

NORTHWESTERN UNIVERSITY

Quantitative analysis of cell fate decisions

A DISSERTATION

SUBMITTED TO THE GRADUATE SCHOOL

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS

for the degree

DOCTOR OF PHILOSOPHY

Field of Chemical and Biological Engineering

By

Sebastian Michal Bernasek

EVANSTON, ILLINOIS

June 2019

© Copyright by Sebastian Michal Bernasek 2019

All Rights Reserved

## ABSTRACT

Quantitative analysis of cell fate decisions

Sebastian Michal Bernasek

Organismal development depends upon countless cell decisions to adopt particular fates at the appropriate time and place. These decisions are executed by systems of biochemical reactions called regulatory networks. Elucidating the general principles underlying the structure and function of these networks is vital to understanding all developmental processes, as well as the diseases that arise when they fail.

Prior studies of regulatory networks, and the decisions they implement, have heavily relied upon qualitative analysis of experimental data. It has since become clear that quantitative strategies are needed to unravel the complexities of systems-level behavior. The research enclosed in this dissertation therefore combines chemical engineering, computer science, statistics, and experimental data to quantitatively explore how regulatory networks reliably coordinate cell fate decisions.

The findings are consolidated into three distinct chapters. The first two are anchored to a common model system of the *Drosophila* larval eye. They deploy an assortment of novel computational tools, mathematical models, and statistical methods to derive meaningful insight from experimental measurements of the processes that govern cell fate decisions

during retinal patterning. The final chapter introduces a mathematical modeling framework to lead the development of an exciting new hypothesis; auxiliary negative regulators enable development to proceed more quickly by mitigating erroneous cell fate decisions when cells are rapidly metabolizing.

Beyond their insights into the mechanics of cell fate decisions, these efforts have spawned several computational tools that may prove valuable to the broader community. All of these resources have been made freely available (see Appendix C), with the hope that their continued development will contribute toward a more quantitative future for developmental biology.

## Acknowledgements

I would like to thank the mentors, friends, family, and coffee growers whose support made this work possible.

## Table of Contents

ABSTRACT	3
Acknowledgements	5
Table of Contents	6
List of Figures	9
List of Tables	12
Chapter 1. Introduction	13
1.1. Molecular origins of cell fate decisions	14
1.2. The power of quantitative analysis	17
1.3. Mathematical modeling of cell fate decisions	19
1.4. Roles for negative feedback in GRNs	22
1.5. Evolutionary drivers of robust cell fate decisions	23
Chapter 2. Automated analysis of mosaic eye imaginal discs in <i>Drosophila</i>	26
2.1. Background on quantitative mosaic analysis	26
2.2. Experimental data	31
2.3. Image segmentation and quantification of nuclear fluorescence levels	32
2.4. Bleedthrough correction	33
2.5. Automated annotation of clones	37
2.6. Manual assessment of annotation performance	47
2.7. Generation of synthetic microscopy data	49

	7
2.8. Synthetic benchmarking of annotation performance	53
2.9. Potential applications and continued development	55
 Chapter 3. Ratiometric control of a transit to differentiation	 57
3.1. Background on the coordination of cell fate decisions	57
3.2. PntGFP expression dynamics in the developing eye	62
3.3. The Pnt-to-Yan ratio varies between cells in different states	62
3.4. Cooperative DNA-binding sensitizes promoters to changes in Pnt-to-Yan ratio	69
3.5. Regulation stabilizes the ratio against varying Pnt and Yan concentrations	74
3.6. Notch signaling lowers the Pnt-to-Yan ratio in progenitor cells	77
3.7. Ras signaling elevates the Pnt-to-Yan ratio in progenitor cells	80
3.8. Ratiometric control as a model for cell fate commitment	83
 Chapter 4. Layered repression synchronizes development with cellular metabolism	 88
4.1. Background on the environmental dependence of developmental tempo	88
4.2. Repressors are less impactful when metabolism is reduced	91
4.3. MicroRNAs are dispensable when metabolism is reduced	95
4.4. Modeling the emergence of developmental errors in variable environments	95
4.5. Protein expression dynamics after partial repressor loss	105
4.6. Effect of full repression loss	108
4.7. Limiting protein synthesis reduces the need for repressors	108
4.8. Implications for the evolution of GRNs	111
4.9. Robustness of modeling results	114
 Chapter 5. Conclusions	 122
5.1. Implications for developmental and systems biology	122
5.2. Avenues for further exploration	126

