are polarized (unevenly distributed) in the oocyte while *hunchback* and *caudal* are evenly distributed along the AP axis. **B)** Upon egg laying proteins are translated. The Bcd protein diffuses from its site of synthesis at the anterior pole to form an exponential antero-posterior gradient. In contrast, Nos diffuses from its site of synthesis at the posterior pole to form a gradient with its highest concentration at the posterior. Bcd inhibits *cad* translation in the anterior, while Nos inhibits *hb* translation in the posterior. These inhibitions result in opposing Caudal and Hunchback protein gradients. The Hunchback gradient is secondarily strengthened by the zygotic transcription of the *hunchback* gene in the anterior nuclei by Bcd. **C)** Summary of gene regulation which establishes the anterior-posterior patterning of the *Drosophila* embryo. Modified with permission from (Gilbert & Barresi, 2016).

The Bcd protein is a homeodomain-containing transcription factor (TF) (Driever & Nüsslein-Volhard, 1988a; Hanes & Brent, 1989), which regulates the anterior organization of the embryo by activating the expression of target genes responsible for anterior development such as *hb* (Driever & Nüsslein-Volhard, 1989). As mentioned above, Bcd also represses *cad* mRNA translation by binding through cad 3'UTR and this leads to the expression of the Cad protein in a posterior gradient (Fig. 5C) (Rivera-Pomar et al., 1996). In embryos from *bcd* mutant females, the head and the thorax do not form (they are replaced by a duplication of the most posterior part of the tail, the telson) indicating that Bcd is necessary for head and thorax formation. Reciprocally, the injection of purified *bcd* RNA in the anterior of embryos from *bcd* mutant females or in the posterior of wild-type embryos induces the formation of anterior structures (Driever, Siegel, & Nusslein-Volhard, 1990). By increasing the amount of Bcd in the embryo using transgenic additional copies of the *bcd* genes in the female, C. Nüsslein-Volhard and W. Driever demonstrated that an increase in Bcd induces a shift of the whole fate map of the embryo towards the posterior and thus that Bcd behaved as a morphogen according to the prediction of the French Flag model (Driever & Nüsslein-Volhard, 1988a; Wolpert, 1969). As shown by allelic series, Bcd has an organizing activity: high concentrations produce the most anterior head structures, slightly lower concentrations produce the segmented head structures, even lower concentrations produce the thorax (Fig. 6 red panel) (Johnston & Nüsslein-Volhard, 1992; Porcher & Dostatni, 2010).

hb is a gap gene that is activated by Bcd in a large domain ranging from the anterior tip to the middle of the embryo, where it has a very sharp and precisely positioned expression border (Crauk & Dostatni, 2005; Porcher et al., 2010). When *hb* transcriptional regulation by Bcd was discovered, it was proposed that its expression pattern resulted from the presence of high-affinity binding sites for Bcd in the *hb* promoter, allowing expression even at low concentrations of Bcd (Driever & Nüsslein-Volhard, 1989; Struhl et al., 1989). The existence of high affinity binding sites for Bcd in the *hb* promoter was proven experimentally by Struhl et al. 1989. Bcd and maternally provided Hb act synergistically to activate *hb* expression zygotically (Simpson-Brose et al., 1994). In turns, Bcd and Hb (maternal and zygotic) act synergistically to promote the transcription of other head gap-genes such as *buttonhead* and *orthodentical* (Fig. 6 red panel) (Gilbert & Barresi, 2016; Simpson-Brose et al., 1994).

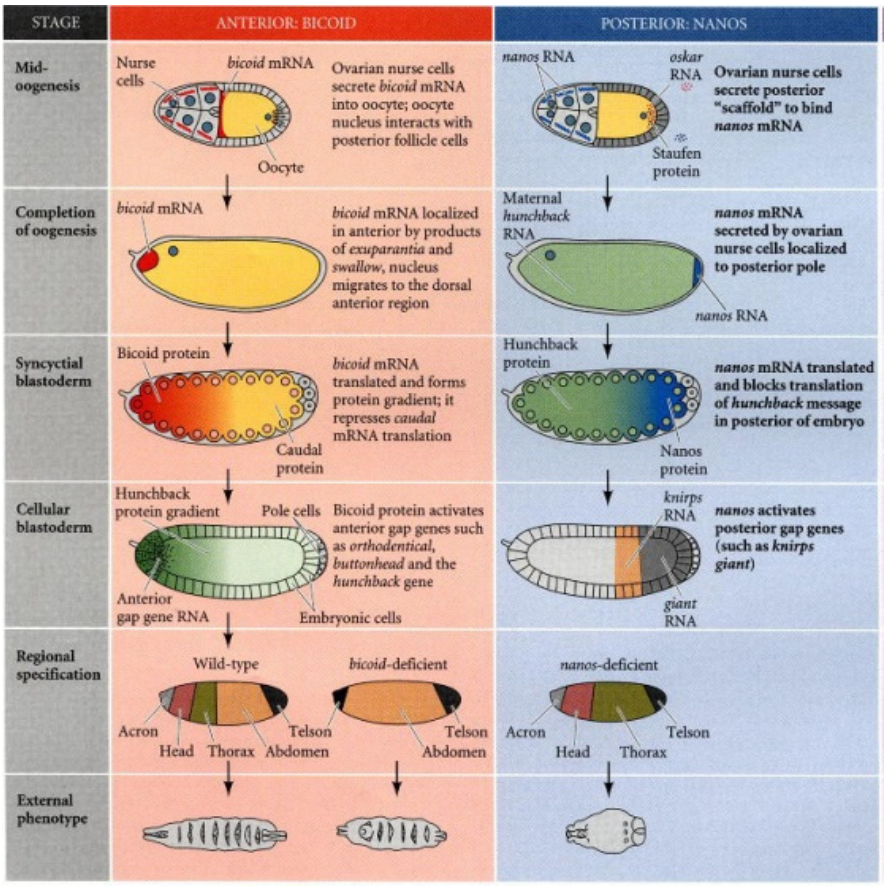


Figure 6. Anterior-Posterior axis development in *Drosophila*: The red panel (left) marks the events happening in the anterior part of the embryo, while the blue panel (right) shows the events happening in the posterior part. The initial asymmetry is established during oogenesis where *bcd* and *nanos* mRNAs are tethered to opposite poles of the oocyte. The pattern is organized by maternal proteins soon after fertilization where Bcd activates *hb* transcription while repressing *cad* mRNA translation in the anterior, and Nos represses *hb* in the posterior. Examples of null mutant experiments show the effect of Bcd and Nos null embryos. Modified with permission from (Gilbert, 2000).

The Nanos posterior gradient represses maternal *hb* mRNA translation in an indirect way (Irish et al., 1989). In the anterior of the embryo, another protein, Pumilio, binds the Nos Response Element in the *hb* 3' UTR allowing normal polyadenylation and translation. In contrast, in the posterior of the embryo Nanos binds Pumilio and prevents it to bind to the *hb* 3' UTR thus preventing the polyadenylation of *hb* RNA, inducing the shortening of *hb* mRNAs poly A tails and preventing their translation (Fig. 5C) (Wreden et al., 1997). As a consequence of their lack of translation in the posterior, *hb* maternal mRNAs are degraded. Importantly, embryos lacking the maternal contribution of *hb* and Nos survive to adulthood. This indicates that the maternal

contribution of *hunchback* is not essential for development and that the main role of Nos in patterning the AP axis is to repress *hb* mRNA translation in the posterior (Irish et al., 1989). The repression of *hb* mRNA translation in the posterior is required for the expression of posterior gap-genes such as *knips* and *giant* (Fig. 6 blue panel) (Gilbert, 2000; Lehmann & Nusslein-Volhard, 1991).

1.2.2 Maternal to Zygotic transition

As we have seen, early embryonic development is under control of the maternal genome (maternal mRNAs and proteins expressed in females during oogenesis). Maternal genes enable the first mitotic divisions, which are extremely rapid, 8 nuclear divisions occurring in about 1 hour. During this period, the chromatin exists in a relatively simple state (no methylation and low acetylation marks) (Lefebvre et al., 2018). Maternal genes also enable the activation of zygotic genes. During this maternal stage, the Zelda pioneer TF has been shown to bind to enhancers of early transcribed developmental genes, favoring their expression (Darbo et al., 2013; Foo et al., 2014; Liang et al., 2008a). This contributes to the first wave of zygotic activation at nc 8 (Fig. 7, light blue) with the transcription of a small set of genes (mostly segmentation genes) which is then followed by a second massive wave of zygotic transcription (Fig. 7, dark blue) starting at nc 13 and involving thousands of genes (Tadros & Lipshitz, 2009). As time progresses, the maternal mRNA pool is progressively degraded (Fig. 7, red), resulting in a switch from maternal to zygotic genome control around the second hour of embryogenesis (Lefebvre et al., 2018).

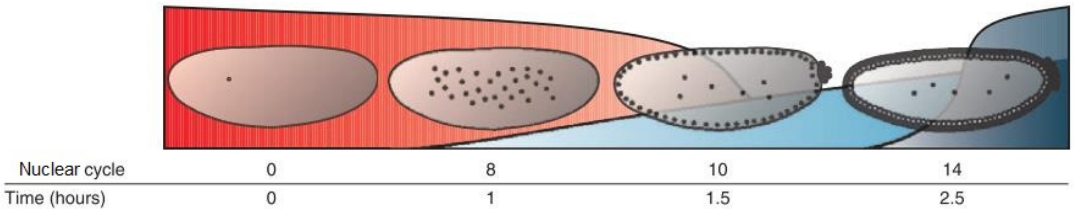


Figure 7. Overview of the maternal to zygotic transition in *Drosophila:* The important embryonic stages are illustrated by their corresponding nuclear cycle and time after fertilization (bottom). The red curve represent maternal mRNAs and their global degradation profiles. The light and dark blue curves illustrate the first and second wave of zygotic genome activation. Modified with permission from (Tadros & Lipshitz, 2009).

1.3 Noise and robustness during embryonic development

Embryonic development is highly dynamic yet very accurate and reproducible between individuals. When we look at the single cell level, gene expression is rather noisy. This noise comes from two different sources: intrinsic noise (from the stochasticity of reaction processes) and extrinsic noise (from environmental fluctuations) (Yang et al., 2018). Therefore, how an embryo can develop so precisely and so robustly despite these significant sources of noise is rather intriguing. Gene-regulatory networks have been proposed to be the answer to that question (Bentovim et al., 2017). The Bicoid – *hunchback* system provides a simple model, which can be used to study the question of noise and robustness during development. In this system, the Bcd exponential decay concentration gradient is thought of as the input, which is transformed into a precise output, the step-like response of the *hb* mRNA or the Hb protein (Gregor et al., 2007a; Porcher et al., 2010).

1.3.1 The Bicoid gradient

The first direct visualization of Bcd gradient was demonstrated using immunohistochemistry and specific antibodies against Bcd (Driever & Nüsslein-Volhard, 1988a; Struhl et al., 1989). Quantification of immunofluorescence allowed calculation of the exponential decay length of the gradient (λ = 100 µm) in the concentration gradient function, $C(x) = Co\ e^{-x/\lambda}$ (Houchmandzadeh et al., 2002). The quantification of the absolute levels of Bcd *in vivo* and the determination of its concentration (Co) remained a challenge.

Bcd's distribution and dynamics were first studied in live embryos by Gregor and collaborators using a Bcd-eGFP fusion expressed under the control of the *bcd* regulatory sequences to visualize the spatial distribution of the protein (Gregor et al., 2007b). These experiments revealed that there was a ~4 fold difference in the nuclear *versus* cytoplasmic concentration of the Bcd protein during interphase (S-phase). They also showed that although Bcd nuclear accumulation disappeared during each mitosis, it was rapidly re-established during each interphase following nuclear envelope formation (Gregor et al., 2007b). Bcd nuclear accumulation and gradient shape could be detected as early as nc 6, but the gradient steady state was only reached at nc 11 (Little et al., 2011). Once the steady state is reached, Bcd nuclear concentration remains constant until

nc 14 with 10% accuracy at any given position along the AP axis between cycles for a given embryo, and with a 10% variation between embryos in the same cycle (Fig. 8A,C) (Abu-Arish et al., 2010; Gregor et al., 2007b; Gregor et al., 2008).

Protein diffusion plays a major role in the formation of morphogen concentration gradients, influencing the range and speed at which they are established. It also influences a morphogen function, limiting the speed and precision at which its concentration can be interpreted (Fradin, 2017). Bcd-eGFP diffusion coefficient (D) was initially measured in live embryos by Fluorescence Recovery After Photobleaching (FRAP) (Gregor et al., 2007b) and was found to be $D = 0.3\ \mu m^2/s$. Given the fast establishment of a steady gradient (~1 hour), this value was too slow to support the hypothesis that the formation of the gradient occurs by simple diffusion (Gregor et al., 2007b). However, Fluorescence Correlation Spectroscopy (FCS) studies later done by the Fradin group showed that these earlier FRAP experiments failed to properly detect rapidly moving Bcd molecules (Abu-Arish et al., 2010; Fradin, 2017; Porcher et al., 2010). In the cytoplasm, the average diffusion coefficient of Bcd as detected by FCS was in fact ~ $7\mu m^2/s$ and 80% of the molecules were moving rapidly (Abu-Arish et al., 2010). This was later confirmed by the Wieschaus's group using photoconvertible proteins (Drocco et al., 2011). In the same work, the lifetime (τ) of the protein was shown to be about $\tau = 50$ min (Drocco et al., 2011). These observations led to the conclusion that the establishment of the Bcd gradient likely occurs via the Synthesis-Diffusion-Degradation model (SDD model) championed by F. Crick (Crick, 1970) and originally proposed in the case of Bcd by Driever and Nüsslein-Volhard (Driever & Nüsslein-Volhard, 1988a). When applied to Bcd gradient formation, this model assumes Bcd synthesis at a constant rate from a point source (the anterior pole where the maternal Bcd mRNAs are concentrated), free diffusion across the embryo (facilitated by the syncytial nature of the fly embryo), and degradation at a constant rate (the degradation is the inverse of Bcd's lifetime). The SDD model predicts that the gradient should be exponential, with a characteristic length ($\lambda = (D\tau)^{1/2}$) on the order of $\lambda = 100\ \mu m$, in good agreement with experimental observations (Fig. 8A,C) (Abu-Arish et al., 2010; Drocco et al., 2011).

Concerning the absolute concentration of Bcd in the embryo, two approaches both using the Bcd-eGFP gradients came to consistent measurements (same order of magnitude). T. Gregor and

coworkers, compared the fluorescence intensity in the embryo with the fluorescence of a solution containing a known concentration of GFP and evaluated the absolute concentration of the fluorescent Bcd-eGFP at ~50 nM at the pole (Gregor, Tank, et al., 2007). Using FCS, the Fradin team, evaluated the concentration at the anterior pole to be ~140 nM (Abu-Arish et al., 2010; Porcher et al., 2010).

1.3.2 The *hunchback* transcriptional response

With increased knowledge of Bcd gradient dynamics, the next question becomes how variability in the Bcd gradient might affect the precision and robustness of target genes expression. Most studies focused on *hb* gene expression, as it is Bcd's major target gene. *hb* is expressed very early (nc 9) (Porcher et al., 2010) and its domain of expression extends over the whole anterior half of the embryo. The border of the Hb protein expression domain is robustly positioned in the middle of the embryo, at x = 0.5 egg length (EL) (Houchmandzadeh et al., 2002). Use of reporters placing *lacZ* under the control of an artificial Bcd-dependent promoter revealed that a simple promoter with just three binding sites for Bcd resulted in an expression border that was precisely positioned with a variability among embryos of only 1.6% egg length (Crauk & Dostatni, 2005). The distance between two nuclei on each side of the border with significantly different probabilities to express *hb* has been shown to be 2% EL (dx ~ 10 µm), looking either at the Hb protein expression domain (Gregor et al., 2007a) or at the *hb* transcription by RNA FISH (Porcher et al., 2010). The change in Bcd concentration over this distance is $\delta c/c = dx/\lambda = 10\%$ (Fig 8B,C)(Gregor et al., 2007b; Porcher et al., 2010). Since the absolute number of Bcd molecules in a single nucleus at the expression border (at 50% EL) has been evaluated to be 700, a nucleus on the anterior side of the *hb* border expressing *hb* thus only contains 70 more Bcd molecules than a nucleus on the posterior side of the *hb* border which is not expressing *hb* (Abu-Arish et al., 2010; Gregor et al., 2007a). Also note that the precision with which a nucleus can establish its position by detecting Bcd concentration, 2% EL, is in the same range as that detected for the position of the Bcd gradient itself (Gregor et al., 2007b). This indicates that all steps in the process, i.e., the establishment of the gradient and the transcription and translation processes are very well controlled (Gregor et al., 2007a; Gregor et al., 2007b; Porcher et al., 2010).

The robust expression of *hb* is acquired, by nc 11 (Porcher et al., 2010), only 30 min after nc 8 which corresponds to both the onset of zygotic transcription and the steady establishment of the Bcd gradient (Gregor et al., 2007a; Gregor et al., 2007b). The robustness of the process contrasts with the stochastic nature of transcription in eukaryotic cells (Raser & O'Shea, 2005). In addition, the rapidity of the process is surprising given that this 30 min period is punctuated by three mitoses, during which the transcription process is interrupted (Porcher et al., 2010). Given this interruption of transcription during mitosis, two hypotheses were proposed: either robust *hb* transcription is achieved during each of the interphases of nc 8 to 14 (which greatly reduces its time length, from about 30 minutes to a few minutes) or the information about the measurement is memorized by each nucleus over mitosis (this hypothesis was called the memory hypothesis) (Porcher et al., 2010).

1.3.3 Precision of the transcriptional response

The precision of the readout is defined as the smallest change in the input resulting in a significantly different output. Using statistical mechanics and assuming that the Bcd TF is free to diffuse in 3D to find its target (its binding site on the *hb* promoter), the maximum precision with which the Bcd concentration can be detected at the *hb* gene, and that can be reached in a given time T, is given by an updated version of the formula originally proposed by Berg and Purcell (Berg & Purcell, 1977; Bialek & Setayeshgar, 2005; Kaizu et al., 2014), $\delta c/c = (4DacT)^{-1/2}$. We see that this precision is limited by the absolute concentration of Bcd molecules at the expression border (c), the size of the Bcd binding site in the promoter region (α) and the diffusion coefficient of the Bcd molecules (D). This formula assumes Bcd as the only input responsible for *hb* response, as well as random arrival times for Bcd molecules at the promoter (Berg & Purcell, 1977).

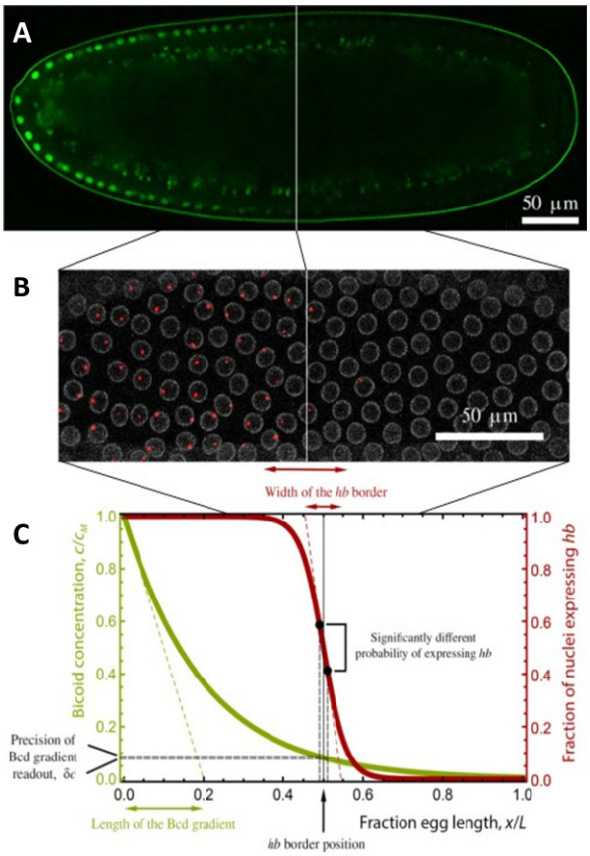


Figure 8. The Bicoid-*hunchback* system to study precision in *Drosophila* embryonic patterning: A) Confocal image of a *D. melanogaster* embryo expressing Bcd-EGFP at nc 12, showing the characteristic Bcd exponential concentration profile mostly visible in nuclei. **B)** Transcriptional activity of the *hb* promoter, visualized by confocal microscopy using the MS2 system (each red dot represents an active *hb* locus, grey circles show nuclear envelops). The channel showing *hb* transcriptional activity has been filtered and thresholded for clarity, and pseudo-colors are used for both channels. **C)** Schematic profile showing how the smooth distribution of the nuclear concentration of the Bcd protein (green) is turned into a sharp *hb* transcriptional response (red). The decay length of the Bcd gradient and precision of the *hb* response corresponds to what has been measured for this system at nuclear cycles 11–13 (decay length of the Bicoid gradient $\lambda \sim 100$ µm ~ 0.2 EL, position of the *hb* border at x=0.5 EL, and readout precision of dx=0.02 EL corresponding to $\delta c/c=0.1$). Reproduced with permission from (Fradin, 2017).

Given that the precision of Bcd readout at the *hb* gene is $\delta c/c \sim 10\%$, that the Bcd diffusion coefficient is $D \sim 7\mu m^2/s$, its concentration in nuclei $c \sim 4.8 \pm 0.6$ molecules/μm^3 (~ 700 molecules per nuclei) at the *hb* expression boundary, and that the effective size of the Bcd binding site in the promoter region can be considered to be about 8 base pairs (~ 3nm), it is possible to calculate the minimum time it should take the system to measure Bcd concentration with the observed precision. This calculation indicates that it should take ~25 minutes for the Bcd-*hb* system to reach the observed precision.

These calculated 25 minutes are much longer than the actual time the embryos have during each single S-phase from nc 8 to 13 (from 5 - 12 min) to measure the Bcd gradient and robustly express *hb*. This raises the question of how the embryos achieve such a precise measurement in such a short period of time? Could processes beyond a simple binding equilibrium of Bcd on the promoter explain the discrepancy between the estimated and measured time to measure Bcd concentration? Or could players other than Bcd be involved in this process?

1.3.4 Live imaging of transcription

To answer these questions, the Dostatni's group and the Gregor's group have developed fluorescent systems to follow the transcription process directly in living embryos. They adapted to the early developing embryo the MS2 approach, which allows the fluorescent tagging of RNAs (Fig. 9A) (Garcia et al., 2013; Lucas et al., 2013). This system relies on the RNA stem-loops of the MS2 bacteriophage and their MS2-coat protein (MCP), which can be fluorescently-tagged (MCP-GFP) (Bertrand et al., 1998). This system, originally developed by R. Singer, had already provided unprecedented access to the dynamics of the transcriptional process in cell culture (Larson et al., 2011; Lenstra et al., 2015) or single cells organisms such as yeast, for instance to study telomere dynamics, which was the focus of my M.Sc. research (Cusanelli et al., 2013).

Using a reporter gene in which the *hb* promoter was placed upstream of an MS2 cassette containing the sequence for 24 MS2 loops, a transgene expressing the MCP-GFP protein and a transgene expressing the nucleoporin (Nup-RFP) as marker for the nuclear envelop, the Dostatni's group showed that Bcd has no effect on the frequency of transcription activity periods (bursts) but increases their duration, suggesting that Bcd allows the maintenance of the flux of polymerase initiating transcription (Lucas et al., 2013).

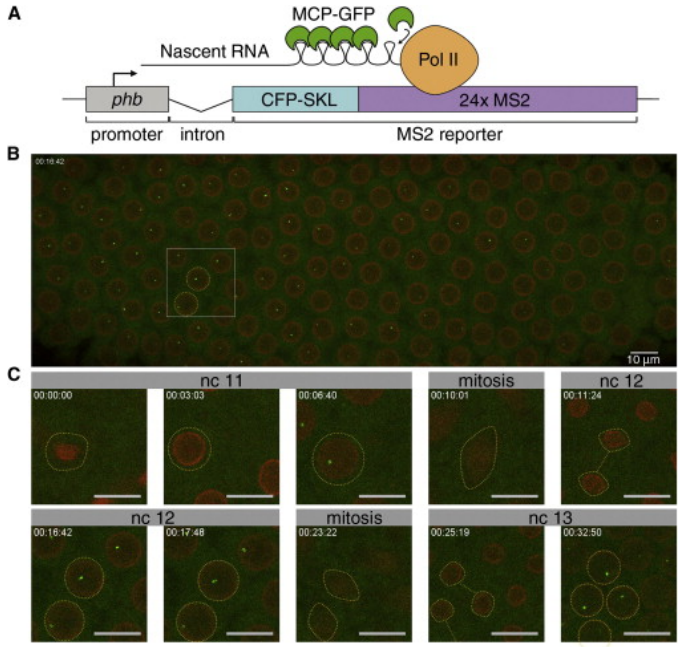


Figure 9. The MS2 system to study transcription in living *Drosophila* embryos: A) Schematic view of the *hb-MS2* transgene. The fluorescent detection of MS2-containing mRNAs was obtained with a second transgene maternally expressing the MCP-GFP at low levels. The fluorescent labeling of Nucleoporin (Nup-RFP) which allows the detection of nuclear envelop was obtained by a third transgene. **B)** A 2D maximal projection snapshot reveals periods of intense transcriptional activity, MCP-GFP proteins are recruited at the nascent MS2-containing mRNA accumulating at the *hb-MS2* locus (green bright spots), while the Nup-RFP proteins localize at nuclear envelop (red). **C)** Snapshots of a nucleus selected at cycle 11 that divides twice in the movie, giving rise to four daughter nuclei at interphase 13. Reproduced with permission from (Lucas et al., 2013).

Even though extremely promising, the *hb-MS2* reporter showed an unexpected expression in the posterior of the embryo (Lucas et al., 2013), which we later found out to be due to unfortunate repetition of Zelda (Zld) binding sites in the MS2 cassette in between the sequence coding the MS2 stem-loop (Lucas et al., 2018). Using a *hb* MS2-reporter with a new MS2 cassette mutated for all the potential Zld binding sites, which recapitulated what was known for endogenous *hb* expression using RNA-FISH, the goal of my thesis was to explore *hb* precise transcription during the very first hours of *Drosphila* development.

1.4 Main project hypothesis and objectives

My project was part of a collaboration between Pr. Cécile Fradin's group at McMaster University (Hamilton, Canada), and Pr. Nathalie Dostatni's group at the Curie Institute (Paris, France). The main objective of this collaboration is to get a better understanding of the mechanisms allowing precise and robust expression of morphogen target genes during the early development of the fruit fly embryo. More specifically, we want to understand how Bcd controls *hb* expression, with a particular emphasis on the fast and precise events taking place at the border separating the anterior region where *hb* is expressed and the posterior region where it is silenced.

Simple models trying to account for *hb* expression using only Bcd as an input fail to explain the spatiotemporal features observed in embryos. In particular, the very rapid measurement of Bcd concentration carried out by each nucleus during each S-phase is difficult to account for. Thus our main hypothesis was that *hb* transcriptional activity can be influenced by other TFs than Bcd, as binding sites for other TFs can be found in the *hb* promoter.

The MS2 approach developed in the Dostatni's lab allows the direct observation, in live embryos, of the transcription of a reporter gene recapitulating the expression of the *hb* gene from RNA FISH experiments. Beyond that, it also provides a unique opportunity to explore the dynamics of the formation of the *hb* expression border. **My first aim** was to use this approach to quantify the transcriptional dynamics of *hb* and determine the time taken for the *hb* gene to precisely measure Bcd concentration and achieve a robust expression.

We also wanted to understand to what extent factors, other than Bcd, might contribute to *hb* expression and how this is achieved. Therefore, **my second aim** was to use the same MS2 approach to start exploring the influence of different TFs on *hb* expression. For this, I used both classical genetics (i.e. mutant embryos missing one or both copies of *hb* maternal and/or zygotic genes) and synthetic biology (i.e. synthetic reporter genes with a minimal promoter region containing binding sites for only a small and well defined number of TFs).

So now let me guide you through what we have uncovered so far on our quest to understanding precision during embryonic development.

Chapter 2
MATERIALS AND METHODS

This chapter describes the methods used during the Ph.D. work. This includes embryo sample preparation and their live imaging, and the image and data analysis of the resulting acquired movies. These methods have been published as two book chapters in the Springer Methods in Molecular Biology Book series – Morphogen Gradients: Methods and Protocols (Perez-Romero et al., 2018; Tran et al., 2018).

The first book chapter, which I sign as a first author, provides the details of embryo preparation and imaging. It is included in its entirety in section 2.1. The second book chapter, which I sign as a second author, covers image and data analysis. The aspects of this second book chapter that are relevant for this thesis are summarized in section 2.2.

2.1 Sample preparation and live imaging

The following book chapter describes the method we developed for imaging transcription in early *Drosophila melanogaster* embryogenesis using *hb* transcription as an example: the *hb* minimal promoter (~700 bp) was placed upstream of an MS2 cassette. It allows the visualization of the fluorescent tagging of the RNA while they are produced at the *hb* promoter and provides access to the spatio-temporal transcriptional dynamics of this promoter during early development. A complete overview of how to prepare embryos for imaging and how to choose the configuration of the confocal laser scanning microscope to image these embryos is given in this chapter. The result of this procedure is a dual-color 3D movie of the transcription of a particular reporter gene in the early fly embryo. Although we standardized the image acquisition in the early stages of my Ph.D., we had to develop methods to check for quality of the microscope setup as output may vary over long periods of time, and this might have an impact on quantitative fluorescence analysis. This quality control was a learning process, but was not systematically used during the thesis. Otherwise all the methods described here are those that were used to acquire that presented in Chapter 3 and Chapter 4 of this thesis.

Live imaging of mRNA transcription in *Drosophila* embryos

Published as a book chapter in Morphogen Gradients: Methods and Protocols, a book from the Methods in Molecular biology series, reproduced with permission (Perez-Romero et al., 2018)

Carmina Angelica Perez-Romero[1,2], Huy Tran[1,3], Mathieu Coppey[4], Aleksandra M. Walczak[3], Cécile Fradin[1,2]*, Nathalie Dostatni[1]*

* Equal contribution

[1]Institut Curie, PSL Research University, CNRS, Sorbonne Université, Nuclear Dynamics, Paris, France. [2]McMaster University, Hamilton, Ontario, Canada. [3]Ecole Normale Supérieure, PSL Research University, CNRS, Sorbonne Université, Physique Théorique, Paris, France. [4]Institut Curie, PSL Research University, CNRS, Sorbonne Université, Physico Chimie, Paris, France.

My contribution:

Optimization of movie acquisition, manuscript writing, and figures production (except for Fig. 1e by Nathalie Dostatni, and Fig. 2 by Cécile Fradin, Table 1 was a joint effort with Cecile Fradin).

Objective of the book chapter:

To provide a comprehensive description of the methods we used to image *hb* transcription in live *Drosophila* embryos, friendly enough for new microscopists and/or new *Drosophila* researchers.

Highlights:

- In-vivo acquisition of transcription is discussed using the MS2 system to track *hb* transcription, and His-RFP as a nuclear marker.
- Embryo collection and maintaining on placing for optimal imaging is explained.
- Confocal microscopy parameters are discussed including their optimization and their importance for image quality.

Chapter 10

Live Imaging of mRNA Transcription in Drosophila Embryos

Carmina Angelica Perez-Romero, Huy Tran, Mathieu Coppey, Aleksandra M. Walczak, Cécile Fradin, and Nathalie Dostatni

Abstract

Live imaging has been used in recent years for the understanding of dynamic processes in biology, such as embryo development. This was made possible by a combination of advancements in microscopy, leading to improved signal-to-noise ratios and better spatial and temporal resolutions, and by the development of new fluorescence markers, allowing for the quantification of protein expression and transcriptional dynamics in vivo. Here we describe a general protocol, which can be used in standard confocal microscopes to image early *Drosophila melanogaster* embryos, in order to learn about the transcriptional dynamics of a fluorescently labeled RNA.

Key words MS2 system, Live imaging, Confocal microscopy, RNA, Embryo

1 Introduction

The Drosophila embryo is a model system to study how cell identity is established at the right place and time during development. It was demonstrated using genetics that most of the patterning along the anteroposterior (AP) axis is set up by the expression of a set of gap genes under the control of morphogen gradients (for a review *see* [1]).

Quantification of gene expression during development was obtained so far from the detection of ribonucleic acid (RNA) on fixed material using fluorescence in situ hybridization (FISH) [2]. However, the rapid development of the embryo and frequent interruptions of gene expression during mitoses result in complex dynamics that cannot be inferred from static observations. Recently, the possibility to fluorescently tag RNA in living cells using well-characterized RNA targets (motifs) for specific RNA binding proteins fused to fluorescent domains provided access to

Cécile Fradin and Nathalie Dostatni contributed equally to this work.

Julien Dubrulle (ed.), *Morphogen Gradients: Methods and Protocols*, Methods in Molecular Biology, vol. 1863,
https://doi.org/10.1007/978-1-4939-8772-6_10, © Springer Science+Business Media, LLC, part of Springer Nature 2018

the dynamics of transcription in real time. These developments combined with advances in confocal microscopy (which allowed reducing signal-to-noise ratio and improving spatiotemporal resolution) and in genome editing to tag the RNA of interest with motifs [3] give access to time dependent measurements of mRNA transcription in living organisms during their development.

Although several systems allowing to visualize RNA have been described [3–6], the MS2 system is probably the most popular one [7–9]. It relies on the MS2 bacteriophage RNA loops that can be added to a target RNA, and its core protein (MCP) that can be fused to a fluorescent protein like green fluorescent protein (GFP) allowing the visualization of the expression of the target RNA and quantify it [4]. Recently, the MS2 system has been used to tag different genes in fruit fly development, which has helped elucidate how gene circuits interact with each other in a quantitative manner [8, 10–15]. For example, we used this system to study the transcriptional response downstream of the Bicoid gradient, through the expression of its target gene *hunchback* [11, 12, 16]. The MS2 system has also been combined with another RNA tagging system (PP7) to tag two different RNA with different fluorescent proteins (GFP and RFP—red fluorescent protein) in the same embryo [17]. In our analyses, MS2 signals were very difficult to observe with epifluorescence and we had to use confocal microscopy to perform optical sectioning such that we could avoid the fluorescence signal coming from the whole thickness of the embryo. With the early fly embryo, our acquisition only involves the first few microns of the periphery of the embryo and thus does not require 2-photon microscopy. Yet, this type of microscopy can be used to image transcription in tissues that require deeper imaging. Finally, light sheet microscopy can also be used even though the technical bottleneck here will be image analysis on the collected data.

Here we describe a protocol to image transcription in living embryos using the MS2 system and standard laser scanning confocal microscopy.

2 Materials

2.1 Genetic Material

The system requires the simultaneous expression of three transgenes:

1. The MS2 reporter transgene: it includes the promoter of interest upstream of an MS2 reporter cassette (in general 24 MS2 loops) fused to the sequence of the iRFP (infrared fluorescent protein) or any coding sequence for a protein that you can easily detect (to insure that the tagged RNA is eventually translated). Importantly, the position of these two sequences relative to each other and to the promoter depends on the

question you ask. Briefly, a stronger MS2 signal will be obtained if the MS2 cassette is placed just downstream of the promoter (best signal-to-noise ratio) but faster dynamics of promoter bursting will be detected when the MS2 cassette is placed further away from the promoter (reviewed in [15]). Also, when designing the MS2 reporter make sure that cryptic binding sites for important transcription factors in your system are not localized in the spacers or stem loop sequences of the MS2 cassette. Finally, it is also possible to tag the 5' UTR, 3' UTR or introns of endogenous genes with the MS2 cassette by using genome editing approaches.

2. The MCP-GFP transgene: it expresses the MCP-GFP fusion protein at relatively low levels for increased signal-to-noise ratio. In the very early fruit fly embryo (syncytial), it was shown that an MCP-GFP without a nuclear localization signal (ΔNLS) improved detection of the specific transcription signal and reduced nuclear GFP aggregates at nuclear cycle 14 [12]. In the fly, there are available UAS-MCP-GFP lines [18] that can be combined with appropriate Gal4 drivers.
3. A transgene allowing the detection of nuclei: this reporter is required to identify the subcellular localization of the MS2 signal. These are for instance nucleoporin fused to RFP (Nup-RFP) to label the nuclear envelop (Bloomington # 35517) [19] or histone fused to RFP to label chromatin (Bloomington # 23650) [20]. In the early embryo, the signal detected from the His-RFP transgene is very strong and regenerates during each nuclear cycle, allowing for easier segmentation during image analysis. Even though it is slightly toxic for the embryo, we prefer it over other nuclear envelop markers such as maternally expressed Nup-RFP, whose signal is weaker and fades away during development.

The analyzed embryos result in general from a cross between females expressing the MCP-ΔNLS-GFP and His-RFP transgenes with males carrying the MS2 reporter. Once crossed, the resulting embryos will express maternally the MCP-GFP and His-RFP proteins and carry only a single MS2 locus, which helps for image analysis (*see* Chap. 11, this book). Embryos are then collected, removed from their chorion and imaged as follows using a confocal Laser Scanning Microscope 780 from Zeiss that enables the visualization of the reporter transgene in the green detection channel and the visualization of the histones in the red detection channel.

2.2 Sample Material

1. Appropriate fly stocks and crosses.
2. Paintbrush.
3. Embryo collection plates (ECP): Space nontreated petri dishes out, so that when mix is ready you can plate it rapidly. Weight

22 g of sucrose, and 14 g of Bacto agar. Add to 300 ml water and boil (using a microwave) until properly mixed and clear. While cool enough to touch, but still clear add 10 ml of pure ethanol 100% and 5 ml of glacial acetic acid. Add 100 ml of grape juice and mix well. Add mix to the petri dishes and let dry. Store in a closed plastic box with a little water to keep the humidity at 4 °C.

4. 10% acetic acid solution: 10% V/V glacial acetic acid in water.
5. Yeast paste: Gradually mix active dry yeast and water until they have a texture similar to cake frosting (sticky but not runny), this will favor embryo laying. Activated yeast will tend to expand so allow space for this to happen and break the air bubbles. The yeast paste can be kept fresh for a few days covered with parafilm and stored at 4 °C.
6. Spatula.
7. Embryo collection cage: To make homemade egg collection chamber you can use a narrow *Drosophila* vial. Carefully cut the bottom part of the vial, making a cylinder. On one end of the cylinder, add a cotton plug to allow for aeration of the chamber yet stop the flies from escaping. Flip the flies into the egg collection chamber and add the embryo collection plate on the other end, trapping the flies in the chamber. Secure the embryo collection plate to the egg collection chamber with tape.
8. Double-sided tape.
9. N-heptane glue: Cut small pieces of double-sided tape and put those inside a small glass bottle (use around 30 cm of double sided tape). Add 10 ml of *n*-heptane in the bottle to cover the tape (under a fume hood). Let the mixture incubate overnight. The heptane should dissolve the glue from the double-sided tape, which will make the heptane sticky. The mixture is good to use until most of the heptane-glue has evaporated. Remember that heptane evaporates fast and is flammable.
10. Fine point precision tweezers.
11. Dissecting microscope.
12. Dissecting needle.
13. Microscope glass coverslip (0.17 mm thickness).
14. Microscope slide.
15. 10 S Voltalef oil.
16. Confocal microscope (in our setup: Laser Scanning Zeiss, LSM 780 integrated with 488 and 585 laser lines with GaASp detection).
17. Heating and cooling chamber.
18. Immersion oil 518 F (23 °C).

19. Software for imaging acquisition.
20. Phosphate buffer saline (PBS, pH 7.4).
21. Fluorescein solution: Dissolve 1 g of fluorescein into 1.5 ml phosphate-buffer saline (PBS, pH 7.4) and vortex the solution. A fluorescein precipitate (red powder) should remain in the solution, indicating that the solution is saturated (this ensures that the concentration of dissolved fluorescein in the stock solution is always exactly the same, even if solvent evaporates). If kept in the dark at room temperature, this saturated fluorescein solution will be stable for at least 2 months. This solution can then be diluted to use for fluorescence standards.

3 Methods

3.1 Collecting Embryos

Virgin females carrying the MCP-GFPΔNLS and His-RFP are crossed with male flies carrying the MS2 reporter. The crosses must be performed 1 day in advance in culture vials to ensure that all females have been fertilized. Since live imaging experiments are usually carried out with a single embryo at a time, small numbers of embryos are sufficient to perform the experiment. However, it is recommended to have at least 10 females to carry out the study. Crosses will have a good embryo yield for about 4 days. Crosses are done in a normal feeding tube (Fig. 1a), and then flies are put to lay embryos on a collection plate (Fig. 1b).

3.1.1 Embryo Laying

1. Take 10 μl of 10% acetic acid solution and spread evenly with a paintbrush on an ECP.
2. With the end tip of the paint brush make a gentle grid of grooves on the ECP agar where flies will preferentially lay their eggs.
3. With a spatula add a small amount of yeast paste to the center of the ECP (Fig. 1a).
4. Take out crosses from narrow *Drosophila* vials and transfer to an embryo collection cage.
5. Rapidly cover the end of the embryo collection cage with the prepared ECP.
6. Safely secure the ECP and eggs collection cage together with tape (Fig. 1b).
7. Put back embryo collection cage into the incubator (*see* **Note 1**).

3.1.2 Chorion Removal and Preparation of the Coverslip for Imaging

Drosophila eggs are around 500 μm in length. They are white and can be seen with the naked eye on the contrasting background of ECPs. However, for chorion removal and preparation of the coverslip higher magnification is needed. Therefore, this procedure

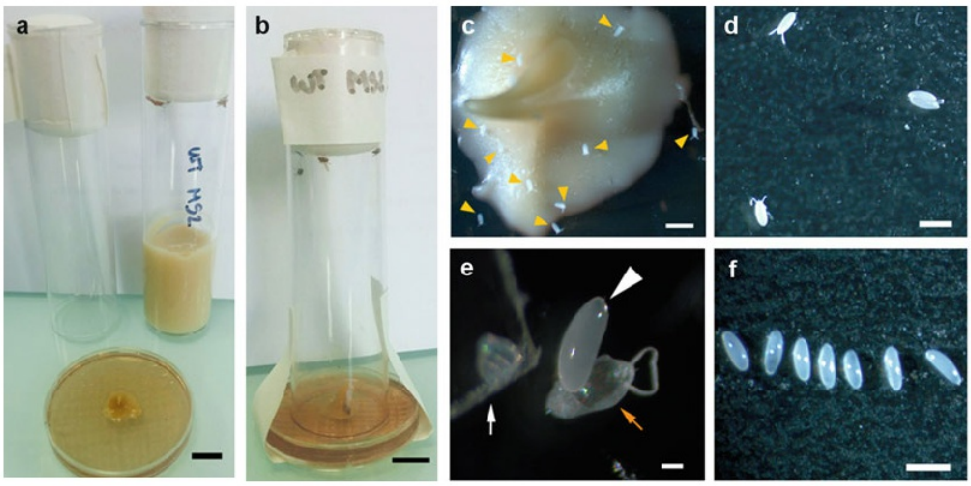


Fig. 1 *Drosophila* embryo collection and chorion removal. (**a**) Setup needed for making an egg laying chamber: an embryo collection plate prepared with 10% acetic acid, yeast paste and groves; an empty feeding tube closed with a plug; and an overnight cross of flies needed for laying. (**b**) Final setup of egg laying chamber once ready should be placed at 25 °C. (**c**) The embryo collection cage is removed from the incubator after the appropriate laying time, and eggs (marked by arrows) that are laid around and on top of the yeast paste (big white blob) are transferred to a slide (**d**) with a double-sided tape for removal of the chorion. (**e**) Embryos are rolled on double sided tape (white arrow) to be released from the chorion (orange arrow). Embryos at nc ~ 9 without a chorion (white arrow head) are transferred to a prepared coverslip with heptane glue. (**f**) The embryos are aligned and a drop of 10S oil is added before imaging. Scale bars: 5 mm (**a** and **b**), 1 mm (**c**), 500 μm (**d** and **f**), and 100 μm (**e**)

should preferably be carried out under a dissecting scope. A microscope with a low magnification objective with 2× or 5× objective can also be used.

1. Take 10 μl of heptane glue with a pipette and spread it over a microscope coverslip making fine lines, let dry (*see* **Note 2**).
2. Take the ECP from the collection cage by tapping gently the inverted cage and rapidly covering the opened cage end with another fresh ECP.
3. Using a precision tweezer or a double sided tape handled with tweezers, take the embryos from the ECP with care and transfer them to the double sided tape slide (Fig. 1c, d, *see* **Note 3**).
4. Remove the chorion of each embryo by gently touching its body, and rolling it over the double sided tape (Fig. 1d, e, *see* **Note 4**). Removal of the chorion by hand avoids the use of bleach (which prevents the embryo from sticking properly on the coverslip).
5. Once the chorion is removed, transfer the embryo to the coverslip, placing it on top of the glue line.

6. Place the embryo in the desired orientation for imaging (Fig. 1f).
7. Using the length of a dissecting needle gently tap the embryo in place to flatten it (*see* **Note 5**).
8. Cover the embryos with a generous drop of 10S Oil (*see* **Note 6**).
9. Secure coverslip into mounting device for imaging.

3.2 Setting and Optimizing Acquisition Parameters for Live Imaging

Here we describe a general protocol for adjusting the acquisition parameters of a Laser Scanning Microscope to obtain optimal spatial and temporal resolution for the imaging of live *Drosophila* embryos (or, in general, large-scale dynamic systems). This protocol should be repeated at the beginning of each different imaging series (e.g., each time a new type of embryo is being imaged). We used a Zeiss LSM 780 microscope and associated Zen Black software to develop this protocol; however, it can easily be adapted to any modern commercial confocal microscope. Tables 1 and 2 contain a brief

Table 1
Primary imaging parameters

Parameter	Symbol	Optimal value	Typical value
Magnification	M	Intermediate (allowing to image sufficiently large sample areas with an appropriate spatial resolution)	40×
Numerical aperture	NA	High (to achieve appropriate spatial resolution)	1.4
Excitation wavelength	λ	Intermediate (low λ generally result in higher signal levels but also more photodamage and more autofluorescence).	488 nm, 561 nm
Excitation power	P	Low (to avoid photobleaching)	3%–3.2% of laser power
Confocal pinhole diameter	D	~2wM = 1 Airy unit (AU)	31.5 μm = 0.99 AU
Detector gain	γ	Intermediate (to achieve both a linear response and sufficient signal level)	Master gain =700 Digital gain = 2
Pixel separation	d	~w	0.15 μm
Image separation	d_Z	~5w	0.5 μm
Pixel dwell time	δ	Short (to allow fast imaging)	0.5 μs
ROI width, height, depth	L_X, L_Y, L_Z		240 μm, 100 μm, 10 μm

Table 2
Derived imaging parameters

Parameter	Relation to other parameters	Typical value
Point-spread function radius	$w \approx \lambda/(2\mathrm{NA})$	488 nm/561 nm 0.17 μm/0.20 μm
Point-spread function half-height	$w_z = 5w$	1 μm
Pixels per line	$n_X = L_X/d$	1200
Lines per image	$n_Y = L_Y/d$	512
Images per z-stack	$n_Z = L_Z/d_Z$	20–30
Image acquisition time	$r > n_X\, n_Y\, \delta$	0.460 ms
Z-stack acquisition time (scan rate)	$R > n_X\, n_Y\, n_Z\, \delta$	10–20 s

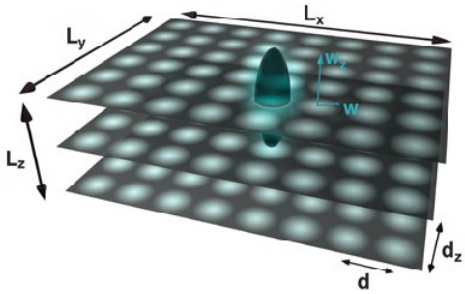


Fig. 2 Schematic representation of the confocal imaging process. The sketch shows the point-spread function at the laser focus (green revolution ellipsoid, with radius w and half-height w_z) as it scans through every pixel (each pixel is represented by a blue disk) in the ROI (width L_X, height L_Y, depth L_Z). The distance between two pixels is d. In this particular example, the confocal imaging results in a z-stack of $n_z = 3$ images, each with pixel dimensions $n_x = 8$ and $n_y = 6$. Also, in this example the images are undersampled ($d > w$)

description of the imaging parameters to be optimized, and give their typical value in our experiments as a reference. Figure 2 gives an illustration of the confocal imaging scheme and a graphic representation of some of these imaging parameters.

1. Select an objective with moderate magnification (e.g., 40×) and high numerical aperture (NA), which usually requires water or oil immersion. For example, the images shown here were acquired with a plan-apochromat 40×/1.4 NA oil immersion objective.
2. Activate the necessary laser lines and select appropriate filter cubes for the fluorophores to be imaged (in our case: GFP, maximally excited at 488 nm and visualized around 520 nm,

and RFP, maximally excited at 586 nm and visualized around 620 nm). The emission filters should have a bandpass as large as possible to maximize fluorescence signal collection, while avoiding cross talk between the two channels.

3. Set the pinhole diameter to 1 Airy unit for an optimal trade-off between resolution (both in the focal plane and along the optical axis) and signal collection. If necessary, the pinhole diameter can later be increased above 1 Airy unit to increase signal collection efficiency (and thereby possibly scan rate), at the cost of spatial resolution and sectioning power (and vice versa).
4. Place a representative sample (of the type that will be studied) on the microscope stage. Quickly select a laser power and detector gain that allows detecting the signal of interest (these two parameters will be adjusted more precisely at a later step).
5. Choose the dimensions of the region of interest (ROI) that is the size of the region in the sample that will be imaged. In general, this will need to be chosen according to your experimental question. Remember, however, that a large ROI can only be imaged at the expanse of spatial and/or temporal resolution.
6. Next choose the number of pixels per line and per row, or, alternatively, the pixel separation (Fig. 2). The Nyquist criterion can be used to determine the optimal pixel separation (*see* **Note 7**). As a rule, the more pixels the better the spatial resolution but the longer the acquisition time. Optimizing pixel separation is thus important, as it will greatly influence scan rate.
7. Choose the span (depth) and increment (separation between two consecutive images) of the z-stack for volumetric acquisition (Fig. 2). Optimizing the distance between images in the z-stack will influence the number of stacks needed for imaging a ROI with a particular depth, and thus affect time resolution. The optimal distance between images in the stack can also be estimated using the Nyquist criterion (*see* **Notes 7** and **8**).
8. The last adjustments to be made concern the pixel dwell time (or alternatively the scan acquisition time, as both are directly related). The pixel dwell time can be adjusted according to your experimental question (so that the dynamics of the process under study can be captured). Shorter pixel dwell times allow to capture faster processes and to minimize photobleaching; however, this is at the cost of signal collection.
9. To optimize the levels of fluorescence signal detected, after a pixel dwell time has been chosen, adjust the laser power and detector gain to achieve a sufficient signal-to-noise ratio to

visualize the process of interest (as estimated by eye from the acquired images). This has to be done separately for both channels. Laser power and detector gain must be adjusted together, until the best compromise is obtained. Remember that the laser power should be kept sufficiently low to minimize phototoxicity (short wavelength light is the most damaging to organisms) and photobleaching, while high detector gain can result in nonlinear detector response.

10. If time resolution is of great importance, after all the other settings have been optimized for this purpose, a bidirectional scanning mode can be selected. This can greatly improve the time resolution; however, it can also cause image artifacts when using high magnification objectives (*see* **Note 9**). Once all the imaging parameters have been optimized, remember to save them (e.g., as an experimental setup in the Zen Black software) in order to be able to use them again in subsequent imaging sessions.
11. As laser power and microscope optical alignment may vary with time, affecting the amount of fluorescence signal collected from embryos, we recommend using a control fluorescence slide to keep a record of the intensity of the signal that can be obtained in this particular instrumental configuration (*see* **Note 10**). This can be used as a benchmark in subsequent experiments.

Once the imaging parameters have been optimized for a particular set of experiments, they should be saved and used consistently in subsequent experiments of the same type, allowing reproducible experiments to be used for quantitative analysis.

3.3 Live Imaging of Drosophila Embryos

Here we describe the basic procedure to record a 3D movie of a live embryo. This procedure generates a series of two-color image stacks. Each image stack is the 3D confocal image of the same region (the region of interest, or ROI) of the embryo at a different time (as illustrated in Fig. 2). Along with the 3D movie, we recommend performing a tile scan of the whole embryo, which will allow for the measurement of the embryo dimensions (antero-posterior and dorso-ventral axes) and exact localization of the ROI within the embryo. Finally, we explain how to use maximum projection to produce a 2D movie that can be then exported as an .avi file and used for quick visualization.

1. Set the microscope incubator chamber to 23 °C, and allow to equilibrate for 30 min (*see* **Note 11**).
2. Turn on the lasers to be used for excitation, and allow their output intensity to equilibrate if necessary.
3. Select the correct objective and carefully add a drop of water or immersion oil if required.

4. Load the confocal imaging parameters saved at the end of the setup procedure described in the previous section, and record an image of a fluorescence calibration slide (*see* **Note 10**). Check that the signal obtained from this slide is similar to that obtained in previous experiments. A signal that is significantly lower (or higher) than usual indicates that there is an issue with the alignment or the settings of the microscope (e.g., incorrect laser intensity, incorrect fluorescence filters, incorrect detector gain, misaligned confocal pinhole), something that should be dealt with before further imaging. Small daily variations in signal, on the other hand, are to be expected and no cause for concern. The intensity recorded for the calibration slide can then be used to normalize the intensity of the images acquired during this session, allowing comparison between the data obtained in different sessions.
5. Place the coverslip with the embryos (prepared as explained in Subheading 3.1) onto the microscope stage. Turn on the bright field (transmitted) illumination and move the region of the sample containing the embryos (that should be visible to the naked eye) in the field of view.
6. Adjust the objective focus until the embryos come into focus.
7. Still using bright field imaging, you can quickly observe each embryo in the sample to determine which ones are in the desired developmental stage (*see* **Note 12**).
8. Once an embryo in the correct developmental stage has been spotted, you can rotate the coverslip to roughly orient the embryo as desired relative to the scanning axis.
9. Start live confocal acquisition to check for the presence of a fluorescence signal in the chosen embryo and confirm its developmental stage (*see* **Note 13**).
10. At this point, the image of the embryo may be digitally rotated to position it properly within the computer screen (e.g., with the antero-posterior axis oriented horizontally).
11. Move the stage to place the region of interest in the center of the field of view.
12. Take a full sagittal plane image of the embryo by acquiring three images (tiles) that can be assembled to form an image of the whole embryo (Fig. 3a). This will allow the measurement of the embryo dimensions, as well as for the exact location of the ROI within the embryo (e.g., along the antero–posterior axis). This is important for quantification and image analysis purposes. This tile scan can also be performed at the end of the experiment (*see* **Note 14**).
13. Choose the ROI (region of interest, Fig. 3) that encompasses the area of interest in the embryo (*see* **Note 15**).

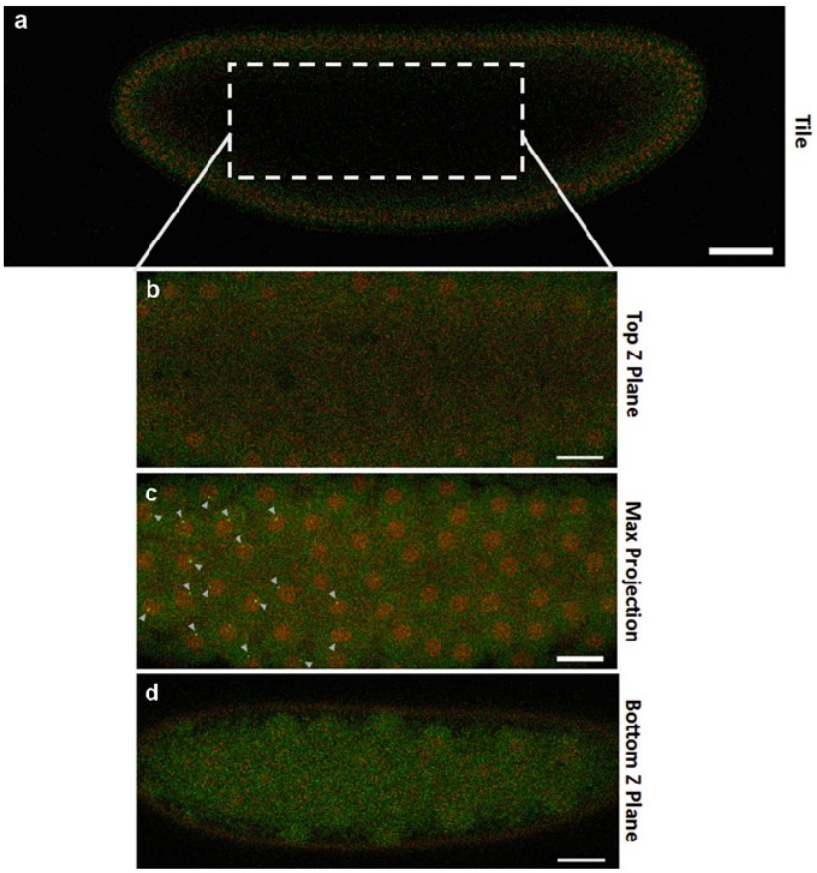


Fig. 3 Live confocal imaging of a Drosophila embryo. (**a**) Tile view of the whole embryo at nc 14, obtained from three adjacent tiles, auto-stitched at the end of the acquisition. The ROI within the embryo is indicated by the dashed line rectangle. Scale bar: 50 μm. **b–d**. Snapshots of different views of the ROI at nc 11. (**b**): Uppermost image in the z-stack, above the layer of nuclei. (**d**): Bottom image in the z-stack, below the layer of nuclei. (**c**): Maximum projection of the 23 images in the z-stack (i.e., all the images acquired between planes (**b**) and (**d**), each separated by a distance $d_Z = 0.5$ μm). Active transcription sites are indicated by white arrowheads. Scale bar: 20 μm (**b–d**)

14. Define the depth of the region to be imaged, by choosing the position of the top and bottom planes in the z-stack (Fig. 3b, d, *see* **Note 16**). The number of images in the z-stack will then be calculated according to the optimal image separation determined during the optimization step (Subheading 3.2).
15. Select continuous acquisition, or a very high amount of imaging cycles. In this way the acquisition can be terminated based on the completion of the biological process under study.

16. Finally, start the acquisition and let the embryo develop without disturbing (*see* **Note 17**).
17. When appropriate, stop the acquisition and save the resulting data (*see* **Note 18**).
18. One can perform a fast image processing of the acquired 3D movie, using maximum projection, in order to generate a 2D movie with a single image per time point (Fig. 3c). This simplified version of the data can then be saved in .avi or .mov format, making it easy to disseminate or to share with collaborators. It also allows for a rapid visualization of the studied process.
    Movies of transcriptional dynamics obtained with flies carrying a MS2 reporter and expressing MCP-GFPΔNLS, as well as a nuclear marker such as His-RFP, can then be analyzed using the *LiveFly* toolbox described in Chap. 11.

4 Notes

1. Regular 12 h day–night incubation cycles are important to have consistent growth and a good yield in embryo laying. Since egg laying occurs preferentially during sunset and dark periods, it can be useful to set up the incubator on an inverted day–night schedule to optimize live imaging experiments during the working hours of the laboratory.
2. Since the heptane evaporates quickly, using too much of heptane glue will create an uneven surface of glue for the embryos to lay. It is thus important to take a small quantity of heptane glue and spread it rapidly in a stripe like manner over the coverslip. The more even the surface, the easier it is to position the embryos and flatten them, which can influence both the quality of imaging and the development of the embryo.
3. The embryos with their chorions will easily stick to one another and to the inside or outside part of the tweezers, making the transfer easier.
4. The chorion will stick to the double sided tape. To release the embryo from the chorion, roll it gently until the chorion breaks. Once the embryo is partially released, move slightly the chorion to release the embryo completely, and using tweezers or a piece of double sided tape handled with tweezers transfer the embryo without chorion to the coverslip. Removal of the chorion by hand should be a gentle process, and it may take time and practice before one is able to do it properly. Also, the younger the embryos, the more fragile they are and the harder it is to remove their chorion without damage.

5. Since the embryo is slightly curved, focusing on its central part may result in both its extremities being slightly out of focus, making them hard to image with high enough resolution. For studies relying on proper imaging of both the central part of the embryos and their anterior and/or posterior end, it is thus important to flatten the embryos. We have found it hard to flatten embryos without affecting their viability and expression. However, the simple step of tapping an embryo with the length (not the tip!) of a dissecting needle helps not only to glue the embryo firmly to the coverslip but also to flatten its ends.
6. The 10S Oil is very viscous, so it is hard to pipette and handle. However, it is important to cover the embryos properly and shortly after removal of their chorion to prevent them from drying. This oil is oxygen permeable such that embryos can continue to develop while under the microscope.
7. The pixel size (sometime referred to as sampling rate), d, is related to the width of the ROI (L_X) and to the number of pixels per line (n_X) through $d = L_X/n_X$. The pixel size can usually either be set directly in the software, or controlled by varying L_X (in the Zeiss LSM, this is done by changing the so-called zoom factor) and/or n_X (that is by changing the image pixel dimensions). Usually, one would use square pixels, with $L_Y/n_Y = L_X/n_X$. For the Nyquist criterion to be met, a minimum of two pixels must be used to image the length of an area with the dimension of the point spread function. In other words, the pixel size, d, should be no larger than the radius of the point-spread function, w (*see* Table 2). If $d > w$, under-sampling occurs, which means that spatial information might be lost. If $d < w$, oversampling occurs, which means that one is sacrificing time resolution without any gain in spatial resolution.
8. The Nyquist criterion is also relevant to the choice of the distance between two images in a z-stack, d_Z. As the point-spread function is elongated along the optical axis, usually by a factor of about 5, the optimal image separation, d_Z, should be around $5d_X$. However, just as for the pixel size, one can choose to trade off spatial resolution (by choosing a larger d_Z) for temporal resolution.
9. To check that the bidirectional scan is not affecting your signal take a small five frame time series of your biological sample. Focus on an area with bright spots that remains immobile over time. Then select a single bright pixel, and check that it does not move along the scanning direction in subsequent frames, and that the image is not blurred. If problems with bidirectional scanning occur, they are usually easily seen from one frame to the next in your biological signal. If problems with

bidirectional acquisitions occur, return to the line scanning mode, which unfortunately has a much lower time resolution.

10. A saturated solution of fluorescein can be used to prepare slides with reproducible fluorophore concentration, for calibration purposes. When needed, dissolve a small amount of the saturated fluorescein solution in fresh PBS. (It is very important to always use the same dilution; it is also very important that the pH of the PBS used for dilution is exactly at 7.4, as fluorescein is a pH-sensitive dye.) Use this diluted solution to prepare a calibration slide (stick a coverslip on a microscope slide with heated parafilm spacers, fill the gap between slide and coverslip with the dye solution, then seal with melted wax or transparent nail polish to avoid evaporation). A fresh calibration slide (usually prepared on the same day) can be used at the beginning of each imaging session for reference: a similar signal obtained from the calibration slides before two different imaging sessions means that a quantitative comparison of the images obtained in both sessions can be made, while decrease in the signal measured from the calibration slide over time may indicate variations in the laser power output or misalignment of the optics (e.g., confocal pinhole alignment). If quantitative image analysis is to be done using the red channel, a similar fluorescence slide standard can be prepared, e.g., using rhodamine. Commercial fluorescence standards such as Argolight (Argo-HM) that work for most wavelengths and are photostable for years are also available, but at a cost.
11. It is important to equilibrate the temperature of the system since changes in temperature can cause focus drift and affect image quality. A temperature control chamber that allows not only heating but also cooling is recommended, since usually heating devices are optimized for 37 °C applications. However, live imaging of *Drosophila* embryos requires a temperature of 23 °C. Typical incubator systems usually have difficulty maintaining the system at this temperature during acquisition, due to the additional heat emanating from the lasers. Therefore, it is recommended to place the microscope in a room kept at a cool temperature, and/or an incubator system with both heating and cooling capabilities, with a feedback to keep the system at a stable 23 °C temperature during imaging, such as the one available from Tokai Hit (INUC-KPP Series).
12. Developmental features of the drosophila embryos can be used to determine their developmental stage (their exact nuclear cycle (nc)). The earliest signal that we have been able to detect with our MS2 system has been around nc 9 or 10. A useful marker for this stage is the emergence of the polar bodies at the posterior side of the embryo in early nc 9. On the other hand,

nc 14 is easily recognized by its very long duration and the size and amount of nuclei in the embryos.

13. Using a *Drosophila* strain expressing either a histone or a nucleoporin fused to RFP can help determine the developmental stage of the embryo during live imaging (by quickly inspecting the red channel), as well as the exact position of nuclei, which is important both during imaging to determine the optimal ROI, and during the segmentation step of image analysis (*see* Chap. 11). If an embryo is found to be at an earlier developmental stage than desired, one can wait for it to develop to the right stage. It is a good idea to keep track of a few embryos at a time in the sample to anticipate which ones might reach the correct stage to be imaged. This can easily be done using a multidimensional acquisition setup to save the position of different embryos. Once one of the embryos is at the desired developmental stage, the multidimensional acquisition can be canceled and the correct embryo selected for imaging.
14. The tile scan should cover the entire length of the embryo, including the ROI in the sagittal plane. If the embryo is horizontal, this can usually be achieved using three horizontal tiles. The tiles can be processed with auto-stitching, or manually stitched. It is then possible, for example, to measure the distance between the ROI and the anterior and posterior poles of the embryo. This might be important for quantification and image analysis when comparing several different embryos. The length of the embryo can also be measured using this tile scan.
15. Once a ROI has been created it can be saved and loaded, allowing to use ROI with the exact same size between experiments. The smaller the ROI, the faster the scan rate, so it is really important to make sure the imaging region is as small as possible.
16. When one is interested in imaging the cortical region of the embryo, where nuclei can be found after nc 8, we recommend to set the bottom and top image planes of the z-stack (rather than just setting the center one). The first (bottom) plane should be set above the lower membrane of the embryo and before the first nuclear signal is seen in the red channel. The last (top) plane should be set up a few microns above the first plane, after most nuclei have been observed (while moving the focus higher in the sample from the first plane). It is a good idea, however, to add one or a few planes up and down to account for focus drift during image acquisition. It is important also to use the same number of images per z-stack between acquisitions, since this will influence the scan rate, which should be

consistent between experiments if quantitative comparisons are to be made.

17. At the beginning of acquisition, make sure that the embryo displays the expected features, for example that foci corresponding to nascent RNA appear in the GFP channel, and that the embryo development appears normal, e.g., by checking the nuclear signal in the RFP channel. Sometimes the chosen embryo might not develop correctly or might not develop at all, in which case another embryo will have to be chosen, or a new slide will have to be prepared, and the procedure repeated.
18. The acquired image stacks can be saved in .czi or .lsm format if using the Zen software. Other formats, for example .tiff can be used with other imaging software. The *LiveFly* toolbox described in Chap. 11 is capable of opening any microscopy format compatible with the Bio-formats LOCI tools [21].

Acknowledgments

The authors thank Patricia Le Baccon and the Imaging Facility PICT-IBiSA of the Institut Curie. This work was supported by a PSL IDEX REFLEX Grant for Mesoscopic Biology (ND, AMW, MC), an Ontario Trillium Scholarship for International Students (CAPR), a Mitacs Global Link Scholarship (CAPR) and an Internal Curie Institute Scholarship (CAPR), ARC PJA20151203341 (ND), a Mayent Rothschild sabbatical Grant from the Curie Institute (CF) and an NSERC discovery grant RGPIN/06362-15 (CF), a Marie Curie MCCIG grant No. 303561 (AMW), ANR-11-LABX-0044 DEEP Labex (ND), ANR- 11-BSV2-0024 Axomorph (ND and AMW) and PSL ANR-10-IDEX-0001-02. Cécile Fradin and Nathalie Dostatni contributed equally to this work. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

References

1. Porcher A, Dostatni N (2010) The bicoid morphogen system. Curr Biol 20:R249–R254. https://doi.org/10.1016/j.cub.2010.01.026
2. Jensen E (2014) Technical review: in situ hybridization. Anat Rec 297:1349–1353. https://doi.org/10.1002/ar.22944
3. Urbanek MO, Galka-Marciniak P, Olejniczak M, Krzyzosiak WJ (2014) RNA imaging in living cells – methods and applications. RNA Biol 11:1083–1095. https://doi.org/10.4161/rna.35506
4. Bertrand E, Chartrand P, Schaefer M, Shenoy SM, Singer RH, Long RM (1998) Localization of ASH1 mRNA particles in living yeast. Mol Cell 2:437–445. https://doi.org/10.1016/S1097-2765(00)80143-4
5. Elf J, Li GW, Xie XS (2011) Probing transcription factor dynamics at the single-molecule level in a living cell. Science 316:1191–1194. https://doi.org/10.1126/science.1141967
6. Nelles DA, Fang MY, O'Connell MR, Xu JL, Markmiller SJ, Doudna JA, Yeo GW (2016) Programmable RNA tracking in live cells with

CRISPR/Cas9. Cell 165:488–496. https://doi.org/10.1016/j.cell.2016.02.054

7. Cusanelli E, Perez-Romero CA, Chartrand P (2013) Telomeric noncoding RNA TERRA is induced by telomere shortening to nucleate telomerase molecules at short telomeres. Mol Cell 51:780–791. https://doi.org/10.1016/j.molcel.2013.08.029
8. Bothma JP, Garcia HG, Ng S, Perry MW, Gregor T, Levine M (2015) Enhancer additivity and non-additivity are determined by enhancer strength in the Drosophila embryo. Elife 4. https://doi.org/10.7554/eLife.07956
9. Desponds J, Tran H, Ferraro T, Lucas T, Perez-Romero CA, Guillou A, Fradin C, Coppey M, Dostatni N, Walczak AM (2016) Precision of readout at the hunchback gene. PLoS Comput Biol 12(12): e1005256. https://doi.org/10.1371/journal.pcbi.1005256
10. Bothma JP, Garcia HG, 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 U S A 111:10598–10603. https://doi.org/10.1073/pnas.1410022111
11. Lucas T, Ferraro T, Roelens B, De Las Heras Chanes J, Walczak AM, Coppey M, Dostatni N (2013) Live imaging of bicoid-dependent transcription in Drosophila embryos. Curr Biol 23:2135–2139. https://doi.org/10.1016/j.cub.2013.08.053
12. Garcia HG, Tikhonov M, Lin A, Gregor T (2013) Quantitative imaging of transcription in living Drosophila embryos links polymerase activity to patterning. Curr Biol 23:2140–2145. https://doi.org/10.1016/j.cub.2013.08.054
13. Lim B, Levine M, Yamakazi Y (2017) Transcriptional pre-patterning of Drosophila gastrulation. Curr Biol 27(2):286–290. https://doi.org/10.1016/j.cub.2016.11.047
14. Esposito E, Lim B, Guessous G, Falahati H, Levine M (2016) Mitosis-associated repression in development. Genes Dev. 30(13): 1503–1508. https://doi.org/10.1101/gad.281188.116
15. Ferraro T, Esposito E, Mancini L, Ng S, Lucas T, Coppey M, Dostatni N, Walczak AM, Levine M, Lagha M (2016) Transcriptional memory in the Drosophila embryo. Curr Biol 26:212–218. https://doi.org/10.1016/j.cub.2015.11.058
16. Ferraro T, Lucas T, Clémot M, De Las Heras Chanes J, Desponds J, Coppey M, Walczak AM, Dostatni N (2016) New methods to image transcription in living fly embryos: the insights so far, and the prospects. Wiley Interdiscip Rev Dev Biol 5:296–310. https://doi.org/10.1002/wdev.221
17. 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
18. Forrest KM, Gavis ER (2003) Live imaging of endogenous RNA reveals a diffusion and entrapment mechanism for nanos mRNA localization in Drosophila. Curr Biol 13:1159–1168. https://doi.org/10.1016/S0960-9822(03)00451-2
19. Katsani KR, Karess RE, Dostatni N, Doye V (2008) In vivo dynamics of Drosophila nuclear envelope components. Mol Biol Cell 19:3652–3666. https://doi.org/10.1091/mbc.E07-11-1162
20. Pandey R, Heidmann S, Lehner CF (2005) Epithelial re-organization and dynamics of progression through mitosis in Drosophila separase complex mutants. J Cell Sci 118:733–742. https://doi.org/10.1242/jcs.01663
21. Linkert M, Rueden CT, Allan C, Burel J-M, Moore W, Patterson A, Loranger B, Moore J, Neves C, Macdonald D, Tarkowska A, Sticco C, Hill E, Rossner M, Eliceiri KW, Swedlow JR (2010) Metadata matters: access to image data in the real world. J Cell Biol 189:777–782. https://doi.org/10.1083/jcb.201004104

2.2 Image and data analysis

The acquired *hb*-MS2/MS2-GFP movies contain a large amount of data from which information on the transcriptional dynamics can be measured and quantified. Although our eyes allow a quick peak at some of this information (like border location), this is just a qualitative way of assessing the data. Therefore, our laboratory developed a protocol allowing for a quantitative analysis of the images acquired as explained in the previous section. This protocol was implemented as a MATLAB program called LiveFly which allows us to process our two-color 3D confocal movies acquired from developing embryos in a semi-automatic way (meaning that user input and oversight is required at different checkpoints). This program delivers quantitative information in the transcriptional activity of each single nucleus in the field of view, which can then be analyzed in different ways.

Although this program was elaborated before my arrival, I contributed to shape it with a robust and user friendly graphical user interphase (GUI) allowing the image processing of movies in a reliable way. Another GUI was later build for analyzing the transcriptional dynamics features obtained from embryos where *hb* transcription has been perturbed. Here is summarized the capabilities of this program that has been published as an accompanying book chapter (Tran et al., 2018), and used to analyze the movies that led to our publications (Desponds et al., 2016; Lucas et al., 2018; Tran et al., 2018).

LiveFly provides an interface for the analysis of two-color 3D confocal movies of *Drosophila* embryos. It is meant for embryos at an early stage of development (nc 10-14), in which nuclei are fluorescently labeled and the transcription of a particular reporter gene has been made visible using the MS2-MCP system. The program performs three different image or data processing procedures: 1) From each movie, nuclei positions are determined, as well as their lineage; 2) Within each nucleus, active transcription sites are identified and transcriptional dynamics information is extracted; 3) Results obtained for different embryos are merged together to increase statistics. The *LiveFly* toolbox provides three different modules that each implement one of these processes, which are detailed in the three sections below.

2.2.1 Nuclei segmentation and tracking

This module extracts the nuclei position and lineage from the nuclei channel, which in our case is the red channel since nuclei were marked with either His-mRPF (Histone-2B) or Nup107-mRFP (Nucleoporin 107). To facilitate this procedure, the nuclei channel is processed in 2D (3D movies are typically ~10 Gb). Therefore, the module first creates, at each time point, a 2D maximum projection of the image stack, which also registers the global change in nuclei movement from frame to frame (frame registration). Next, nuclei are automatically detected using user specified parameters (size, threshold). A mask of the detected nuclei is created, which can be manually corrected by the user by adding/deleting nuclei in each frame through the movie (nuclei segmentation, Fig. 10A). The segmented nuclei are then automatically tracked from frame to frame, and an individual ID number is given to each nucleus ID each nuclear cycle. Genealogical information is also stored (i.e. for each nucleus, the ID number of the mother is saved). At this step, the user has to verify that tracking was indeed done correctly, and that the ID given to nuclei correspond between frames. Tracking accuracy was greatly increased by adding frame registration, which helped with the correct detection of most nuclei in each frame and minimized the tracking errors especially during drastic nuclei movements in mitosis. Mitosis poses another challenge, as nuclei divide in waves, starting from the two embryo poles (Farrell & O'farrell, 2014). Therefore, nuclei at the anterior pole may produce MCP-MS2 spots earlier due to either earlier chromatin decondensation or earlier reentrance of Bcd into the nucleic space. To correct for this, we identify the time of birth of each nucleus after mitosis, defined as the time when the segregation from its sibling is complete (Fig. 10B). This time is then set as t = 0 for that particular nucleus at this given nuclear cycle and it will be used as such for the MS2 time traces that the nucleus possibly expresses.

After the position of each nucleus has been determined in each frame, and their genealogy has been established along with their time of birth, the next step is to extract the transcriptional dynamics information.

2.2.2 Spot detection

This module extracts the fluorescent intensity of active MCP-GFP transcription loci in each nucleus as a function of time (MS2 time traces). These time traces contain information on the

transcriptional dynamics of the corresponding gene (in our case a single *hb* locus). For each nucleus identified in the previous step (from the 2D maximum projection), the full 3D image stack in the MCP-GFP channel is used for spot detection. The approximate spot location is first determined using a thresholding method where the threshold value can be set and tested by the user for best signal detection of all transcription loci throughout the entire movie. The detected spot is then fitted with a 3D Gaussian function, from which the spot exact position and intensity are extracted (Fig. 10C). The obtained information is then compiled into a tabular output file, which combines the results of the nuclei segmentation (ID number, mother ID number, position, radius) with the results of the spot detection (intensity of the signal detected at the transcription site as a function of time, see Fig. 10C).

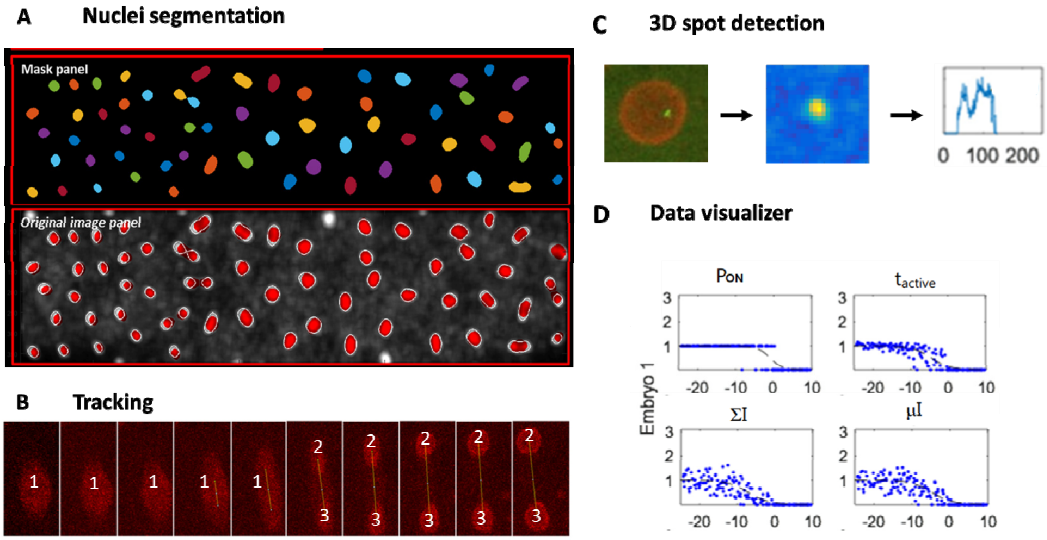


Figure 10. Overview of *LiveFly*: A) Nuclei segmentation example, with the upper panel showing the mask created by the program which overlays nicely with the nuclei seen in the original maximum projection image shown in the bottom panel. The user can adjust the segmentation parameters or perform manual corrections at this point. **B)** Nuclei tracking allows following the nuclei though space and time, as well as recording their genealogy after mitosis occur. **C)** Spot detection example: First a nucleus expressing the *MS2* reporter locus 3D MCP-GFP is shown, followed by the MCP-GFP signal thresholded for detection, which when performed over time for each image in which the nucleus is present, finally gives us the transcriptional dynamics information for the particular locus. **D)** The transcriptional data for each analyzed movie can be seen in our data visualizer under the form of graphs showing, for P_{ON}, t_{active}, ΣI, μI as a function of position along the AP axis. These various parameters can be fitted to a Hill curve, which allow us to measure the Hill coefficient, the plateau of expression as well as the border of expression. Modified with permission from (Tran et al., 2018), (Lucas et al., 2018) and personal data.

2.2.3 Data visualizer

This module allows merging the data obtained for multiple embryos (as long as they have the same genetic background), and to visualize the transcription information contained in fluorescent traces produced at the MS2 reporter locus.

The program can extract several features of the transcriptional traces, such as: time of first spot appearance counted from the time of birth from each individual nucleus (t_{init}, called the transcription initiation time), the total spot appearance duration (t_{active}), the integrated spot intensity (ΣI) used as relative measure for the total amount of mRNA produced at a given locus, the mean spot intensity (μI) used as a relative measurement for the mRNA production rate, wether transcription was ON at any time during that cycle (activity = 1) or not (activity = 0) give us the probability for nuclei to be ON (P_{ON}). All these features are displayed as a function of the nucleus position along the AP axis (Fig. 10D) (Lucas et al., 2018; Tran et al., 2018).

To compare and merge results from several embryos of the same genetic background, the features are first normalized by their value at the anterior pole to account for variability due to data acquisition process. In addition, to take into account variations in the embryo's growth rate, time is normalized for each cycle by the average length of the cycle.

Potential variations in the amount of maternal *bcd* mRNA in the embryo's anterior pole may lead to variations in Bcd concentrations at a given position in different embryos and thus variations in boundary position of the expression pattern. To take into account these variations likely due to the input (the Bcd gradient) if wanted by the user, the embryos' AP axis can be aligned by their respective border position based on the P_{ON} feature. In order to do this, the border position is set by inspecting where along the AP axis P_{ON} equals 0.5. The embryos' AP axes are then shifted to have their respective P_{ON} border at position 0% EL.

Additionally, this module can create a graph of the pattern based on these features, and characterize the *hb* expression pattern's steepness. The features are first normalized by their expected mean average value at the anterior pole to remove differences coming from the data acquisition process. The steepness of the features can then be obtained by least-square fitting of

the feature value along the AP axis in each embryo and nuclear cycle with a sigmoid function, Hill curve (Lucas et al., 2018; Tran et al., 2018):

$$f([X]) = \frac{1}{1 + e^{H(X-X_0)/L}}.$$

where X is the position along the AP axis, X0 is the pattern's border position ($f(X_0)=0.5$) and H is the pattern steepness (i.e. the Hill coefficient) for the feature of interest. L is the decay length of the Bcd gradient, which is fixed at its known value of ~100 µm or 20 % EL (Fradin, 2017; Gregor, Wieschaus, et al., 2007).

Note that this alignment was done for WT embryos that have a very constant output (Chapter 3). However, no alignment was performed when working with mutants, as they display more diverse phenotype and a less clear border, which we decided to first analyze without constrains (Chapter 4). Therefore, in any case, to determine the border position and the plateau value that P_{ON} reaches at the anterior part of the embryo, the experimentally determined P_{ON} profile was first smoothed using a moving average filter with a width corresponding to 7% EL, then the plateau value was assimilated to the maximum value reached by the smoothed profile, and the border position as the position at which the smoothed P_{ON} profile reached exactly one half of the plateau value.

Chapter 3

HOW LONG DOES IT TAKE FOR THE *hunchback* GENE TO PRECISELY MEASURE Bicoid CONCENTRATION?

This chapter contains our first valid observations of the transcriptional dynamics of the *hb* promoter, accessible for the first time thanks to the MS2 reporter system. Our main paper, Lucas et al. 2018 (reproduced in its entirety in section 3.1) reports on the time necessary to establish a precise response. Two closely related articles illustrate more complex ways in which the information contained in the MS2 movies can be further exploited. The first, Desponds et al. 2016 (summarized in section 3.2) focuses on the temporal analysis of the transcriptional traces in order to establish the dynamics state of the *hb* promoter (poisson, two states, cycle). The second, Tran et al. 2018, summarized in section 3.3 describes our efforts to build a model of activation consistent with our experimental data.

Although I am not the first author on any of these articles, I have significantly contributed to this work. I have acquired most of the movies in both (Lucas et al., 2018) and (Desponds et al., 2016). Movie processing was a joint effort with Dr. Lucas. Also, I participated in the discussions leading to the development of the models described in (Tran, Desponds, et al., 2018). We have included here the article to which I contributed the most (Lucas et al., 2018), and summarized the findings reported in the other two, which are of a more theoretical nature (Desponds et al., 2016; Tran, Desponds, et al., 2018). All three articles showcase the potential of the live MS2 data acquired during my PhD, and they establish the need to consider the contribution of TFs other than Bcd to explain the rapid acquisition of the *hb* response "sharpness" which will be the object of Chapter 4.

3.1 How long does *hb* precise transcription take?

Lucas T, Tran H, Perez Romero CA, Guillou A, Fradin C, Coppey M, Walczak AM, Dostatni N. "3 minutes to precisely measure morphogen concentration". *PLOS Genet*. 2018;14(10):e1007676. doi:10.1371/journal.pgen.1007676.

This paper gives an answer to the long standing question of the time necessary for *hb* transcription to "measure" Bcd input and transform it into a domain of expression with a steep border, a step-like output we can observe for the transcription of its mRNA.

We measure the time it takes for *hb* to form its precise border during nuclear cycles 11-13 using a *hb-MS2* reporter recapitulating the endogenous expression of the *hb* gene. Using modeling we show that only 6 Bcd binding sites (number of binding sites for Bcd described in the *hb* promoter) alone are not able to explain the fast establishment of the sharp expression border observed experimentally, opening the possibility that either there are more binding sites for Bcd in the *hb* promoter or that, in addition to Bcd, other TFs contribute to the process.

3 minutes to precisely measure morphogen concentration

Published as a research article in PLOS Genetics and reproduced with permission

Tanguy Lucas[1,*], Huy Tran[1,2,*], Carmina Angelica Perez Romero[1,3], Aurélien Guillou[1], Cécile Fradin[1,3], Mathieu Coppey[4], Aleksandra M. Walczak[2], and Nathalie Dostatni[1]

* Equal contribution

[1]Institut Curie, PSL Research University, CNRS, Sorbonne Université, Nuclear Dynamics, Paris, France. [2]Ecole Normale Supérieure, PSL Research University, CNRS, Sorbonne Université, Physique Théorique, Paris, France. [3]Dept. of Physics and Astronomy, McMaster University, Hamilton, Ontario, Canada. [4]Institut Curie, PSL Research University, CNRS, Sorbonne Université, Physico Chimie, Paris, France.

My contribution:

Fly keeping, ensuring right microscopy setup and conditions for live imaging embryos, movie acquisitions parameters optimization. I acquired all the movies mentioned in the manuscript and movie processing was a joint effort with Dr. Tanguy Lucas. I reviewed the manuscript.

Research Objective:

To measure the time taken by *hb* to precisely measure Bcd concentration.

Research Highlights:

- We built and use a novel MS2 reporter without Zelda binding sites that recapitulates endogenous *hb* expression as seen in fixed embryos using RNA FISH.
- The time period required for *hb* expression to reach maximum efficiency increases with the distance from the anterior pole, consistent with a Bcd dose-dependent activation process.
- We show that it takes 3 minutes at each interphase of nc 11-13 for *hb* to measure Bcd concentration and to establish a steadily positioned and steep expression boundary.

PLOS | GENETICS

RESEARCH ARTICLE

3 minutes to precisely measure morphogen concentration

Tanguy Lucas[1☯], Huy Tran[1,2☯], Carmina Angelica Perez Romero[1,3], Aurélien Guillou[1], Cécile Fradin[1,3], Mathieu Coppey[4], Aleksandra M. Walczak[2], Nathalie Dostatni[1*]

1 Institut Curie, PSL Research University, CNRS, Sorbonne Université, Nuclear Dynamics, Paris, France, **2** Ecole Normale Supérieure, PSL Research University, CNRS, Sorbonne Université, Physique Théorique, Paris, France, **3** Dept. of Physics and Astronomy, McMaster University, Hamilton, Ontario, Canada, **4** Institut Curie, PSL Research University, CNRS, Sorbonne Université, Physico Chimie, Paris, France

☯ These authors contributed equally to this work.
* nathalie.dostatni@curie.fr

Check for updates

OPEN ACCESS

Citation: Lucas T, Tran H, Perez Romero CA, Guillou A, Fradin C, Coppey M, et al. (2018) 3 minutes to precisely measure morphogen concentration. PLoS Genet 14(10): e1007676. https://doi.org/10.1371/journal.pgen.1007676

Editor: Michael B. Eisen, University of California Berkeley, UNITED STATES

Received: April 26, 2018

Accepted: September 5, 2018

Published: October 26, 2018

Copyright: © 2018 Lucas et al. This is an open access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: RNA FISH data and the time lapse movies are accessible at the following link: http://xfer.curie.fr/get/NjnfH8faAlt/LUCAS_2018_DATA.zip in the form of a Zipped folder of ~ 50 Go. All other relevant data are within the paper and its Supporting Information files.

Funding: This work was supported by a PSL IDEX REFLEX Grant for Mesoscopic Biology (ND, AMW, MC), ANR- 11-BSV2-0024 Axomorph (ND and AMW), ARC PJA20151203341 (ND), ANR-11-LABX-0044 DEEP Labex (ND), an Ontario Trillium Scholarship for International Students (CAPR), a

Abstract

Morphogen gradients provide concentration-dependent positional information along polarity axes. Although the dynamics of the establishment of these gradients is well described, precision and noise in the downstream activation processes remain elusive. A simple paradigm to address these questions is the Bicoid morphogen gradient that elicits a rapid step-like transcriptional response in young fruit fly embryos. Focusing on the expression of the major Bicoid target, *hunchback* (*hb*), at the onset of zygotic transcription, we used the MS2-MCP approach which combines fluorescent labeling of nascent mRNA with live imaging at high spatial and temporal resolution. Removing 36 putative Zelda binding sites unexpectedly present in the original MS2 reporter, we show that the 750 bp of the *hb* promoter are sufficient to recapitulate endogenous expression at the onset of zygotic transcription. After each mitosis, in the anterior, expression is turned on to rapidly reach a plateau with all nuclei expressing the reporter. Consistent with a Bicoid dose-dependent activation process, the time period required to reach the plateau increases with the distance to the anterior pole. Despite the challenge imposed by frequent mitoses and high nuclei-to-nuclei variability in transcription kinetics, it only takes 3 minutes at each interphase for the MS2 reporter loci to distinguish subtle differences in Bicoid concentration and establish a steadily positioned and steep (Hill coefficient ~ 7) expression boundary. Modeling based on the cooperativity between the 6 known Bicoid binding sites in the *hb* promoter region, assuming rate limiting concentrations of the Bicoid transcription factor at the boundary, is able to capture the observed dynamics of pattern establishment but not the steepness of the boundary. This suggests that a simple model based only on the cooperative binding of Bicoid is not sufficient to describe the spatiotemporal dynamics of early *hb* expression.

Mitacs Global Link Scholarship (CAPR) and an Internal Curie Institute Scholarship (CAPR), a Mayent Rothschild sabbatical Grant from the Curie Institute (CF) and an NSERC discovery grant RGPIN/06362-15 (CF), a Marie Curie MCCIG grant No. 303561 (AMW) and PSL ANR-10-IDEX-0001-02. - https://www.psl.eu/en/research/funding-opportunities/psl-calls-proposals - http://www.agence-nationale-recherche.fr/en/ - https://www.fondation-arc.org/espace-chercheur - https://grad.uwo.ca/current_students/student_finances/ots.html - https://www.mitacs.ca/en/programs/globalink - https://enseignement.curie.fr/en/content/senior-researcher-sabbatical - http://www.nserc-crsng.gc.ca/index_eng.asp -https://ec.europa.eu/research/participants/portal/desktop/en/opportunities/fp7/calls/fp7-people-2013-cig.html The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors have declared that no competing interests exist.

Author summary

During development, the first thing that an embryo needs to know is the orientation of its body. We study how the head-to-tail axis forms in the fruit fly embryo. To position the axis, the embryo relies on proteins called morphogens broadcasting instructions to other genes, so that cells know, depending on where they are along the axis, what they should become. The concentration of the Bicoid morphogen is much higher in the head of the embryo and lower towards the tail. Bicoid activates *hunchback*, which divides the embryo in two parts. By visualizing this process in real time in living embryos, we see the *hunchback* gene activated very efficiently in the head part, where each nucleus expresses the gene, whereas it is not expressed at all in the tail part. Intriguingly, the boundary separating the two domains is precisely positioned and becomes very steep in no more than three minutes. We use a modeling approach to understand how this is achieved so rapidly. Given the parameters of the system, we find that although our model is able to reproduce the fast dynamics of the process, it fails to reproduce the steepness of the boundary suggesting that a more complex approach is needed to capture the additional mechanisms involved.