	8
5.3. Perspective on quantitative biology	134
References	138
Appendix A. Supporting information for Chapter 3	165
A.1. Genetics	165
A.2. Immunohistochemistry	166
A.3. Quantification of expression levels	168
A.4. Conversion of distance to time	169
A.5. Computation of moving averages and confidence intervals	169
A.6. Alignment of expression data	170
A.7. Analysis of <i>yan</i> clones	171
A.8. Parameterization of equilibrium binding model	172
A.9. Visualization of relative Pnt and Yan expression in <i>Notch</i> mutant discs	173
A.10. Analysis of periodic spatial patterns in <i>Notch</i> mutant discs	173
A.11. Supplementary figures	174
Appendix B. Supporting information for Chapter 4	179
B.1. Model system	179
B.2. Genetics	179
B.3. Analysis of mutant phenotypes	182
B.4. Quantification of sfGFP-Sens in the wing disc	185
B.5. Quantification of Yan-YFP dynamics in the eye	187
B.6. Simulation procedure	189
B.7. Dependence of model parameters on metabolic conditions	190
Appendix C. Computational tools for quantitative biologists	194
Appendix D. Reproduction data and code	196

## List of Figures

2.1	Perturbing gene expression via mitotic recombination.	28
2.2	Conventional versus quantitative mosaic analysis.	30
2.3	Example clones in the larval fly eye.	32
2.4	Automated correction of fluorescence bleedthrough in the larval eye.	34
2.5	Characterization of bleedthrough contribution strengths.	37
2.6	Training a clone annotation model.	39
2.7	Label assignment using a trained clone annotation model.	40
2.8	Automated unsupervised annotation of clones in the larval eye.	48
2.9	Comparison of automated annotation with manually assigned labels.	48
2.10	Simulated growth of a synthetic cell culture.	51
2.11	Tunable generation of synthetic microscopy data.	52
2.12	Synthetic benchmarking of automated annotation performance.	54
2.13	Spatial context is most informative for large and ambiguous clones.	55
3.1	Dynamics of Pnt and Yan expression during eye development.	63
3.2	Pnt-to-Yan ratios differ between cellular states.	68
3.3	Dynamics of expression variability in progenitors and R cells.	69
3.4	Simple two-species competitive binding model	70
3.5	Thermodynamic model of transcription factor DNA binding.	73

		10
3.6	Cooperative DNA-binding sensitizes promoters to the Pnt-to-Yan ratio.	75
3.7	P:Y ratio is stabilized against varying Pnt and Yan concentrations.	78
3.8	Notch signaling lowers the Pnt-to-Yan ratio in progenitor cells.	81
3.9	Pnt and Yan expression dynamics in <i>EGFR<sup>ts</sup></i> eye discs.	82
3.10	Ras signaling elevates the Pnt-to-Yan ratio in progenitor cells.	84
3.11	Conceptual models for regulation of the Pnt-to-Yan ratio.	87
4.1	IPC ablation slows development by restricting energy metabolism.	92
4.2	Eye developmental defects are rescued by slower energy metabolism.	93
4.3	Reducing energy metabolism rescues sensory bristle patterning.	94
4.4	All microRNAs are dispensable when energy metabolism is slowed.	96
4.5	A dynamic model of gene regulation during cell fate determination.	97
4.6	Block diagram depiction of the mathematical model.	102
4.7	Simulated errors are less frequent when metabolism is reduced.	104
4.8	Reduced biosynthesis diminishes the frequency of developmental errors.	105
4.9	Expression dynamics are less affected when metabolism is reduced.	107
4.10	Reduced metabolism cannot compensate for complete loss of repression.	109
4.11	Simulated errors are less frequent when protein synthesis is reduced.	110
4.12	Reducing ribosome number rescues sensory organ development.	111
4.13	Reducing ribosome number diminishes <i>sfGFP-sens</i> overexpression.	112
4.14	Robustness of reduced metabolism simulations to parameter values.	116
4.15	Robustness of reduced metabolism simulations to threshold definition.	116
4.16	Robustness of reduced metabolism simulations to model assumptions.	120

		11
4.17	Reduced metabolism desensitizes protein dynamics to repression.	121
4.18	Robustness of ribosomopathy simulations to model assumptions.	121
5.1	The race from embryo to adulthood.	125
5.2	Simplified framework for modeling the loss of an auxiliary repressor.	130
5.3	Suppression depends on parameter sensitivity to the environment.	131
5.4	Phase diagram of auxiliary promoter loss.	132
5.5	Phase diagram of auxiliary repressor loss.	133
A.1	PntGFP expression during eye development.	175
A.2	PntGFP and Yan expression dynamics in additional cell types.	176
A.3	<i>yan</i> null clones in the eye.	177
A.4	Spatial analysis of Pnt and Yan expression in $N^{ts}$ eye discs.	178
B.1	Mutation of the four identified miR-7 binding sites in the <i>yan</i> transcript.	181
B.2	Evaluation of protein overexpression metric.	191

## List of Tables

4.1	Elementary steps of gene expression	98
4.2	Elementary steps of gene regulation	98
B.1	Mutant and transgenic alleles	180
B.2	Reaction rate parameter dependence on environmental conditions	191
B.3	Feedback strength dependence on environmental conditions	192
B.4	Two-state model dependence on environmental conditions	192
B.5	Hill kinetics model dependence on environmental conditions	193
D.1	Reproduction data	196
D.2	Reproduction code	196

## CHAPTER 1

### Introduction

The natural world presents a stunning variety of multicellular organisms. We tend to distinguish them by their physiological traits. After all, we recognize penguins by their unusual stature, black and white fur, long beak, and impressive ability to wobble around on ice. These stereotyped features are a culmination of many complex cellular processes, collectively known as *development*.