Introduction

Morphogens are at the origin of complex axial polarities in many biological systems. In these systems, positional information is proposed to be provided by morphogen concentrations, which allow each cell to measure its position along the embryo's axes and turn on expression of target genes responsible for the determination of its identity. Although the existence of these gradients is now well established [1], the quantitative details of their functioning (i.e. how small differences in morphogen concentrations are precisely and robustly interpreted into a threshold-dependent step-like response) remains largely debated [2].

To address this question, we study the Bicoid morphogen, which specifies cell identity along the antero-posterior (AP) axis of the fruit fly embryo [3]. The Bicoid concentration gradient reaches its maximum value at the anterior pole [4] and is distributed steadily in an exponential gradient along the AP axis after one hour of development [4–6] (Fig 1A). Bicoid is a homeodomain transcription factor that binds DNA. Bicoid binding sites are found in the regulatory sequences of Bicoid target genes and are both necessary [7–9] and sufficient [10–12] for Bicoid-dependent expression. Changes in Bicoid dosage induce a shift of the expression boundary of these genes along the AP axis [9, 13] indicating that Bicoid provides concentration-dependent positional information to the system [14]. In young syncytial embryos (nuclear cycles 9 to 13), the major Bicoid target gene *hunchback* (*hb*) is expressed under Bicoid control in a large domain spanning the anterior half of the embryo [15]. At cycle 14, *hb* is also expressed in a narrow posterior domain [16] and in the parasegment 4 [17] and even later during development in the nervous system [18]. Expression of endogenous *hb* is initiated at two different promoters, the distal promoter P1 and the proximal promoter P2 [19]. P1 is responsible for maternal and late blastoderm expression in parasegment 4 and in the posterior stripe [16, 17]. P2 mediates the early Bcd-dependent expression of *hb* [17, 19]. At nuclear cycle 14, expression of endogenous *hb* is also controlled by two distal enhancers, the shadow enhancer and the stripe enhancer, which both contribute to the robustness of *hb* expression [20, 21].

Here, we focus exclusively on the most early zygotic expression of *hb*, which is Bicoid-dependent, driven by the P2 promoter and occurring from nc11 to nc13. Just 30 min after the onset of zygotic transcription and the steady establishment of the Bicoid gradient, endogenous

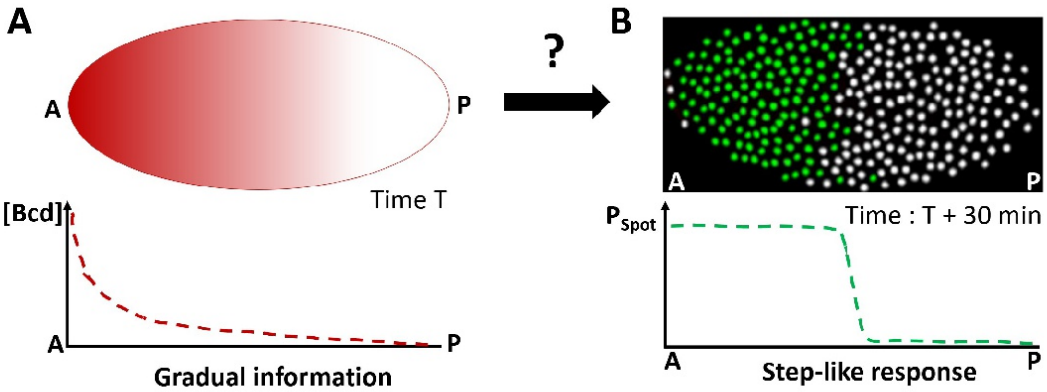


Fig 1. The Bicoid system transforms the gradual information contained in the Bicoid concentration gradient into a step-like response in 30 min. A) At nc 8 (T = 1 hr), the Bicoid exponential gradient (red) is steadily established [5, 6] with its highest concentration at the anterior pole (A) and its lowest concentration at the posterior pole (P). The first hints of zygotic transcription are detected by RNA FISH marking the onset of zygotic transcription [15]. **B)** At nc 11 (T = 1 hr 30 min), the main Bicoid target gene, *hunchback* (*hb*), is expressed within a large anterior expression domain. *hb* expression is schematized here from RNA FISH data [15]: nuclei where ongoing transcription at the *hb* loci is detected are shown in green and nuclei silent for *hb* are shown in white.

https://doi.org/10.1371/journal.pgen.1007676.g001

hb already exhibits a step-like expression pattern [15]. At this developmental time period, this pattern is characterized by an anterior domain containing almost exclusively *hb* transcriptionally active nuclei and a posterior domain containing exclusively *hb* silent nuclei [3] (Fig 1B). The boundary separating expressing and non-expressing nuclei is very steep despite the stochastic nature of transcription in eukaryotic cells [22], the short interphase duration (~5 min) and the subtle difference in the Bicoid concentration (10%) on either side of the boundary [15]. How the *hb* pattern so rapidly acquires such a steep boundary with high levels of expression in the whole anterior domain is unclear. It could involve a purely quantitative threshold-dependent process, in which Bicoid, acting as a direct transcription activator, is the main source of positional information. Alternatively, additional mechanisms including activation by maternal Hb [12, 15, 23] or posterior inhibitors such as those acting later during development at cycle 14 to set the position of the boundary along the AP axis [7, 24, 25] could be considered. In any case, the mechanism involved should account for a Bicoid dose-dependent effect on the positioning of the boundary along the AP axis. Also, given the early timing of development, most actors in this process are likely to be already present at the onset of zygotic transcription and therefore maternally provided.

To shed light on the formation of the *hb* expression boundary during these early steps of development, we have previously adapted the MS2-MCP approach to developing fly embryos [26]. This approach allows the fluorescent tagging of RNA in living cells and provides access to the transcription dynamics of an MS2 reporter locus [27, 28]. In our first attempt, we placed the *hb* P2 proximal promoter region (~ 750 bp) upstream of an MS2 cassette containing 24 MS2 loops and analyzed expression of this reporter (*hb-MS2*) in embryos expressing the MCP-GFP protein maternally [26]. The *hb-MS2* reporter was expressed very early as robustly as endogenous *hb* in the anterior half of the embryo. However, unlike endogenous *hb*, the reporter, which only encompasses the *hb* P2 proximal promoter region (~750 bp), was also expressed in the posterior, though more heterogeneously and more transiently than in the anterior. The different expression of the *hb-MS2* reporter and endogenous *hb* in the posterior suggested that, the 750 bp of the *hb* P2 proximal promoter were not sufficient to recapitulate

the endogenous expression of *hb* at the onset of zygotic transcription. The most obvious interpretation of these discrepancies was that the *hb-MS2* reporter was missing key cis-response elements allowing repression of endogenous *hb* by an unknown repressor in the posterior [26].

Here, we first show that the homogenously distributed Zelda transcription factor is responsible for the expression of the *hb-MS2* reporter in the posterior. BAC recombineering indicated that the MS2 cassette itself mediates Zelda posterior expression and *in silico* analysis reveals the presence of about 36 putative Zelda binding sites in the MS2 cassette. A new reporter (*hb-MS2ΔZelda*), placing the *hb* P2 proximal promoter (~750 bp) upstream of a new MS2 cassette (MS2-ΔZelda), in which those unfortunate Zelda binding sites have been mutated, faithfully recapitulates the early expression of endogenous *hb* observed by RNA FISH [15]. Thus, the *hb* P2 proximal promoter (~750 bp) is sufficient for a robust step-like expression of the reporter. Quantitative analysis of the MS2 time traces of this new reporter reveals a transcription process dividing the anterior of the embryo in a saturating zone with stable features and a limiting zone closer to the boundary, with more variable features. A high probability for the promoter to be ON (P_{ON}) is reached faster in the anterior where the concentration of Bicoid is higher than close to the boundary where Bicoid concentration is lower. In each interphase, full step-like response is established in not more than three minutes, after which the expression boundary is locked at a given position along the AP axis. To understand this observed dynamics, we used a simple model of position readout through the binding/unbinding of a transcription factor to N operator sites on the *hb* promoter [29]. The model that best fits the data is able to capture the very fast dynamics of establishment of the boundary. However, high steepness of the experimental pattern (coefficient of the fitted Hill function $N_{Hill} \sim 7$, based on various features of the MS2 time-traces) is not achievable assuming only $N = 6$ Bicoid binding sites, which is the number of known Bicoid binding sites in the canonical *hb* promoter [8]. This indicates that a simple equilibrium model only taking into account a single activator and varying degrees of cooperativity between the binding of several molecules on the *hb* promoter in the reporter, from which a steep *hb* pattern can emerge [29], is not sufficient to capture its activity and fit the data. It suggests that additional mechanisms are required to define the steepness of the boundary.

Results

Zelda induces posterior expression of the *hb-MS2* transgene at early cycles

In our first attempt of using the MS2 system to study the transcriptional response downstream of Bicoid, we placed the *hb* P2 proximal promoter region (~ 750 bp) upstream of a classical MS2 cassette [30]. The reporter (*hb-MS2*) was expressed in the posterior of the embryo and did not recapitulate expression of endogenous *hb* [26]. The Zelda transcription factor is a major regulator of the first wave of zygotic transcription in fruit fly embryos [31] and is involved in the transcriptional regulation of the *hb* gene [32]. To determine how Zelda contributes to the expression of the *hb-MS2* reporter, we analyzed expression of the reporter by live imaging (S1 Movie) and double RNA FISH (using *hunchback* and *MS2* probes, Fig 2A) in embryos expressing various amounts of maternal Zelda. As expected [26], wild-type embryos show high expression of the *hb-MS2* reporter both in the anterior and the posterior domain ([26] and Fig 2B, top panel), with an average of 80% of expressing nuclei dispersed through the posterior domain (Fig 2B, bottom panel). In embryos from *zelda* heterozygous mutant females ($zld^{Mat-/+}$), expression of the *hb-MS2* reporter is reduced by 5 fold in the posterior (with only 15% of active nuclei). In embryos from *zelda* mutant germline clones, completely devoid of Zelda maternal contribution ($zld^{Mat-/-}$), expression boundaries of endogenous *hb* and *hb-MS2* reporter are shifted towards the anterior (Fig 2D). Moreover, posterior expression of the *hb-*

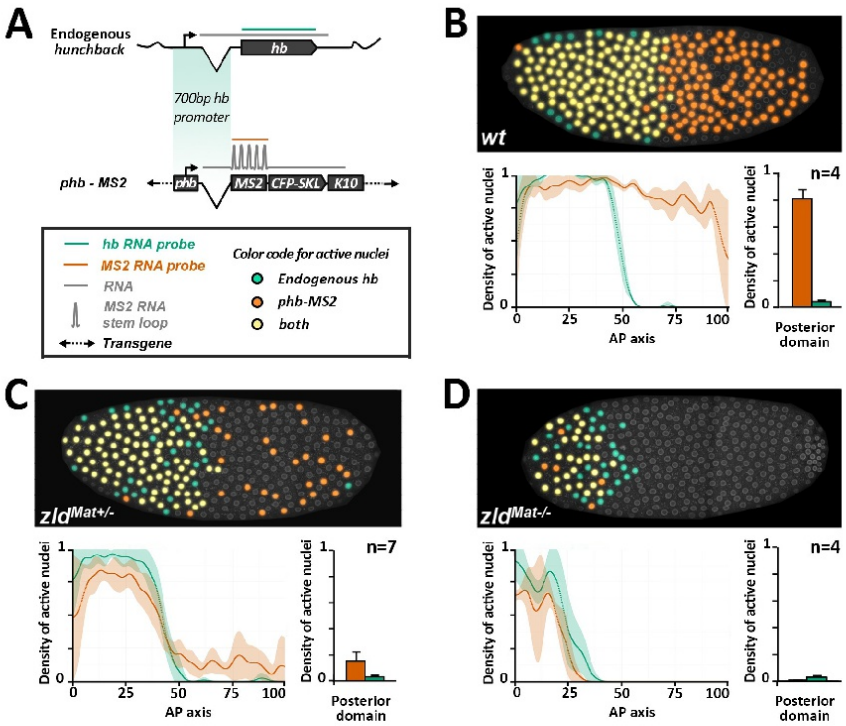


Fig 2. Posterior expression of the *hb-MS2* reporters is Zelda dependent. A) Dual RNA FISH using a *hb* probe (green) and an *MS2* (orange) probe on wild-type embryos carrying one copy of the *hb-MS2* reporter, placing the *MS2-CFP-SKL-K10* cassette under the control of the 750 bp canonical promoter of *hb* [26]. **B-D)** Embryos are wild-type (**B**), from heterozygous mutant females for Zelda (**C**) or germline mutant clones for Zelda (**D**). **Top panels:** Expression map of cycle 11 embryos after segmentation of nuclei and automated processing of FISH staining. Nuclei expressing only *hb* are labelled in green, nuclei expressing only the *hb-MS2* reporter are labelled in orange and nuclei expressing both *hb* and the *hb-MS2* reporter are labelled in yellow. **Bottom panels:** On the left, density of active nuclei for either *hb* (green) or the *hb-MS2* reporter (orange) along the AP axis with the anterior pole on the left (0) and the posterior pole on the right (100). On the right, density of active nuclei for the *hb-MS2* reporter (orange) and *hb* (green) in the posterior domain. For each embryo, the position of the expression boundary is calculated as the position of the maximal derivative of the active nuclei density curve. Mean values were calculated for n embryos and error bars correspond to standard deviation.

https://doi.org/10.1371/journal.pgen.1007676.g002

MS2 reporter is reduced to less than 1% of posterior nuclei (Fig 2D). We confirm thus that Zelda maternal proteins contribute to the early expression of endogenous *hb* [32] and conclude that posterior expression of the *hb-MS2* reporter in early embryos is mostly due to Zelda.

The *MS2* cassette carries the cis-acting sequence responsible for *hb-MS2* posterior expression

To understand the discrepancies of expression between endogenous *hb* and the *hb-MS2* reporter at early cycles, we first aimed at identifying the minimal sequence sufficient to recapitulate early expression of the *hb* locus. While a transgene carrying 18 kb of the *hb* locus was shown to recapitulate *hb* anterior expression at nc13 and nc14, its expression at earlier cycles had not been documented [20, 21, 33]. RNA FISH, using a *hb* probe on embryos carrying a

single insertion of the *hb*-18kb BAC, reveals ongoing transcription at the three *hb*-encoded loci and indicates that expression of the *hb*-18kb BAC and endogenous *hb* largely overlap in the majority of anterior nuclei (Fig 3B). No RNA FISH signals are detected within the posterior domain of these embryos indicating that the *hb*-18kb BAC encompasses all the regulatory sequences to spatially control *hb* expression during early nuclear cycles.

Taking advantage of BAC recombineering, we generated *MS2-hb-18kb* transgenes carrying insertions of the MS2 cassette either in the 5'UTR within the intron of *hb* (*5'MS2-hb-18kb*) or in the 3'UTR of *hb* (*3'MS2-hb-18kb*). Expression of these new MS2 reporters was assessed by live imaging (S2 Movie) and double RNA FISH using a *hb* probe and an *MS2* probe. In the anterior of nc11 embryos either homozygous for the *5'MS2-hb-18kb* transgene (Fig 3C) or heterozygous for the *3'MS2-hb-18kb* transgene (Fig 3D), the *hb* probe reveals at most four spots of ongoing transcription : one at each of the two endogenous *hb* and two (C) or one (D) at the *MS2-hb-18kb* loci. In most cases, two of the *hb* spots co-localize with two (C) or one (D) *MS2* spot(s), which specifically label ongoing transcription at the *MS2-hb-18kb* loci. In these embryos, we also detect *hb* and *MS2* spots in posterior nuclei (Fig 3C and 3D), which co-localize for most of them, thus revealing ongoing transcription at the *MS2-hb-18kb* loci. Posterior expression of the *MS2-hb-18kb* loci is also detected in living embryos expressing the MCP-GFP (S2 Movie). Altogether, these data strongly argue that posterior expression of the *MS2-hb-18kb* transgenes is mediated by the MS2 cassette and suggest that Zelda-dependent posterior expression of the *hb-MS2* reporter is mediated by cis-acting sequences in the MS2 cassette.

The *hb* canonical promoter is sufficient to recapitulate early zygotic expression of *hb*

Given the trans-acting effect of Zelda on the posterior expression of the *hb-MS2* reporter (Fig 2) and the enhancer-like behavior of the MS2 cassette for posterior expression (Fig 3), we searched for potential Zelda binding sites in the MS2 sequence, using the ClusterDraw2 online algorithm [16] and Zelda position weight matrix [34]. The canonical Zelda binding site is a heptameric motif (CAGGTAG, Fig 4A) over represented in the enhancers of pre-cellular blastoderm genes [32, 35]. Strikingly, in the sequence of the MS2 cassette we find the motif CAGGTCG (a single mismatch with the canonical Zelda site) repeated 12 times and the motifs TAGGTAC (two mismatches) and TAGGCAA (three mismatches) each repeated 12 times (Fig 4B). This *in silico* analysis indicates that the MS2 sequence contained a total of 36 potential Zelda binding sites all located within linkers between MS2 loops (Fig 4B). Although we cannot be conclusive about the affinity strength of these various binding sites for Zelda, all of them share high similarity with TAGteam motifs [32]. We thus engineered a new MS2 cassette mutating the 36 putative Zelda binding sites of our original *hb-MS2* reporter and inserted this new MS2-ΔZelda cassette under the control of the *hb* canonical promoter as in the original *hb-MS2* reporter. Expression of this new *hb-MS2ΔZelda* reporter was assessed in living embryos expressing the MCP-GFP protein (S3 Movie). We do not detect any MS2 spots in the posterior of the embryo (Fig 5A) indicating that unlike our original *hb-MS2* reporter [26], the new *hb-MS2ΔZelda* reporter is expressed exclusively in the anterior at early cycles 10 to 11 as detected for endogenous *hunchback* by RNA FISH [3].

As RNA FISH are performed on fixed embryos whereas the MS2 data are obtained from live material, we wondered whether these two different approaches provide consistent quantification of *hb* transcriptional activity when focusing on the "steepness" of the expression boundary. Therefore, the probability to be active for *hb-MS2ΔZelda* loci at a given position along the AP axis and at a given time ($P_{SPOT}(t)$) was extracted from movie snapshots and compared to RNA FISH data of endogenous *hb* expression [3]. To compare data from several

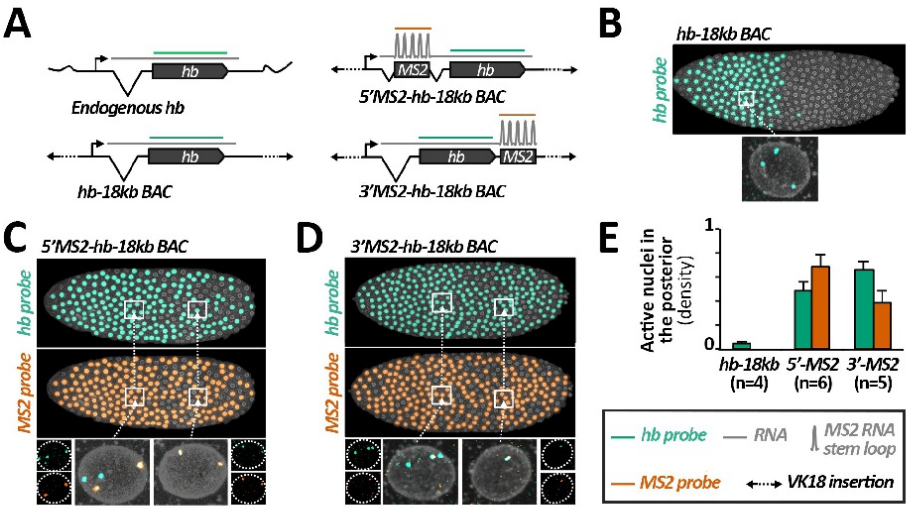


Fig 3. Posterior expression of the *hb-MS2* reporters is mediated *in cis* by the MS2 cassette. A) Structures of the *hb* expressing loci and relative positioning of the *hb* probe (green) or the *MS2* probe (orange) hybridizing to the transcripts. The *hb-18kb BAC* encompasses 18 kb of chromosome III spanning the *hb* locus. The 1.3kb MS2 cassette is inserted in the *hb-18kb BAC* within the intron of *hb*, 0.7kb downstream the transcription start site (*5'MS2-hb-18kb* BAC, top right) or within the 3'UTR of *hb*, ~ 3 kb downstream of the transcription start site (*3'MS2-hb-18kb* BAC, bottom right). All the BACs are inserted at the same position (VK18, chromosome II) within the fly genome. **B)** Top: Expression map of a typical nc11 embryo heterozygous for the *hb-18kb BAC* insertion. As highlighted on one example (bottom), nuclei in the anterior exhibit three sites of ongoing transcription detected with the *hb* probe: two of these sites correspond to expression at the endogenous *hb* loci and one of them corresponds to expression at the *hb-18kb* BAC locus. No expression is detected in the posterior. **(C-D)** Embryos are homozygous for the *5'MS2-hb-18kb* BAC insertion (C) or heterozygous for the *3'MS2-hb-18kb* BAC insertion (D). Top: Expression map of a typical nc11 embryo detected with the *hb* probe (green) and with the *MS2* probe (orange). Bottom: close-up of single nuclei in the anterior (left) and the posterior (right). **E)** Mean density of active nuclei in the posterior detected by the *hb* probe (green) or by the *MS2* probe (orange) on nc11 embryos carrying the *hb-18kb BAC* (*hb-18kb*), the *5'MS2-hb-18kb* BAC (*5'MS2*) or the *3'MS2-hb-18kb* BAC (*3'MS2*) insertion. Mean values are calculated for n embryos and error bars correspond to standard deviation.

https://doi.org/10.1371/journal.pgen.1007676.g003

embryos, embryos were aligned fixing the origin of the AP axis when the probability for a locus to experience transcription at any time during the interphase (P_{ON}) is equal to 0.5 (for the definition of the boundary see details in S3 Text). This embryo alignment allows us to

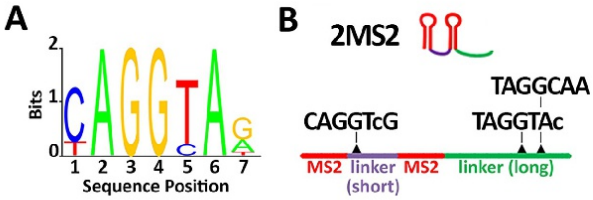


Fig 4. Zelda putative binding sites in the MS2 cassette. A) Position weight matrix of the Zelda binding site [34]. The critical positions in the canonical Zelda site C**AGG**TAG are highlighted in bold. **B)** The 24 MS2 stem loop repeat corresponds to 12 tandem repetitions of the two MS2 repeats shown. In this sequence, the MS2 loops (red) are separated by either a short (purple) or a long (green) linker. The short linker contains one CAGGTCG sequence which harbors a single but critical mismatch with the canonical Zelda site. The long linker contains one TAGGTAC and one TAGGCAA which harbor respectively two or three permissive (not critical) mismatches with the canonical Zelda site.

https://doi.org/10.1371/journal.pgen.1007676.g004

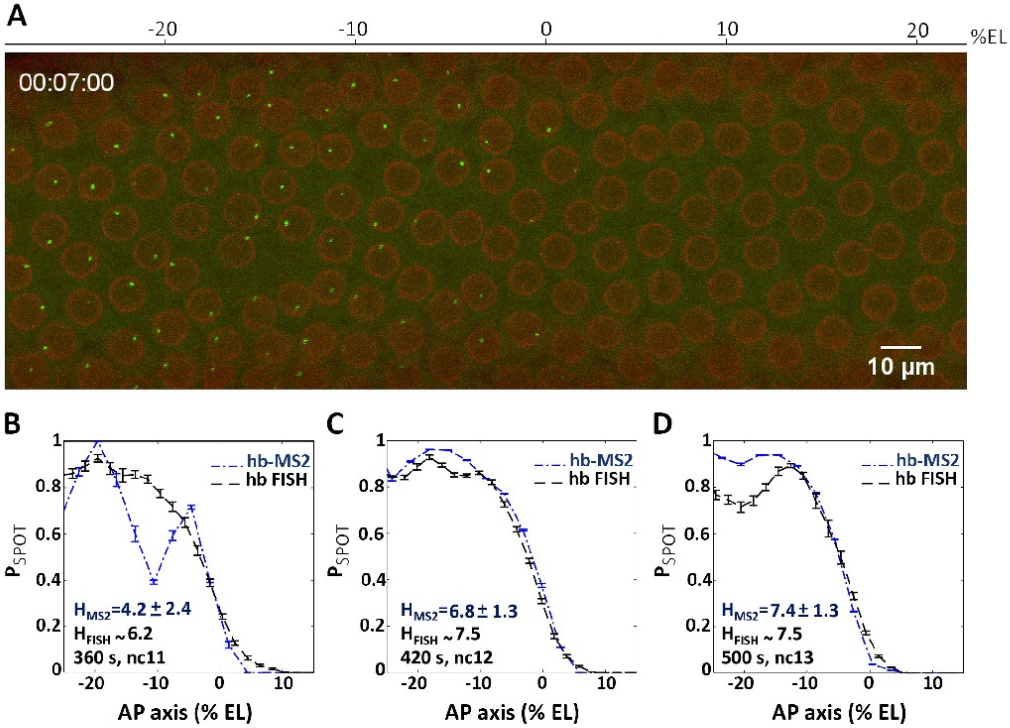


Fig 5. The *hb* canonical promoter expresses the *MS2ΔZelda* cassette in an anterior domain with a steep posterior boundary. (A) A 2D maximum projection snapshot from Movie3 (SI) was taken at ~7 minutes after the onset of nc12 interphase. In the green channel, MCP-GFP proteins recruited by the nascent MS2-containing mRNA can be seen accumulating at the *hb-MS2ΔZelda* loci (bright spots). In the red channel, the mRFP-Nup proteins localized at the nuclear envelope delineate nuclei. (B-D) Probability of active *hb-MS2ΔZelda* loci along the AP axis (dashed blue lines with error bars), extracted from snapshots of 6 movies near the end of each nuclear cycle: 360 s after the onset of nc11 interphase (B), 420 s after the onset of nc12 interphase (C) and 500 s after the onset of nc13 interphase (D). In each panel the probability of active endogenous *hb* loci (black dashed lines with error bars) extracted from the FISH data from [3] is also shown. Hill coefficients (H) are indicated in blue (*hb-MS2* reporter) or in black (endogenous *hb* from FISH data). In B, the difference between the FISH signal and the probability of active *hb-MS2ΔZelda* loci is due to small number of nuclei and large variability distance between them at nc11 which limit statistics.

https://doi.org/10.1371/journal.pgen.1007676.g005

compensate for input noise (variability in the Bicoid gradient), which was shown to be of about 2–3% EL [6], and focus only on the output noise (noise in the transcriptional response). Throughout the paper, we refer to average measures of expression by computing the probability of the locus to be ON as a function of position along the AP axis and as a function of time. We named this time-dependent probability to be ON, P_{SPOT} (t). P_{SPOT}(t) is thus a measure of the instantaneous gene activity at a given position along the AP axis and at a given time within the nuclear cycle. We first assigned a value of 1 to a nucleus that expresses the MS2-MCP gene above a certain threshold (see methods) at any time during the nuclear cycle. P_{SPOT}(t) is calculated by averaging over nuclei at a given position along the AP axis. We also use a cumulative P_{SPOT} named P_{ON}, which is used only for embryo alignment and which is a cumulative statistic per nuclear cycle. P_{ON} indicates the probability for a nucleus at a given position to experience transcription during nuclear cycle. A value of 1 is assigned to each nucleus experiencing transcription during the cycle

and P_{ON} is obtained by averaging over the nuclei at a given position. $P_{SPOT}(t)$ can decrease within a nuclear cycle if nuclei stop expressing, whereas P_{ON} cannot decrease.

As shown in Fig 5B–5D, the curves plotting the mean spot appearance $P_{SPOT}(t)$ as a function of position along the AP axis are similar when extracted either from the RNA FISH data (dashed line, [3]) or the MS2 movie snapshots (blue lines) with Hill coefficients varying from 4.5 (nc11) to ~ 7 at nc12 and nc13. Thus, the *hb-MS2ΔZelda* reporter is expressed at early cycles in an anterior domain with a boundary as steep as the boundary of the endogenous *hb* expression domain. These experiments show that the *hb* canonical promoter is sufficient to faithfully recapitulate the early zygotic expression of endogenous *hb*. This is in contradiction with our first interpretation [26] deduced from the expression of the original *hb-MS2* misleading reporter with its high number of unexpected Zelda binding sites. Nevertheless, the now almost perfect match between $P_{SPOT}(t)$ activity along the AP axis of the FISH data of endogenous *hb* expression and of the data extracted from movies of the *hb-MS2ΔZelda* reporter at the end of the interphase, ensures that the analysis of this reporter expression dynamics can help understand how the step-like expression of *hb* arises.

The *hb-MS2ΔZelda* reporter shows varying transcription kinetics in the anterior domain

The MS2 movies provide access to the transcription dynamics of each *hb-MS2 ΔZelda* single locus in all the visible nuclei of developing embryos. Key features are extracted from the time traces of the MS2-GFP spots (Fig 6A), choosing as the origin for time the onset of interphase for that particular nucleus (see details in S2 Text and S4 Fig). These features include: *i)* the initiation time (t_{init}) which measures the time period from the onset of interphase to the first detection of the MS2 transcription signal at the locus, *ii)* the time period during which the locus is activated (t_{active}) and *iii)* the time period at the end of interphase during which the locus is turned off (t_{end}). From the time traces (Fig 6A), the integral activity (ΣI) integrates the area under the trace and provides a relative measure of the total amount of mRNA produced. The average mRNA production rate (μI) is calculated by dividing ΣI by t_{active}. Features are obtained from 5 (nc11), 8 (nc12) and 4 (nc13) embryos. Embryos were aligned spatially fixing the origin of the axis at boundary position (P_{ON}) at nc12 and the origin of time was calculated for each nuclei as the origin of the respective cycle (see S2 Text and S4 Fig). As shown in Fig 6, several of these features exhibit different behaviors depending of the position along the AP axis. Notably, t_{init} and ΣI appear more variable when loci are located close to the boundary than in the most anterior part (Fig 6B–6G). Similarly, in a region of about 10% EL at the boundary, the mRNA production rate drops from the constant value reached in the anterior region to 0 at the boundary (Fig 6H–6J). Thus, the dynamics of the transcription process at the *hb-MS2ΔZelda* reporter exhibit two distinct behaviors: in the anterior of the embryo, time trace features are stable likely reflecting a maximum PolII loading rate and saturated levels of Bicoid; in a region of ~ 10% EL anterior to the expression boundary, time trace features are fluctuating reflecting limiting amounts of Bicoid. The pattern steepness corresponds to a Hill coefficient of ~7 to 8 of the regulation function (see S3 Text). The MS2 live imaging gives us the opportunity to really decipher the dynamics of the *hb* expression compared to the FISH that was just giving us $P_{SPOT}(t)$ for one arbitrary time point (probability for *hb* loci to be active at a given position along the AP axis and at the time of embryo fixation).

At each nuclear cycle, the *hb* expression pattern reaches steady-state in 3 minutes

The temporal dynamics of the *hb* pattern establishment at the scale of the whole embryo (along the AP axis) were extracted from the MS2 movies: the probability for the locus to be

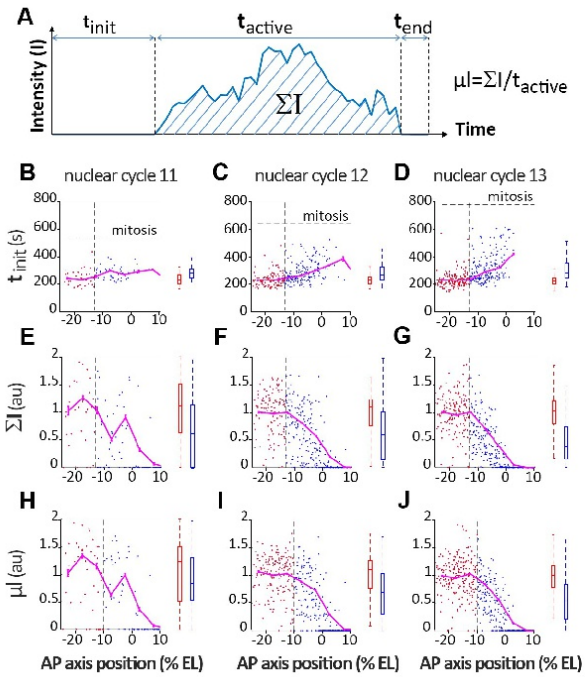


Fig 6. Distribution of time trace features along AP axis. (A) Description of the time trace features. Initiation time (t_{init}, B-D), normalized integral spot intensity (ΣI, E-G) and production rate (μI, H-J) are indicated as a function of position along the AP axis with the origin fixed for each embryo at the position of the expression boundary at nc12. Embryos were at nc11 (B, E and H), nc12 (C, F and I) and nc13 (D, G and J). The vertical dashed line indicates when the significant change in the feature distribution from the anterior pole is first detected in nc13, using the Kolmogorov-Smirnoff test (p-value 0.05). This line separates the *hb* expression domain into 2 zones: the anterior zone which exhibits a saturating process (red box) and the boundary zone which exhibit a more stochastic activity characterized by more variability (blue box). On the right of each panel (B-J), the distribution of the two populations is shown (median ± 25% and min/max). Example of MS2-MCP time traces are given in S1 Fig, S2 Fig and S3 Fig. Data were obtained from 5 (nc11), 8 (nc12) and 4 (nc13) embryos. All nc11 and nc12 traces were recorded for the whole duration of the cycle. Embryos were aligned spatially fixing the origin of the axis at the boundary position (P_{ON} = 0.5) at nc12 and the origin of time was calculated for each nuclei as the origin of the respective nuclear cycle (see S2 Text and S4 Fig). For t_{init}, the data points are for expressing nuclei only, whereas for ΣI and μI both expressing and non-expressing nuclei were taken into account.

https://doi.org/10.1371/journal.pgen.1007676.g006

ON as a function of the position along the AP axis and time in the cycle ($P_{SPOT}(t)$) can be visualized in S4 Movie and is plotted in the form of kymographs on pulled embryos as described above (Fig 7A–7C). At each cycle, spots can appear as early as ~ 150 s after mitosis. Given the interruption of transcription during mitosis, this limit of 150 s corresponds to the period required to re-establish transcription during the interphase and likely includes genome decondensation, the time it takes for the Bicoid protein to be imported in the nucleus after mitosis, and the time it takes for the MS2 system to produce a signal that is above background. After this period, expression rapidly turns on in the anterior to reach the plateau value where all nuclei express the reporter ($P_{SPOT}(t)$ ~ 1). The time period to reach the plateau is shorter at the most anterior position in the field of view imposed by the movie recording (~ - 25% EL)

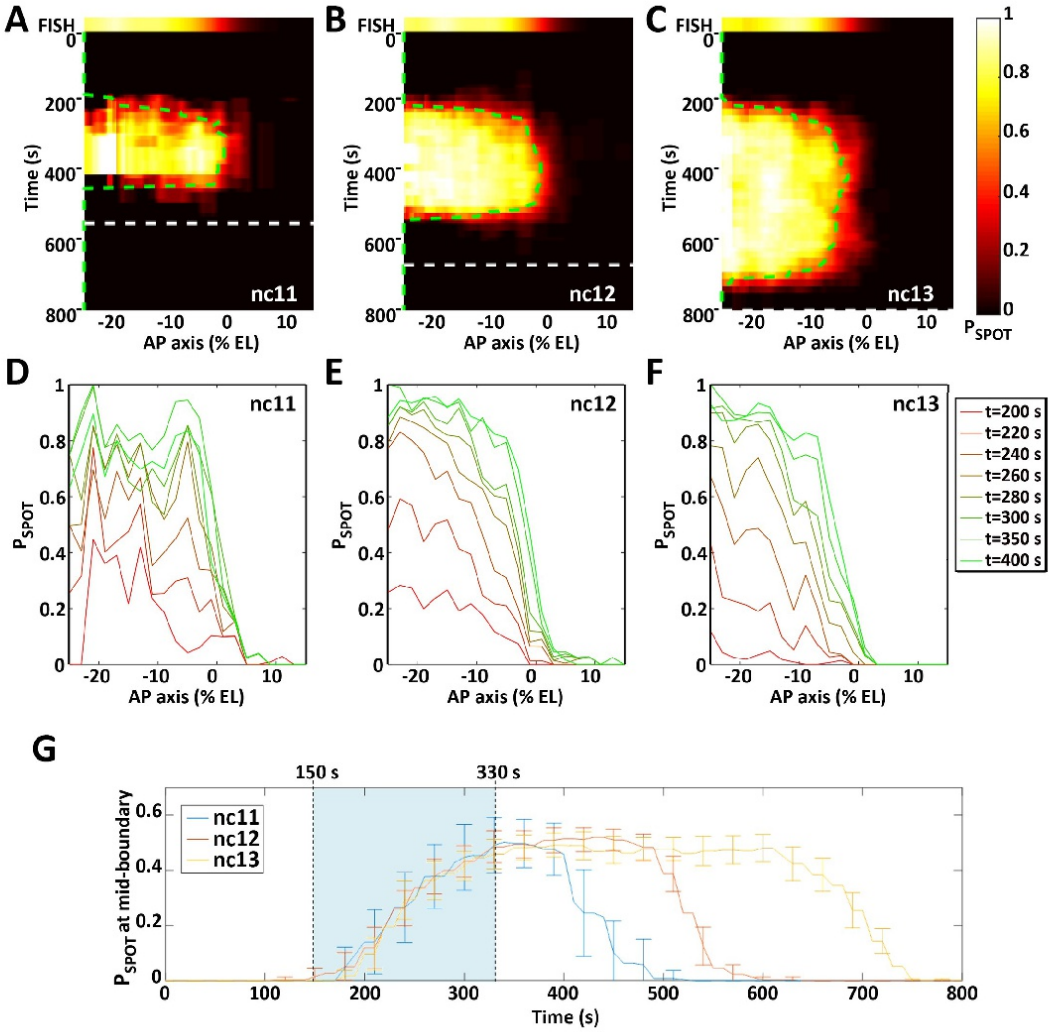


Fig 7. The dynamics of *hb-MS2ΔZelda* expression pattern. A-C: The probability for a given locus to be ON ($P_{SPOT}(t)$) is indicated by a heat map (color scale on the right where $P_{SPOT}(t) = 0$ is black and $P_{SPOT}(t) = 1$ is white) horizontally as a function of position along the AP axis (0% EL positioned where $P_{ON} = 0.5$ at nc12) and vertically as a function of time (s) fixing the origin at the onset of interphase for each nucleus (see details in S2 Text and S4 Fig). The top panel shows the endogenous *hb* expression pattern as measured from RNA FISH (15). For each cycle (**A**: nc11; **B**: nc12; **C**: nc13), the end of interphase (onset of the next mitosis) is indicated by a dashed line (white). The green dashed line indicates the position of the expression boundary ($P_{SPOT}(t) = 0.5$) over time. **D-F:** $P_{SPOT}(t)$ as a function of position along the AP axis for different times during the interphase of nc11 (**D**), nc12 (**E**) and nc13 (**F**). **G:** $P_{SPOT}(t)$ as a function of time (s) at the mid-boundary position (where $P_{SPOT}(t)$ reaches a steady value of 0.5). The first hints of transcription are observed at the mid-boundary position ~ 150 s after the onset of the interphase (lower limit of the light blue zone) and steady state is reached at ~ 330 s (higher limit of the light blue zone). The boundary formation reaches steady state in ~ 180 s. Data were obtained from 5 (nc11), 8 (nc12) and 4 (nc13) embryos. Embryos were aligned spatially by fixing the origin of the axis (0% EL) at the boundary position ($P_{ON} = 0.5$) at nc12 and the origin of time was calculated for each nuclei as the origin of the respective nuclear cycle (see S2 Text and S4 Fig).

https://doi.org/10.1371/journal.pgen.1007676.g007

and increases with the distance to the anterior pole (Fig 7D–7F) consistent with a dose-dependent activation process of *hb* by Bicoid [4, 8]. Then, the expression pattern is steadily established, which includes fixing the boundary position and steepness (see S4 Text for measurement details). This stable state lasts for a period of time which varies with the length of the cycle until *hb* expression rapidly gets switched off simultaneously at all positions along the AP axis: at 400 s at nc11 (Fig 7A & 7G), 500 s at nc12 (Fig 7B & 7G) and 750 s at nc13 (Fig 7C & 7G). Importantly, the dynamics of the boundary positioning is the same at the three nuclear cycles considered (Fig 7G) and the steady state of boundary positioning is reached rapidly: ~330 s after the onset of the cycle corresponding to ~ 180 s after the first hints of transcription at this position (Fig 7G). Dynamics of pattern steepness and average spot intensity also exhibit similar behaviors (see S5 Text and S9 Fig, respectively). Thus, expression of the *hb-MS2ΔZelda* reporter allows us to directly observe position-dependent transcriptional activation, which is consistent with Bicoid dose-dependent transcriptional activation. Also, it demonstrates that the steady-state of positional measurement is reached in no more than 3 min at each cycle with very similar dynamics between cycles. Once the steady state is reached, the *hb* boundary is fixed around the position ~ -5% EL.

Modeling the Bicoid-mediated dynamics of the *hb* expression pattern

To better understand how the stable response downstream of Bicoid is achieved, we build a stochastic model of *hb* expression regulation by the Bicoid transcription factor (TF), coupled with a stochastic transcription initiation process assuming random arrival of RNA polymerases when the gene is activated. The mechanism of *hb* expression regulation through the cooperative binding of multiple TFs to the promoter was originally proposed in 1989 to explain how the shallow Bicoid gradient could give rise to an expression pattern with a steep boundary [8, 9]. It was subsequently proposed [36] that within an equilibrium binding model, a pattern steepness quantified by a Hill coefficient of N requires a promoter with at least N TF binding sites. Here, we consider a model with N TF binding sites, where the promoter state P_i is decribed by the probability of having i TF bound at a given time (Fig 8A). Transitions between the states occur via binding and unbinding events of TF to the sites. We assume an equilibrium binding model, in which all transitions are reversible and the binding sites are identical. As a result the state of the promoter is described solely by the number of bound TF, and not their position on the *cis* regulatory array. We assume that gene expression is activated only when all N binding sites are bound by TF. In the regime of parameters corresponding to a high pattern steepness, this "all-or-nothing" assumption is shown to have very limited impact on the pattern dynamics (see [29] and our companion paper, [37]). Following this activation, RNA polymerases can arrive to the promoter in a Poisson process and initiate the downstream transcription initiation process. Our simplified model does not account for a non constant polymerase arrival rate [38].

We investigate the *hb* pattern dynamics by solving a stochastic time dependent master equation (Eq 7 in S6 Text) and considering binding rates that vary as a function of TF concentration *[TF]*: at the boundary position, TF concentrations are lower than in the anterior region [5, 39] and it takes more time for a diffusing TF to reach binding sites at the promoter than in the anterior region where TF concentration is higher. The value of the binding rate constants is further limited by the value of the diffusion coefficient (see S6 Text, 1). In our model, the cooperativity between the binding of TFs is modeled implicitly by the value of the unbinding rates of the TF to the binding sites: higher cooperativity corresponds to higher stability of the promoter state with N bound TF, which is modeled through a lower unbinding rate. The unbinding rate is limited at the boundary so that P_{SPOT} (t) = 1/2. Meanwhile, the unbinding

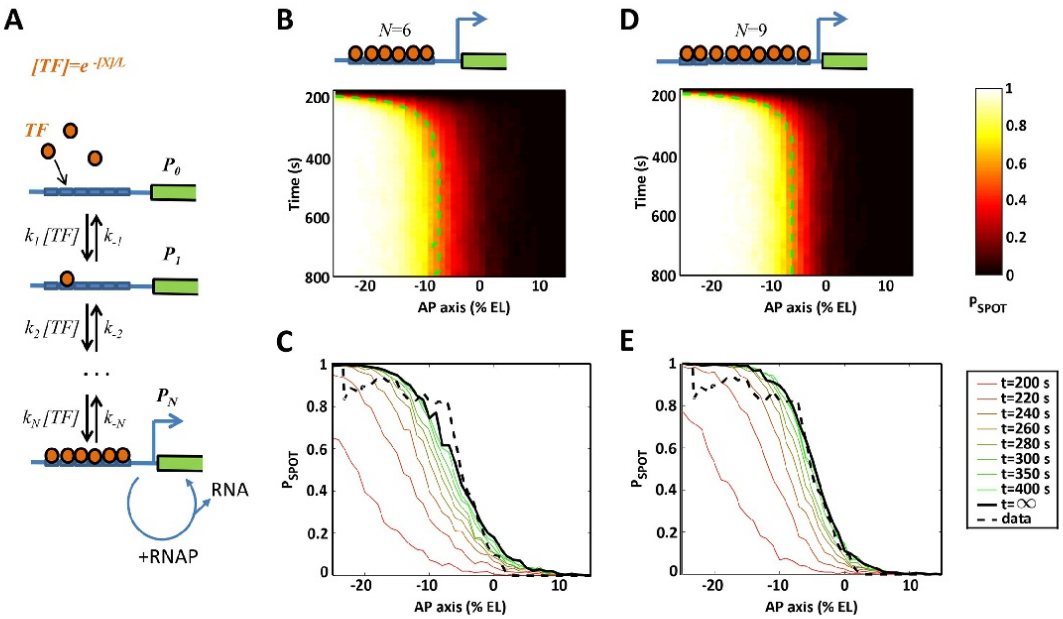


Fig 8. Modeling transcriptional regulation by the Bicoid transcription factor through interactions with the *hb* promoter operator sites. A) A model of regulation by Bicoid transcription factor (TF) binding to multiple binding sites on the *hb* promoter coupled with stochastic transcription initiation. Transcription initiation is allowed only when the binding sites are fully bound. During this window, RNAP can randomly bind to the promoter and initiate transcription to produce mRNA. **B or D**) The model prediction for the probability of an active transcription locus ($P_{SPOT}(t)$, colorbar) as a function of time in the nuclear cycle and position along the AP axis for a model with 6 (**B**) or 9 (**D**) Bicoid binding sites. **C or E**) The simulated pattern evolution of $P_{SPOT}(t)$ along the AP axis over time (colored line), shown with the pattern predicted at steady state (solid black line) and the stable pattern extracted from the data in nc13 (dashed black line, as in Fig 7F). Panels **C** and **E** represent cuts in time of panels **B** and **D**, respectively. The kinetic parameters were chosen so as to match the observed *hb* pattern steepness and formation time at the boundary ($P_{SPOT}(t) = 0.5$) (see S2 Table). The value of $P_{SPOT}(t)$ is calculated from 200 trajectories per AP position.

https://doi.org/10.1371/journal.pgen.1007676.g008

rate constants are kept constant along the AP axis (independant of Bicoid concentration) but differ depending on the occupancy state of the promoter to account for copperativity. At time t = 0, all binding sites are free of TF.

Motivated by the 6 known Bcd binding sites on the *hb* promoter [8, 17], we first consider a model with 6 binding sites ($N = 6$). The binding and unbinding rates are chosen (Eq 5 in S6 Text) to achieve the closest steepness and establishment period to the *hb* pattern observed in the movies (Fig 7A–7C). The model fits are performed on data pulled from all embryos and nuclear cycles. The gene expression pattern dynamics, shown as the probability ($P_{SPOT}(t)$) of a nucleus having an active locus (bright spot) as a function of position along the AP axis and as a function of time during the nuclear cycle, shows a good qualitative agreement between the model and the data. In Fig 8B, similarly to Fig 7A–7C, initially there is no expression since the TF are not bound to the promoter. Active transcription loci first appear near the anterior pole, where the activator concentration is the highest, and then transcription activation propagates quickly to the mid-anterior region. After a certain amount of time, the steep expression pattern becomes stable, with the boundary (dashed green line) located near the mid-embryo region where the *hb* boundary is located. However, using the canonical number of binding sites

($N = 6$) results in a steepness of the boundary lower than that observed in the experimental data (Fig 8C). Increasing the binding sites number from 6 to 7, 8, 9 or 10 allows us to match the observed boundary steepness to the experimental data (H~7). Yet, we have shown in a companion manuscript [37] that the steeper the patterns are, the longer the time needed for them to establish. Specifically, obtaining the observed Hill coefficient of H~7 in 3 minutes is not possible with 7 or 8 binding sites but requires 9 binding sites (Fig 8E). Increasing N beyond 9 did not result in a significantly better fit (see S6 Text, 5).

Discussion

In this study we show that the removal of 36 putative binding sites for the transcription factor Zelda (unfortunately present in the sequence of the MS2 cassette) reveals a new temporal dynamics of the *hb* canonical promoter at the onset of zygotic transcription. Unlike our original *hb-MS2* reporter [26], the new *hb-MS2ΔZelda* reporter faithfully reproduces the early zygotic expression of the endogenous *hb* observed with RNA FISH [3, 26]. It indicates that the 750 bp of the *hb* locus, including 300 bp of the proximal enhancer, the P2 promoter and the intron, are sufficient to reproduce the endogenous expression of the *hb* gene in the early nuclear cycles (11 to 13). The dynamics of establishment of the *hb* pattern at these early stages of development is thus properly captured by the new *hb-MS2ΔZelda* reporter.

These MS2 movies provide access to the *hb* pattern dynamics, which was not perceivable in previous *in situ* experiments on fixed embryos. *hb* expression first occurs in the anterior then proceeds to the boundary region. A difference in the activation time (initiation time) following mitosis is observed even within the anterior region, allowing us to visualize for the first time position-dependent activation of *hb* and thus likely dose-dependent activation by Bicoid. Analysis of the MS2 time traces indicates that the transcription process is more variable among nuclei at the boundary of the expression domain, where Bicoid concentration is low and probably limiting, than in the anterior where the concentration of Bicoid is high. Tailor-made analysis of the time traces allowed us to extract different kinetic parameters of promoter activity in these two regions and to demonstrate that transcription of the *hb-MS2ΔZelda* reporter is bursty (described by a two state model) in both regions [40]. This anti-correlation between the relative variability of mRNA production among nuclei and the Bicoid concentration (position along the AP axis) supports the idea that Bicoid interactions with the *hb* promoter are rate-limiting processes contributing to "bursty" transcription. Nevertheless, rate-limiting interaction of Bicoid with DNA is not the sole factor contributing to bursty transcription, as it is also observed in the anterior region with very high Bicoid concentration [40]. Thus, despite extremely fast transcription initiation imposed by the 5 min interphase and frequent mitoses, bursty transcription is clearly observed at the onset of zygotic transcription in fly embryos.

The MS2 movies indicate that the *hb* boundary is established within 3 minutes at each nuclear cycle with a very high steepness (H ~7). How the steepness and positioning of the boundary are reached so rapidly is unclear. Our data (Fig 6 and Fig 7) indicate that it takes more time for *hb* expression to reach steady-state levels in the boundary region and that Bicoid is thus likely to be a rate-limiting factor in the formation of the *hb* pattern around the boundary region. The time scale of interactions (i.e. binding and unbinding) of Bicoid with the *hb* promoter is critical in determining how quickly the accurate *hb* response is established. Therefore, the interactions between Bicoid molecules and the *hb* promoter need to be modeled explicitly, rather than implicitly. We propose an equilibrium model of transcription regulation via the binding/unbinding of transcription factors to the operator sites of the target promoter.

This model can account for various types and degrees of cooperativity between the binding of TFs, from which a steep *hb* pattern can emerge [29].

Using this model, we show in a tandem paper that accommodating a steep gene expression pattern within the considered model requires very slow promoter dynamics and thus would result in a very long pattern formation time [37]. Considering this trade-off, the steepness and the pattern formation time observed from the movies are even more intriguing. The most relevant model with 6 binding sites, motivated by previous work [8], fits well the pattern dynamics but fails to reproduce the observed steepness of the boundary in such a short time period of 3 min. The failure of the 6 binding sites model to completely reproduce the experimental data indicates that the assumption that 6 binding sites for Bicoid are sufficient for the observed response is wrong and that additional mechanisms have to be included in the model to enhance the steepness of the boundary to the observed level. It was recently proposed that energy expenditure, encoded as non-equilibrium binding of the TF, can allow a Hill coefficient greater than the number of binding sites (up to 11 with 6 sites) [29]. Alternatively, different regulatory scenarios could play a role. First, we found that increasing the number of binding sites up to 9 allows a significantly better fit of the model with the data than with 6 binding sites. It is thus possible that the 750 bp of the *hb* gene that are sufficient to elicit the pattern contain more than 6 Bicoid binding sites. As recently pointed out, the importance of low affinity binding sites might be critical to confer specificity and robustness in expression [37] and a closer analysis of the *hb* regulatory sequence looking for potential weak binding sites for Bicoid might clarify this point. A second possibility is the involvement of other transcription factors distributed as gradients that could bind to the *hb* promoter and contribute to the increase in the steepness of the boundary. The Hb maternal protein, which is also expressed as an anterior to posterior gradient and able to bind to the endogenous *hb* promoter [38], contributes to the *hb* expression process by allowing expression at lower Bicoid concentration thresholds [12] and faster activation [15]. Although this has not yet been investigated, maternal Hb might also contribute to sharpen the *hb* boundary in a timely manner. Alternatively, maternal repressors expressed as gradients in the posterior region or downstream of the boundary could also contribute to the steepness of the *hb* boundary. Among potential candidates are Caudal, which is expressed as a posterior to anterior gradient but so far has been described as a transcriptional activator in fly embryos [41], or Capicua, a transcriptional repressor at work in the center of the AP axis, where the *hb* boundary forms [7].

From the movies, the dynamics of establishment (before reaching steady state) of the *hb* pattern appears to be invariant at the three nuclear cycles considered (nc11, nc12 and nc13). At these three nuclear cycles, it takes ~180 s from the detection of the first MS2-MCP spots to the establishment of the pattern near the mid-embryo position (Fig 7G). This invariance in the dynamics of establishment indicates that there are no dramatic changes in the regulation of the transcription process during the three cycles suggesting that Bicoid remains one of the main patterning factors of *hb* transcription dynamics at the boundary region in these stages of development. The other transcription factors involved in *hb* expression [15, 19], if any, would need to be stably maintained during the three cycles. As mentioned above, one possible factor is the Hb protein itself, which was shown to contribute to the expression of the *hb* expression pattern with both a maternal and a zygotic contribution [12, 15]. A likely hypothesis is that up to nc13, the zygotic Hb protein production is balanced by maternal Hb protein degradation, thus leading to a stabilized Hb gradient over the three considered cycles.

Previous observations of Bicoid diffusion in the nucleus space pointed out that it requires at least 25 minutes for a single Bicoid binding site to sense the Bicoid concentration with 10% accuracy, a level observed using FISH and protein staining experiments [5, 15, 39]. This estimation is based on the Berg-Purcell limit in the precision of concentration sensing of diffusing

molecules via surface receptors [42–44]. It should be noted that the Berg-Purcell limit applies when the interactions between the binding sites and the TF are independent. This limit for the Bicoid/*hb* system results in a non-steep gene expression pattern ($H \sim 1.31$, see S2 Table). In-depth analysis of the model in the tandem paper [45] shows that increasing the pattern steepness slows down the switching rate between the gene's active and inactive states, due to the required cooperativity between the binding sites [29]. Consequently, our fitted models (with $N = 6$ or $N = 9$ binding sites) result in much higher errors in the integrated readout when compared to the model of independent binding sites (S12 Fig), and require a significantly longer integration time to achieve a specific precision level. Thus, the answer to the precision of the *hb* pattern achieved in such a short interphase duration remains elusive. In the future, systematic studies using synthetic promoters with a varying number of Bicoid binding sites and quantitative analyses of promoter dynamics captured with the MS2-MCP system will help characterize not only the cooperativity of Bicoid binding sites but also the kinetics of the downstream processes once the gene is activated. Added to this, the Bicoid search time for its binding sites on *hb* [15, 39] also needs to be revisited. The employed value (~4 s) in our model is estimated assuming a 3D search process inside the nucleus space but if Bicoid can slide along DNA in search for the sites, the process can be ~100 times faster [46]. Such a possibility is compatible with fluorescence correlation spectroscopy measurements of Bcd-eGFP motion [15, 47] and the recent observed clustering of the Bicoid molecules across the embryo [48]. The advent of single molecule tracking methods [48, 49] represent a promising approach to further shed lights on the mechanism of this process.

Importantly, despite these very rapid and precise measurements of Bicoid concentration along the AP axis, the expression process itself shows great variability in the total amount of mRNA produced during the interphase per nucleus ($\delta mRNA/\langle mRNA \rangle \sim 150\%$ in the boundary region) [40]. It is thus difficult to gauge to which degree errors in sensing Bicoid concentration contribute to the variability of the total mRNA produced at the boundary [40]. This also raises the question of understanding how precision in the downstream processes required for embryo segmentation is achieved at the scale of the whole embryo. If the embryo is capable of spatially averaging *hb* expression between nuclei at the same AP position (for example by the diffusion and nuclear export of its mRNA and nuclear import of its protein) [25], this will help the system distinguishing between the anterior and posterior region based on the *hb* transcription pattern alone. However, spatial averaging has a limit, due to the very short time available, the limited number of nuclei and the finite diffusion coefficient of the *hb* gene products [25]. It is likely that nuclei need to integrate *hb* gene expression over the nuclear cycle to reduce the noise in the *hb* readout. In nc11, the pattern collapses soon after reaching steady-state due to the next mitosis round. Therefore, any integrations of gene expression in nc11 and earlier (interphase duration shorter than ~400 s) are likely to lead to bias the pattern boundary to the anterior. Only starting at nc12 is the interphase duration long enough to reliably produce a steep pattern with the border around the mid-boundary position, based on the number of mRNA produced per nucleus (ΣI).

Materials and methods

Drosophila stocks

The original reporter *hb-MS2* and the Nup-RFP and MCP-GFP transgenes, both inserted on the second chromosome, were from [26]. All transgenic stocks generated in this study were obtained by BestGene. The *zld*^{294}*FRT19A* chromosome was a gift from C. Rushlow [32] and female germline clones were induced using heat-shock FLP recombinaison [50] with a *OvoD1hsFLPFRT19A* chromosome (# 23880, Bloomington). All stocks were raised at 25˚C.

Plasmids and BACs

The *hb-18kb-BAC* spanning the *hb* locus was the BAC CH322 55J23 obtained from PACMAN BAC libraries [51]. The *24xMS2 cassette* was inserted in the 5'UTR within the *hb* intron (*5'MS2-18kb-BAC*) or in the 3'UTR (*3'MS2-18kb-BAC*) using BAC recombineering [52] (details in S1 Text). The plasmid used to generate the new *hb-MS2ΔZelda* reporter was generated by replacing in the *pCasPeR4-hb-MS2* construct from [26], the MS2 cassette by a new MS2 sequence synthetically generated by Genscript in which all putative Zelda binding sites (*24xMS2-SL-ΔZelda*) had been mutated. The sequence coding for the CFP in the *pCasPeR4-hb-MS2* construct from [26] was also replaced by the sequence coding the iRFP (Addgene 45457) in which a unique putative Zelda binding site had been mutated (See S1 Text for more information).

The MS2-SL-ΔZelda sequence

Sequence-based binding-site cluster analysis was performed using the online software Cluster-Draw2 v2.55 at line.bioinfolab.net/webgate/submit.cgi. PWM for Zelda has been generated from [34] using the following Zelda binding sites: CAGGTAG; CAGGTAA; TAGGTAG; CAGGTAC; CAGGTAT; TAGGTAA; CAGGCAG and CAGGCAA. All heptamers detected as being a potential Zelda binding sites have been mutated and the new sequence synthesized by Genscript. The sequence of the new MS2 ΔZelda cassette is given in S1 Text.

RNA FISH and analysis

Embryo fixation and RNA *in-situ* hybridization were performed as describe in [3]. Briefly, RNA probes were generated using T7/T3 *in-vitro* transcription kit (Roche). *hb* RNAs were labelled with digoxigenin-tagged anti-sense probes detected with a sheep anti-dig primary antibody (1/1000 dilution, Roche) and donkey anti-sheep Alexa 568 secondary antibody (1/400 dilution, Invitrogen). MS2 containing RNAs were labelled with biotin-tagged anti-sense probes detected with a mouse anti-bio primary antibody (1/400 dilution, Roche) and chicken anti-mouse Alexa 488 secondary antibody (1/400 dilution, Invitrogen). Embryos were incubated 10 min with DAPI for DNA staining and 10 min in WGA-Alexa 633 (1/500 dilution, Molecular Probes) for nuclear envelop staining. Fixed embryos were mounted in VectaShield (Vector) and imaged in 3D (~20Z x 0.45μm) with a XY resolution of 3040*3040, 8bits per pixels, 0.09μm pixel size, 1 airy unit using a Zeiss LSM780 confocal microscope with a Zeiss 40x (1.4 NA) A-Plan objective. A full embryo 3D RAW image is composed on three 3D RAW Images that were stitched using Stitch Image Grid plugins from FIJI with 10% overlap and the linear blending fusion method. Image processing and analyzing were performed as describe [26]. Embryos of the proper stage in the nc11 interphase were selected according a threshold based on the nuclear area above 80 μm^2 (corresponding to late interphase) as in [3]. The expression map of endogenous *hb* or *MS2* transgenes have been manually false colored on FIJI and flatten on the nuclear channel.

Live embryo imaging

Imaging conditions were comparable to those outlined in [26, 53]. Embryos were collected 1h after egg laying, dechorionated by hand, fixed on the cover slip using heptane-dissolved glue, and immersed in to halocarbon oil (VWR). Mounted embryos were imaged at ~ 25˚C on a Zeiss LSM780 confocal microscope with a Zeiss 40x (1.4 NA) A-Plan objective. Image stacks of the lower cortical region of the embryo close to the middle of the AP axis (pixel size 0.2 μm, 0.54 μs pixel dwell time, 8 bits per pixels, confocal pinhole diameter 92 μm, distance between consecutive images in the stack 0.5 μm, ~1200x355 pxl, ~30 z-stacks) were collected

continuously. The GFP and RFP proteins were excited with a small fraction of the power output of a 488nm and a 568nm laser, 1.2% and 2% respectively. Images were acquired using the ZEN software (Zeiss). For each embryo, a tiled image of the midsection of the whole embryo was obtained, by stitching 3 separate images, from which the position of the anterior and posterior poles could be inferred.

Image processing and data extraction

Live imaging processing was performed in MATLAB as in [26]. Following imaging, movies are checked manually to verify all the nuclei included in data analysis are fully imaged in their depth and incompletely imaged nuclei (mostly nuclei at the periphery of the imaging field) are excluded. Nuclei segmentation is performed in a semi-automatic manner using our own software [54]. Only nuclei that exist throughout the nuclear interphase are used for the analysis. The MS2 spot detection was performed in 3D using a thresholding method. An average filter was applied before thresholding on each Z of the processed time point for noise reduction. MS2 spots were detected by applying a threshold equal to ~2 fold above background signal and only the spots composed of at least 10 connected voxels were retained. The intensity of the 3D spot is calculated as the sum of the voxel values of each Z-stack. At the end of nc13, some MCP aggregates may cause false spot detections: they are less bright and their shape is more spread compare to the MS2-MCP spots. The aggregates are eliminated automatically by raising the spot detection threshold without affecting the detection of the MS2 spots. We manually checked each movie to ensure the correct spot detection. The data from a segmented movie indicates for each nucleus its segmentation profile, identifier number and the intensity trace of the detected spot over time.

During mitosis, nuclei divide in waves, usually from the embryo poles. Therefore, nuclei at the anterior pole may produce MCP-MS2 spots earlier due to either earlier chromatin decondensation or earlier reentrance of Bicoid into the nucleic space [6]. We correct for this by realigning all the intensity traces by choosing the origin for time for each trace when the two sibling nuclei are first separated (see S2 Text and S4 Fig)

Stochastic model of *hb* expression regulation

The general model of transcription regulation through transcription factor (TF) binding/unbinding to the operator sites (OS) is a based on a graph-based linear framework [29, 55, 56]. We introduce a single time scale, which is the TF search time for a single operator site at the boundary $t_{bind} \sim 4s$. The details of the model are described in S6 Text and S7 Text.

Supporting information

S1 Text. Supplementary molecular biology.
(PDF)

S2 Text. Determining the origin of time for each time trace.
(PDF)

S3 Text. Embryo alignment.
(PDF)

S4 Text. Quantifying the steepness of *hb* expression pattern.
(PDF)

S5 Text. Evolution of the transcription pattern steepness over time.
(PDF)

S6 Text. Models of transcription factor (TF) sensing and transcription.
(PDF)

S7 Text. Comparing the noise in positional readout between models.
(PDF)

S8 Text. Supporting References.
(PDF)

S1 Fig. Examples of individual spot intensity over time in nuclear cycle 11. The time traces are sorted by their respective nuclei position, which is shown in the boxes in % EL (Position 0 corresponding to the middle of the embryo). Nuclei with position beyond -3.1% EL (last shown nucleus) have no spots. Horizontal axis: time in seconds. Vertical axis: spot fluorescent intensity in arbitrary units.
(PDF)

S2 Fig. Examples of individual spot intensity over time in nuclear cycle 12. The time traces are sorted by their respective nuclei position, which is shown in the boxes in % EL (Position 0 corresponding to the middle of the embryo). Nuclei with position beyond 6.6% EL (last shown nucleus) have no spots. Horizontal axis: time in seconds. Vertical axis: spot fluorescent intensity in arbitrary units.
(PDF)

S3 Fig. Examples of individual spot intensity over time in nuclear cycle 13. The time traces are sorted by their respective nuclei position, which is shown in the boxes in % EL (Position 0 corresponding to the middle of the embryo). Nuclei with position beyond -1.5% EL (last shown nucleus) have no spots. Horizontal axis: time in seconds. Vertical axis: spot fluorescent intensity in arbitrary units.
(PDF)

S4 Fig. Determining the nuclei birth time. (A) Examples of frame-by-frame monitoring of two sibling nuclei after mitosis. The time interval between frames is 13.05 s. The yellow line is drawn automatically to connect the two siblings' centroids once division is detected. (B) The distance between the centroids (blue line) and its derivative (red line) over time. The nuclei's birth is set as the time when the speed of segregation between the sibling nuclei decreases to near-zero. (C) Examples of the nuclei birth time t_{birth} along AP axis. Shown is the t_{birth} extracted from the movies (blue circles) and from the fitted model in Eq 1 (red dashed line). The y axis (t_{birth}) is shifted so as the two mitotic waves from the two poles meet at $t_{meet} = 0$.
(PDF)

S5 Fig. Fitting the trace feature patterns in nuclear cycle 11. The fitted curves (dashed black lines) are shown with data points (blue dots). Each data point corresponds to a single trace feature value. The horizontal axis is the AP axis in % EL.
(PDF)

S6 Fig. Fitting the trace feature patterns in nuclear cycle 12. The fitted curves (dashed black lines) are shown with data points (blue dots). Each data point corresponds to a single trace feature value. The horizontal axis is the AP axis in % EL.
(PDF)

S7 Fig. Fitting the trace feature patterns in nuclear cycle 13. The fitted curves (dashed black lines) are shown with data points (blue dots). Each data point corresponds to a single trace's

feature value. The horizontal axis is the AP axis in % EL.
(PDF)

S8 Fig. Time evolution of the pattern steepness H(t) over time. Shown for nc 11 (blue solid line), nc 12 (red solid line) and nc 13 (yellow solid line) along with the margins of errors (p-value = 0.05). Also shown (dashed lines) are the Hill coefficients extracted from FISH data in for the respective cycles. The coefficients from FISH in nc12 and nc13 are almost identical.
(PDF)

S9 Fig. Dynamics of *hb-MS2ΔZelda* expression (mean spot intensity) as a function of time. **A-C**: The average spot intensity I(t) is indicated by a heat map (color scale on the right) horizontally as a function of position along the AP axis (0% EL positioned P_{ON} boundary at nc12) and vertically as a function of time (s) fixing the origin at the onset of interphase for each nucleus (see details in S2 text and S4 Fig). For each cycle (**A**: nc11; **B**: nc12; **C**: nc13), the end of interphase (onset of next mitosis) is indicated by a dashed line (white). The green dashed line indicates the position of the expression boundary (I(t) equals half the average spot intensity at the anterior pole) over time. **D:** I(t) as a function of time (s) at mid-boundary position. The first hints of transcription are observed at mid-boundary position ~ 170 s after the onset of interphase (lower limit of the light blue zone) and steady state is reached at ~ 350 s. Boundary formation reaches steady state in ~ 180 s. Data were obtained from 5 (nc11), 8 (nc12) and 4 (nc13) embryos. Embryos were aligned spatially fixing the origin of the axis at boundary position (P_{ON}) at nc12 and the origin of time was calculated for each nuclei as the origin of its respective cycle (see S2 Text and S4 Fig).
(PDF)

S10 Fig. MS2 binding site configuration. *L(t) is* the number of MS2-MCP binding sites on a nascent RNA at time *t* after its transcription initiation.
(PDF)

S11 Fig. Examples of simulated trajectories at mid-embryo for *N* = 6. (A) The number of bound TF molecules to the promoter over time. (B) The gene transcriptional state given the number of bound molecules to the promoter. The gene is turned ON when the promoter is fully bound by TF. (C) Occurrences of transcription initiation events $I_{RNAP}(t)$, corresponding to (B). (D) Transcription loci intensity *I(t)*, corresponding to (C).
(PDF)

S12 Fig. Comparing the relative noise in the positional readout $\delta mRNA/\langle mRNA \rangle$ as a function time. For fitted model with *N* = 6 (blue line), fitted model with *N* = 9 (red line) and "no cooperativity" model (yellow line).
(PDF)

S1 Table. Parameters of fitting the sigmoid function with time trace features in nc11-13. The position of the feature pattern border (X_0, in units of % EL), the pattern steepness (*H*) and their respective confidence interval (in brackets) for the locus activity (P_{active}), the time period during which the locus is activated (t_{active}), the integral transcription activity (ΣI) and the mean transcription rate (μI). The data are inferred from all aligned embryos in the respective cycles.
(PDF)

S2 Table. Fitted models of varying number of operator sites *N* and the model of 6 independent sites. Shown is the p-value of the likelihood ratio test between the fitted model of *N* and *N-1* OS. Also shown is the Bayesian Information Criterion (BIC) for each model. The data for

the fitting is pulled from all embryos in all nuclear cycles.
(PDF)

S1 Movie. Live imaging of transcription dynamics of *hb-MS2 (Lucas et al. 2013)* subject to varying doses of Zelda. (A) Zelda hetero (Zld Mat +/-): expression of the *hb-MS2* reporter in embryos from *zld*^{294} heterozygous females. (B) Zelda GLC (Zld Mat -/-): expression of the *hb-MS2* reporter in *zld*^{294} germline clone embryos. The movies have two channels: MCP-GFP channel (green) for monitoring the dynamics of nascent mRNA production and NUP-RFP (red) for nuclei detection. The capture frame is from -25% to 25% of embryo length. The anterior pole is on the left side of the frame.
(AVI)

S2 Movie. Live imaging of the transcription dynamics with the MS2 cassette inserted at 5'-UTR and 3'-UTR. The MS2 cassette *(Lucas et al. 2013)* is placed at the (A) 5'-UTR within the intron of the *hb* gene (*5'MS2-hb-18kb*) and the (B) 3'-UTR of *hb* gene (*3'MS2-hb-18kb*). The movies have two channels: MCP-GFP channel (green) for monitoring the dynamics of nascent mRNA production and NUP-RFP (red) for nuclei detection. The capture frame is from -25% to 25% of embryo length. The anterior pole is on the left side of the frame.
(AVI)

S3 Movie. Live imaging of transcription dynamics of *hb-MS2ΔZelda*. Expression of the *hb-MS2ΔZelda* reporter in wild-type embryos. The movie has two channels: MCP-GFP channel (green) for the monitoring of nascent mRNA production dynamics and NUP-RFP (red) for nuclei detection. The capture frame is from -25% to 25% of embryo length. The anterior pole is on the left side of the frame.
(AVI)

S4 Movie. The transcription pattern dynamics in nuclear cycle 13. The movie shows the transcription patterns, represented by the probability of spot appearance $P_{SPOT}(t)$ along AP axis, at a given time after the onset of nuclear interphase. The data are from MS2 movies of *hb-MS2ΔZelda* reporter expression in wild-type embryos. Data are shown for each of the 4 individual embryos (color lines in left panel) and pulled (dashed blue line with error bars in the right panel). Also shown is in the transcription pattern extracted from FISH (dashed black line with error bars) from (15).
(AVI)

Acknowledgments

The authors thank Patricia Le Baccon, Mickael Garnier and the Imaging Facility PICT-IBiSA of the Institut Curie, BestGene Inc for transgenics and C. Rushlow and the Bloomington stock center for fly stocks.

Author Contributions

Conceptualization: Tanguy Lucas, Huy Tran, Cécile Fradin, Mathieu Coppey, Aleksandra M. Walczak, Nathalie Dostatni.

Data curation: Tanguy Lucas, Huy Tran, Carmina Angelica Perez Romero, Mathieu Coppey, Aleksandra M. Walczak.

Formal analysis: Tanguy Lucas, Huy Tran, Mathieu Coppey, Aleksandra M. Walczak.

Funding acquisition: Cécile Fradin, Mathieu Coppey, Aleksandra M. Walczak, Nathalie Dostatni.

Investigation: Tanguy Lucas, Huy Tran, Aurélien Guillou, Cécile Fradin, Aleksandra M. Walczak, Nathalie Dostatni.

Methodology: Tanguy Lucas, Huy Tran, Carmina Angelica Perez Romero, Aurélien Guillou, Cécile Fradin, Mathieu Coppey, Nathalie Dostatni.

Project administration: Cécile Fradin, Mathieu Coppey, Aleksandra M. Walczak, Nathalie Dostatni.

Resources: Tanguy Lucas, Huy Tran, Carmina Angelica Perez Romero, Aurélien Guillou.

Software: Huy Tran, Carmina Angelica Perez Romero, Aleksandra M. Walczak.

Supervision: Cécile Fradin, Mathieu Coppey, Aleksandra M. Walczak, Nathalie Dostatni.

Validation: Tanguy Lucas, Huy Tran, Carmina Angelica Perez Romero, Cécile Fradin, Mathieu Coppey, Aleksandra M. Walczak, Nathalie Dostatni.

Visualization: Tanguy Lucas, Huy Tran, Carmina Angelica Perez Romero, Cécile Fradin, Mathieu Coppey, Aleksandra M. Walczak, Nathalie Dostatni.

Writing – original draft: Tanguy Lucas, Huy Tran, Carmina Angelica Perez Romero, Cécile Fradin, Mathieu Coppey, Aleksandra M. Walczak, Nathalie Dostatni.

Writing – review & editing: Tanguy Lucas, Huy Tran, Carmina Angelica Perez Romero, Cécile Fradin, Mathieu Coppey, Aleksandra M. Walczak, Nathalie Dostatni.

References

1. Rogers KW, Schier AF. Morphogen gradients: from generation to interpretation. Annu Rev Cell Dev Biol. 2011; 27:377–407. https://doi.org/10.1146/annurev-cellbio-092910-154148 PMID: 21801015
2. Stathopoulos A, Iber D. Studies of morphogens: keep calm and carry on. Development. 2013; 140 (20):4119–24. https://doi.org/10.1242/dev.095141 PMID: 24086076
3. Porcher A, Dostatni N. The Bicoid morphogen system. Current Biology 2010; 20(5):R249–R54. https://doi.org/10.1016/j.cub.2010.01.026 PMID: 20219179
4. Driever W, Nusslein-Volhard C. A gradient of bicoid protein in Drosophila embryos. Cell. 1988; 54 (1):83–93. PMID: 3383244
5. Abu-Arish A, Porcher A, Czerwonka A, Dostatni N, Fradin C. Fast mobility of Bicoid captured by flourescent correlation spectroscopy: implication for the rapid establishment of its gradient Biophysical Journal. 2010; 99:L33–L5. https://doi.org/10.1016/j.bpj.2010.05.031 PMID: 20712981
6. Gregor T, Wieschaus EF, McGregor AP, Bialek W, Tank DW. Stability and Nuclear Dynamics of the Bicoid Morphogen Gradient. Cell. 2007; 130(1):141–52. https://doi.org/10.1016/j.cell.2007.05.026 PMID: 17632061
7. Chen H, Xu Z, Mei C, Yu D, Small S. A System of Repressor Gradients Spatially Organizes the Boundaries of Bicoid-Dependent Target Genes. Cell. 2012; 149(3):618–29. https://doi.org/10.1016/j.cell.2012.03.018 PMID: 22541432
8. Driever W, Nusslein-Volhard C. The bicoid protein is a positive regulator of hunchback transcription in the early Drosophila embryo. Nature. 1989; 337(6203):138–43. https://doi.org/10.1038/337138a0 PMID: 2911348
9. Struhl G, Struhl K, Macdonald PM. The gradient morphogen bicoid is a concentration-dependent transcriptional activator. Cell. 1989; 57(7):1259–73. PMID: 2567637
10. Crauk O, Dostatni N. Bicoid Determines Sharp and Precise Target Gene Expression in the Drosophila Embryo. Current Biology. 2005; 15(21):1888–98. https://doi.org/10.1016/j.cub.2005.09.046 PMID: 16271865
11. Ronchi E, Treisman J, Dostatni N, Struhl G, Desplan C. Down-regulation of the Drosophila morphogen bicoid by the torso receptor-mediated signal transduction cascade. Cell. 1993; 74(2):347–55. PMID: 8343961

12. Simpson-Brose M, Treisman J, Desplan C. Synergy between the hunchback and bicoid morphogens is required for anterior patterning in Drosophila. Cell. 1994; 78(5):855–65. PMID: 8087852
13. Driever W, Nüsslein-Volhard C. The bicoid protein determines position in the Drosophila embryo in a concentration-dependent manner. Cell. 1988; 54(1):95–104. PMID: 3383245
14. Wolpert L. Positional information and the spatial pattern of cellular differentiation. Journal of Theoretical Biology. 1969; 25(1):1–47. PMID: 4390734
15. Porcher A, Abu-Arish A, Huart S, Roelens B, Fradin C, Dostatni N. The time to measure positional information: maternal Hunchback is required for the synchrony of the Bicoid transcriptional response at the onset of zygotic transcription. Development. 2010; 137(16):2795–804. https://doi.org/10.1242/dev.051300 PMID: 20663819
16. Margolis JS, Borowsky ML, Steingrimsson E, Shim CW, Lengyel JA, Posakony JW. Posterior stripe expression of hunchback is driven from two promoters by a common enhancer element. Development. 1995; 121(9):3067–77. PMID: 7555732
17. Wimmer EA, Carleton A, Harjes P, Turner T, Desplan C. Bicoid-independent formation of thoracic segments in Drosophila. Science. 2000; 287(5462):2476–9. PMID: 10741965
18. Kambadur R, Koizumi K, Stivers C, Nagle J, Poole SJ, Odenwald WF. Regulation of POU genes by castor and hunchback establishes layered compartments in the Drosophila CNS. Genes & development. 1998; 12(2):246–60.
19. Schroder C, Tautz D, Seifert E, Jackle H. Differential regulation of the two transcripts from the Drosophila gap segmentation gene hunchback. The EMBO journal. 1988; 7(9):2881–7. PMID: 2846287
20. Perry MW, Boettiger AN, Levine M. Multiple enhancers ensure precision of gap gene-expression patterns in the Drosophila embryo. Proceedings of the National Academy of Sciences. 2011; 108(33):13570–5.
21. Perry Michael W, Bothma Jacques P, Luu Ryan D, Levine M. Precision of Hunchback Expression in the Drosophila Embryo. Current Biology. 2012; 22(23):2247–52. https://doi.org/10.1016/j.cub.2012.09.051 PMID: 23122844
22. Raser JM, O'Shea EK. Noise in Gene Expression: Origins, Consequences, and Control. Science. 2005; 309(5743):2010–3. https://doi.org/10.1126/science.1105891 PMID: 16179466
23. Datta RR, Ling J, Kurland J, Ren X, Xu Z, Yucel G, et al. A feed-forward relay integrates the regulatory activities of Bicoid and Orthodenticle via sequential binding to suboptimal sites. Genes & development. 2018; 32(9–10):723–36.
24. Manu Surkova S, Spirov AV, Gursky VV, Janssens H, Kim A-R, et al. Canalization of Gene Expression in the *Drosophila* Blastoderm by Gap Gene Cross Regulation. PLoS Biol. 2009; 7(3):e1000049. https://doi.org/10.1371/journal.pbio.1000049 PMID: 19750121
25. Ochoa-Espinosa A, Yu D, Tsirigos A, Struffi P, Small S. Anterior-posterior positional information in the absence of a strong Bicoid gradient. Proceedings of the National Academy of Sciences of the United States of America. 2009; 106(10):3823–8. https://doi.org/10.1073/pnas.0807878105 PMID: 19237583
26. Lucas T, Ferraro T, Roelens B, De Las Heras Chanes J, Walczak A, Coppey M, et al. Live imaging of Bicoid-dependent transcription in Drosophila embryos Current Biology. 2013; 23(21):2135–9. https://doi.org/10.1016/j.cub.2013.08.053 PMID: 24139736
27. Bertrand E, Chartrand P, Schaefer M, Shenoy SM, Singer RH, Long RM. Localization of ASH1 mRNA Particles in Living Yeast. Molecular Cell. 1998; 2(4):437–45. PMID: 9809065
28. Ferraro T, Lucas T, Clemot M, De Las Heras Chanes J, Desponds J, Coppey M, et al. New methods to image transcription in living fly embryos: the insights so far, and the prospects. Wiley Interdisciplinary REviews Developmental Biology. 2016; 5(3):296–310. https://doi.org/10.1002/wdev.221 PMID: 26894441
29. Estrada J, Wong F, DePace A, Gunawardena J. Information Integration and Energy Expenditure in Gene Regulation. Cell. 2016; 166(1):234–44. https://doi.org/10.1016/j.cell.2016.06.012 PMID: 27368104
30. Fusco D, Accornero N, Lavoie B, Shenoy SM, Blanchard J-M, Singer RH, et al. Single mRNA Molecules Demonstrate Probabilistic Movement in Living Mammalian Cells. Current biology: CB. 2003; 13(2):161–7. PMID: 12546792
31. Blythe Shelby A, Wieschaus Eric F. Zygotic Genome Activation Triggers the DNA Replication Checkpoint at the Midblastula Transition. Cell. 2015; 160(6):1169–81. https://doi.org/10.1016/j.cell.2015.01.050 PMID: 25748651
32. Liang H-L, Nien C-Y, Liu H-Y, Metzstein MM, Kirov N, Rushlow C. The zinc-finger protein Zelda is a key activator of the early zygotic genome in Drosophila. Nature. 2008; 456(7220):400–3. https://doi.org/10.1038/nature07388 PMID: 18931655

33. Bothma JP, Garcia HG, Ng S, Perry MW, Gregor T, Levine M. Enhancer additivity and non-additivity are determined by enhancer strength in the Drosophila embryo. eLife. 2015; 4:e07956.

34. Nien CY, Liang HL, Butcher S, Sun Y, Fu S, Gocha T, et al. Temporal coordination of gene networks by Zelda in the early Drosophila embryo. PLoS Genet. 2011; 7(10):e1002339. https://doi.org/10.1371/journal.pgen.1002339 PMID: 22028675

35. ten Bosch JR, Benavides JA, Cline TW. The TAGteam DNA motif controls the timing of Drosophila pre-blastoderm transcription. Development. 2006; 133(10):1967–77. https://doi.org/10.1242/dev.02373 PMID: 16624855

36. Lopes FJP, Spirov AV, Bisch PM. The role of Bicoid cooperative binding in the patterning of sharp borders in Drosophila melanogaster. Developmental Biology. 2012; 370(2):165–72. https://doi.org/10.1016/j.ydbio.2012.07.020 PMID: 22841642

37. Tran H, Desponds J, Perez Romero CA, Coppey M, Fradin C, Dostatni N, et al. Precision in a rush: trade-offs between positioning and steepness of the hunchback expression pattern. PLoS Computational Biology. In press.

38. Coulon A, Ferguson ML, de Turris V, Palangat M, Chow CC, Larson DR. Kinetic competition during the transcription cycle results in stochastic RNA processing. eLife. 2014; 3:e03939.

39. Gregor T, Tank DW, Wieschaus EF, Bialek W. Probing the Limits to Positional Information. Cell. 2007; 130(1):153–64. https://doi.org/10.1016/j.cell.2007.05.025 PMID: 17632062

40. Desponds J, Tran H, Ferraro T, Lucas T, Perez Romero C, Guillou A, et al. Precision of Readout at the hunchback Gene: Analyzing Short Transcription Time Traces in Living Fly Embryos. PLOS Computational Biology. 2016; 12(12):e1005256. https://doi.org/10.1371/journal.pcbi.1005256 PMID: 27942043

41. Juven-Gershon T, Hsu J-Y, Kadonaga JT. Caudal, a key developmental regulator, is a DPE-specific transcriptional factor. Genes & development. 2008; 22(20):2823–30.

42. Berg HC, Purcell EM. Physics of chemoreception. Biophysical Journal. 1977; 20(2):193–219. https://doi.org/10.1016/S0006-3495(77)85544-6 PMID: 911982

43. Bialek W, Setayeshgar S. Physical limits to biochemical signaling. PNAS. 2005; 102(29):10040–5. https://doi.org/10.1073/pnas.0504321102 PMID: 16006514

44. Kaizu K, de Ronde W, Paijmans J, Takahashi K, Tostevin F, ten Wolde Pieter R. The Berg-Purcell Limit Revisited. Biophysical Journal. 2014; 106(4):976–85. https://doi.org/10.1016/j.bpj.2013.12.030 PMID: 24560000

45. Tran H, Desponds J, Perez-Romero CA, Coppey M, Fradin C, Dostatni N, et al. Precision in a rush: hunchback pattern formation in a limited time. To be submitted and deposited on the Bioarxiv shortly. in prep.

46. Slutsky M, Mirny LA. Kinetics of Protein-DNA Interaction: Facilitated Target Location in Sequence-Dependent Potential. Biophysical Journal. 2004; 87(6):4021–35. https://doi.org/10.1529/biophysj.104.050765 PMID: 15465864

47. Fradin C. On the importance of protein diffusion in biological systems: The example of the Bicoid morphogen gradient. Biochimica et Biophysica Acta (BBA)—Proteins and Proteomics. 2017; 1865(11, Part B):1676–86.

48. Mir M, Reimer A, Haines JE, Li X-Y, Stadler M, Garcia H, et al. Dense Bicoid hubs accentuate binding along the morphogen gradient. Genes & Development. 2017; 31(17):1784–94.

49. Xu H, Sepúlveda LA, Figard L, Sokac AM, Golding I. Combining protein and mRNA quantification to decipher transcriptional regulation. Nature Methods. 2015; 12:739. https://doi.org/10.1038/nmeth.3446 PMID: 26098021

50. Chou T-b, Perrimon N. The Autosomal FLP-DFS Technique for Generating Germline Mosaics in Drosophila melanogaster. Genetics. 1996; 144(4):1673–9. PMID: 8978054

51. Venken KJT, Carlson JW, Schulze KL, Pan H, He Y, Spokony R, et al. Versatile P[acman] BAC libraries for transgenesis studies in Drosophila melanogaster. Nat Meth. 2009; 6(6):431–4.

52. Warming S, Costantino N, Court DL, Jenkins NA, Copeland NG. Simple and highly efficient BAC recombineering using galK selection. Nucleic Acids Research. 2005; 33(4):e36–e. https://doi.org/10.1093/nar/gni035 PMID: 15731329

53. Perez-Romero CA, Tran H, Coppey M, Walczak AM, Fradin C, Dostatni N. Live imaging of mRNA transcription in Drosophila Embryos. Methods Mol Biol. In press; "Morphogen Gradients: Methods and Protocols" edited by J. Dubrulle.

54. Tran H, Perez-Romero CA, Ferraro T, Fradin C, Dostatni N, Coppey M, et al. LiveFly—a toolbox for the analysis of transcription dynamics in live Drosophila embryos. Methods Mol Biol. In press;"Morphogen Gradients: Methods and Protocols" edited by J. Dubrulle.

55. Gunawardena J. A linear framework for time-scale separation in nonlinear biochemical systems. PloS one. 2012; 7(5):e36321. https://doi.org/10.1371/journal.pone.0036321 PMID: 22606254

56. Lloyd-Price J, Gupta A, Ribeiro AS. SGNS2: a compartmentalized stochastic chemical kinetics simulator for dynamic cell populations. Bioinformatics (Oxford, England). 2012; 28(22):3004–5.

3.2 *hb* Promoter state

The data contained in the *hb* MCP-GFP movies can be analyzed in diverse ways. Here I present a summary of the work that we have performed to infer the state of the *hb* promoter from the intensity traces of individual transcription sites, described in detail in:

Desponds J, Tran H, Ferraro T, Lucas T, Perez Romero CA, Guillou A, Fradin C, Coopey M, Dostatni N, Walczak AM. "Precision of readout at the hunchback gene". *PLoS Comput Biol*. 2016;(October):1-35. doi:10.1101/063784.

My contribution:

I acquired all the movies mentioned in the manuscript and used them to generate the time traces that were analyzed to compare transcription models. Movie processing was a joint effort between Dr. Tanguy Lucas and myself. I made some minimal contributions to manuscript editing.

Research Objective:

To infer the dynamics of the promoter state from the short traces of nascent transcript dynamics in living embryos using the *hb-MS2* reporter gene.

Research Highlights:

- An inference approach based on autocorrelation functions allows the quantification of transcription dynamics.
- Comparing the predictions made using various models of transcription (Poisson, two-state and cycle models) to our experimental MS2 traces, we concluded that the *hb* promoter dynamics likely follows a two-state model.
- We found that the *hb* promoter ON rate (k_{on}) is higher at the embryo's anterior, where Bcd concentration is higher, than at the boundary, whereas its OFF rate (k_{off}) remains the same throughout the expression domain.

In this manuscript, we hypothesized that one could infer the state of the *hb* promoter (ON or OFF, Fig. 11 A,B) from the fluctuations (bursts) in the MS2 traces (Fig. 11D). This information was extracted by analyzing the autocorrelation function of the MS2 traces (Fig. 11E). This paper explores the use of such autocorrelation functions, using the data acquired in WT embryos, to check which model best describes the transcription activation of the *hb* promoter.

We explored three different models in this study that can be described using a traffic light analogy (car = polymerase, green light = ON, orange/red light = OFF). In the Poisson model, the light is always green and cars arrive sparsely and randomly at a certain average rate. In the two-state model, the light switches between red and green, while in the cyclic model, the light changes from green to orange to red. Cars have to stop when the light is orange or red light, and are only able to pass when the light is green (Fig. 11A). The waiting times between light switching follow an exponential distribution.

Autocorrelation functions allow us to determine the degree of correlation existing in a signal fluctuating over time (Fig. 11E). The resulting curve describes the persistence of the signal, which in our case contains information about the persistence of the promoter states (ON/OFF). The autocorrelation curve decays with a characteristic relaxation time, which is related to how long the promoter stays in its ON/OFF states. From this decay time, and given the probability of the promoter to be ON calculated from the average trace intensity, we can extract the promoter switching rates k_{on} and k_{off} for two-state or cycle promoter models.

Simulations had shown that the autocorrelation function is affected by both the transcription elongation time and the short duration of the fluorescent time traces typically acquired in this developmental phase (Fig. 11F) (Ferraro et al., 2016). The paper introduces *home-made* mathematical modeling to overcome these problems, allowing the quantification of *hb* promoter dynamics of transcription accurately (i.e. the switching rates k_{on} and k_{off} between active and inactive transcription periods) (Fig. 11E) (Ferraro et al., 2016).

We fitted the data to different promoter state models and showed that all three considered models were consistent with our data, although the two-state and cycle models gave almost identical fits which were visibly better fits to the data (Fig. 11G). After examining this data more closely with

theoretical predictions of relative mRNA production variability (δmRNA/mRNA) for the two state and the Poisson models, we were able to conclude that the *hb* promoter follows a two-state model (Fig. 11J). Next we calculated the promoter switching rates for this model, and found out that in the boundary region of the embryo k_{on} is much smaller than in the anterior, while k_{off} has a similar range in the anterior and at the boundary (Fig 11K). This behavior can be attributed to a Bcd dependent response, as in the anterior Bcd concentration is higher allowing the activation of the transgene, whereas at the boundary lower Bcd concentrations result in smaller activation rates.

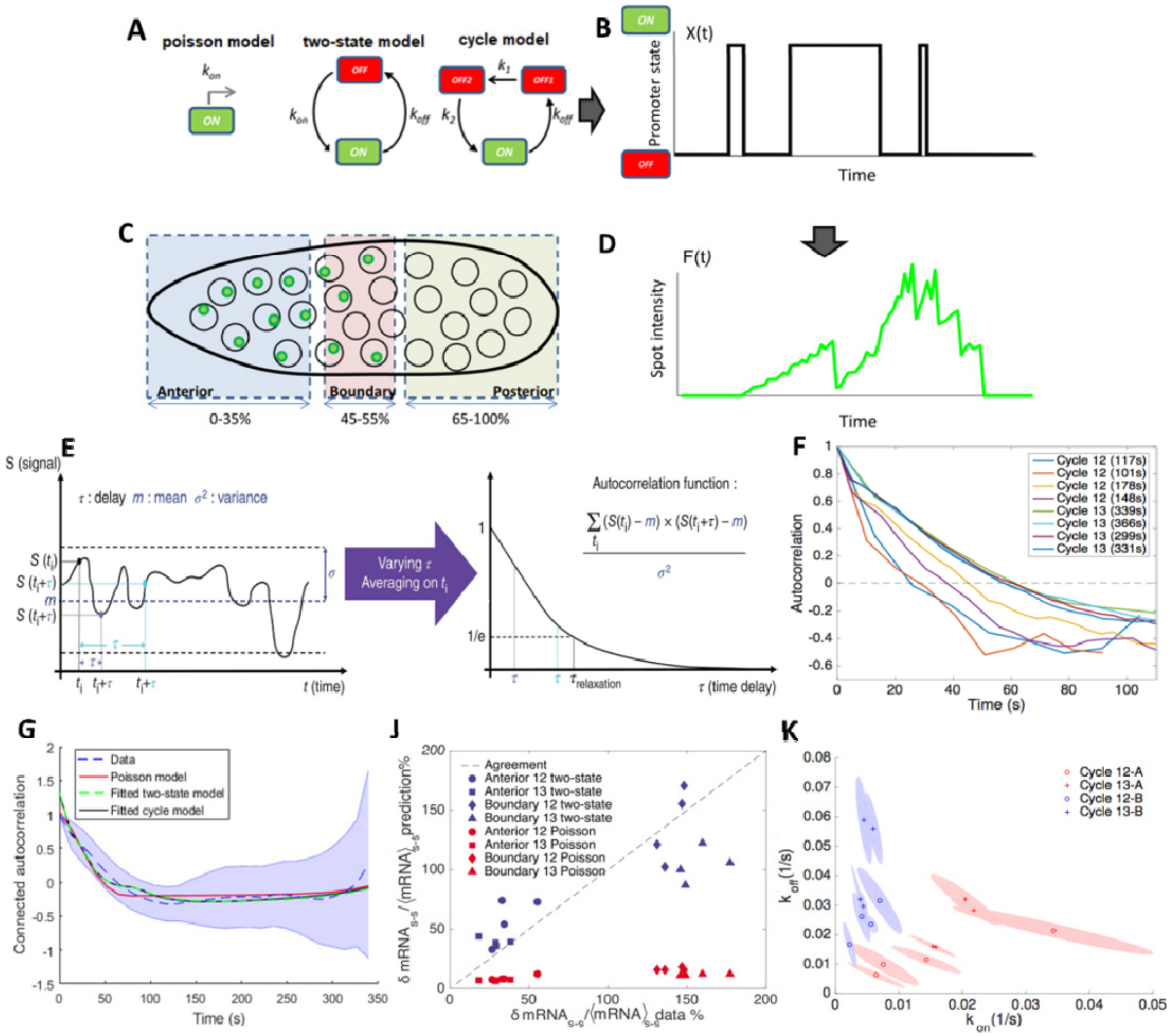


Figure 11. Summary of results from autocorrelation analysis: A) Three models of transcription dynamics are considered: Poisson, two-state and cycle models. **B)** Example of the promoter state

dynamics (ON or OFF) as a function of time. **C)** Transcription events on individual nuclei's positioned in the embryo Anterior - A, Boundary region - B or Posterior **D)** Example of a MS2 trace. **E)** An autocorrelation function can be calculated from a fluctuating noisy signal, which allows the extraction of promoter dynamics information through its relaxation time. Taken from (Ferraro et al., 2016). **F)** Example of auotcorrelation curves for traces in nc 12 and nc 13. **G)** Example of autocorrelation curve in the boundary region at nc 13 (blue) fitted to different promoter models: Poisson (red), two-state (green) and cycle (black). **J)** Comparison of relative error in the mRNA produced theoretically (y-axis) vs data (x-axis), at different nuclear cycles using two different promoter models: Poisson (red), and two-state (blue). **K)** Inferred two-state promoter k_{off} (y-axis) and k_{on} (x-axis) values in the embryo's anterior (red) and boundary region (blue) for nc 12 (circle) and nc 13 (cross). Modified with permission from (Desponds et al., 2016).

3.3 Modeling *hb* transcription

Here I present our efforts to reconcile different theoretical binding models of activation of the *hb* promoter by Bcd with the data I acquired by live imaging of *hb* transcription. This work is detailed in the paper: Tran H, Desponds J, Perez Romero CA, Coppey M, Fradin C, Dostatni N, Walczak AM. "Precision in a rush: trade-offs between positioning and steepness of the *hunchback* expression pattern". *PLoS Comput Biol*. 2018;14(10):e1006513.

My contribution:

I acquired and processed all the movies mentioned in the manuscript and used to generate the time traces that were analyzed to compare activation models. I made some minimal contributions to manuscript editing and review.

Research Objective:

To explore different activation models for the *hb* gene, and explain how it can form such a steep border in such a short time via its array of Bcd binding sites.

Research Highlights:

- The limited readout time imposed by nuclear cycles short duration in the early embryo affects genes ability to read positional information and achieve precise expression.
- Explored the trade-off between the ability of a regulatory system based on cooperative Bcd binding to the promoter region to produce a steep boundary and to minimize expression variability between different nuclei in the same region along the AP axis.
- Comparing our theoretical results with the *hb* transcription spatio-temporal dynamics extracted from our movies, we discuss possible regulatory strategies, suggesting an important role for additional binding sites, gradients or non-equilibrium binding and modified transcription factor search strategies.

Our main objective was to explore how the *hb* gene measures the Bcd input concentration and forms a steep precise border in a limited time, assuming that activation is caused by Bcd binding to the array of binding sites present in the *hb* promoter region. We found out that the limited readout time, imposed by nuclear cycle length, affects the genes ability to read positional information along the antero-posterior axis and to achieve precise expression. Therefore we explored the trade-off between the ability of a regulatory system to produce a steep boundary and to minimize expression variability between different nuclei along the embryo. We compared our theoretical results with *hb* transcription observed with the MS2 system.

From live imaging of *hb* transcription (Lucas et al., 2018), we showed that *hb* expression has a stable steep binary expression pattern established within the three minute interphase of nuclear cycle 11. In order to better understand the trade-off between biologically short cell cycles, steepness, readout error, we studied different models of gene expression regulation where transcription is controlled by the binding and unbinding of the Bcd transcription factor (TF) to multiple operator sites on the *hb* promoter. We consider equilibrium-binding models with different expression rules, non-equilibrium models (Estrada et al., 2016) and equilibrium models with two TF gradients.

When visualizing *hb* expression, one can see the border of the expression domain from two different perspectives (Fig. 12A): 1) Its steepness, characterized by the Hill coefficient (H), which is estimated by fitting a Hill function to the *hb* expression profile over the embryo length (Fig. 12B). 2) Its readout error, calculated by comparing the response of nuclei at a similar position along the AP axis (Fig. 12C). It has been proposed that the steep boundary in *hb* expression might be the result of cooperative binding of Bcd at its different TF binding sites (Driever et al., 1989b; Estrada et al., 2016; Gregor et al., 2007a). However, none of these studies questions whether such a steep boundary is achievable within the limited nuclear cycle time window, during which Bcd concentration is read. Using equilibrium models at steady state limited by nc length, we uncover a trade-off between the readout error and steepness of the expression pattern at the boundary: 1) the steeper the boundary, the larger the readout error (Fig. 12D). 2) For short nuclear cycles, there are few promoters switching, with high readout error for

all H values. 3) Longer nuclear cycles overcome this problem, with steeper boundaries achievable with less readout error (Fig. 12D).

We propose a feature "positional resolution" ΔX of individual nuclei at the *hb* expression border to describe how well nearby nuclei can discern inputs such as Bcd (Fig. 12E) on the two sides of the expression boundary. The positional resolution is affected by the trade-off mentioned above where low steepness has minimal error and similar expression in nuclei output (e.g. mRNA), while a high steepness has a high error in nuclei output (Fig. 12E). We explored this using equilibrium models at steady state and concluded that optimal steepness (H*) decreases with cell cycle time (Fig. 12F). Our theoretical data also indicated that, at early nc 11 the embryo can best discriminate readouts when producing a very high H, while at later nc12-13 a moderate H value between 2-5 is preferred (Fig. 12G). However, this model was not able to explain the high H ~ 7 values observed in our live imaging experiments at nc12-13.

Therefore, we decided to explore a hybrid equilibrium/non-equilibrium model, which was able to achieve high H* values (≤ 8); however, the best positional resolution is still achieved with $2 \leq H \leq 5$, which does not agree with the experimental values (Fig. 12H). All of the models we used so far used Bcd as the only input, therefore we explored if having an equilibrium model with two opposing gradients, one anterior activator (Bcd) and one possible posterior repressor (Cad). It can achieve a steepness as high as H ~ 7, but the best positional resolution is still achieved with H ~ 4.

This experiments let us conclude that considering Bcd as the sole input of *hb* transcriptional regulation is a very narrow view unlikely to explain the reality of this complex biological system. Therefore, it will be interesting to consider other transcription factors and their regulation on this system, so we can start building up a more relevant model that can recapitulate the precision of *hb* expression patterns.

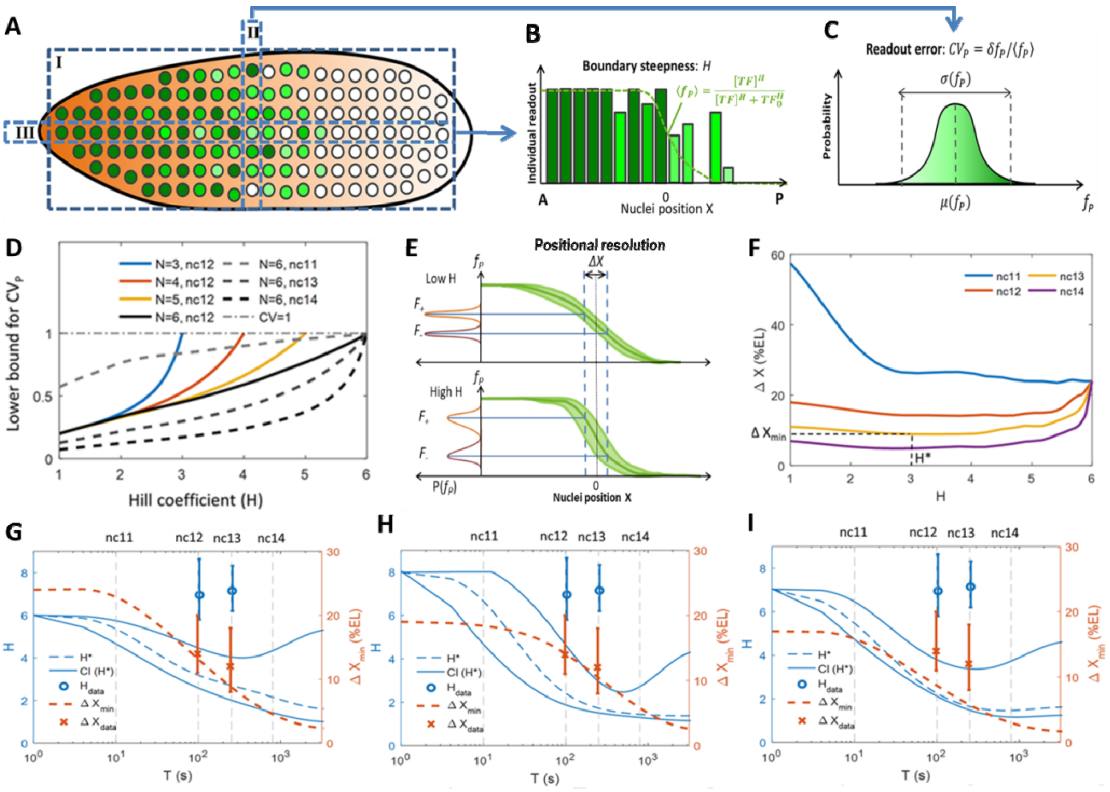


Figure 12. Summary of results from modeling analysis: A) Embryo illustration, expressing different concentrations of *hb* mRNA at the nuclei (dark green = high expression), we can see the embryo from 3 perspectives: whole embryo (I), a column of nuclei at a similar AP position (II), and a row of nuclei along the AP axis. **B)** From III we can extract where the boundary is located and its sharpness by fitting a Hill curve. **C)** From II we can extract the readout error, how different *hb* expression is in nuclei at similar input concentrations. **D)** Equilibrium model with different binding sites (N, colored lines) at steady state limited by nc length (dashed lines), standard variation of readout error (y-axis), Hill coeffient (x-axis). **E)** Illustration of how positional resolution ΔX can describe nearby nuclei where at low H, low error, similar expression; at high H, high error, different expression. **F)** Equilibrium models at steady state at different nc, which give smallest readout error at optimal H*. **G)** Comparison of results find in F with experimental data. **H)** Hybrid equilibrium/non-equilibrium model, N=6 with 3 non-equilibrium and 3 equilibrium. **I)** Two opposing gradient model. **G-I)** H (y-axis left) at different times/nc (x-axis), ΔX (y-axis right), minimal achievable ΔX (orange dashed line), optimal H* yielding minimal ΔX (blue dashed line), tolerance interval of positional resolution (solid blue lines), data ΔX (orange crosses), data H (blue circles). Modified with permission from (Tran, Desponds, et al., 2018)

Chapter 4

CONTRIBUTION OF OTHER TRANSCRIPTION FACTORS TO THE DYNAMICS OF *hunchback* TRANSCRIPTION

A synthetic biology approach to understand the rapid and precise dynamics of *hunchback* transcription

This part of my PhD work is presented in the form of a manuscript in progress:

Carmina Angelica Perez Romero[1,2], Huy Tran[1,3], Mathieu Coppey[4], Aleksandra M. Walczak[1], Cecile Fradin[1,2], Nathalie Dostatni[1]

[1]Institut Curie, PSL Research University, CNRS, Sorbonne Université, Nuclear Dynamics, Paris, France. [2]McMaster University, Biochemistry Department, Hamilton, Ontario, Canada. [3]Ecole Normale Supérieure, PSL Research University, CNRS, Sorbonne Université, Physique Théorique, Paris, France. [4]Institut Curie, PSL Research University, CNRS, Sorbonne Université, Physico Chimie, Paris, France.

My contribution:

Acquisition of all movies mentioned in the manuscript, movie processing, data analysis, preparation of figures, and manuscript writing.

Research Objective:

To explore the role played by different TFs in *hb* transcriptional dynamics using a synthetic promoter and mutant analysis approach.

Research Highlights:

- Bcd binding sites alone are able to provide transcriptional activation of the promoter but are not sufficient to reproduce the dynamics of *hb* transcription.
- Maternal Hunchback contributes in early nc 11, 12 to efficient transcription, while zygotic Hunchback plays the same role during later nc 13, 14. Both work together in setting a precise border.
- Zelda is crucial for efficient transcription and activation of *hb* expression
- Caudal may act as a negative regulator of *hb* transcription through counter-repression.

A synthetic biology approach to understand the rapid and precise dynamics of *hunchback* transcription

1. ABSTRACT

In embryos, cell differentiation often occurs downstream of spatial concentration or activity gradients of molecules called morphogens. These molecules control the expression of sets of target genes that determine cell identity. A simple model to study the functioning of morphogens is the Bicoid system in the early fruit fly embryo which allows the rapid step-like transcriptional expression of its major target gene *hunchback*. Here we use a live imaging approach to study how, in addition to Bicoid, different transcription factors affect the dynamics of *hunchback* transcription. We used a two-pronged approach which consisted in *i)* exploring the transcription dynamics of synthetic reporters containing, in addition to Bicoid sites, binding sites for the Caudal or Hunchback transcription factors and *ii)* exploring the transcription dynamics of a *hunchback* reporter (recapitulating the expression of endogenous *hunchback*) in mutants for the maternal and/or zygotic contribution of Hunchback or Zelda. We show that in contrast to Hunchback and Caudal, Bicoid is sufficient for weak transcriptional activation. In contrast, Hunchback cannot activate transcription on its own, but in presence of Bicoid it increases the probability for transcription to be activated in the anterior expression domain. Caudal limits the effect of Hunchback and may act as a plausible repressor. Our loss-of-function analysis indicates a role for the maternal contribution of Zelda and Hunchback in early activation of *hunchback*, while zygotic contribution of Hunchback also acts slightly later on. Our work delineates the importance of transcription factors, other than Bicoid, in the *hunchback* gene-regulatory network.

2. INTRODUCTION

The first two hours of fruit fly embryo development are crucial for defining body segmentation and cell identity[1,2]. During this time a complex cascade of gene expression is initiated in the

embryo, allowing each nucleus to measure its spatial position in the body and fine-tune the gene expression program responsible for its identity[3–5].

Along the antero-posterior (AP) axis of the fly embryo, the first patterning event is induced by the Bicoid (Bcd) morphogen, whose RNA is expressed in nurse cells during oogenesis and anchored at the anterior pole of the oocyte. Upon egg-laying, the translation of the protein proceeds but only at the anterior pole. Unlike the RNA, the protein can diffuse towards the posterior pole and its distribution in the embryo reaches a steady state exponential decay concentration gradient by nuclear cycle (nc) 8, 1 hour after egg laying[6]. The Bcd protein is a homeodomain containing transcription factor (TF) which activates numerous target genes in distinct anterior domains. Among the Bcd target genes, *hunchback* (*hb*) is expressed in the whole anterior of the embryo. RNA FISH experiments indicated that the posterior boundary of the *hb* expression domain is very steep and already precisely positioned at cycle 11, i.e., only 30 min after the Bcd gradient reaches steady state and only 30 min after the onset of zygotic transcription[7]. Importantly, during these 30 min, three nuclear cycles occur (nc 9, nc 10 and nc 11). Thus this time period is punctuated by three mitoses. As the transcription process stops during mitosis, the available time for the *hb* locus to “read" the concentration of the Bcd gradient is limited to the interphase of each nuclear cycle and is thus strikingly short.

The very short time in which the establishment of a precise *hb* transcription border is achieved challenges our current understanding of how a gene “measures" the concentration of a TF. An often-invoked argument to estimate the minimum time necessary for the *hb* locus to precisely measure Bcd concentration comes from statistical mechanics[1,3,8]. This argument is based on the assumption that Bcd molecules (the input) randomly reach the regulatory sequences of the gene by diffusion and are the only factor influencing the transcription status of the *hb* locus (the output). The precision of the readout is defined as the smallest change in the input resulting in a significantly different output. The maximum achievable precision is then given by:

$$\delta c/c = (4DacT)^{-1/2}$$

(an adaptation of the formula originally derived by Berg and Purcell[9–11]). In this scenario, precision is limited by the absolute concentration of Bcd molecules at the expression border (c),

the apparent size of the promoter region (a), the diffusion coefficient of the Bcd molecules (D), and the time available to generate the output (T). The actual precision of the readout of the Bcd concentration by the *hb* gene has been measured by both immuno-fluorescent detection of the Hb protein[12] and RNA FISH[7] to be $\delta c/c = 10\%$. The value of the free Bcd diffusion coefficient (D), although controversial at first, is now widely accepted to be ~7μm^2/s[13,14]. The Bcd concentration (c) has been measured by several groups independently to be ~700 molecules per nuclei at the *hb* expression boundary[6–8]. Finally, the effective size of the Bcd binding site in the promoter region (a) has been considered to be the size of a single binding site (8 base pairs, or ~3 nm). Given these parameters, according to the Berg and Purcell model, it should take at least ~25 minutes for the system to measure the Bcd concentration with the observed precision and form an expression boundary with the observed steepness. In reality the embryos achieve this precision in a much shorter time, as the total duration of the nc 11 interphase is only ~9 min. In fact, we have recently demonstrated, using the MS2-system to visualize the expression of the *hb* promoter (P2) in living embryos, that the steep *hb* boundary is established in only 3 min at each of the nuclear cycles 11, 12 or 13[15].

These experimental results indicate that it is very likely that the Berg and Purcell model[6,7,10,12] used to derive the theoretical limit to precision mentioned above is too simplistic in this particular case. Indeed, one should keep in mind that one of the major assumptions of this model is that there is a single input (here Bcd molecules diffusing in 3D and binding to a single binding site) that determines the output (here *hb* expression). Even models of increasing complexity taking into account cooperativity in Bcd binding to the six known Bcd binding sites in the promoter region are not sufficient to explain how such a steep *hb* boundary can be establish in so little time[16,17]. This leads us to question the assumption that Bcd is indeed the only required input for *hb* expression. The fact that the *hb* P2 promoter carries binding sites for several other possible TFs playing important roles in early embryonic development (Hb itself, Caudal, Zelda) suggested that multiple input are involved in the process, even if Bcd remains the primary one. We therefore set out to explore the influence of these different factors using both a bottom-up approach (by using synthetic promoters with increasing numbers of binding sites for various TFs

in order to recover the expression of the *hb* promoter) and a top-down approach (by using a loss-of-function approach and mutant fly lines missing some of those TFs).

3. RESULTS

The regulatory sequences of the *hb* gene (Fig. 1A) include a proximal promoter (P2) that controls early activation of the gene by Bcd, a distal promoter (P1) known to allow maternal *hb* expression[18], a shadow enhancer which also responds to Bcd activation and helps redefine the readout of P2 at early nc14[19,20], and finally a stripe enhancer responsible for late nc14 expression of *hb* in a posterior stripe[19,21]. P2 contains 6 binding sites for Bcd (3 strong affinity binding sites and 3 low affinity binding sites), but also one binding site for Hb, one binding site for Caudal (Cad) identified using an in silico approach and two Zelda (Zld) binding sites (Fig. 1B). We have shown recently that 750 bp including 500 bp of P2 and the downstream intron of *hb* are sufficient to recapitulate the known early expression of *hb* detected by FISH[15].

To explore how different players might affect *hb* expression and the dynamics of its border precision, we decided to build on our previous work using the MS2 approach to visualize transcription in live embryos (Section 3.1[15]). We employed a two-pronged strategy. First, we constructed synthetic MS2 reporters with a defined number of binding sites for TFs of interest (Bcd, Hb, and Cad). This allowed us to study the contribution of each TF to early transcription along the AP axis, both by itself and in combination with the others. Second, we used a genetic strategy to reduce or eliminate the maternal and/or zygotic contribution of some of the proteins known to bind the *hb* enhancer: Zld, considered as a "pioneer" factor in early fruit fly embryos[22], and Hb, which has been shown to act synergistically with Bcd in anterior development[7,23]. For this second set of experiments, we studied the expression of an MS2 reporter that contains the P2 promoter and the *hb* intron[15] (Fig. 1B, sequence and exact position of binding sites are given in Sup. Fig. 1), but lacks all of the other enhancer regions of the endogenous *hb* gene.

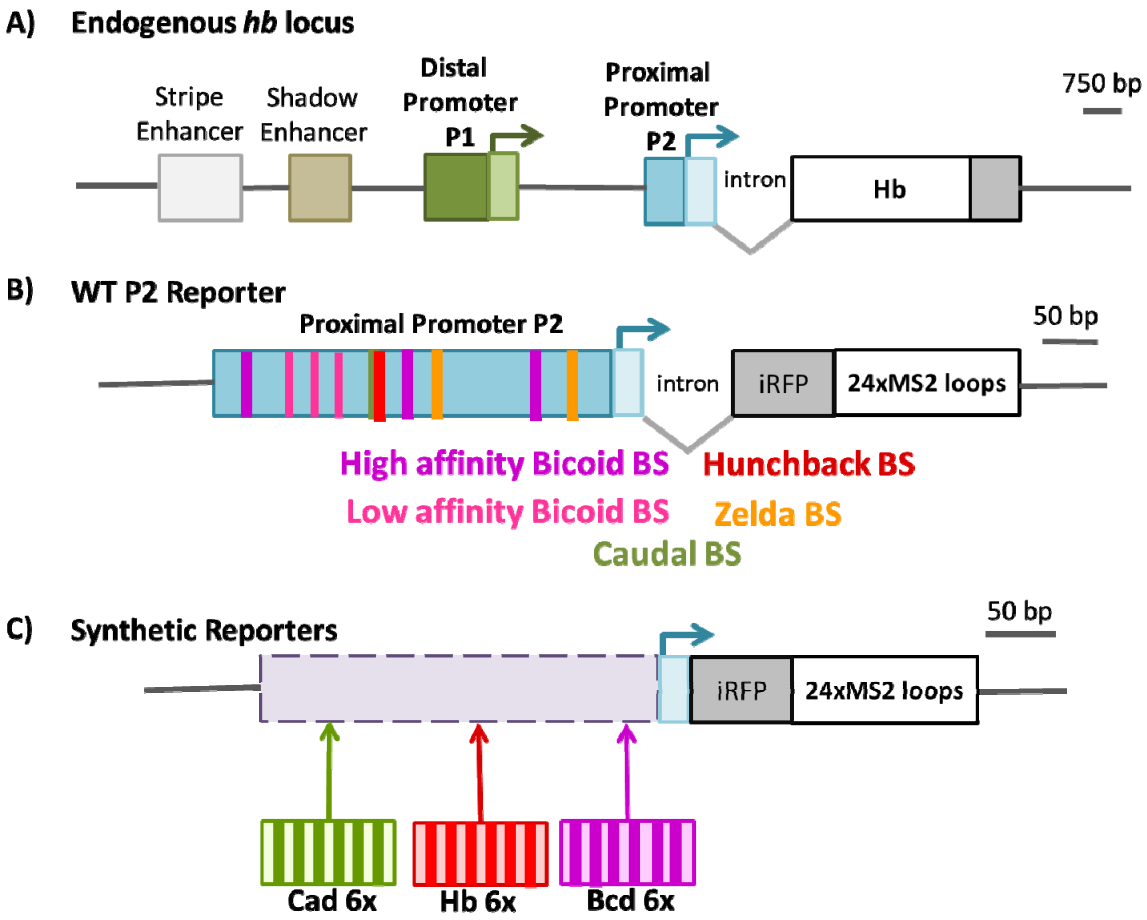


Figure 1. Schematics of the *hunchback* locus and the reporter constructs designed for this study: A) The endogenous *hb* locus: which consists, from 5' to 3', of the stripe and the shadow enhancers active at nc14, the maternally expressed P1 promoter, and the P2 promoter which controls early *hb* transcription and is the subject of our study. The P2 promoter with its transcriptional start site is located upstream of an intron and two exons. **B) The WT P2 reporter:** which consists of the first 500 bp of the P2 proximal promoter including the first non-coding exon and the intron of *hb* placed upstream of the sequence coding for the iRFP and the 24xMS2-loops. The P2 promoter carries 3 high (purple) and 3 low (pink) affinity binding sites for Bcd, 1 Hb binding site (red), 2 Zld binding sites (orange), and 1 potential Cad binding site (green). Of note, the *hb* intron which is included in this reporter, carries potential binding sites for Zld, Cad and Hb which are not included in this schematic but are shown in Supplementary Fig. 1. **C) The synthetic reporters:** were designed to contain 6 binding sites for each TF either alone or in combination. Binding sites were for Cad (Cad 6x), Hb (Hb 6x) or Bcd (Bcd 6x, strong binding site).

3.1 Using synthetic biology to study hunchback transcription

To understand the role of various TFs in the dynamics of *hunchback* transcription in early embryos, we first employed a synthetic biology approach. We created synthetic reporters containing a minimal promoter and various combinations of TF binding sites, allowing

expression of a mRNA carrying the sequence of the fluorescent protein iRFP and the MS2 cassette (24 MS2 RNA stem loops, Fig. 1C) [15,24]. To allow for quantitative comparisons, all the synthetic reporters were inserted at the same position in the fly genome using the AttB insertion site located in VK33[25]. In living embryos, which maternally express the GFP fluorescently-tagged MS2 coat protein (MCP) with no nuclear localization signal (MCP-GFPnoNLS)[26], the binding of the MCP-GFP protein to the MS2 mRNA stem loops allows visualizing the nascent mRNA expressed from the synthetic reporter. To identify nuclei and follow mitoses, embryos also express a RFP fluorescently-tagged histone H2B (His-RFP). This system provides quantitative information about transcriptional dynamics (section 3.2[16], 3.3[17]). To begin with, we decided to study the effect of Bcd, Hb and Cad binding sites on reporter expression. For this purpose, the selected TF binding sites were introduced in groups of 6, either alone or in combination, just upstream of the promoter (Fig. 1C). The exact sequence of each binding site was identical from that found in the P2 promoter and multimerized (see methods and Sup. Fig.1). The "simple" reporters (Bcd6x, Hb6x, Cad6x) contained respectively 6 strong Bcd, 6 Hb or 6 Cad binding sites. To study potential interactions between these different TFs, these binding sites were combined to obtain the following reporters: Hb6x Bcd6x, Cad6x Bcd6x and Cad6x Hb6x Bcd6x. Finally, as a control, we also built a synthetic reporter without any binding sites. We verified that this reporter was not expressed at all in the early embryo (n = 4, data not shown). Therefore, any transcription detected for a given reporter would be mediated by TF binding sites added to its sequence.

3.1.1 The wild-type P2 enhancer drives efficient *hb* transcription with a robust and precise border throughout nuclear cycles 11 to 14

First, to have a point of comparison, we imaged embryos carrying the WT P2 reporter through nc 10 to early nc 14 at 23°C using a Zeiss LSM 780 setup as explained in detail in Perez et al. 2018 (section 2.1[28]), and Lucas et al. 2018 (section 3.1[15]). After acquisition, movies where analyzed using our in house LiveFly toolbox written for Matlab, which allows nuclei segmentation and tracking using the RFP channel, followed by spot detection in the EGFP channel, as explained in Tran et al, 2018 (section 2.2[29]). This treatment allowed determination, *i)* for each nucleus and at each time point, of whether reporter transcription was ON or OFF, and *ii)* at each cycle, the fluorescent time trace of each individual transcription locus for which a MCP-GFP spot was

detected to extract its key spatio-temporal features (Fig. 2A). For an expressing locus, the initiation time (t_{init}) measures the time interval between the beginning of interphase of this particular nucleus and the first detection of the transcription signal. Similarly, the end of transcription (t_{end}) is defined as the period of time between the last time when the fluorescent signal is detected and the beginning of the next mitosis. The active time (t_{active}) is the total time during which the fluorescent signal is detected at the locus. The integral activity (ΣI) is calculated as the integral of the fluorescent trace during t_{active} and gives a relative measure of the total amount of mRNA produced at this locus. A relative measure of the average mRNA production rate (μI) is calculated as the ratio of ΣI and t_{active}.

A first useful way to look at the result of these analyses is to examine the probability for nuclei to express the reporter at a given position along the AP axis and at a given time, which we refer to as P_{spot} (x,t). P_{spot}(x,t) is best visualized in the form of a kymograph, that is a 2D plot where P_{spot}(x,t) is represented using a color scale ranging from 0 to 1 as a function of both space (where the position, x, is counted from the anterior pole and given as a percentage of egg length) and time (where the time, t, is counted from the beginning of interphase). Kymographs of P_{spot} (x,t) for each studied reporter (obtained by averaging the data obtained for nuclei within a certain interval of positions along the AP axis for several different embryos, as explained in Methods) are shown in Figs. 2B-F. At each nuclear cycle, these kymographs provide a visual comparison of the spatio-temporal expression of the different reporter genes[15].

Another interesting parameter to consider is the probability for nuclei found at a given position along the AP axis to express the reporter during a particular cycle, which we refer to as P_{ON}. This parameter is calculated by looking at the state of each nucleus over the whole nuclear cycle (as soon as a locus is ON, even if for a very short time, it is considered as an active nucleus in the calculation of P_{ON}). The profile of P_{ON} along the AP axis is shown in Fig. 2G for all the studied reporters at all studied nuclear cycles. The average values of t_{init}, t_{active}, μI and ΣI as a function of position along the AP axis are also shown, in Fig. 2H-K. They are established using only the features of expressing nuclei.

From each P_{ON} profile, we can determine the plateau value that P_{ON} reaches in the anterior part of the embryo, as well as the position of the *hb* expression border. To determine these values,

the experimentally determined P_{ON} profile was first smoothed using a moving average filter with a width corresponding to 7% EL, then the plateau value was assimilated to the maximum value reached by the smoothed profile (P_{ON}^{max}, shown in Fig. 3A), and the border position as the position at which the smoothed P_{ON} profile reached exactly one half of the plateau value (x_B, shown in Fig. 3B). We also quantify the steepness of the expression border by fitting $P_{ON}(x)$ to a sigmoidal function, as described in Tran et al. 2018[17]. More specifically, we assume that the response of the system (P_{ON}) depends on the input (Bcd concentration, c) according to a Hill equation, $P_{ON} = P_{ON}^{max}/[1+(c/c_{1/2})^H]$, where $c_{1/2}$ is the Bcd concentration at which P_{ON} falls to half its maximum value. We further assume that Bcd forms a perfect exponential concentration gradient of length λ, $c(x) = c_{max}e^{-x/\lambda}$, where c_{max} is value of the Bicoid concentration at the anterior pole. This leads to $P_{ON}(x) = P_{ON}^{max}/[1+((c_{max}/c_{1/2})e^{-x/\lambda})^H]$. Fitting this function to the data allows retrieving the Hill coefficient, H, which can be interpreted as a measure of how efficiently subtle differences in Bcd concentration are measured in this system (Fig. 3C).

A striking feature of the *hb* expression dynamics when using the WT P2 reporter (Movie 1) is that t_{init} is similar for all the studied nuclear cycles, with t_{init} ~350s (Fig. 2B). This may reflect the time needed for genome decondensation, nuclear import of Bcd and other TFs and recruitment of the transcription machinery. Note that t_{init} also includes the time necessary for the MCP-EGFP signal to build up (until it can be detected) at the transcription site while the gene has already started to be transcribed: the MS2 cassette is positioned towards the 3' end of the mRNA, meaning that the whole reporter needs to be transcribed before the MS2 loops can form and recruit MCP-EGFP proteins[30]. With the WT P2 reporter, the border position is established as early as nc 11, near the middle of the embryo (Fig. 2B & 2G, nc11). It remains in a similar position up to nc 13 (Fig. 2B & 2G, nc11 to nc13). At nc 14, we notice a ~4% egg length (EL) shift of the border position towards the anterior pole (Fig. 2B,G, Fig. 3B), which has been previously noticed in our FISH[7] and live imaging[15] studies. This could be attributed to a decrease in Bicoid availability, due to progressive degradation of the limited maternally supplied Bcd source. We can see from the kymographs that nuclei express this construct very efficiently (Fig. 2B,G, Fig. 3A), and with a very sharp border, especially at nc 13 and 14 when the Hill coefficient is maximum (Fig. 3C). At the border, one can also notice a drop in the value of t_{init} (Fig. 2H), t_{active} (Fig. 2I), the amount of RNA produced in total (Fig. 2K) and the rate at which it

is produced (Fig. 2J). This is consistent with previously observations[6,12,15], which shows that *hb* transcription follows a Bcd dose-dependent activation process at the border region, whereas in most of the anterior region expression seems to be dose-independent[6,7,12,16].

3.1.2 Bicoid alone elicits gene transcription but not efficiently

We next compared the expression of the synthetic reporter carrying 6 strong Bicoid binding sites (Movie 2) with that of the WT P2 reporter. The first striking observation in the case of the Bcd6x is the low number of nuclei with an active locus in the anterior of the embryo, indicated by low values of P_{spot} (i.e. low brightness of the kymographs in Fig. 2C) and the much lower value of P_{ON} compared to the WT (Fig. 2G, compare the orange and the bleu curves). This is consistent with very low levels of RNA production (~4.5 fold reduction compared to the WT reporter, Fig 2J,K) and very short t_{active} (~3.5 fold shorter, Fig 2I) throughout the different nuclear cycles. Of notice instead of the clear step-like pattern observed for the expression of the WT reporter, the P_{ON} of Bcd6x decays gradually from its highest point at the anterior when going towards the posterior (Fig. 2G, Fig. 3C). Thus, the expression behavior of the Bcd6x reporter reflects more directly the changes in Bcd concentration throughout the anterior region, than observed with the WT P2 reporter. Importantly, at nc 13, P_{ON} is at most 0.49 in the most anterior part of the embryo for the Bcd6x reporter, whereas it reaches 0.9 for the WT P2 reporter (Fig. 2G, Fig. 3A). Similarly, the time necessary to initiate transcription is much longer with the Bcd6x reporter than with the WT reporter at each nuclear interphase (Fig. 2H). As a consequence, while the expression border for the Bcd6x reporter is established as early as nc 11 just like for the WT reporter, it is shifted by ~6% EL towards the anterior (Fig. 2G, Fig. 3B). It remains at this position until nc 14, when it shifts by an additional ~4% EL towards the anterior. In conclusion, the analysis of the Bcd6x reporter expression indicates that 6 copies of the strong Bcd binding sites are sufficient to drive transcription activation in the anterior of the embryo. However, the expression border of the Bcd6x reporter is positioned more anteriorly (about 7% EL) than that of the WT P2 reporter, and with more nuclei-to-nuclei variation in transcription outcome since less than 50% of the loci transcribes the Bcd6x reporter in the anterior region. Therefore, other TFs are likely required, in addition to Bcd, to achieve the robust transcriptional output of the WT reporter.

3.1.3 Hunchback reduces the stochasticity of Bcd-dependent transcription

We next wanted to determine the role of Hb binding sites on both the transcription process *per se* and Bcd-dependent transcription. Therefore, we first imaged expression of a synthetic reporter carrying 6 Hb binding site (Hb6x). Although six embryos carrying this reporter were imaged, no MCP-GFP spots were detected (data not shown). This shows that, on its own, and unlike Bcd, Hb is not sufficient to activate transcription.

To determine the role of Hb on Bcd-dependent transcription, we imaged the expression of a synthetic reporter carrying 6 Hb binding sites in addition to 6 Bcd binding sites (Hb6x Bcd6x). The first surprising observation for Hb6x Bcd6x (Fig. 2D), was that from nc 11 to nc 14 P_{ON} much higher than in the case of Bcd6x and close to the levels observed with the WT P2 promoter (Fig. 2G, Fig. 3A). The steepness of the expression border is higher than for Bcd6x (Fig. 3C). Finally, t_{init} (Fig. 2H) and t_{active} (Fig. 2I) are similar to those observed for the WT P2 reporter (compare the yellow and blue curves). Also while the border of the Bcd6x Hb6x reporter is established at the same position as the WT P2 reporter at nc 11, it is first shifted towards the anterior ~2% EL at nc 12 & nc 13, and further shifted by ~8.5% EL towards the anterior at nc14 (Fig. 2G, Fig. 3B). In comparison with the Bcd6x reporter, t_{active} is only ~4 fold higher with Bcd6x Hb6x during all cycles (Fig. 2I, compare orange and yellow curves). This is accompanied by a ~2.5 fold increase in the rate of RNA production of Bcd6x Hb6x when compared to Bcd6x (Fig. 2J).

In conclusion, the presence of 6 Hb binding sites in combination with 6 Bcd binding sites allows the reporter to be expressed almost as efficiently as the WT P2 reporter, with the probability for the loci to be ON reaching a plateau of 0.9 in a step-like pattern at nc 13 (Fig. 2G, Fig. 3A). Thus, the presence of the Hb binding sites allows the transcription process to be more efficient and less stochastic than with Bcd binding sites alone. This suggests that Hb contributes to Bcd-dependent transcription by increasing the probability to be ON in the domain of expression.

3.1.4 Insights on the role of Caudal in Anterior patterning

We next inquired on the role of Cad binding sites on the expression of the synthetic reporter. We first analyzed expression of a synthetic reporter carrying only 6 Cad binding sites (Cad6x). Although four embryos were imaged in this case, no MCP-GFP spots could be detected (data not shown). This indicates that Cad alone is not sufficient to activate transcription. To determine if addition of 6 Hb binding sites could affect the transcriptional potential of 6 Cad binding sites alone, we imaged 3 embryos with the Cad6x Hb6x reporter. Similarly to the Cad6x and Hb6x reporters, the Cad6x Hb6x reporter was completely silent as no MCP-GFP spots were detected (data not shown). We therefore conclude that neither Cad nor Hb alone or in combination are sufficient to activate transcription.

To analyze the relationship between Cad and Bcd binding sites, we imaged expression of the Cad6x Bcd6x reporter. The expression of the Cad6x Bcd6x reporter is slightly stronger than the expression of the Bcd6x reporter (Fig. 2E,G, Fig. 3A). However, the values for t_{init} (Fig. 2H), t_{active} (Fig. 2I) and the rate of mRNA production (Fig. 2J,K) remain very low and similar for both constructs throughout the studied nuclear cycles (compare the purple and orange curves). The position of the Cad6x Bcd6x border is also very similar to that of Bcd6x (Fig 3B). Similarly, the P_{ON} are very similar for both reporters (Fig 2G, compare the red and purple curves). Altogether, these data suggest that Caudal does not significantly contribute to Bcd-dependent transcription.

Lastly, we decided to test all these actors together and imaged expression of the Cad6x Hb6x Bcd6x reporter. A first observation of the kymographs indicates that the expression pattern of the Cad6x Hb6x Bcd6x reporter is generally similar to that of the Hb6x Bcd6x reporter (compare Fig. 2F with Fig. 2D). However, when comparing the P_{ON} values obtained with the Cad6x Hb6x Bcd6x reporter with those obtained for the Hb6x Bcd6x reporter, one can see a slight reduction when the Cad sites are present which is more pronounced at nc 13 and nc 14 (compare the yellow and green, Fig. 2G & Fig. 3A). Similarly, when the Cad sites are present, t_{init} is longer (Fig. 2H), t_{active} is shorter (Fig. 2I) and the RNA production levels are lower (Fig. 2J,K). The Hill coefficients of the Cad6x Hb6x Bcd6x and the Hb6x Bcd6x reporter are very similar, and lower than with the WT PT reporter (Fig. 3C, compare the yellow and green plots with the blue one). The P_{ON} of the Cad6x Hb6x Bcd6x reporter remains higher than the P_{ON} of Bcd6x reporter from

nc11 to nc13 but similar at nc14 (Fig. 2G, Fig. 3A, compare the red and green). The position of the expression domain border of the Cad6x Hb6x Bcd6x reporter is very anterior similarly to that of the Bcd6x and Cad6x Bcd6x reporter (Fig 3B, compare the red, purple and green plots). Thus the posterior shift observed at nc11 and nc12 for the position of the border when adding the 6 Hb binding sites in the synthetic reporter with the 6 Bcd sites, is compensated by the further addition of Cad binding sites. These data indicate Cad can act as a possible repressor in this system.

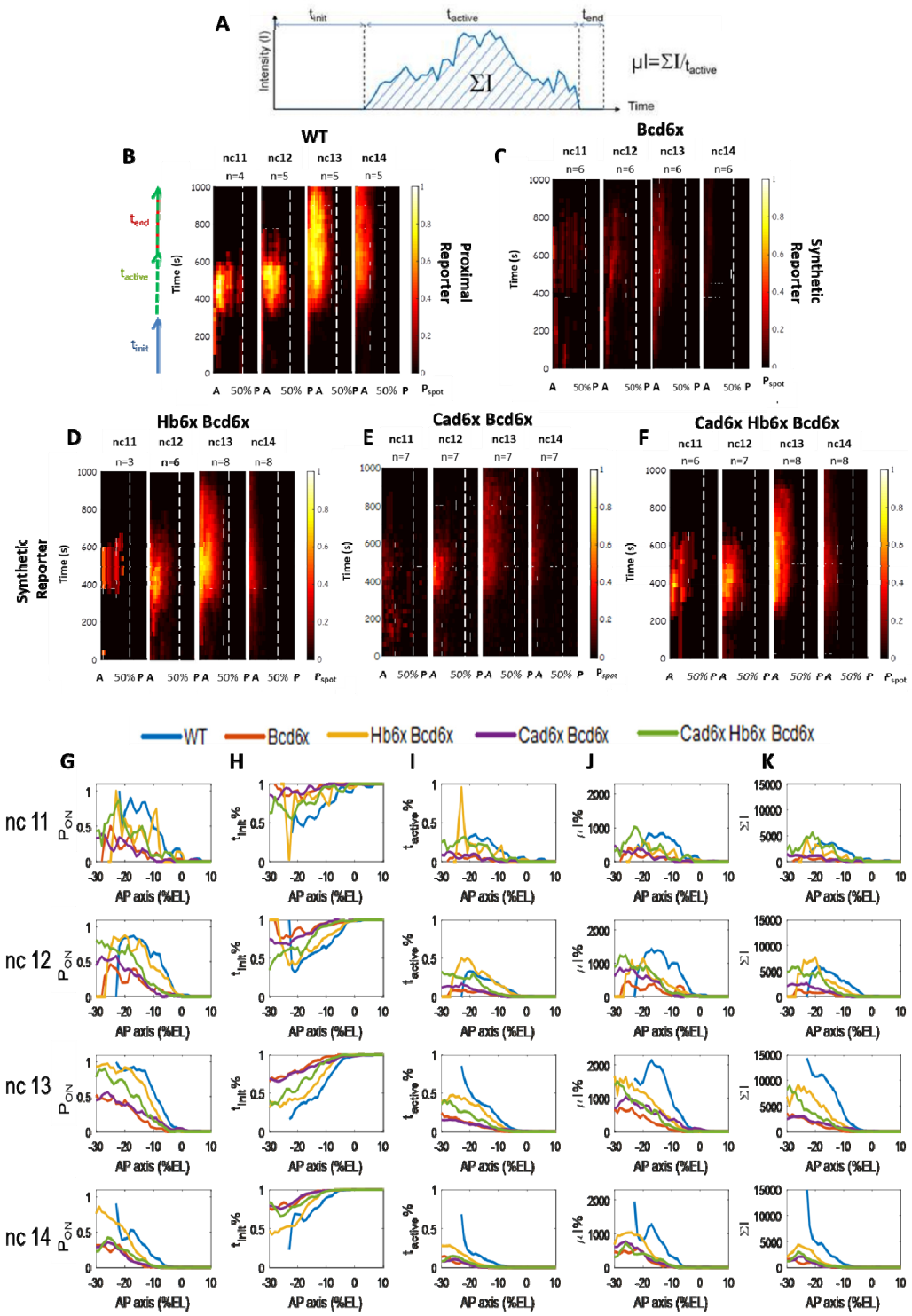


Figure 2. Spatiotemporal expression features obtained for different synthetic reporters with binding sites for Bcd, Hb and Cad in live early fly embryos: A) Schematic of a single time trace

representing the MCP-GFP fluorescent signal as a function of time measured at a single transcription site, which illustrates t_{init}, t_{active}, t_{end}, μI, and ΣI.). **B-F)** Kymographs summarizing the transcriptional dynamic for each reporter, showing the value of P_{spot} ($P_{spot} = 0$ is black and $P_{spot} = 1$ is white) as a function of both position along the AP axis (the middle of the embryo, i.e. at 50% egg length, is marked by a dashed white line) and time (time zero is the onset of interphase for each nucleus). Data from several embryos at each nuclear cycles 11, 12, 13, and 14 were pulled together (see details in section 2.2) and used to produce kymographs for each reporter (the number of embryos is indicated above each kymograph): WT **(B),** Bcd6x **(C)**, Hb6x Bcd6x **(D)**, Cad6x Bcd6x **(E)**, Cad6x Hb6x Bcd6x **(F).** The data for each reporter were then used to extract transcriptional features: P_{ON} **(G)**, t_{init} **(H)**, t_{active} **(I)**, μI **(J)**, ΣI **(K).** Color code for each reporter is indicated on the top of the G to K panels: WT (blue), Bcd6x (red), Hb6x Bcd6x (yellow), Cad6x Bcd6x (purple), Cad6x Hb6x Bcd6x (green).

These data indicate that Bicoid and Hunchback may work in synergy (compare Bcd6x and Hb6x Bcd6x in Fig. 3C), likely through two independent processes: Bcd providing positional information measure through concentration thresholds and Hb increasing the probability for a loci to be ON in the expression domain. When we compare expression of the Cad6x Hb6x Bcd6x and the Hb6x Bcd6x reporters, the presence of Cad binding sites slightly reduces the probability of the reporter to be ON (Fig. 3A) and reduces the shift in the position of the border induced by the presence of Hb binding sites (Fig. 3B). This effect of Cad is only observed in the presence of Hb binding sites, raising the possibility that the two proteins might have antagonist activities and possibly work during the same step in the activation process. Thus Caudal could act as a repressor in a normal WT context.

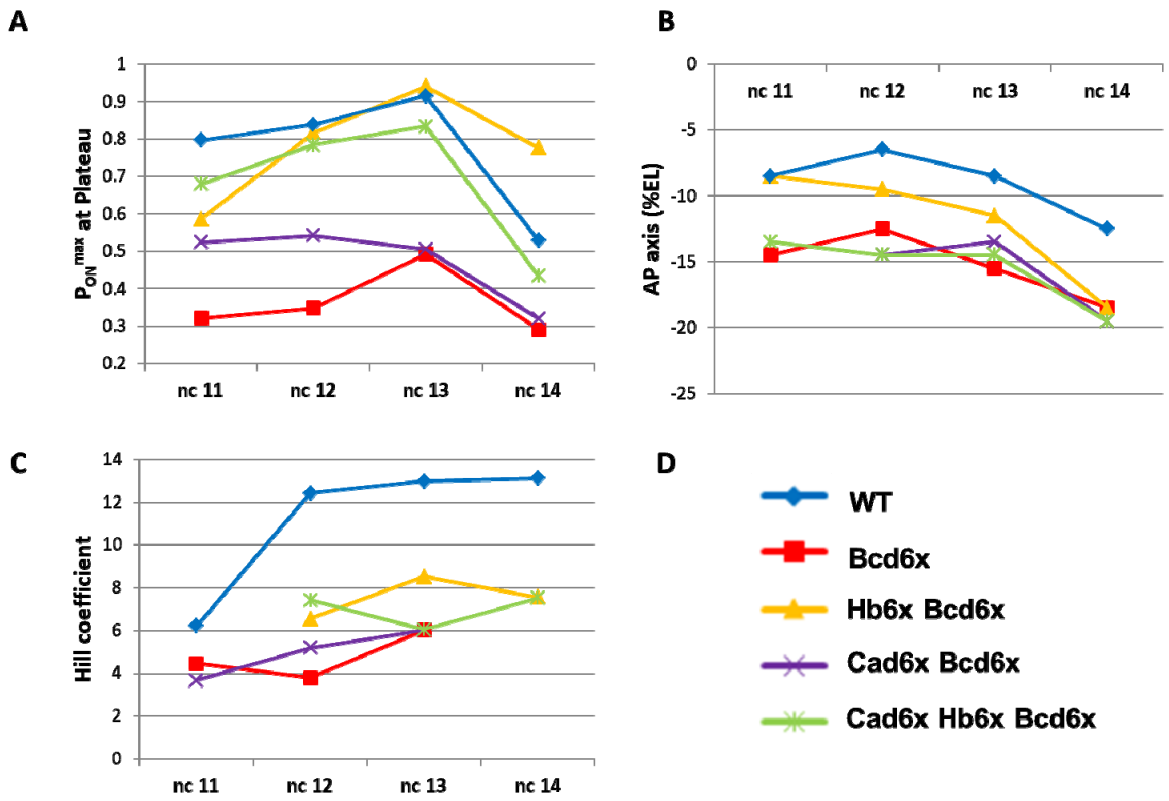


Figure 3. Characteristics of the expression border for the WT P2 reporter and the synthetic reporters: A) P_{ON}^{max} at plateau of expression, for each reporter: WT, Bcd6x, Hb6x Bcd6x, Cad6x Bcd6x, Cad6x Hb6x Bcd6x against each nuclear cycle is plotted, y-axis being the P_{ON}^{max} expression level (1 being ON, 0 being OFF). **B)** Border position $x_{1/2}$ indicated as a % of EL (the origin being in the middle of the AP axis) for each reporter at each nuclear cycle. **C)** The transcription features were fitted to a Hill curve, which allow us to measure the Hill coefficient for each reporter at each nuclear cycle.

3.2 In-silico Analysis of the P2 WT reporter reveals other possible players involved in Hunchback transcription and how Hunchback and Caudal may interact

Imaging of synthetic reporter expression shows how expression efficiency is increased by adding Bcd and Hb binding sites to the system. In combination with Hb binding sites, Cad binding sites account for a slight decrease in P_{ON} expression during later nuclear cycles particularly at nc14. We were intrigued by this result of Cad acting through Hunchback and wondered if it could be an indication that Cad might act as a repressor in the WT P2 promoter. Interestingly, the Cad site that we used in the synthetic reporters (5'-TTTATG-3') share 5 out of 6 base pairs in common

with the Hb site (5‘-AGTTTTTTGA-3’). We felt that the similarities between the two binding sites were worth exploring a bit further in the WT P2 promoter or more generally in the early TFs network (Fig. 1B, Sup. Fig. 1).

Therefore we first looked at the DNA-binding motifs in the *Drosophila* genome for both Cad and Hb to check for sequence similarities. We used the Fly Factor Survey website (which contains access to several motif databases)[31]. We were able to extract DNA-binding motifs and their position weights matrixes (PWMs) for both TFs, from databases built using different strategies: either a Bacterial one-hybrid system (B1H) to bait specific DNA-binding motifs[32], or DNase footprinting[33] followed by sequencing to get the TFs binding sequence motifs using Solexa Ilumina technology[33], or traditional Sanger sequencing[32] (Figure 4). At first sight one can appreciate that the Hb and Cad sequences share some similarities and could overlap (Fig. 4 blue squares). This first rapid analysis, however, does not weight the probability of each nucleotide found in the sequence. Therefore, we decided to perform a similarity comparison between the PWMs of the Cad and Hb DNA-binding motifs using MACRO-APE[34]. This analysis allows us to show that their sequence similarity is in fact very low ~2.7% (Fig. 4). Interestingly, a recent study found that at nc14, Cad prevents repression by Hb through binding upstream of Hb binding sites in the even-skipped stipe 2 enhancer[35]. Even though in this later case Hb acts as a repressor whereas Cad prevents this repressive activity, it will be interesting to study the possibility of similar mechanism acting in our WT P2 and synthetic reporters using direct DNA binding experiments as well as point mutations, in overlapping binding sites for Cad or Hb.

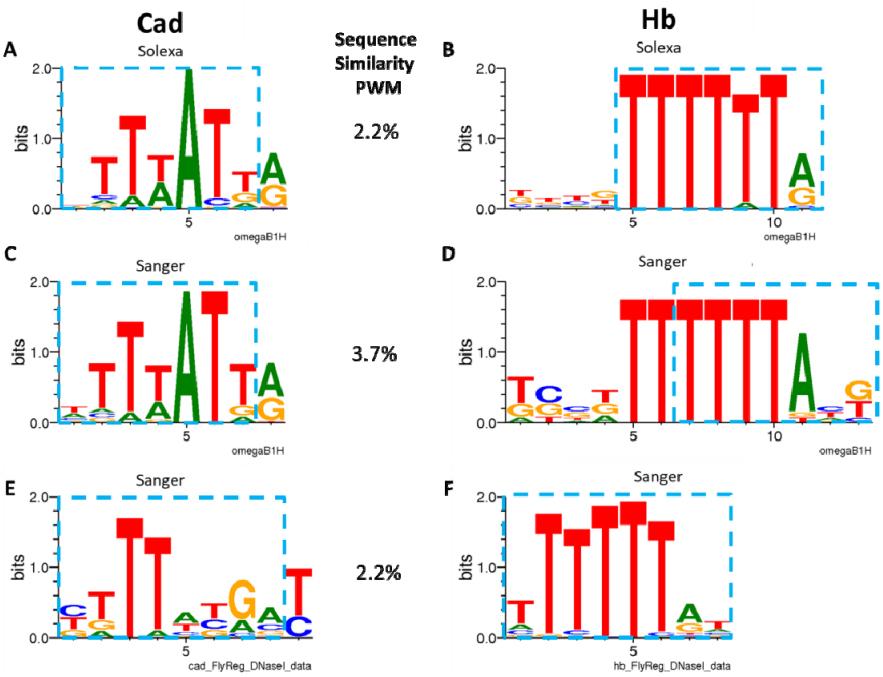


Figure 4. Caudal and Hunchback DNA-binding motifs share similarities: (A, B) DNA-binding motifs for Cad and Hb using the B1H system and Solexa sequencing[32]. **(C, D)** DNA-binding motifs for Cad and Hb using the B1H system and Sanger sequencing[32]. **(D, E)** DNA-binding motifs for Cad and Hb using DNase footprinting and Sanger sequencing[33]. The overlapping sequence similarities that can be seen by alignment of the sequence can be seen in blue squares. The similarity between each pair of position weight matrix (PWM) using the same methods can be found in the middle of the image[34].

This analysis inspired us to continue using in-silico tools to further analyze our P2 WT reporter sequence, and find binding sites for known Transcription Factor (TF) motifs in *Drosophila melanogaster*. To do this we used the PROMO virtual laboratory which allows the identification of putative TF binding sites (TRANSFAC database) in a DNA sequence, using their PWM[36] and confirmed some of them using TF sites from an older database (ooTFD)[37]. This analysis gives us a list of putative TFs that could bind to the P2 promoter and *hb* intron. To curate this data, we looked for each, the function of each TF using FlyBase[38] and for their expression profile during development using FlyMine[39], since we are interested in TFs that may affect *hb* expression in very early embryonic development. A summary of the possible players that we could further explore using our synthetic promoter can be seen in Table 1 (the position of the corresponding binding sites in the WT P2 promoter can be found in Sup. Fig. 2, and other motifs bound by proteins acting only later in development can be found in Sup. Table 1). Of note, the design of

our synthetic reporters is consistent with the 6 Bcd and 6 Hb binding sites: there are 5 binding sites for Hb and one binding site for Cad in the *hb* intron that are not depicted in Figure 1.

Transcription factors[36]	Number of Binding sites[33]	Function[38]	Expression during development[39]
Bicoid*	6	axis specification	00-04 hour, M
Hunchback*	6	axis specification	00-06 hour, M, Z
Zelda	3	genome activator	00-12 hour, M, Z
Caudal*	2	axis specification	00-06 hour, M, Z
Tailles (Tll)	8	pattern specification	00-06 hour, Z
Fushi tarazu (Ftz)	3	pattern specification	00-06 hour. Z
Paired (Prd)	12	partitioning by pair rule gene	00-06 hour, Z
Dorsal (Dl)	9	axis specification	00-06 hour, M, Z
Mother against Dpp (Mad)	12	pattern specification	00-24 hour, M, Z
Deformed (Dfd)	4	head segmentation	00-18 hour, Z
Giant (Gt)*	1	axis specification	00-06 hour, Z
Zeste (Z)	2	chromatin silencing; + gene expression; + transcription	00-18 hour

Table 1. In-silico identification of TF binding sites in P2 promoter and *hb* intron: Putative TFs binding sequences were looked using PROMO virtual laboratory. The binding motifs confirmed using TFsitescan are marked by a *. The number of binding sites motifs found is shown next. The known developmental function and the time of activity during development shown in the last two columns are taken from Flybase and Flymine, respectively. M: maternal expression, Z: zygotic expression.

3.3 Using null-mutants to understand the role of maternal and zygotic Hunchback in *hunchback* transcription

The Maternal-to-Zygotic Transition (MZT) is a major process in early embryonic development which allows the embryo to switch control from maternal inputs to its own embryonic program resulting in the expression of its own (zygotic) genome. Although we know that the Zygotic-Genome-Activation (ZGA) starts as early as nc 8 with the first detection of the zygotic genome transcription, most of the global changes in transcriptional output measured by RNA-seq studies are first detected in early nc 14[40]. However, for individual transcription factors involved in this activation, the question of how and when the transition from maternal to zygotic control happens has not yet been thoroughly investigated. We decided to explore the role of the Hb protein on *hb* transcription throughout early embryonic development (nc11-nc14) and in particular the balance between the roles played by maternal and zygotic contribution of Hb. We again systematically imaged the expression of our *hb* reporter in different *hb* mutant contexts to characterize the interaction of the Hb protein with its own P2 promoter (Fig. 5). We aimed to answer two long-standing questions: Until when does the maternally provided Hb protein act on *hb* transcription? When does zygotic Hb start acting on the activation of its own transcription?

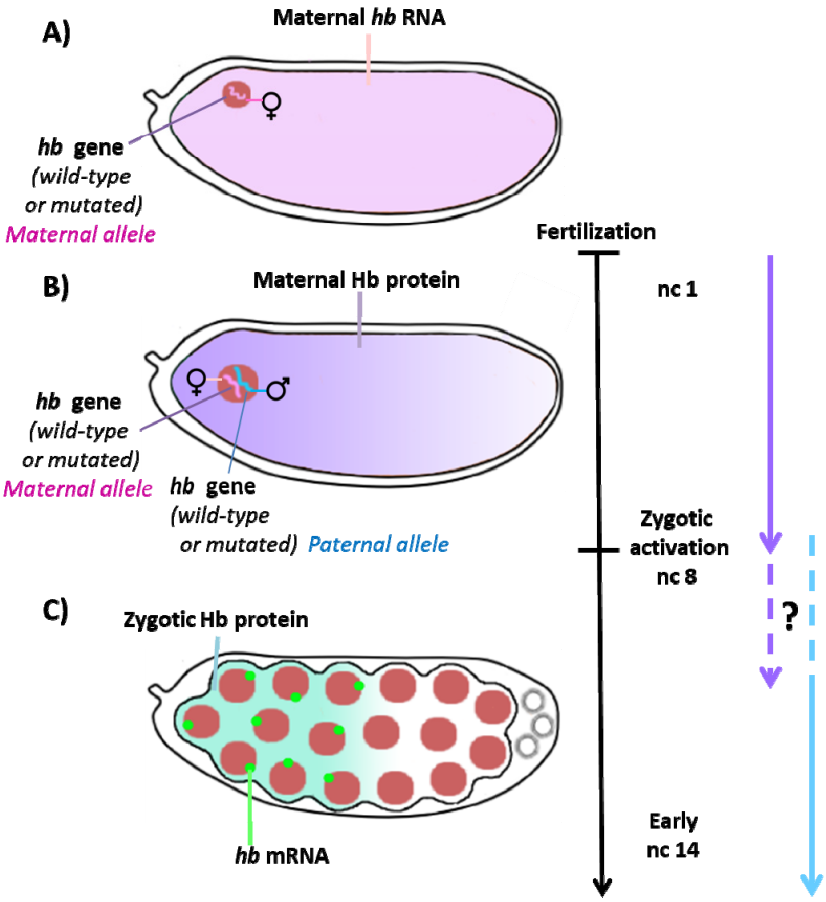


Figure 5. Schematic of Maternal and Zygotic Hunchback contribution in early fly embryonic development: A) *hb* is a maternally supplied RNA and homogeneously distributed along the whole length of the mature oocyte (in pink). In our study, it can be contributed by WT (+/+) females, heterozygous for a null allele hb^{FB} (+/-) females or germline clones induced mitotically in heterozygous $FRThb^{FB}/FRTovoD1$ females. In this later case, the maternal genotype of the diploid cell that will give rise to the oocyte is hb^{FB}/hb^{FB} (-/-). On the female chromosomes (in light pink), the *hb* maternal allele can either be wild-type allele (+) or a mutated allele (hb^{FB},-). **B)** After fertilization, the female and male pronuclei fuse together to form a diploid nucleus with one maternal (in pink) and one paternal (in bleu) genome: this is nc 1. In our study, the genotype of the male was either WT (+/+) or heterozygous for a null allele hb^{FB} (+/-) and the paternal allele were either all (+) in the case of WT males or 50% (+) and 50% (-) in the case of hb^{FB} (+/-). Shortly after egg laying, the maternally supplied RNAs are translated into proteins. Due to posterior repression by Nanos, Hb maternal protein is mostly expressed in the anterior (in purple). **C)** At nc 8 (after 7 division of the diploid nuclei), transcription of the genome of the embryo starts from maternally loaded RNA polymerase and transcription factors, among which the maternal Hb protein. This leads to the expression of the Hb protein from the embryo genome (the zygotic contribution of Hb). The maternal (in purple on the far right) and zygotic Hb (in blue on the far right) proteins are distributed in the same domain of the embryo, they are identical and their timing of expression partially overlap. It is thus difficult to know when exactly zygotic Hb protein starts playing a role in its own transcription. At nc14, the Hb protein (in light blue) is expressed in a domain with a defined border of expression in the middle of the embryo, as well as its *hb* mRNA (in green).

3.3.1 Crosses used to study maternal and zygotic effect of Hb in activating *hb* transcription

For this study we used several genotypes to produce embryos with different levels of maternal (M) and zygotic (Z) Hb ranging from the WT full dose M_1Z_1 to $M_{1/2}Z_{1/2}$ and down to M_0Z_0 (1 meaning +/+, ½ meaning +/- and 0 meaning -/-). As Z_0 (-/-) is a lethal mutation in late embryogenesis[41], these embryos could be analyzed when young but it would impossible to generate M_0 living females and this is the reason why we had to use germline mutant clones (as explained below). All the mutants used in the study are listed in Table 1.

Name	Genotype	n
M_1Z_1	M +/+ Z +/+	5
$M_1Z_{1/2}$	M +/+ Z +/-	2
$M_{1/2}Z_1$	M -/+ Z +/+	6
$M_{1/2}Z_{1/2}$	M -/+ Z -/+	4
$M_{1/2}Z_0$	M -/+ Z -/-	2
$M_0Z_{1/2}$	M -/- Z -/+	7
M_0Z_0	M -/- Z -/-	3

Table 1. Nomenclature for the mutants used for this study: In the first column, the short name of each mutant is shown, followed by the corresponding genotype in the second column and the number of embryos imaged for each mutant (n) in the third column. Embryos are listed in order of decreasing Maternal and Zygotic Hb dosage.

To obtain the mutants shown in Table 1, several crosses where performed:

- **Cross 1 -** With the help of FRT/FLP induced mitotic recombination in the germline, selected by the loss of the *ovoD1* mutation which induces female sterility, we produced hb^{FB} germline clones in heterozygotes $FRThb^{FB}/FRTovoD1$ females: the hb^{FB} allele, now named $hb^{[15]}$, is a null mutant of the *hb* gene obtained in 1987 by R. Lehmann and C. Nusslein-Volhard[41]. $FRThb^{FB}/FRTovoD1$ females were crossed to a WT male and the only embryos emerging from this cross arise from germline clone oocytes, induced by heat shock at 37°C of the larvae at L2 and L3 stages. These embryos all lacked the maternal contribution of *hb* and were heterozygous for zygotic *hb* with an hb^{FB} maternal allele and a WT paternal allele: they were thus all $M_0Z_{1/2}$ (M-/- Z-/+).

	♂ + / +
♀ - / -	100% M -/- Z -/+

- **Cross 2 -** The same females as in cross 1 were also crossed with *hb*^{FB}/+ males. These embryos emerging from the heat shocked larvae, all lacked the maternal contribution of Hb and carried a *hb*^{FB} maternal allele. 50% of them carried a WT paternal allele and were thus $M_0Z_{1/2}$ (M-/- Z-/+) and 50% carried the *hb*^{FB} paternal allele and were thus M_0Z_0 (M-/- Z-/-).

	♂ - / +
♀ - / -	50% M -/- Z -/- 50% M -/- Z -/+

- **Cross 3** - We crossed WT female flies with *hb*^{FB}/+ males to generate embryos which receive the full contribution of maternal Hb (M_1) and a wild-type maternal allele. Since the paternal genotype was *hb*^{FB}/+, 50% of these embryos received a wild-type paternal allele and were thus M_1Z_1 or a mutant paternal allele and were thus ($M_1Z_{1/2}$). We only imaged 2 embryos of this last genotype and more embryos need to be analyzed.

	♂ - / +
♀ + / +	50% M +/+ Z +/- 50% M +/+ Z +/+

- **Cross 4** - Next we crossed heterozygous *hb*^{FB}/+ females to WT males: this cross generates embryos which all received a wild-type paternal allele and half the dose of maternal Hb (they were all $M_{1/2}$) but 50% of them received a wild-type maternal allele and are thus $M_{1/2}Z_1$ while the remaining 50% received a mutated maternal allele and are thus $M_{1/2}Z_{1/2}$.

	♂ + / +
♀ - / +	50% M -/+ Z -/+ 50% M -/+ Z +/+

- **Cross 5** - We crossed hb^{FB}/+ females to hb^{FB}/+ males to generate embryos receiving half the dose of maternal Hb and which will be either zygotically WT ($M_{1/2}Z_1$, M-/+ Z+/+, 25%), heterozygous ($M_{1/2}Z_{1/2}$, M-/+ Z+/-, 50%) or homozygous ($M_{1/2}Z_0$, M-/+ Z-/-, 25%).

	♂ - / +
☿ - / +	25% M -/+ Z -/- 50% M -/+ Z -/+ 25% M -/+ Z +/+

For each of these crosses (except for cross 1), the challenge was then to identify the genotypes of the different embryos emerging from a single cross. The embryos with different genotypes emerging from a cross are undistinguishable at early stages of development, but for each cross we were hoping by imaging and analyzing enough embryos that we will be able to detect significant differences in *hb-MS2* reporter transcription to distinguish different genotypes.

Indeed, for cross 4, we were able to confidently distinguish between two populations of embryos ($M_{1/2}Z_{1/2}$ and $M_{1/2}Z_1$), due to significant differences in expression levels as evidenced in the corresponding kymographs and values of P_{ON} (Sup. Fig. 4). With cross 5, we were able to acquire enough movies to tease out three populations of embryos ($M_{1/2}Z_0$, $M_{1/2}Z_{1/2}$, $M_{1/2}Z_1$), although more embryos need to be analyzed before we can be fully confident about this sorting procedure (Sup. Fig. 5). For cross 3 only two embryos were imaged which had similar kymographs and P_{ON} expression values that were different from the WT population, which allow us to conclude that the two imaged embryos are very likely $M_1Z_{1/2}$. However, since two remains very low for statistics, more embryos will need to be imaged to have enough confident in this result (Sup. Fig. 3). Finally, for cross 2, the M_0Z_0 embryos were obtained but were extremely sick with a lot of nuclear catastrophes and a lot of aberrant chromosomal separation: it was not possible to image them, while a single embryo developed that had a similar phenotype to the one shown by the $M_0Z_{1/2}$, however we consider a single embryo not to be statically relevant to include in this study (although it can be further explore in future).

A summary of all the crosses generated with their respective genotypes, and number of embryos imaged for each genetic context can be found in Table 2.

	♂ + / +	♂ - / +
♀ - / -	Cross 1 100% M -/- Z -/+ n=7	Cross 2 50% M -/- Z -/- dying n=3 50% M -/- Z -/+ n=1
♀ - / +	Cross 4 50% M -/+ Z -/+ n=2 50% M -/+ Z +/+ n=4	Cross 5 25% M -/+ Z -/- n=2 50% M -/+ Z -/+ n=2 25% M -/+ Z +/+ n=2
♀ + / +	WT 100% M +/+ Z +/+ n=5	Cross 3 50% M +/+ Z +/- n=2 50% M +/+ Z +/+ n=0

Table 2. Summary of the genetic crosses and genotypes obtained. The female genotype is provided in the first column, while male genotype is provided in the first row, + indicates a WT allele, while - indicates the $hb^{FB.}$ null allele. The notation "female -/-" is a misuse of language : these females are *FRThb*^{FB}/*FRTovoD1* and the only living oocytes that they produces are germline clone mutants emerging from hb^{FB} (-/-) germ cells which are able to develop since they lack the female sterile *ovoD1* allele. The probability of each genotype appearance in the embryos studied is shown as a %, and the number of embryos studied for each cross is indicated by n.

3.3.2 The transition between maternal to zygotic Hb control of the *hb* gene transcription occurs at nuclear cycle 13

The kymograph for the $M_0Z_{1/2}$ embryos (obtained from a cross 1 between *FRThb*^{FB}/*FRTovoD1* females induced for germline clones and WT males) shows a very weak expression of the reporter (Fig. 6A). Lack of maternal Hb delays the transcription of the reporter gene until later nuclear cycles, with very little P_{ON} observed in nc 11 and 12 (see light blue curve in Fig. 6G, Fig. 7A). P_{ON}^{max} is extremely weak at nc11 and nc12 (Fig. 7A) and this emphasizes the importance of the Hb maternal protein for early activation (Fig. 6G-I). We also observe that P_{ON} reaches its maximum of 0.43 at nc13 and is below this value at nc14 (Fig. 6G, Fig. 7A), indicating that the effect of zygotic Hb on reporter gene expression is only detected at nc13 and after. This late activation may be due to the low level of total Hb (no maternal Hb content and only half the dose of zygotic Hb) which could significantly interfere with the capacity of the context to efficiently activate *hb* transcription, as evidenced by the much lower levels of P_{ON} throughout the nuclear cycles compared to the WT (~2 fold difference at nc 13 and ~10 fold at earlier cycles, Fig. 6G & Fig. 7A). The same observation holds true for the duration of the transcription and initiation time (Fig. 6I-J) and the average and total amount of RNA produced (Fig. 6J-K). Of note, the border

position is set at nc 12 at ~-12.5% EL and stays similar throughout all the nuclear cycles, including nc 14 (Fig. 4F), in contrast with the more dynamic border observed for the WT (Fig 7B).

Only very few embryos were obtained for the $M_{1/2}Z_0$ genetic background (Sup. Fig. 5), and their kymographs only show an extremely weak expression of the reporter (Fig. 6B), with levels similar to the ones seen for the $M_0Z_{1/2}$ embryos above (Fig. 6A). However, it is interesting to notice that a single copy of maternal Hb remains sufficient for the reporter to be ON at nc12 and nc 13 before missing drastically in nc 14 as seen in the P_{ON} graphs (Fig. 6G, green curve). The P_{SPOT} at plateau values at nc 12 and 13 are around 0.3 and show a decrease of ~3 fold when compared to WT expression (Fig. 7A). t_{init} and t_{active} of the reporter expression in these $M_0Z_{1/2}$ mutants are similar to the ones of $M_{1/2}Z_0$ (Fig. 6H,I, compare light blue and green plots), while the amount of RNA produced is slightly higher in the $M_{1/2}Z_0$ genetic background (Fig. 6K, light blue and green plots). In the $M_{1/2}Z_0$ genetic context, the border position is set at nc 12 at the same position of WT in nc 11, after the border recedes toward the anterior in nc 13 before returning to its original set position in nc 14 (Fig 7B, green plots).

Kymograph of *hb-MS2* expression in the $M_{1/2}Z_{1/2}$ genetic background (Fig. 6C) indicates that a copy of maternal Hb and a copy of zygotic Hb are not enough for efficient activation of *hb* transcription, as seen by the ~2-3 fold decay of P_{ON} expression levels compared to the WT (compare purple curve to dark blue curves Fig. 6G, Fig.7A). It is interesting to notice that the values of P_{ON}, t_{init}, t_{active} obtained for this mutant remain similar throughout the nuclear cycles to those of $M_0Z_{1/2}$ (Fig. 6G-I, compare the purple and the light blue plots). The border position in the $M_{1/2}Z_{1/2}$ genetic background is: it is set at -8% EL towards the anterior at nc12 and it remains at this position at nc13 to only slightly shifted towards the anterior at nc14 (Fig. 7B, purple plot). This result shows that reducing the dose of maternal and zygotic Hb by half does not greatly modify the position of the border.

Expression in the $M_{1/2}Z_1$ genetic background highlights the contribution of a full copy of zygotic Hb for *hb* activation, which is closer to expression in wild-type (Compare Fig. 6D & Fig. 6F). The P_{ON} profiles (Fig. 6G) show that the probability to be ON (yellow curve) improves through the nuclear cycles reaching its peak efficiency at nc 13 and are very similar to WT (dark blue

curve). Although, P_{ON}, tinit and t_{active} levels are similar to those of the WT in the different nuclear cycles (Fig. 6G-I), the RNA production levels remain significantly lower compared to those of the WT (Fig. 6J-K). This is an indication that when it is on, expression is either weaker or more "bursty" (i.e. with periods of inactivity between the t_{init} and t_{end}) in the $M_{1/2}Z_1$ than in the WT. More thorough analysis of the traces might help distinguish these two possibilities. Surprisingly, the border position of *hb-MS2* reporter expression in this genetic background is slightly shifted towards the posterior of the embryo by ~2% EL at each nuclear cycle when compared to the wild-type background (Fig. 7B, compare the yellow and dark blue plots) and we don't really have an explanation for this observation.

From the kymograph of $M_1Z_{1/2}$ mutants (Fig. 6E), we can detect expression at nc11 and this indicates that maternal Hb is active at early nuclear cycles: it helps to maintain strong expression up until nc 13. The profiles (red plot) show that the transcription activation levels in the $M_1Z_{1/2}$ genetic background are similar to those of the WT and $M_{1/2}Z_1$ throughout all studied nuclear cycles (Fig. 6G for P_{ON} and Fig. 7A for P_{ON}^{max}, compare the dark blue, the red and yellow plots). The observed values of t_{init} and t_{active} indicate that maternal Hb, supplemented by one copy of zygotic Hb, is sufficient for sustained activation of *hb* early on (Fig. 6H,I). In contrast, a single dose of zygotic Hb is not enough to have effective activation at nc 14 although initiation time remain similar as in a WT background (Fig. 6H,I). Also, the total RNA produced remains significantly lower than in the WT genetic background but similar to the rest of the mutants (Fig. 6K). As above, the fact that the relative amount of RNA produced is very different in the $M_1Z_{1/2}$ and WT genetic background while the other features are similar might be an indication of how bursty the traces are. A thorough analysis of those traces, in particular the measurements of the kinetic parameters of promoter function (k_{ON} & k_{OFF})[16] will shed light on this question.

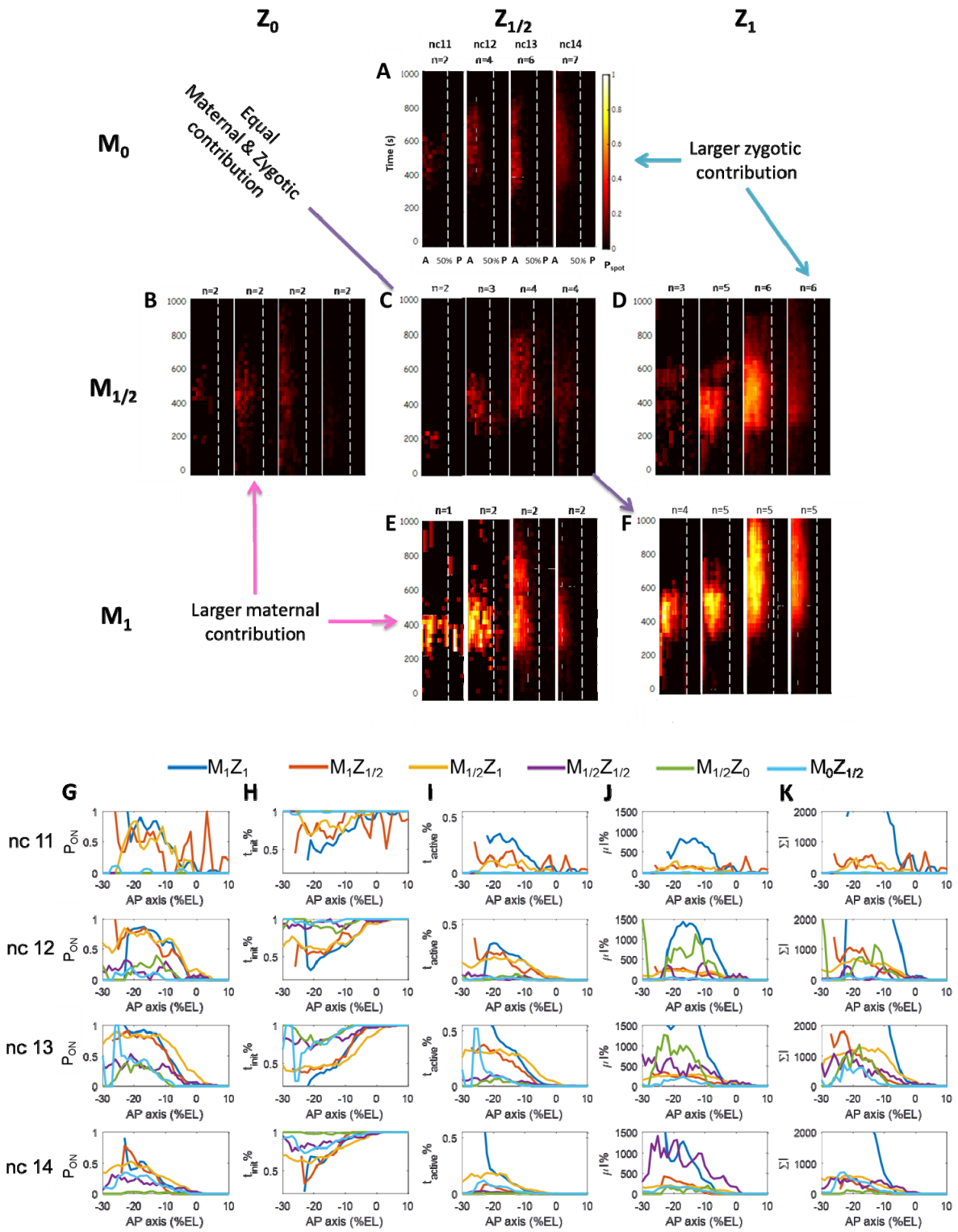


Figure 6. Contribution of Maternal and Zygotic Hunchback in early development the *hb-MS2* reporter expression: A-F) Kymographs summarizing the transcriptional dynamics observed for the *hb-*

MS2 reporter in different mutant backgrounds. Heat maps indicate the level of P_{Sspot}, the probability for a locus to be ON (as in Figure 1), they are aligned to show the difference in concentration dependence of maternal and zygotic Hb. $M_0Z_{1/2}$ **(A),** $M_{1/2}Z_0$ **(B)**, $M_{1/2}Z_{1/2}$ **(C)**, $M_{1/2}Z_1$ **(D)**, $M_1Z_{1/2}$ **(E)** and M_1Z_1 **(F)**. The pulled data were used to extract each transcriptional feature: P_{ON} **(G)**, t_{init} **(H)**, t_{active} **(I)**, μI **(J)**, ΣI **(K).** Color code for each genetic background is indicated on the top of the G to K panels: M_1Z_1 (blue), $M_1Z_{1/2}$ (red), $M_{1/2}Z_1$ (yellow), $M_{1/2}Z_{1/2}$ (purple), $M_{1/2}Z_0$ (green) and $M_0Z_{1/2}$ (light blue).

In conclusion we can see that the switch from maternal to zygotic control happens in between nc12 and nc13 (depending on the feature analyzed). This analysis allowed to tease out the importance of maternal Hb in early activation of *hb* transcription and the importance of zygotic Hb for later nuclear cycles where the embryo pursues its own developmental program towards gastrulation. Our analysis also reveal some redundancy between the partially overlapping maternal *vs* zygotic contribution of Hb : this can be really highlighted when analyzing the P_{ON}^{max} in Fig 7A, which show very similar expression behavior of the *hb-MS2* reporter expression in WT (dark blue), $M_1Z_{1/2}$ (red) or $M_{1/2}Z_1$ (yellow) genetic backgrounds which all express at least 1.5 doses of Hb. In contrast, in genetic backgrounds expressing less than 1.5 the dose of Hb, $M_0Z_{1/2}$ (light blue), $M1_{/2}Z_0$ (green) or $M_{1/2}Z_{1/2}$ (purple), expression of the hb-MS2 reporter exhibits much lower values of the probability to be ON. This observation is further highlighting the role of Hb in increasing the probability for the promoter to be ON, as observed in the first part of our analysis with the synthetic reporters. It is also intriguing to see that both maternal and zygotic Hb seems to play opposite roles in setting the border position. Finally, it is interesting to see that the Hill coefficient does not seem to change greatly between mutants (Fig. 7C). Finally, a more careful study of these data including statistics and an increased number of embryos for more robust statistics and modeling to study the sharpness of the border is required to strengthen our analysis.

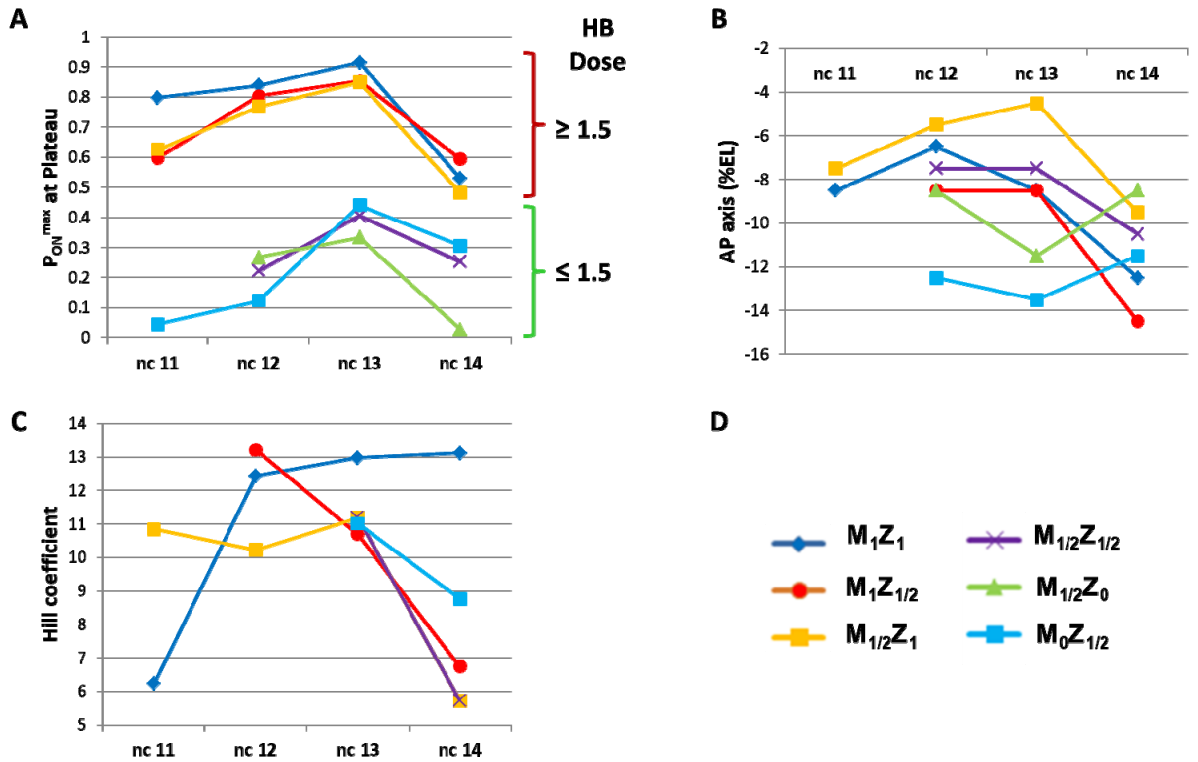


Figure 7. Characteristics of the *hb-MS2* reporter expression border in genetic backgrounds expressing various levels of maternal and/or zygotic Hb: A) P_{ON}^{max} at plateau of expression of the *hb-MS2* reporter for each genetic backgrounds at most nuclear cycles. **B)** Border position $x_{1/2}$ of the *hb-MS2* reporter (% of EL placing the origin in the middle of the AP axis with negative values towards the anterior) in each genetic background at most nuclear cycles. **C)** The transcription features were fitted to a Hill curve, which allow to measure the Hill coefficient for each *hb-MS2* reporter for each genetic backgrounds at most nuclear cycles. Color code for genetic backgrounds: M_1Z_1 (dark blue), $M_1Z_{1/2}$ (red), $M_{1/2}Z_1$ (yellow), $M_{1/2}Z_{1/2}$ (purple), $M_{1/2}Z_0$ (green) or $M_0Z_{1/2}$ (light blue).

3.4 Zelda is important for early *hunchback* activation and efficient transcription

In the last decade, evidence has accumulated suggesting that maternally-deposited Zld acts as a "pioneer factor" in *Drosophila*: Zld binds to a large number of enhancers or promoters of developmental genes, priming them for rapid zygotic transcription upon activation by a specific differentiation signal[22,42,43]. Zld was shown to bind to the *hb* P2 promoter and to be required for its early expression[22]. As shown in Chapter 3.1, Zld binding sites located in the original MS2 cassette allowed expression of the *hb-MS2* reporter in the posterior of the embryo[15].

In order to gain quantitative information on the role of Zld in *hb* transcription, we imaged embryos from females heterozygous (Zld M-/+) for the *zld*^{294} null allele[22] (amorphic allele,

molecular evidence) and from *zld*^{294} germline clones induced using FRT/FLP mitotic recombination and which allow producing embryos completely devoid of any *zld* maternal contribution (Zld M-/-). Importantly, even though there is a strong maternal contribution of Zld in the embryo, the phenotype is completely penetrant and not rescued by the zygotic contribution[22], and this is in contrast to what we observed for Hb (see above).

When we analyzed progeny of heterozygous females mutant for Zld, *i.e.* when one dose of maternal Zld is removed from the embryo (Zld M-/+), we can see in the kymographs that *hb* expression is extremely low at nc 11, increases during the nc12 and nc13 and decreases at nc 14 (Fig. 8B). P_{ON} levels are ~1.5 fold lower than those seen in the WT reporter (compare orange and blue curves in Fig. 8D) and a similar trend can be seen for t_{init} (Fig. 8E), t_{active} (Fig. 8F), RNA production rate (Fig. 8G) and RNA production levels (Fig. 8H). In Zld M-/+ embryos, the *hb-MS2* reporter has a P_{ON}^{max} at plateau lower than in WT (Fig 9A) and a more anterior border of expression (Fig. 9B). It is shifted of ~3% EL towards the posterior at nc 13 when P_{ON} reaches a maximum level (compare Fig. 9A with Fig. 9B). The sharpness of the border remains lower than that of WT as seen from the value of the Hill coefficient (Fig. 9C), indicating that Zld contributes to the sharpening of the border.

In *zld* germline clones, *i.e.* when all maternal Zld content was removed from the embryo (Zld M-/-), the situation becomes even clearer and confirms the importance of Zld in *hb* expression. This is particularly obvious early on, as no expression can be detected at nc 11, and even later on as only very little expression is detected (Fig. 8C). The P_{ON} values are reduced ~2 fold compared to Zld M-/+ (compare yellow and orange curves, Fig 8D) and by ~4 fold compared to the WT (yellow and blue curves, Fig 8D, Fig. 9A). We observe the same trend for the other transcription features, showing a systematic reduction in activation: it starts later on average at each cycle (t_{init} is higher compare yellow curve with red and blue ones in Fig. 8E), it does not last as long (t_{active} is lower compare yellow curve with red and blue ones in Fig. 8F) and the average RNA production rate is lower (compare yellow curve with red and blue ones in Fig. 8G). At nc12, the border position is set very anteriorly at ~-23.5% EL (Fig. 9B) and then it is shifted of ~9% EL towards the posterior at nc 13 where it remains until nc 14 (Fig. B).

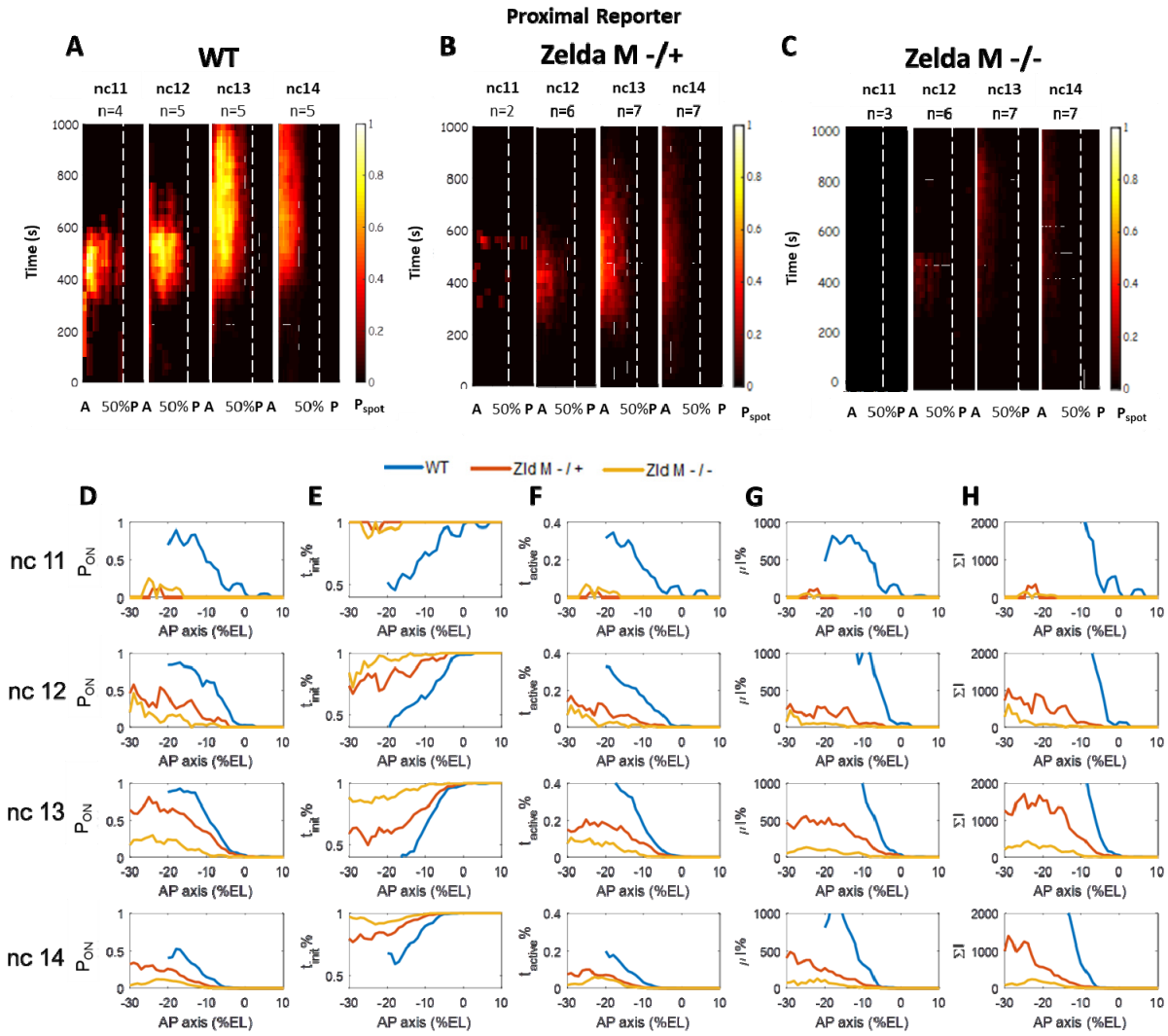


Figure 8. Contribution of Zld to the *hb*-MS2 reporter expression: A-E) Kymographs summarizing the transcriptional dynamics for the *hb-MS2* reporter in Zelda mutant backgrounds: WT **(A)**, Zld M-/+ **(B)**, Zld M-/- **(C).** The data were used to obtain the following transcriptional feature: P_{ON} **(D),** t_{active} **(E)** t_{active} **(F)**, μI **(G)**, ΣI **(J).** Color code for panel **D** to **H** is blue for the WT, orange for Zld M-/+, yellow for Zld M-/-.

In conclusion, this analysis indicates that maternal Zelda is required for an efficient early transcription of *hb* transcription. It appears required both to increase the probability for a locus to be ON in the expression domain (P_{ON}) and for most of the transcription features to reach their WT values. It should be noticed however that the behavior of the expression of the *hb-MS2* reporter in the presence of one dose of maternal Zld is much weaker than expected, given that the

expression of endogenous *hb* is almost not affected at nc11 in the same genetic background[15]. The reason for this huge difference is not clear but likely reflects the difference of regulatory sequences between the *hb-MS2* reporter and the endogenous gene. Also, the posterior shift of the border at nc13 and nc14 for the *hb-MS2* reporter in absence of maternal Zld is not clear neither. The possibility that it might be due to the zygotic contribution of Zld will have to be tested.

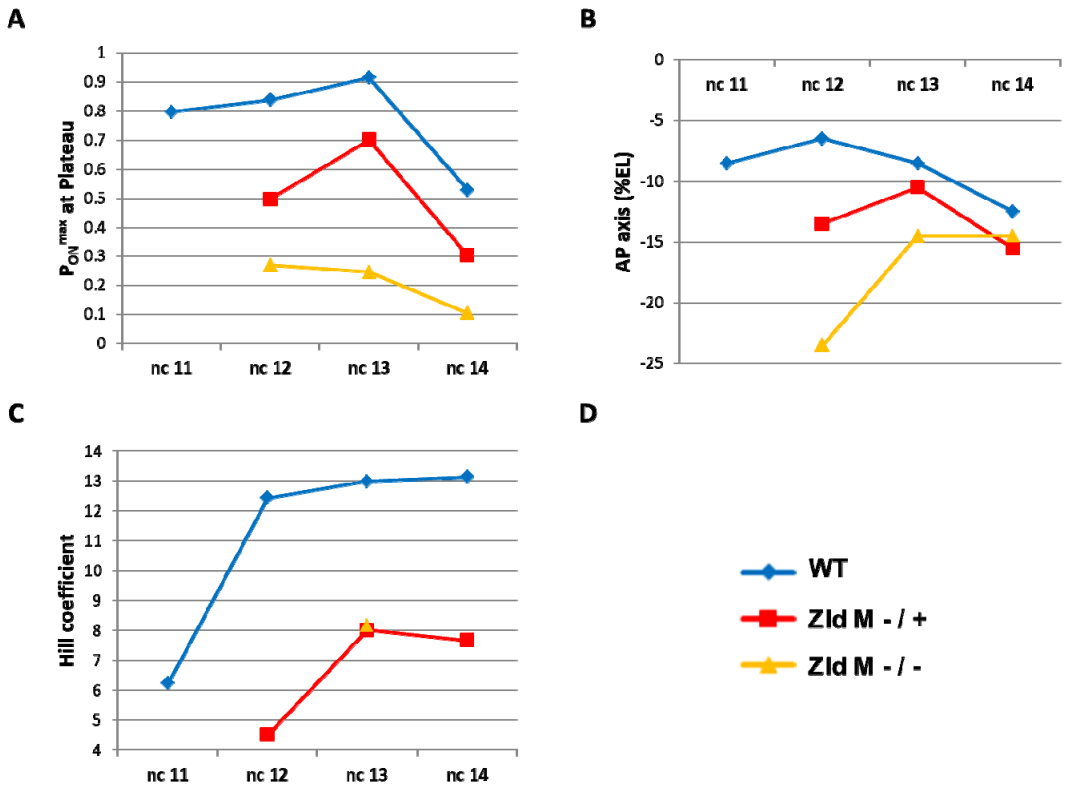


Figure 9. Characteristics of the *hb-MS2* reporter expression border in genetic backgrounds expressing various levels of maternal Zelda: A) P_{ON}^{max} at plateau of expression of the *hb-MS2* reporter in WT, Zld M-/+, and Zld M-/- genetic backgrounds is indicated through the nuclear cycles. **B)** Border position of the *hb-MS2* reporter expression domain is indicated for the various nuclear cycles. **C)** Hill coefficient for each *hb-MS2* reporter for each genetic backgrounds at most nuclear cycles. Color code for genetic backgrounds: WT (blue), Zld M-/+ (red), Zld -/- (yellow).

4. CONCLUSION

In conclusion, using synthetic MS2 reporters combined with genetics provided a better understanding of how different transcription factors play a role in *hb* regulation. Bcd is sufficient for transcription activation but alone only lead to inefficient transcription: only few loci turned ON in the expression domain and Bcd-dependent transcription bears stochastic features. Zld and Hb both increase the number of loci turned ON by Bcd in the expression domain and thus contribute to reduce the stochasticity of Bcd-dependent transcription. Our genetic analysis allowed determining that maternal Hb is important for early activation (nc11 and nc12), while zygotic Hb had a similar role in later activation (nc13 and nc14). The relay from maternal to zygotic contribution of Hb is passed on during nc12 and nc13. Finally, the role of Cad is for the moment less clear: it might be a repressor of *hb* which surprisingly requires the binding of Hb to the promoter (Figure 10). We hope our system will be found useful by the research community to further the study of other developmental genes and to better understand the gene-regulatory network in complex biological systems such as embryos.

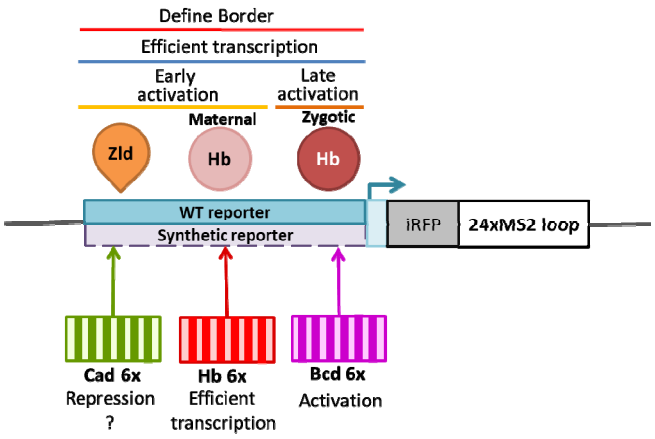


Figure 10. Summary of our findings on *hb* transcription: Our WT *hb* P2 reporter is shown in blue, with a transcription initiation site, the sequence coding for a fluorescent protein marker (iRFP) followed by the sequence of 24xMS2 stem loops. We analyzed the expression of this reporter in genetic backgrounds modifying the level of expression of different TFs such as: Zld (orange), maternal Hb (light red) and zygotic Hb (dark red). This analysis indicates that Zld and maternal Hb contribute to early activation, while zygotic Hb is important for activation at later nuclear cycle. Both Zld and Hb contribute to border positioning and efficient expression. From our synthetic promoter studies we were able to conclude that Bcd is sufficient for activation, Hb for efficient transcription and Cad might act as a repressor.

5. MATERIAL AND METHODS

Drosophila stocks

The transgenic MS2 reporters used were as follows: WT P2 reporter (*hb*-MS2) is a random P-element insertion on the 3rd chromosome and correspond to the insertion analyzed in[15] and synthetic reporters were all form Pacman vectors inserted on the 3rd chromosome in the VK33 landing site[25]. A transgene expressing the MCP-GFP without nuclear localization signal (MCP-GFPnoNLS)[26] was used for better signal to noise ratio. A transgene expressing the Histone 2Av fused to mRFP under the control of its own promoter inserted on the second chromosome (#23651 from Bloomington) was used a marker to follow the nuclei. For mutant analysis, female germline clones were induced using heat-shock FLP recombination (3 hours at 37°C at L2 larva stage)[44]: for *zld* mutants, we use *Ovo*^{D1} *hs-FLP FRT19A* chromosome (#23880, Bloomington) with the *zld*^{294} *FRT19A* chromosome carrying a null allele of *zld* (a gift from C. Rushlow[22]); for *hb* mutants, we use *hs-FLP* insertion on the first chromosome p{hsFLP}22 (#8862, Bloomington), the *Ovo*^{D1} *FRT82B* chromosome (#2149, Bloomington) and the *hb*^{FB} allele[41] recombined with the FRT recombination sequence in 82B by Dr. Tanguy Lucas. All stocks were raised at 25°C. All the stocks and crosses used for this study are shown in Table 4 for the synthetic reporters, and Table 5 for the mutant analysis.

For synthetic analysis:

♀	♂	Construct
w- ; His-RFP/Cyo; MCP-GFPnoNLS	; ; *hb*-MS2/TM3[15]	**WT**
	w- ; ; Bcd6x-MS2_VK33	**Bcd6x**
	w- ; ; Hb6x-MS2_VK33	**Hb6x**
	w- ; ; Hb6xBcd6x-MS2_VK33	**Bcd6x Hb6x**
	w- ; ; Cad6x-MS2_VK33	**Cad6x**
	w- ; ; Cad6xHb6x-MS2_VK33	**Cad6x Hb6x**

	w- ; ; Cad6xHb6xCad6x-MS2_VK33	**Cad6x Hb6x Bcd6x**
	w- ; ; empty_VK33	**Empty vector**

Table 4. Stocks and crosses used for Synthetic promoter study. Females (column 1) were crossed with males (column 2) and embryos (named in column3) emerging form the cross were imaged and analyzed.

For mutant analysis:

♀	♂	Construct
yw *hs*-FLP ; His-RFP,MCP-GFPnoNLS ; *hb^{FB} FRT82 / FRT82B Ovo^{D1}* **Induced for germline clones (3hrs at 37°C at L2 stage)**	w-; ; *hb*-MS2/TM3^{15}	**$M_0Z_{1/2}$** (the chromosome carrying the *hb-MS2* also carries wild-type *hb*)
	w- ; ; *hb^{FB}FRT82BNeo^{e},hb*-MS2/TM6B	**M_0Z_0** (the chromosome carrying the *hb-MS2* also carries the *hb^{FB}* allele)
w- ; His-RFP,MCP-GFPnoNLS ; *hb^{FB}FRT82BNeo^{e}*/TM6B		**$M_{1/2}Z_0$ / $M_{1/2}Z_{1/2}$** (the chromosome carrying the *hb-MS2* also carries the *hb^{FB}* allele)
	w-; ; *hb*-MS2/TM3^{15}	**$M_{1/2}Z_{1/2}$ / $M_{1/2}Z_1$**(the chromosome carrying the *hb-MS2* also carries wild-type *hb*)
w- ; His-RFP,MCP-GFPnoNLS ; MKRS/TM6B	w- ; ; *hb^{FB}FRT82BNeo^{e},hb*-MS2/TM6B	**$M_1Z_{1/2}$** (the chromosome carrying the *hb-MS2* also carries the *hb^{FB}* allele)
zld^{294} FRT19A/ Ovo^{D1}hs*-FLP *FRT19A	w-; ; *hb*-MS2/TM3^{15}	**Zld Maternal - / -**

; His-RFP,MCP-GFPnoNLS ; Induced for germline clones (3hrs at 37°C at L2 stage)		
***zld*^{294}*FRT19A*/FM7 ; His-RFP,MCP-GFPnoNLS ;**	; ; *hb*-MS2/TM3[15]	**Zld Maternal - / +**

Table 5. Stocks and crosses used for Mutant analysis study. Females (column 1) induced for germline clones or not (as indicated) were crossed with males (column 2) and embryos (named in column3) emerging form the cross were imaged and analyzed. Importantly, in the case of the female induced for germline clones, it is only after mitotic recombination when *OvoD1* is exchanged with the second hb^{FB} or zld^{294} alleles that the females will produce eggs. If the induction for germline clones was not efficient, these females do not lay eggs because the *ovoD1* induced sterility. Of note, if induction for germline clones was efficient, the embryo produced carries a mutated maternal allele.

MS2 reporters

In the WT P2 reporter, the *hb* P2 promoter and intron (-300 to +445) is placed upstream of the iRFP fluorescent protein coding sequence followed by the ΔZld-MS2 cassette with 24 stem loops (iRFP-24MS2)[15]. We have shown that this reporter is sufficient to recapitulate the known early expression of *hb* detected by FISH[7,15].

In the synthetic reporters, the iRFP-24xMS2[15] cassette was placed under the control of a minimal *hsp70* promoter (-40 to +70)[45] and the sequence of the various binding sites were inserted by cloning just upstream of the promoter. These included 6 strong Bcd sites (GGGATTA, separated by 8 randomized bp), 6 Hb sites (AAAAAAC, separated by 8 randomized bp) and 6 Cad sites (TTTATG, separated by 2 randomized bp). When combining the synthetic reporters a spacer of 21 random bp between each TFs motif was left. An empty vector without any insertion of binding sites was used as control for unspecific expression. These synthetic reporters, inserted in Pacman[25] vectors, were engineered and prepared by Aurélien Guillou & Laurent Belkadi, and inserted at the VK33 site by Bestgene.

Live imaging

Live imaging was performed as by me, as explained in Section 2.1[15,28]. Females and males were crossed overnight, then put in a collection cage with grape juice agar plates. After an hour of egg laying, embryos were collected, manually dechorionated, glued to a coverslip using homemade heptane glue, flattened by slight tapping with a dissecting needle and then embedded with Halocarbon oil (VWR). Mounted embryos were imaged at ~ 23°C using an incubator/cooler from Tokai Hit (INUC-KPP Series) on a Zeiss LSM780 confocal microscope with a Zeiss 40x (1.4 NA) A-Plan objective. Image stacks (pixel size 0.2 µm, pixel dwell time 0.54 µs, 8 bits per pixels, confocal pinhole diameter 92 µm, distance between consecutive images in the stack 0.5 µm, ~1200x355 pxl, ~30 z-stacks) of the lower cortical region of the embryo close to the middle of the AP axis were collected continuously. The GFP and RFP proteins were excited with a small fraction of the power output of a 488nm and a 568nm laser. Images were acquired using the ZEN black software (Zeiss). For each embryo, a tiled image of the midsection of the whole embryo was obtained, by stitching 3 separate images, from which the position of the anterior and posterior poles could be inferred.

Image processing and data extraction

Live imaging processing was performed in MATLAB as in Section 2.2[15,29]. Following imaging, movies were checked manually to verify all the nuclei included in data analysis are fully imaged in their depth and incompletely imaged nuclei (mostly nuclei at the periphery of the imaging field) are excluded. Nuclei segmentation was performed in a semi-automatic manner using our own software LiveFly[29]. Only nuclei that exist throughout the nuclear interphase were used for the analysis. The MS2 spot detection was performed in 3D using a thresholding method. An average filter was applied before thresholding each Z plane of the processed time point for noise reduction. MS2 spots were detected by applying a threshold equal to ~2 fold above background signal and only the spots composed of at least 5 connected voxels were retained. The intensity of the 3D spot was calculated as the sum of the voxel values of each Z-stack. We manually checked each movie to ensure the correct spot detection. The data from a segmented movie indicated for each nucleus its segmentation profile, identifier number and the intensity trace of the detected spot over time. During mitosis, nuclei divide in waves, usually from the embryo poles. Therefore, nuclei at the anterior pole may produce MCP-MS2 spots earlier due to either earlier

chromatin decondensation or earlier reentrance of Bcd into the nucleic space[6]. We correct for this by realigning all the intensity traces by choosing the origin for time for each trace when its own nucleus was first separated from its sibling[15].

During the late stages of acquisition of this study, the microscope unfortunately had a significant reduction in fluorescence intensity (~4 fold or more) and although it was repaired its output was never as strong as before. Therefore, we decided to change laser power intensity while keeping the rest of the parameters constant since they might influence quantification. We subtracted the background intensity from the newly imaged WT embryos and normalized their MS2 spot intensity (over all nuclei and all nc) to that of embryos collected previous to the problem with the microscope. Using this correction, we found that the P_{ON} parameter was not affected, but other parameters such as t_{active}, μI, ΣI needed normalization to be quantitatively informative and comparable to previous measurements (Sup. Fig. 7). Therefore after this analysis we normalized the rest of the acquired data in the same way.

This problem brought into focus the importance of having imaging standards to check that facility microscopes continue to be at their best for imaging overtime. We thus repurposed the common fluorescein standard used for calibration of microscopes for fluorescence correlation spectroscopy (FCS)[7,13], to avoid this kind of common problems in the future. A detailed explanation of how this standards are made can be found in Perez-Romero et al. 2018[28].

In-silico analysis

Cad and Hb DNA-binding motifs and PWMs were find the Fly Factor Survey (http://mccb.umassmed.edu/ffs/)[31]. Similarity comparison between two PWMs was done using the online MACRO-APE (http://opera.autosome.ru/macroape/compare/new)[34].

The search of putative TF binding sites in the WT *hb* P2 promoter and intron, where carried using the PROMO virtual laboratory (http://alggen.lsi.upc.es/)[36] and confirmed using TFsitescan (http://www.ifti.org/)[37]. The TFs function was taken from FlyBase (http://flybase.org)[38] and their expression profile during development from FlyMine (http://www.flymine.org)[39].

6. REFRENCES

1. Wieschaus, E. Positional Information and Cell Fate Determination in the Early Drosophila Embryo. 1–13 (2016). doi:10.1016/bs.ctdb.2015.11.020

2. Jaeger, J. The gap gene network. *Cell. Mol. Life Sci.* **68,** 243–74 (2011).

3. Yang, Z., Wu, X., Yang, N. & Liu, F. Noise transmission during the dynamic pattern formation in fly embryos. *Quant. Biol.* **6,** 15–29 (2018).

4. Green, J. B. A. & Sharpe, J. Positional information and reaction-diffusion: two big ideas in developmental biology combine. *Development* **142,** 1203–11 (2015).

5. Briscoe, J. & Small, S. Morphogen rules: design principles of gradient-mediated embryo patterning. *Development* **142,** 3996–4009 (2015).

6. Gregor, T., Wieschaus, E. F., McGregor, A. P., Bialek, W. & Tank, D. W. Stability and nuclear dynamics of the bicoid morphogen gradient. *Cell* **130,** 141–52 (2007).

7. Porcher, A. *et al.* The time to measure positional information: maternal hunchback is required for the synchrony of the Bicoid transcriptional response at the onset of zygotic transcription. *Development* **137,** 2795–2804 (2010).

8. Fradin, C. On the importance of protein diffusion in biological systems: The example of the Bicoid morphogen gradient. *Biochim. Biophys. Acta - Proteins Proteomics* **1865,** 1676–1686 (2017).

9. Berg, H. C. & Purcell, E. M. Physics of chemoreception. *Biophys. J.* **20,** 193–219 (1977).

10. Bialek, W. & Setayeshgar, S. Physical limits to biochemical signaling. *Proc. Natl. Acad. Sci. U. S. A.* **102,** 10040–5 (2005).

11. Kaizu, K. *et al.* The Berg-Purcell Limit Revisited. *Biophys. J.* **106,** 976–985 (2014).

12. Gregor, T., Tank, D. W., Wieschaus, E. F. & Bialek, W. Probing the Limits to Positional Information. *Cell* **130,** 153–164 (2007).

13. Abu-Arish, A., Porcher, A., Czerwonka, A., Dostatni, N. & Fradin, C. High mobility of bicoid captured by fluorescence correlation spectroscopy: implication for the rapid establishment of its gradient. *Biophys. J.* **99,** L33-5 (2010).

14. Drocco, J. A., Grimm, O., Tank, D. W. & Wieschaus, E. Measurement and perturbation of morphogen lifetime: effects on gradient shape. *Biophys. J.* **101,** 1807–15 (2011).

15. Lucas, T. *et al.* 3 minutes to precisely measure morphogen concentration. *PLOS Genet.* **14,** e1007676 (2018).

16. Desponds, J. *et al.* Precision of readout at the hunchback gene. *PLoS Comput. Biol.* 1–35 (2016). doi:10.1101/063784

17. Tran, H. *et al.* Precision in a rush: Trade-offs between reproducibility and steepness of the

hunchback expression pattern. *PLOS Comput. Biol.* **14,** e1006513 (2018).

18. Margolis, J. S., Borowsky, M., Shim, C. W. & Posakony, J. W. A Small Region Surrounding the Distal Promoter of the hunchback Gene Directs Maternal Expression. *Dev. Biol.* **163,** 381–388 (1994).

19. Perry, M. W., Boettiger, A. N. & Levine, M. Multiple enhancers ensure precision of gap gene-expression patterns in the Drosophila embryo. *Proc. Natl. Acad. Sci. U. S. A.* **108,** 13570–5 (2011).

20. Perry, M. W., Bothma, J. P., Luu, R. D. & Levine, M. Precision of hunchback expression in the Drosophila embryo. *Curr. Biol.* **22,** 2247–52 (2012).

21. Margolis, J. S. *et al.* Posterior stripe expression of hunchback is driven from two promoters by a common enhancer element. *Development* **121,** 3067–77 (1995).

22. Liang, H.-L. *et al.* The zinc-finger protein Zelda is a key activator of the early zygotic genome in Drosophila. *Nature* **456,** 400–3 (2008).

23. Simpson-Brose, M., Treisman, J. & Desplan, C. Synergy between the hunchback and bicoid morphogens is required for anterior patterning in Drosophila. *Cell* **78,** 855–65 (1994).

24. Lucas, T. *et al.* Live imaging of bicoid-dependent transcription in Drosophila embryos. *Curr. Biol.* **23,** 2135–9 (2013).

25. Venken, K. J. T., He, Y., Hoskins, R. A. & Bellen, H. J. P[acman]: a BAC transgenic platform for targeted insertion of large DNA fragments in D. melanogaster. *Science* **314,** 1747–51 (2006).

26. Bothma, J. P. *et al.* Enhancer additivity and non-additivity are determined by enhancer strength in the Drosophila embryo. *Elife* **4,** (2015).

27. Tran, H. *et al.* Precision in a rush: trade-offs between positioning and steepness of the hunchback expression pattern. *plos co* **provisiona,** 305532 (2018).

28. Perez-Romero, C. A. *et al.* Live Imaging of mRNA Transcription in Drosophila Embryos. in 165–182 (Humana Press, New York, NY, 2018). doi:10.1007/978-1-4939-8772-6_10

29. Tran, H. *et al.* LiveFly: A Toolbox for the Analysis of Transcription Dynamics in Live Drosophila Embryos. in 183–195 (Humana Press, New York, NY, 2018). doi:10.1007/978-1-4939-8772-6_11

30. Ferraro, T. *et al.* New methods to image transcription in living fly embryos: the insights so far, and the prospects. *Wiley Interdiscip. Rev. Dev. Biol.* **5,** 296–310 (2016).

31. Brodsky Michael, Wolfe Scot, Zhu Julie, O. J. Fly Factor Survey. Available at: http://mccb.umassmed.edu/ffs/. (Accessed: 14th October 2018)

32. Noyes, M. B. *et al.* Analysis of Homeodomain Specificities Allows the Family-wide Prediction of Preferred Recognition Sites. *Cell* **133,** 1277–1289 (2008).

33. Bergman, C. M., Carlson, J. W. & Celniker, S. E. Drosophila DNase I footprint database: a systematic genome annotation of transcription factor binding sites in the fruitfly, Drosophila

melanogaster. *Bioinformatics* **21,** 1747–1749 (2005).

34. Vorontsov, I. E., Kulakovskiy, I. V & Makeev, V. J. Jaccard index based similarity measure to compare transcription factor binding site models. *Algorithms Mol. Biol.* **8,** 23 (2013).

35. Vincent, B. J. *et al.* Hunchback is counter-repressed to regulate even-skipped stripe 2 expression in Drosophila embryos. *PLOS Genet.* **14,** e1007644 (2018).

36. Farré, D. *et al.* Identification of patterns in biological sequences at the ALGGEN server: PROMO and MALGEN. *Nucleic Acids Res.* **31,** 3651–3 (2003).

37. Ghosh, D. Object-oriented transcription factors database (ooTFD). *Nucleic Acids Res.* **28,** 308–10 (2000).

38. Gramates, L. S. *et al.* FlyBase at 25: looking to the future. *Nucleic Acids Res.* **45,** D663–D671 (2017).

39. Lyne, R. *et al.* FlyMine: an integrated database for Drosophila and Anopheles genomics. *Genome Biol.* **8,** R129 (2007).

40. Lefebvre, F. A., Benoit Bouvrette, L. P., Bergalet, J. & Lécuyer, E. Biochemical Fractionation of Time-Resolved Drosophila Embryos Reveals Similar Transcriptomic Alterations in Replication Checkpoint and Histone mRNA Processing Mutants. *J. Mol. Biol.* (2017). doi:10.1016/j.jmb.2017.01.022

41. Lehmann, R. & Nüsslein-Volhard, C. hunchback, a gene required for segmentation of an anterior and posterior region of the Drosophila embryo. *Dev. Biol.* **119,** 402–417 (1987).

42. Harrison, M. M., Li, X.-Y., Kaplan, T., Botchan, M. R. & Eisen, M. B. Zelda Binding in the Early Drosophila melanogaster Embryo Marks Regions Subsequently Activated at the Maternal-to-Zygotic Transition. *PLoS Genet.* **7,** e1002266 (2011).

43. Haines, J. E. & Eisen, M. B. Patterns of chromatin accessibility along the anterior-posterior axis in the early Drosophila embryo. *PLOS Genet.* **14,** e1007367 (2018).

44. Chou, T. & Perrimon, N. The Autosomal FLP-DFS Technique for Generating Germline Mosaics in Drosophila melanogaster. *Genetics* **144,** (1996).

45. Ronchi, E., Treisman, J., Dostatni, N., Struhl, G. & Desplan, C. Down-regulation of the Drosophila morphogen bicoid by the torso receptor-mediated signal transduction cascade. *Cell* **74,** 347–55 (1993).

7. SUPPLEMENTARY FIGURES

CTGTCGACTCCTGACCAACG**TAATCCCCA**TAGAAAACCGGTGGAAAATTCGCAGCTCGCT**G**

CTAAGCTGGCCATCC**GCTAAGCT**CCCGGATCAT**CCAAATCC**AAGTGCGCATA**ATTTTTTG**T

TTCTGCTC**TAATCCAGA**ATGGAT**CAAGAGCG**CAATCCTCAATCCGCGATCCGTGATCCTCG

ATTCCCGACCGATCCGCGACCTGTACCTGACTTCCCGTCACCTCTGCCCATCTAATCCCTTG
ACGCGTGCATCCGTCTACCTGAGCGATATATAAACTAATGCCTGTTGCAATTGTTCAGTCAG
TCACGAGTTTGTTACCACTGCGACAACACAACAGAAGCAGCACCAATAATATACTTGCAAAT
CCTTACGAAAATCCCGACAAATTTGGAATATACTTCGATACAATCGCAATCATACGCACTGA
GCGGCCACGAAACGGTAGGATATTGTTAGCCATTACCAAGTGTCTCCATTTTGAACACAAAA
TCACTCAAATCGCCTTCAGGGGGTGGGTGCCGCCCAGCCACCCCTGACGTATTTTTTGTTAG
GGGTGGTGCCGCAAGCACACCAAAAAAAGAGAAAAAAAAAATAAAAGCGAGGAAAAATAA
AATGAAAAACAAGCGGAAAAAAAGAGGAAAAAACTCGACGCAGGCGCAGTGCATGAATGA
ATAAATGAATATGCCCACTAACCCCACTCTCTCTGTTTTCTTATCCATTACAGCCGTCTAGAG
CCGCCAAGGGATCT

Zelda **Low Affinity Bicoid** **High Affinity Bicoid** **Huncback**

TATA box ***hunchback* intron** **putative Caudal**

Supplementary Figure 1. Sequence of WT P2 promoter and the *hb* intron: In be, the TATA box of the P2 promoter and highlighted in grey the *hb* intron. Potential binding sites for each transcription factor studied in here marked by different colors: Zelda (yellow), low affinity Bicoid binding sites (pink), high affinity Bicoid binding sites (purple), Hunchback in (red), and the putative Caudal binding site (highlighted in green).

Transcription factors[36]	Number of Binding sites[31,36]	Function[38]	Expression during development[39]
Adf-1	3	neurons development	adult fly
TF B	1	general TF	all the time
Ecdysone-induced protein 74EF (E74EF)	3	endocytosis and pupal development	late pupal stages
DTF-1	1	not known	not known

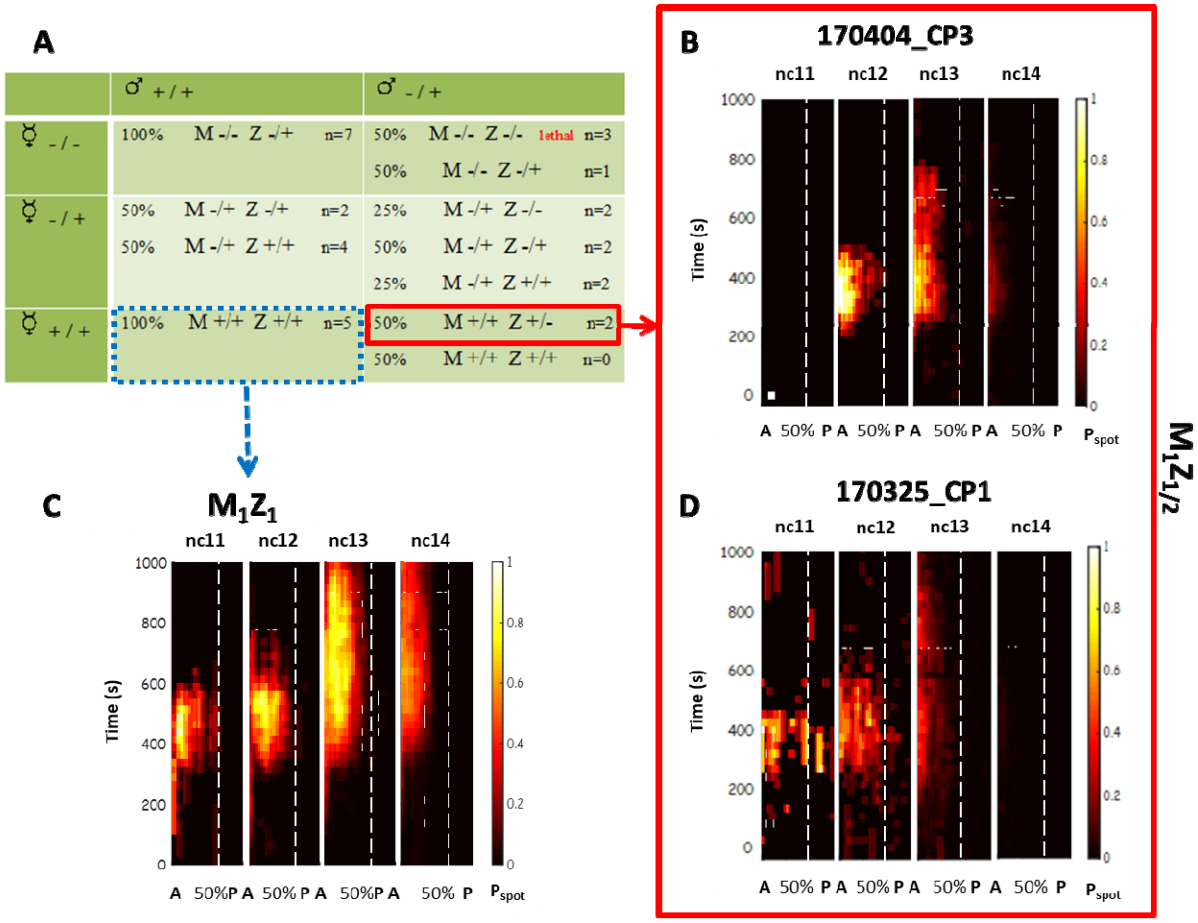
Tramtrack (Ttk)	1	nervous system, trachea and eye development	6-18 hour embryo
Glia cell missing (GCM)	1	neurons development	4-14 hour embryo
E2F	2	general TF	all time
Doublesex (Dsx)	3	reproductive system development	early pupal stage and adult flies

Supplementary Table 1. Transcription Factors Binding Sites in the P2 promoter and the *hb* intron which have only a late expression during development: Putative transcription factors were found using PROMO virtual laboratory, the number of binding sites motifs found is show next. The relevant function at the time of development of each factor is taken from Flybase and Flymine, as one can see this list concerns the ones that are at later stages of development which are out of our study scope.

CTGTCGACTCCTGACCAACGTAATCCCCATAGAAAACCGGTGGAAAATTCGCAGCTCGCTG
CTAAGCTGGCCATCCGCTAAGCTCCCGGATCAT**CCAAATC**CAAGTGCGCA*TAATTTTT*TGTTT
CTGCTCTAATCCAGAATGGATCAAGAG**CGCAATCCTCAATC**CGCGATCCGTGATCCTC**GAT**
TCCCGACCGATCCGCGACCTGTACCTGACTTCCCGTCACCTCTGCCCATCTAATCCCTTGAC
GCGTGCATCCGTCTACCTGAGCGATATATAAA*CTAATG*CCTGTTGCAATTGTTCAGTCAGTCA
CGAGTTTGTTACCACTGCGACAACACAACAGAAGCAGCACCAATAATATACTT**GCAAATC**C
TTACGAAAATCCCGACAAATTTGGAATATACTTCGA**TACAATCGCAATC**ATACGCACTGAG
CGGCCACGAAACGGTAGGATATTGTT*AGCCATTACC*AAGTGTCTCCATTTTGAACA**CAAAAT**
CACTCAA**ATC**GCCTTCAGGGGGTGGGTGCCGCCCAGCCACCCCTGACGTATTTTTTGTTAGG
GGTGGTGCCGCAAGCACACCAAAAAAAGAGAAAAAAAAAATAAAAGCGAGGAAAAATAA
AATGAAAAACAAGCGGAAAAAAAGAGGAAAAAAACTCGACGCAGGCGCAGTGCATGAATG
AATAAATGAATATGCCCACTAACCCCACTCTCTCTGTTTTCTT*ATCCATTA*CAGCCGTCTAG
AGCCGCCAAGGGATCT

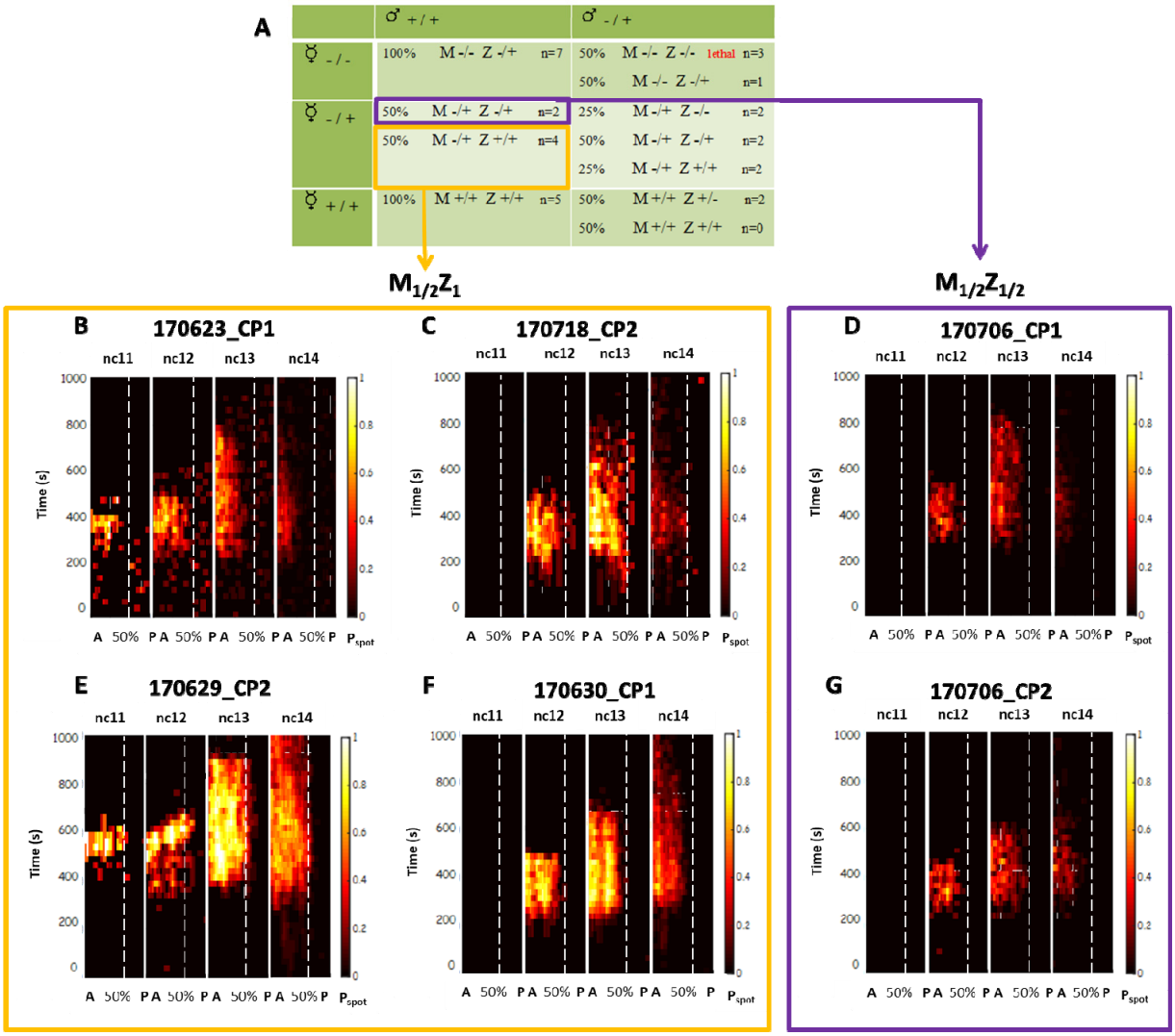
Tailless Fushi-tarazu **Paired** Dorsal **Mother against Dpp** ***Deformed*** **Giant** Zeste

Supplementary Figure 2. Putative binding sites for other player in the WT P2 promoter and intron: Sequence of the WT P2 promoter sequence and of the *hb* intron (highlighted in grey), with the potential binding sites find for other transcription factors identified with an in-silico analysis using PROMO. Bindings sites marked by different colors: Tailless (yellow), Fushi-tarazu (red), Paired (underlined), Dorsal (dark blue), Mother against Dpp (underline light blue), Deformed (in italics), Giant (underlined pink) and Zeste (green).



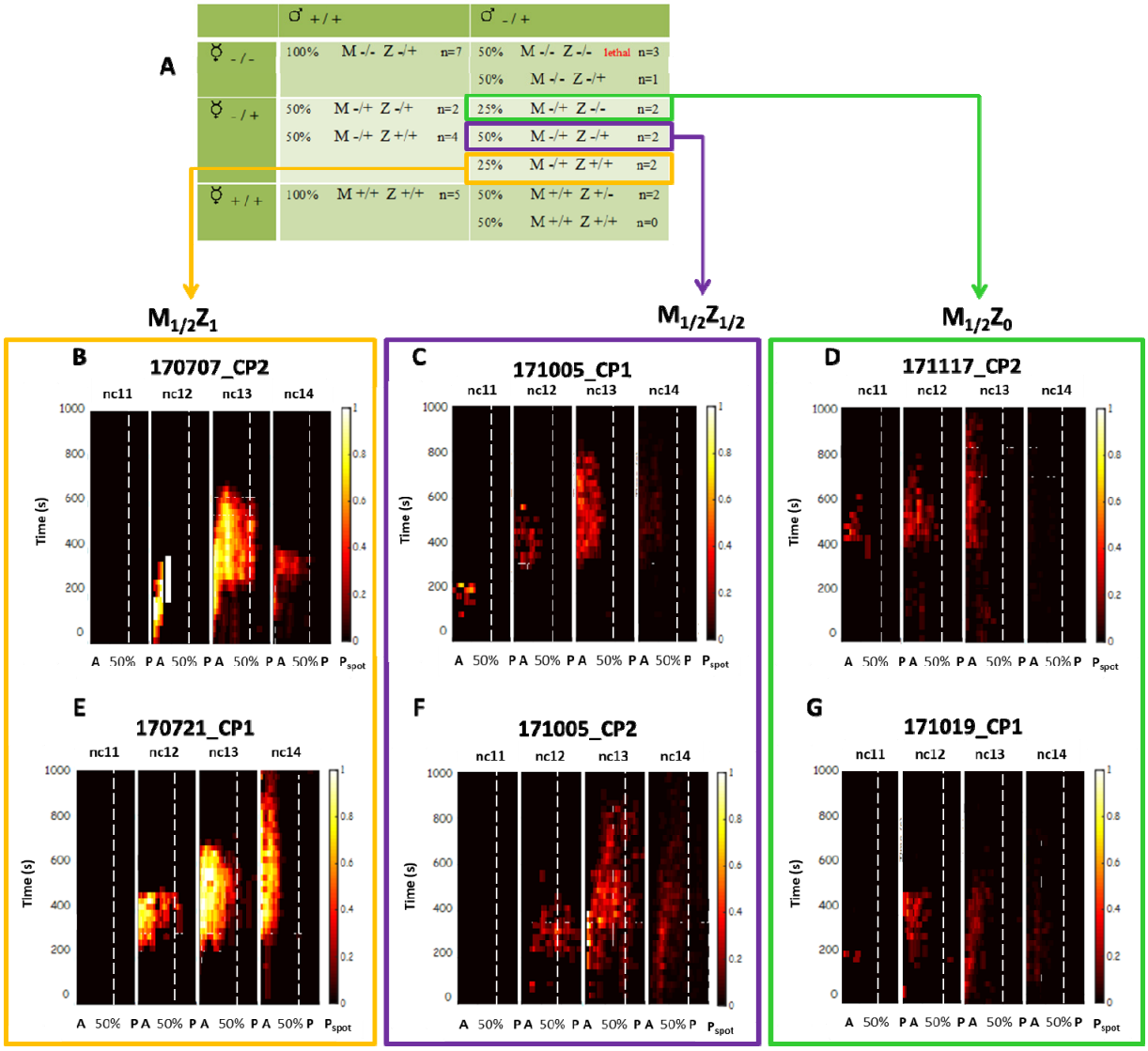
Supplementary Figure 3. Dynamics of *hb-MS2* expression in the two imaged embryos that emerged from the cross of WT females with *hb*^{FB} /+ males: The females also carried the His-RFP and MCP-GFP transgenes and the males carried the hb-MS2 reporter. **A)** This cross should generate 50% of WT embryos (M_1Z_1, blue) and 50% of heterozygous for zygotic *hb* ($M_1Z_{1/2}$, red). Only two embryos were analyzed for this study and allowed to generate the individual kymographs shown in panels **B** and **D.** The kymographs indicate much weaker expression than for the WT context M_1Z_1 embryo **(C)**. The kymographs shown in panels B and D are very similar to each other and different from the kymograph of a WT embryo (C). We

conclude from this comparison that the two embryo **(B,D)** are likely $M_1Z_{1/2}$ (see red box). The data from these two embryos were pulled (see Section 2.2) to generate the curves shown in this study. Note that here, we need more data from additional embryos. Also, as the identification was obtained by eye, we need to set-up a quantitative and statistical analysis allowing to separate the different classes of movies unambiguously.



Supplementary Figure 4. Dynamics of *hb-MS2* expression of the six imaged embryos that emerged from the cross of hb^{FB}/+ females with WT males: The females also carried the His-RFP and MCP-GFP transgenes and the males carried the *hb-MS2* reporter. **A)** This cross should generate embryos expressing half the dose of maternal Hb that will be either 50% heterozygous for zygotic *hb* ($M_{1/2}Z_{1/2}$) or 50% WT for zygotic *hb* ($M_{1/2}Z_1$). Six embryos for this cross where analyzed and allowed to generate the six

individual kymographs shown in panels **B** to **G**. Among the six kymographs, two groups emerge: four are very similar to each other **(B,C,E,F)** and the two remaining **(D,G)** are also very similar to each other. We conclude from this analysis that the four **(B,C,E,F)** kymographs likely arise from four $M_{1/2}Z_1$ (yellow box) while the two **(D,G)** kymographs likely arise from $M_{1/2}Z_{1/2}$ embryos (purple box). Data from embryos of each of the two groups were pulled together to generate the curves shown in this study. Note that here, we need more data from additional embryos. Also, as the identification was obtained by eye, we need to set-up a quantitative and statistical analysis allowing to separate the different classes of movies unambiguously



Supplementary Figure 5. Dynamics of *hb-MS2* expression of the six imaged embryos that emerged from the cross of *hb*^{FB}/+ females with heterozygous *hb*^{FB}/+ males: A) This cross should generate embryos expressing half the dose of maternal Hb that will be either zygotically WT ($M_{1/2}Z_1$, yellow, 25%), heterozygous ($M_{1/2}Z_{1/2}$, purple, 50%) or homozygous ($M_{1/2}Z_0$, green, 25%). Six embryos for this cross were analyzed and allowed to generate the six individual kymographs shown in panels **B** to **G**. Among the six kymographs, three groups emerge each composed of similar kymographs: **B** and **E** kymographs show strong expression and likely arise from $M_{1/2}Z_1$ (yellow box)**.** **C** and **F** show intermediate expression and likely arise from $M_{1/2}Z_{1/2}$ embryos (purple box). The two remaining **D** and **G** show weak expression likely arise from $M_{1/2}Z_0$ (green box). Data form embryos of each of the two groups were pulled together to generate the curves shown in this study. Similarly as above, note that we need more data from additional embryos. Also, as the identification was obtained by eye, we need to set-up a quantitative and statistical analysis allowing to separate the different classes of movies unambiguously

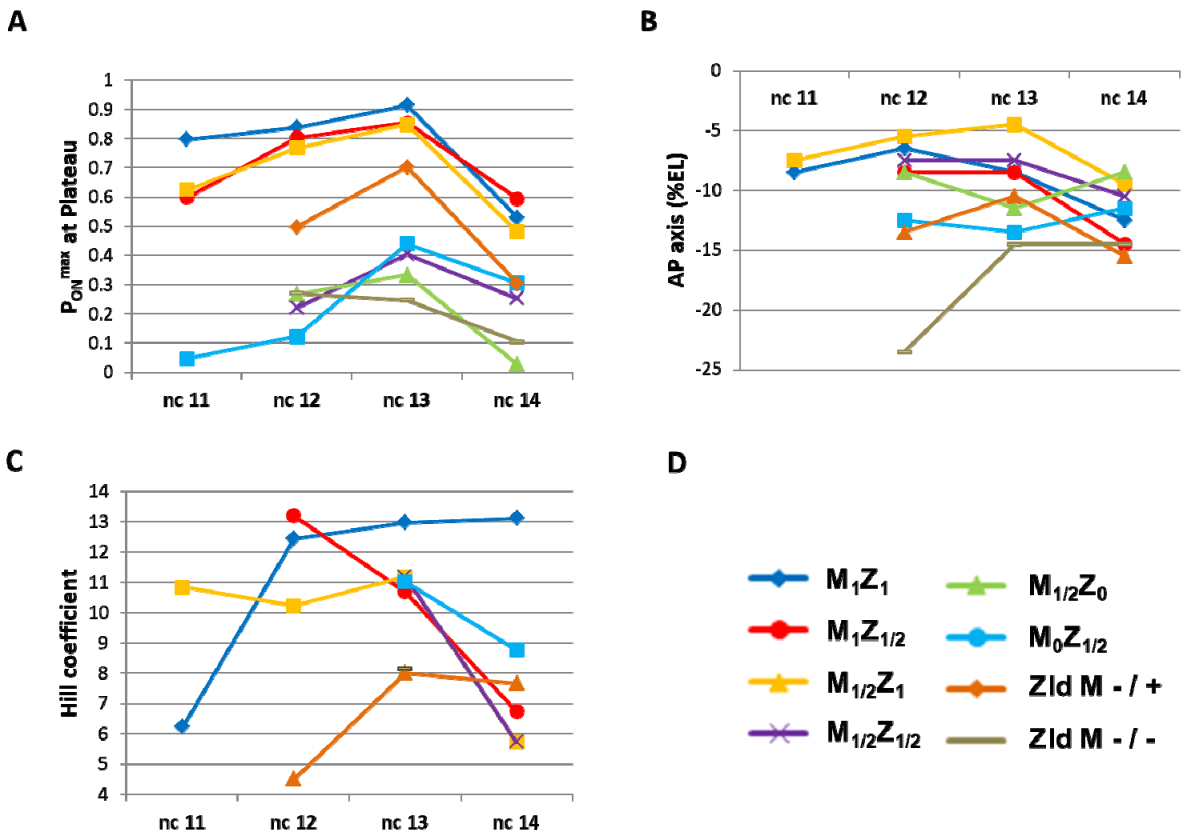


Figure 6. Characteristics of the *hb-MS2* reporter expression border in genetic backgrounds with different levels of Hb and Zld protein expressions: A) P_{ON}^{max} at the plateau levels of each reporter at the border through the nuclear cycles: WT (M_1Z_1), $M_1Z_{1/2}$, $M_{1/2}Z_1$, $M_{1/2}Z_{1/2}$, $M_{1/2}Z_0$, $M_0Z_{1/2}$, Zld M-/+, and Zld M-/-. **B)** Border position of each reporter through the nuclear cycles: the averaged border position of pulled data for each reporter is plotted, the y-axis being the AP axis (% of EL) and the x-axis each nuclear cycle imaged 11, 12, 13, 14. **C)** Hill coefficient for each construct. Color code: dark blue (WT, M_1Z_1), red ($M_1Z_{1/2}$), yellow ($M_{1/2}Z_1$), purple ($M_{1/2}Z_{1/2}$), green ($M_{1/2}Z_0$), light blue ($M_0Z_{1/2}$), orange (Zld M-/+), brown (Zld M-/-).

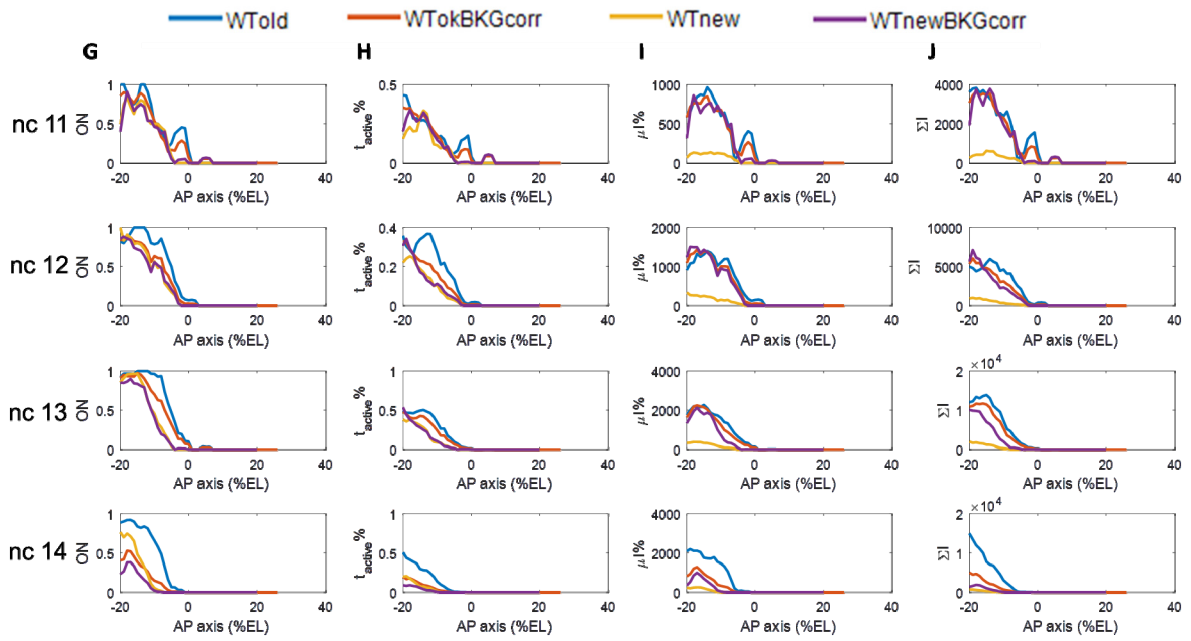


Figure 7. Summary of results before and after normalization of MS2 spot intensity. These are the results coming from WT embryos acquired before the microscope problem (**WTold** in blue), after the microscope problem (**WTnew** in yellow). Plus the background corrected and normalized MS2 spot intensity values for embryos after the microscope problem (**WTnewBKGcorr** in purple) and the pulled embryos of all the WT embryos old and the new ones corrected (**WTokBKGcorr** in red), which was as the WT standard for the rest of the study. The pulled data were used to extract each transcriptional feature: **P_{ON} (A), t_{active} (B), μI (C), ΣI (D).**

Chapter 5

DISCUSSION AND PERSPECTIVES

Over the last 3 decades the different proteins and genes involved in early *Drosophila melanogaster* patterning have been identified (Driever & Nüsslein-Volhard, 1989a, 1989b; Driever & Nüsslein-Volhard, 1988a, 1988b; Lehmann & Nüsslein-Volhard, 1987) and we now qualitatively understand most of the interaction network linking them with each other. On the other hand, we also know that the enhancer/promoter sequences of a gene are of the outmost importance in determining the timing and efficiency of its transcription. However, we still know little about the mechanisms by which these sequences control transcription and how their perturbation might affect the networks that they regulate. At this level, one of the main challenges remains in understanding the synergy between different TFs involved in the transcription process of particular genes. Experimental approaches that have been used to study the interaction of TFs with promoter regions include site specific mutagenesis (Driever & Nüsslein-Volhard, 1988a, 1988b; Lehmann & Nüsslein-Volhard, 1987), genome-wide TF occupancy assays (Combs & Eisen, 2017; Kvon et al., 2014; Satija & Bradley, 2012) and computational tools which predict possible binding sites for known TF motifs (Blanco et al., 2007; Ghosh, 2000). They revealed the importance of the number, position, and affinity of TF binding sites in a regulatory sequence (Margolis et al., 1995) for the transcription process. However, the question remains of how to quantify the effect of each TF in the transcriptional output in regard to all the noise (natural fluctuations in the process) and complexity found in a biological sample.

Synthetic reporters

Our synthetic reporter approach defines a proof of concept of how such approaches can be used to better understand the role that each TF play in the transcription process and how they cooperate with other TFs affect the transcriptional output. Our results show that, as expected, Bcd mostly sets the position of the expression border, likely in a concentration-dependent manner. The interaction of Bcd with the promoter is sufficient for transcriptional activation.

However, the probability to express the locus remains low and this indicates some stochasticity in the decision to turn it ON. In contrast, Hb cannot activate transcription on its own, but increases transcription efficiency once the promoter has been switched ON by Bcd. Cad might be a repressor reducing Hb function in some way. We have planned to extend this approach to study the role played by Zld in *hb* transcription, by inserting 6 Zld binding sites in our synthetic reporter. This will allow us to further dissect the role of Zld in the transcription process by itself and with the other transcription factors. The reporter fly lines are ready but the data acquisition and analysis still need to be carried out.

Our synthetic reporter approach also opens the possibility of playing with the number of TFs binding sites available, the spacing between them, their position, their orientation, and changes to their affinity for the TF. For instance, it will be interesting to find out how different numbers of Bcd binding sites (3, 9, 12, 15) perform on their own. Notably, our modeling analysis (Section 3.1) indicates that 6 binding sites for Bcd are not sufficient to reproduce the dynamics of the hb-MS2 reporter expression and that 9 binding sites might be better (Lucas et al., 2018). The comparison between the Bcd6x and the Bcd9X reporter will allow testing this model. Another possibility could be to make a minimal synthetic promoter where we use the exact same number/spacing/affinity/orientation of the TF found in the WT P2 enhancer however in a synthetic enhancer context, this way we can see if the DNA surrounding the binding sites is important for efficient, precise and robust expression. This should allow determining if the sequences outside of the binding site might play a role in the activity of the promoter.

The synthetic reporter system that we used with the MS2 tagging system for RNA can be converted to a PP7 system to be able to study transcription driven from two different enhancers (as has been done for different RNAs (Cusanelli et al., 2013; Lenstra et al., 2015)), or from one single enhancer (as done for *snail* (Fukaya, Lim, & Levine, 2016)), then with the recent invention of the transcription-activator like effectors (TALEA) system which allows detection and modulation of transcription enhancer (Crocker & Stern, 2013; Crocker, Tsai, & Stern, 2017; Tsai et al., 2017). Although this might be difficult to set experimentally, one could think of combining two or three of these systems to study a gene-regulatory network cascade over time, for example: a gap gene such as *hb,* one of its target pair-rule gene such as *even-skipped* and

finally a segment polarity gene such as *wingless*. On the other hand, one could also think of studying several players in a single gene-regulatory network at a specific time scale. For example, a similar study to ours could be done, but expanding to the Posterior, this way we could tag a morphogen such as Bcd, and two of its target genes like *hb* which it regulates in a positive way, and *cad* RNA which it prevents from being translated. Even though, this would generate a huge amount of data that will require adequate treatment, we could use the advances in light-sheet microscopy in this kind of study to allow us to image the whole embryo at high resolution.

In conclusion, these kinds of experiments could allow dissecting the so called "cis-regulatory code" for the gene-regulatory network (Yáñez-Cuna, Kvon, & Stark, 2013) involved in *hb* precise and robust expression in early fly embryonic development.

In-silico analsyis

None of the synthetic reporters that we so far analyzed completely reproduced *hb* WT expression levels and border position. It will thus be interesting to look for other possible TFs which might play a role in *hb* transcription. Although we already performed some of this analysis, we could look again in more detail for the TFs binding sites in the WT P2 promoter or intron with other in-silico analysis software like PATSER (Tsai et al., 2017) or SiteOut (Crocker et al., 2017) which allow looking for low affinity binding sites (by changing different query parameters). This way we could start adding new TF binding sites to the synthetic reporter, or mutating binding sites for specific TFs from the WT reporter, and perform a comprehensive analysis of their effect on *hb* spatiotemporal activation. The goal of such studies, combined with theoretical modeling, could lead to the creation of a 100% synthetic promoter which fully reproduces *hb* WT expression. Such a model system of gene activation could in turn lead to a better understanding of the gene-regulatory networks in early fly development.

Mutant analysis

Cad has been shown to counter-repress Hunchback to regulate *even-skipped* stripped 2 expression in later embryonic development (Vincent et al., 2018). Given the possibility that Cad could act as a possible repressor of Hb, it will be interesting to study expression of the *hb-MS2* reporter in mutants for Cad: as for Hb and Zld in this paper, this will require to obtain germline

clones but might help understand the possible role of maternal and/or zygotic Cad on *hb* expression. Given the potential interactions between Cad and Hb (revealed by the fact that the Cad repressive effect requires Hb binding sites in the reporter), double mutants of zygotic Hb/Cad will be interesting to study, looking for genetic interactions. In an early stage of this study, we have acquired data for a Zygotic Cad deficiency mutant in our WT P2 reporter but with the previous nuclear marker (Nup-RFP) which makes the image analysis more challenging. When we changed to the His-RFP reporter, we saw some changes in the t_{init} time for the *hb-MS2* expression in the WT genetic background. The reason for these differences is not yet clear but the analysis of the *hb-MS2* expression in genetic backgrounds modifying the levels of Cad will have to be done using our current system, which works best for imaging and analysis.

It is interesting to note that the t_{init} measured for the WT P2 reporter in this set of experiments was ~300s, longer than the value we had previously measured, which was around ~ 200s (Lucas et al., 2018). Of note, in this current study the imaging was done at a temperature of 23°C, whereas in the previous study it was 25°C (Lucas et al., 2018). This slight change in temperature might explain part of the difference. A parallel increase can be seen in the Hill coefficient. In this study, the Hill coefficient varies from 6 (nc 11) to 12 (nc 13), while in our previous study it varied from 4.5 (nc 11) to 7 (nc 13) (Lucas et al., 2018). Although we will have to look closely at these variations and evaluate thoroughly their statistical significance, one possibility is that they are due in part to the differences in the markers used to track the nuclei: in previous studies (Lucas et al., 2018), we used a red fluorescent nuclear pore protein (NUP-RFP) which was easily photo-bleached and did not renew over the cell cycles. This weak signal was not optimal for automated treatment of the movies and therefore, we decided to change to a stronger nuclear marker, a red fluorescent Histone-H2B (His-RFP) (Bothma et al., 2015; Ferraro et al., 2016; Garcia et al., 2013; Morisaki et al., 2016). This new marker turned out to help the image analysis and we decided to use it for all the new studies including the studies presented in this manuscript. Since the two markers do not label exactly the same material (nuclear envelop vs chromatin), the Matlab code used to precisely identify the beginning of the cycle after each mitosis had to be adjusted and it is likely that the two methods do not exactly position the origin of time in the cycle at the same time. In addition, although His-RFP is well established marker which has been used in many studies (Bothma et al., 2015; Ferraro et al., 2016; Garcia et al., 2013; Morisaki et

al., 2016), the endogenous H2B is normally expressed in the lines we used. This means that H2B is overexpressed, which might be slightly toxic and explain the difference in transcription dynamics observed between this study and our previous one. This can gives us an example of the importance of chromatin structure in transcription regulation and playing with other chromatin factors could lead to understand how condensation and its de-condensation, as well as tethering to the nuclear membrane could help us understand the role of epigenetic factors in early development, which is highly under explored.

Our results confirmed the critical role for Zld in efficient transcription of *hb* (Liang et al., 2008). In fact recent studies have shown that Zld might be able to cooperate to some degree with Bcd (Haines & Eisen, 2018; Park et al., 2018) to help with further activation of gap-genes including *hb* in the anterior-region but the detailed mechanisms of this interaction have not be unraveled. The combination of our mutant analysis with synthetic reporters including Zld binding sites might allow to better understand at which step in the transcription process is Zld cooperating with Bcd. We can also analyze double mutants for both Zld/Hb and see what the effect on the expression of the hb-MS2 reporter is. It would also be interesting to use the fly lines from E. Wieschaus's laboratory (Hannon, Blythe, & Wieschaus, 2017), which have different degrees of flat Bcd concentration (no gradient), to determine the Bcd concentration threshold required for *hb* expression and how using a flat Bcd input instead of an exponential one could affect *hb* expression.

Transcription memory hypothesis

Another way to reach fast sharpness during rapid embryo development which was barely discussed but was explored during my studies is the possibility of transcriptional memory (Crauk & Dostatni, 2005; Porcher et al., 2010; Ferraro et al., 2016). This hypothesis propose that activated loci recall Bcd concentration from one S-phase to the next, by keeping track of its transcriptional status across mitosis and promoting rapid reactivation of gene expression profiles on early S-phase (Fig. 13). Using the expression of the *hb-MS2* reporter in a WT background, we looked for correlations in transcription among loci belonging to the same nuclear lineage in order to address the question of transcriptional memory in our system. However, so far, it was very

difficult, mostly because of very low statistical power to separate spatial (loci position and Bcd concentration), and temporal correlation (lineage and familial history).

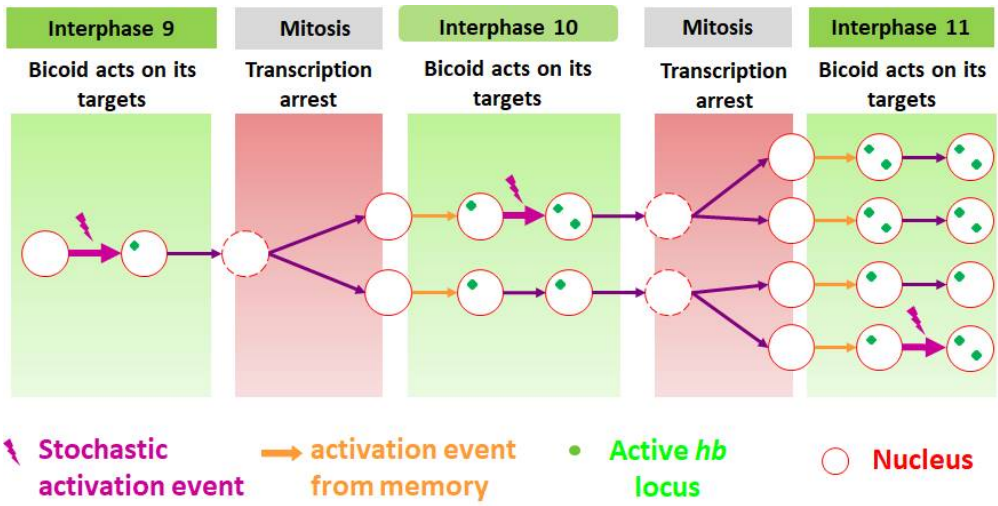


Figure 13. The memory hypothesis: The memory hypothesis is formed of two propositions: (1) Activation by random Bcd binding: an unactive *hb* locus can sense the local Bcd concentration by stochastically binding to approaching Bcd molecules, and a sufficient number of random Bcd binding events leads to the activation of transcription; and (2) Activation by memory: this active state is memorized by the promoter region, and transmitted to daughter nuclei at the next cycle.

Such a mechanism has been shown to contribute to the expression of the *snail* gene at nc14 in fly embryos (Ferraro et al., 2016). In some of our synthetic reporter movies as well as with the *hb-MS2* reporter in some mutants, we indeed see some spots persisting after mitosis and passing to one of the daughters. These quite rare events, seem to only occur over the early nuclear mitosis 10 to 11, and 11 to 12. It will be interesting to revisit this question with the mutant/synthetic data were it might be easier to tease out real memory effects, and now that we have more movies that can help with the statistical distinction between memory/no memory scenarios.

Finally, it will also be interesting to re-explore Bcd diffusion (Fradin, 2017) to tease out differences between the fast and slow moving population in both the nucleus (Porcher et al., 2010) and cytoplasm (Abu-Arish et al., 2010). By using mutants that reduce Bcd binding in the nucleus, like using Hira mutants which create haploid embryos, and Cad mutants to reduce binding partners on the cytoplasm.

Conclusion

Using live imaging of the *hb* P2 promoter with the MS2 system, we were able to see that its transcription is bursty and likely reflects its functioning as a two-state promoter (Desponds et al., 2016). At each nuclear interphase, *hb* P2 transcription is first observed in the anterior, rapidly spreads towards the posterior before reaching a steady state with the setting of the expression border, as expected from a dose-dependent activation response by Bcd. We showed that it takes only 3 minutes after transcription is first detected in the anterior for the border of expression to be precisely stablished with a high steepness (Lucas et al., 2018). With modeling we were able to show that 6 binding sites of Bcd alone are not able to explain the precise boundary in such a short time (Lucas et al., 2018; Tran et al., 2018). Either more binding sites of Bcd (Lucas et al., 2018), a repressor gradient or investigating other transcription factors as inputs could help reconcile the data to the models.

Therefore, next we decided to introduce our synthetic promoter strategy combined with the analyses of the *hb-MS2* expression in various mutant genetic backgrounds to try to better understand how efficient and precise transcription is achieved in 3 minutes downstream of the *hb* P2 promoter. This analysis indicate that Zld and Hb play important roles in the expression of the *hb-MS2* reporter, and that maternal Hb is active at early nc 11-12 while zygotic Hb acts at later nc 13-14. Our synthetic reporter approach showed that Bcd is able to activate *hb* transcription on its own but is not very efficient and results in a stochastic activation of the loci. In contrast, Hb on its own is unable to activate transcription but when combined to Bcd plays an important role in reducing this stochasticity, allowing to achieve a more steep border and efficient transcription. Our work also suggests that Cad might act as a posterior repressor gradient. However, it still remains to be explored how these various cis-acting binding sites and trans-acting factors can mechanistically contribute to the establishment of a precise border in 3 min. Therefore, it will be interesting to develop similar models to the ones presented for our WT promoter (Desponds et al., 2016; Tran et al., 2018) and that will reconcile all the data acquired from this last study. To find the minimal TF combination needed to achieve efficient readout of a morphogen gradient in only 3 minutes and help understand *hb* efficient and precise transcription.

As always in research: “One never notices what has been done; one can only see what remains to be done”. - Marie Curie.

Afterword

Our journey began by trying to understand how a single cell knows its place in the blueprint of life. I hope you find my PhD quest to answer how cells measure inputs and translate them into precise outputs as a small grain of sand added to the endless search for understanding that drives us and science. Hopefully the next time you see that little fly in your kitchen you might just wonder what new things that little pest will help reveal next. Thank you for bearing with me and letting me share this journey with you.

"Every new beginning comes from some other beginning's end". - Seneca

I would like to share some personal thoughts that came to me as I progressed though this journey, as this PhD opened my eyes to the great importance of open research and scientific collaboration. It also showed me the great potential of the scientific community in voicing our concerns and knowledge to the public community. In this age of fake-news and miss information, I believe we as scientists need to play a more crucial role in helping share our facts and scientific knowledge, not only in our fields as we traditionally have been doing by publishing papers, but also by making our research reachable to the general public, by using individual creativity and our communication skills.

This is rather important as many of the people that make decisions on how much money science should get aren't always are very close to our areas or knowledge and it's important for them to understand the impact of our research. Plus it is important for us to become politically active and give our voices and concerns as a community for policy changes, and how money should be spent in our governmental system, this is because if we don't do it, our representatives will get misinformation from fake news. Which we as a community should really start considering as a true threat to science based knowledge in our times.

Finally I think if we want to better ourselves and the community, we have to continue confronting our own biases (racial, sexist, abelist, homophobic, etc) so we can continue to have a more open community where scientific knowledge can be pursued without diminishing anyone's role in it.

References

Abu-Arish, A., Porcher, A., Czerwonka, A., Dostatni, N., & Fradin, C. (2010). High mobility of bicoid captured by fluorescence correlation spectroscopy: implication for the rapid establishment of its gradient. *Biophysical Journal*, *99*(4), L33-5. https://doi.org/10.1016/j.bpj.2010.05.031

Bastock, R., & St Johnston, D. (2008). Drosophila oogenesis. *Current Biology : CB*, *18*(23), R1082-7. https://doi.org/10.1016/j.cub.2008.09.011

Bentovim, L., Harden, T. T., & DePace, A. H. (2017). Transcriptional precision and accuracy in development: from measurements to models and mechanisms. *Development (Cambridge, England)*, *144*(21), 3855–3866. https://doi.org/10.1242/dev.146563

Berg, H. C., & Purcell, E. M. (1977). Physics of chemoreception. *Biophysical Journal*, *20*(2), 193–219. https://doi.org/10.1016/S0006-3495(77)85544-6

Bertrand, E., Chartrand, P., Schaefer, M., Shenoy, S. M., Singer, R. H., & Long, R. M. (1998). Localization of ASH1 mRNA Particles in Living Yeast. *Molecular Cell*, *2*(4), 437–445. https://doi.org/10.1016/S1097-2765(00)80143-4

Bialek, W., & Setayeshgar, S. (2005). Physical limits to biochemical signaling. *Proceedings of the National Academy of Sciences of the United States of America*, *102*(29), 10040–5. https://doi.org/10.1073/pnas.0504321102

Blanco, E., Guigó, R., & Messeguer, X. (2007). Multiple non-collinear TF-map alignments of promoter regions. *BMC Bioinformatics*, *8*(1), 138. https://doi.org/10.1186/1471-2105-8-138

Bothma, J. P., Garcia, H. G., Ng, S., Perry, M. W., Gregor, T., & Levine, M. (2015). Enhancer additivity and non-additivity are determined by enhancer strength in the Drosophila embryo. *ELife*, *4*. https://doi.org/10.7554/eLife.07956

Combs, P. A., & Eisen, M. B. (2017). Genome-wide measurement of spatial expression in patterning mutants of Drosophila melanogaster. *F1000Research*, *6*, 41. https://doi.org/10.12688/f1000research.9720.1

Crauk, O., & Dostatni, N. (2005). Bicoid determines sharp and precise target gene expression in the Drosophila embryo. *Current Biology : CB*, *15*(21), 1888–98. https://doi.org/10.1016/j.cub.2005.09.046

Crick F. (1970). Diffusion in embryogenesis.

Crocker, J., & Stern, D. L. (2013). TALE-mediated modulation of transcriptional enhancers in vivo. *Nature Methods*, *10*(8), 762–767. https://doi.org/10.1038/nmeth.2543

Crocker, J., Tsai, A., & Stern, D. L. (2017). A Fully Synthetic Transcriptional Platform for a Multicellular Eukaryote. *Cell Reports*, *18*, 287–296. https://doi.org/10.1016/j.celrep.2016.12.025

Cusanelli, E., Perez-Romero, C. A., & Chartrand, P. (2013). Telomeric Noncoding RNA TERRA Is Induced by Telomere Shortening to Nucleate Telomerase Molecules at Short Telomeres. *Molecular*

Cell, *51*(6), 780–791. https://doi.org/10.1016/j.molcel.2013.08.029

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

Desponds, J., Tran, H., Ferraro, T., Lucas, T., Perez Romero, C., Guillou, A., … Walczak, A. M. (2016). Precision of readout at the hunchback gene. *PLoS Computational Biology*, (October), 1–35. https://doi.org/10.1101/063784

Driever, W., & Nüsslein-Volhard, C. (1988a). A gradient of bicoid protein in Drosophila embryos. *Cell*, *54*(1), 83–93. https://doi.org/10.1016/0092-8674(88)90182-1

Driever, W., & Nüsslein-Volhard, C. (1988b). The bicoid protein determines position in the Drosophila embryo in a concentration-dependent manner. *Cell*, *54*(1), 95–104. https://doi.org/10.1016/0092-8674(88)90183-3

Driever, W., & Nüsslein-Volhard, C. (1989). The bicoid protein is a positive regulator of hunchback transcription in the early Drosophila embryo. *Nature*. https://doi.org/10.1038/337138a0

Driever, W., Siegel, V., & Nusslein-Volhard, C. (1990). Autonomous determination of anterior structures in the early Drosophila embryo by the bicoid morphogen. *Development*, *109*(4).

Driever, W., Thoma, G., & Nusslein-Volhard, C. (1989). Determination of spatial domains of zygotic gene expression in the Drosophila embryo by the affinity of binding sites for the bicoid morphogen. *Nature*, *340*(6232), 363–367. https://doi.org/10.1038/340363a0

Drocco, J. A., Grimm, O., Tank, D. W., & Wieschaus, E. (2011). Measurement and perturbation of morphogen lifetime: effects on gradient shape. *Biophysical Journal*, *101*(8), 1807–15. https://doi.org/10.1016/j.bpj.2011.07.025

Estrada, J., Wong, F., DePace, A., Gunawardena, J., Ackers, G. K., Johnson, A. D., … Mirny, L. A. (2016). Information Integration and Energy Expenditure in Gene Regulation. *Cell*, *166*(1), 234–44. https://doi.org/10.1016/j.cell.2016.06.012

Farrell, J. A., & O'farrell, P. H. (2014). From Egg to Gastrula: How the Cell Cycle Is Remodeled During the Drosophila Mid-Blastula Transition. https://doi.org/10.1146/annurev-genet-111212-133531

Ferraro, T., Esposito, E., Mancini, L., Ng, S., Lucas, T., Coppey, M., … Lagha, M. (2016). Transcriptional Memory in the Drosophila Embryo. *Current Biology : CB*, *26*(2), 212–8. https://doi.org/10.1016/j.cub.2015.11.058

Ferraro, T., Lucas, T., Clémot, M., De Las Heras Chanes, J., Desponds, J., Coppey, M., … Dostatni, N. (2016). New methods to image transcription in living fly embryos: the insights so far, and the prospects. *Wiley Interdisciplinary Reviews. Developmental Biology*, *5*(3), 296–310. https://doi.org/10.1002/wdev.221

Foo, S. M., Sun, Y., Lim, B., Ziukaite, R., O'Brien, K., Nien, C.-Y., … Rushlow, C. A. (2014). Zelda Potentiates Morphogen Activity by Increasing Chromatin Accessibility. *Current Biology*, *24*(12), 1341–1346. https://doi.org/10.1016/J.CUB.2014.04.032

Fradin, C. (2017). On the importance of protein diffusion in biological systems: The example of the Bicoid morphogen gradient. *Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics*, *1865*(11), 1676–1686. https://doi.org/10.1016/J.BBAPAP.2017.09.002

Fukaya, T., Lim, B., & Levine, M. (2016). Enhancer Control of Transcriptional Bursting. *Cell*, *166*(2), 358–368. https://doi.org/10.1016/j.cell.2016.05.025

Garcia, H. G., Tikhonov, M., Lin, A., & Gregor, T. (2013). Quantitative Imaging of Transcription in Living Drosophila Embryos Links Polymerase Activity to Patterning. *Current Biology*, *23*(21), 2140–2145. https://doi.org/10.1016/j.cub.2013.08.054

Ghosh, D. (2000). Object-oriented transcription factors database (ooTFD). *Nucleic Acids Research*, *28*(1), 308–10. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/10592257

Gilbert, S. F. (2000). Early Drosophila Development. Sinauer Associates. Retrieved from http://www.ncbi.nlm.nih.gov/books/NBK10081/

Gilbert, S. F., & Barresi, M. J. (2016). *Developmental Biology* (11 ed). Sunderland Massachusetts: Sinauer Associatesina.

Gregor, T., McGregor, A. P., & Wieschaus, E. F. (2008). Shape and function of the Bicoid morphogen gradient in dipteran species with different sized embryos. *Developmental Biology*, *316*(2), 350–8. https://doi.org/10.1016/j.ydbio.2008.01.039

Gregor, T., Tank, D. W., Wieschaus, E. F., & Bialek, W. (2007). Probing the Limits to Positional Information. *Cell*, *130*(1), 153–164. https://doi.org/10.1016/j.cell.2007.05.025

Gregor, T., Wieschaus, E. F., McGregor, A. P., Bialek, W., & Tank, D. W. (2007). Stability and nuclear dynamics of the bicoid morphogen gradient. *Cell*, *130*(1), 141–52. https://doi.org/10.1016/j.cell.2007.05.026

Haines, J. E., & Eisen, M. B. (2018). Patterns of chromatin accessibility along the anterior-posterior axis in the early Drosophila embryo. *PLOS Genetics*, *14*(5), e1007367. https://doi.org/10.1371/journal.pgen.1007367

Hanes, S. D., & Brent, R. (1989). DNA specificity of the bicoid activator protein is determined by homeodomain recognition helix residue 9. *Cell*, *57*(7), 1275–83. https://doi.org/10.1016/0092-8674(89)90063-9

Hannon, C. E., Blythe, S. A., & Wieschaus, E. F. (2017). Concentration dependent chromatin states induced by the bicoid morphogen gradient. *ELife*, *6*, e28275. https://doi.org/10.7554/eLife.28275

He, L., Wang, X., & Montell, D. J. (2011). Shining light on Drosophila oogenesis: live imaging of egg development. *Current Opinion in Genetics & Development*, *21*(5), 612–619. https://doi.org/10.1016/j.gde.2011.08.011

Houchmandzadeh, B., Wieschaus, E., & Leibler, S. (2002). Establishment of developmental precision and proportions in the early Drosophila embryo. *Nature*, *415*, 798–802.

Irish, V., Lehmann, R., & Akam, M. (1989). The Drosophila posterior-group gene nanos functions by

repressing hunchback activity. *Nature*, *338*(6217), 646–648. https://doi.org/10.1038/338646a0

Jaeger, J. (2011). The gap gene network. *Cellular and Molecular Life Sciences : CMLS*, *68*(2), 243–74. https://doi.org/10.1007/s00018-010-0536-y

Johnston, D. S., & Nüsslein-Volhard, C. (1992). The origin of pattern and polarity in the Drosophila embryo. *Cell*, *68*(2), 201–219. https://doi.org/10.1016/0092-8674(92)90466-P

Kaizu, K., de Ronde, W., Paijmans, J., Takahashi, K., Tostevin, F., & ten Wolde, P. R. (2014). The Berg-Purcell Limit Revisited. *Biophysical Journal*, *106*(4), 976–985. https://doi.org/10.1016/J.BPJ.2013.12.030

Kerszberg, M., & Wolpert, L. (2007). Specifying positional information in the embryo: looking beyond morphogens. *Cell*, *130*(2), 205–9. https://doi.org/10.1016/j.cell.2007.06.038

Kvon, E. Z., Kazmar, T., Stampfel, G., Yáñez-Cuna, J. O., Pagani, M., Schernhuber, K., … Stark, A. (2014). Genome-scale functional characterization of Drosophila developmental enhancers in vivo. *Nature*, *512*(7512), 91–95. https://doi.org/10.1038/nature13395

Larson, D. R., Zenklusen, D., Wu, B., Chao, J. A., & Singer, R. H. (2011). Probing Transcription Factor Dynamics at the Single-Molecule Level in a Living Cell. *Science*, *316*(5828), 1191–1194. https://doi.org/10.1126/science.1141967

Lefebvre, F., Lécuyer, É., Lefebvre, F. A., & Lécuyer, É. (2018). Flying the RNA Nest: Drosophila Reveals Novel Insights into the Transcriptome Dynamics of Early Development. *Journal of Developmental Biology*, *6*(1), 5. https://doi.org/10.3390/jdb6010005

Lehmann, R., & Nusslein-Volhard, C. (1991). The maternal gene nanos has a central role in posterior pattern formation of the Drosophila embryo. *Development*, *112*(3).

Lehmann, R., & Nüsslein-Volhard, C. (1987). hunchback, a gene required for segmentation of an anterior and posterior region of the Drosophila embryo. *Developmental Biology*, *119*(2), 402–417. https://doi.org/10.1016/0012-1606(87)90045-5

Lenstra, T. L., Coulon, A., Chow, C. C., & Larson, D. R. (2015). Single-Molecule Imaging Reveals a Switch between Spurious and Functional ncRNA Transcription. *Molecular Cell*, *60*(4), 597–610. https://doi.org/10.1016/j.molcel.2015.09.028

Liang, H.-L., Nien, C.-Y., Liu, H.-Y., Metzstein, M. M., Kirov, N., & Rushlow, C. (2008). The zinc-finger protein Zelda is a key activator of the early zygotic genome in Drosophila. *Nature*, *456*(7220), 400–3. https://doi.org/10.1038/nature07388

Little, S. C., Tkačik, G., Kneeland, T. B., Wieschaus, E. F., & Gregor, T. (2011). The Formation of the Bicoid Morphogen Gradient Requires Protein Movement from Anteriorly Localized mRNA. *PLoS Biology*, *9*(3), e1000596. https://doi.org/10.1371/journal.pbio.1000596

Lucas, T., Ferraro, T., Roelens, B., De Las Heras Chanes, J., Walczak, A. M., Coppey, M., & Dostatni, N. (2013). Live imaging of bicoid-dependent transcription in Drosophila embryos. *Current Biology : CB*, *23*(21), 2135–9. https://doi.org/10.1016/j.cub.2013.08.053

Lucas, T., Tran, H., Perez Romero, C. A., Guillou, A., Fradin, C., Coppey, M., … Dostatni, N. (2018). 3 minutes to precisely measure morphogen concentration. *PLOS Genetics*, *14*(10), e1007676. https://doi.org/10.1371/journal.pgen.1007676

Margolis, J. S., Borowsky, M. L., Steingrímsson, E., Shim, C. W., Lengyel, J. A., & Posakony, J. W. (1995). Posterior stripe expression of hunchback is driven from two promoters by a common enhancer element. *Development (Cambridge, England)*, *121*(9), 3067–77. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/7555732

Martinez-Arias, A., & Stewart, A. (2002). Molecular Principle of Animal Development. In *Oxford Uni. Press. 1 edition.* (p. Chapter 9).

Morgan, T. H. (1927). *Experimental embryology.* --. New York : Columbia University Press,. https://doi.org/10.5962/bhl.title.82143

Morisaki, T., Lyon, K., DeLuca, K. F., DeLuca, J. G., English, B. P., Zhang, Z., … Stasevich, T. J. (2016). Real-time quantification of single RNA translation dynamics in living cells. *Science*, *352*(6292).

Park, J., Estrada, J., Johnson, G., Ricci-Tam, C., Bragdon, M., Shulgina, Y., … DePace, A. H. (2018). Dissecting the sharp response of a canonical developmental enhancer reveals multiple sources of cooperativity. *BioRxiv*, 408708. https://doi.org/10.1101/408708

Perez-Romero, C. A., Tran, H., Coppey, M., Walczak, A. M., Fradin, C., & Dostatni, N. (2018). Live Imaging of mRNA Transcription in Drosophila Embryos (pp. 165–182). Humana Press, New York, NY. https://doi.org/10.1007/978-1-4939-8772-6_10

Porcher, A., Abu-Arish, A., Huart, S., Roelens, B., Fradin, C., & Dostatni, N. (2010). The time to measure positional information: maternal hunchback is required for the synchrony of the Bicoid transcriptional response at the onset of zygotic transcription. *Development (Cambridge, England)*, *137*, 2795–2804. https://doi.org/10.1242/dev.051300

Porcher, A., & Dostatni, N. (2010). The bicoid morphogen system. *Current Biology : CB*, *20*(5), R249-54. https://doi.org/10.1016/j.cub.2010.01.026

Raser, J. M., & O'Shea, E. K. (2005). Noise in gene expression: origins, consequences, and control. *Science (New York, N.Y.)*, *309*(5743), 2010–3. https://doi.org/10.1126/science.1105891

Rivera-Pomar, R., Niessing, D., Schmidt-Ott, U., Gehring, W. J., & Jacklë, H. (1996). RNA binding and translational suppression by bicoid. *Nature*, *379*(6567), 746–749. https://doi.org/10.1038/379746a0

Satija, R., & Bradley, R. K. (2012). The TAGteam motif facilitates binding of 21 sequence-specific transcription factors in the Drosophila embryo. *Genome Research*, *22*(4), 656–65. https://doi.org/10.1101/gr.130682.111

Simpson-Brose, M., Treisman, J., & Desplan, C. (1994). Synergy between the hunchback and bicoid morphogens is required for anterior patterning in Drosophila. *Cell*, *78*(5), 855–65. https://doi.org/10.1016/S0092-8674(94)90622-X

Struhl, G., Struhl, K., & Macdonald, P. M. (1989). The gradient morphogen bicoid is a concentration-

dependent transcriptional activator. *Cell*, *57*(7), 1259–1273. https://doi.org/10.1016/0092-8674(89)90062-7

Tadros, W., & Lipshitz, H. D. (2009). The maternal-to-zygotic transition: a play in two acts. *Development (Cambridge, England)*, *136*(18), 3033–42. https://doi.org/10.1242/dev.033183

Tran, H., Desponds, J., Perez Romero, C. A., Coppey, M., Fradin, C., Dostatni, N., & Walczak, A. M. (2018). Precision in a rush: Trade-offs between reproducibility and steepness of the hunchback expression pattern. *PLOS Computational Biology*, *14*(10), e1006513. https://doi.org/10.1371/journal.pcbi.1006513

Tran, H., Perez-Romero, C. A., Ferraro, T., Fradin, C., Dostatni, N., Coppey, M., & Walczak, A. M. (2018). LiveFly: A Toolbox for the Analysis of Transcription Dynamics in Live Drosophila Embryos (pp. 183–195). Humana Press, New York, NY. https://doi.org/10.1007/978-1-4939-8772-6_11

Tsai, A., Muthusamy, A. K., Alves, M. R., Lavis, L. D., Singer, R. H., Stern, D. L., & Crocker, J. (2017). Nuclear microenvironments modulate transcription from low-affinity enhancers. *ELife*, *6*, e28975. https://doi.org/10.7554/eLife.28975

Turing, A. (1952). The Chemical Basis of Morphogenesis. *Royal Society*, 37–72. Retrieved from http://rstb.royalsocietypublishing.org/Downloadedfrom

Vincent, B. J., Staller, M. V., Lopez-Rivera, F., Bragdon, M. D. J., Pym, E. C. G., Biette, K. M., … DePace, A. H. (2018). Hunchback is counter-repressed to regulate even-skipped stripe 2 expression in Drosophila embryos. *PLOS Genetics*, *14*(9), e1007644. https://doi.org/10.1371/journal.pgen.1007644

Wolpert, L. (1969). *Positional Information and the Spatial Pattern of Cellular Differentiationt. Theoret. Biol* (Vol. 25). Retrieved from https://ac.els-cdn.com/S0022519369800160/1-s2.0-S0022519369800160-main.pdf?_tid=0bff6730-a73b-4cda-a2e2-179a32e7d188&acdnat=1535658964_e9d3b9b74c26191df4c0aaf4c7c4596b

Wreden, C., Verrotti, A. C., Schisa, J. A., Lieberfarb, M. E., & Strickland, S. (1997). Nanos and pumilio establish embryonic polarity in Drosophila by promoting posterior deadenylation of hunchback mRNA. *Development*, *124*(15).

Yáñez-Cuna, J. O., Kvon, E. Z., & Stark, A. (2013). Deciphering the transcriptional cis-regulatory code. *Trends in Genetics*, *29*(1), 11–22. https://doi.org/10.1016/j.tig.2012.09.007

Yang, Z., Wu, X., Yang, N., & Liu, F. (2018). Noise transmission during the dynamic pattern formation in fly embryos. *Quantitative Biology*, *6*(1), 15–29. https://doi.org/10.1007/s40484-018-0135-8

Copyright Acknowledgements

All images and articles used in this thesis were reprinted with permission of the publishers as acknowledged in each section.