Development begins with a single cell, with subsequent growth and division events driving progression toward adulthood. Cells acquire increasingly specific roles and functions as growth proceeds, ultimately giving rise to stereotyped adult morphologies. This process, known as lineage restriction, demands that each cell decides to pursue the correct fate at the appropriate time and place [1].

Cell fate decisions are remarkably robust, collectively yielding consistent phenotypes amidst the vast array of conditions encountered in natural environments. They are so reliable that we often take them for granted. After all, it is hard to imagine a scenario in which we might mistake another human for a penguin. However, they can and do make mistakes with dire consequences for human health [2–4]. Researchers therefore continue to study these complex processes in the hope that they might one day be able to control their behavior; either to exploit them in novel biotechnologies, or rectify the diseases that arise when they fail. Research efforts are predominantly motivated by two fundamental questions. First, how do cells make decisions? Second, how do they make them reliably?

This dissertation addresses subtle aspects of both questions by combining chemical engineering, computer vision, and statistics to extract meaningful insight from experimental data. The remaining sections of this chapter serve to prime the reader with the context needed to situate the presented findings within the broader literature.

### 1.1. Molecular origins of cell fate decisions

Early experiments in the common fruit fly demonstrated that developmental processes are encoded in the genome [5, 6]. Genetics were therefore believed to provide a predefined road map for the journey from embryo to adulthood, inspiring researchers throughout the twentieth century to probe the roles of individual genes in coordinating adult phenotypes [7–9]. Most of their efforts embraced a common philosophy; break something and see what happens. In *Drosophila*, the genes themselves bear the legacy of this approach, as they are predominantly named after the phenotypes that emerge in their absence. Perturbing *eyeless* or *wingless* may now seem rudimentary, but these types of genetic perturbations were vital to the discovery of genetic components and architectures that dictate cell fate decisions in all organisms [9–11]. They revealed that some genes confer pleiotropic functions across several stages of development, while others are limited to a single context [10, 12, 13]. They also showed that some genes are only essential for proper development when others are absent, making them unnecessary under normal conditions [14–16]. Moreover, *Drosophila* continues to be a prominent model system for studying developmental processes today, owing to its conveniently short life cycle, wealth of prior knowledge, and deep library of available genetic machinery [17–19].

Among these tools, gene-specific reporters augment traditional genetic perturbations by providing localized readouts of transcript and protein abundance. Researchers can now

break something and see what happens to specific components of the developmental program. Reporters have proven particularly useful for monitoring the activities of transcription factors; proteins that bind the promoter region of other genes in order to modulate their expression. Multiple transcription factors may interact with each other, allowing for the assembly of gene regulatory networks (GRN) that integrate upstream signaling cues to elicit specific changes in gene expression [20, 21].

GRNs thus arm cells with a chemical mechanism to orchestrate cell fate decisions in space and time. A prominent example occurs during the earliest stages of *Drosophila* embryogenesis, where spatial morphogen gradients drive variegated expression of the Gap genes [22, 23]. The expressed proteins trigger subsequent developmental events in a concentration-dependent manner, inducing localized cascades of GRN activity that ultimately give rise to distinct morphological segments [24]. The embryonic landscape of Gap gene expression thereby defines a template for later stages of patterning. Developmental success is contingent upon GRNs tightly controlling the spatial precision of the template over time [25]. Thus, segment polarity definition exemplifies a broader role of GRNs; they confer positional information to inform downstream cell fate decisions [26].

The resolution of positional information is thought to be refined over time. This assertion is in part based on experimental studies of retinal patterning in the *Drosophila* eye, a setting with an enduring experimental legacy that continues to garner attention [11, 17]. It remains popular in part because the temporal history of cell fate decisions is visibly encoded in the eye imaginal disc. During the third larval instar, a wave of differentiation steadily progresses across a disordered pool of multipotent cells [27, 28]. Differentiating cells propagate this morphogenetic furrow (MF) by relaying extracellular cues downstream. Decades of experiments have steadily revealed how cells interpret the signals and commit

to particular fates [29]. Fate-specific reporters have shown that R8 photoreceptor neurons are recruited from pools of multipotent cells at the leading edge of the furrow [30]. The first cell to differentiate uses paracrine signaling to inhibit differentiation among its neighbors, resulting in a repeated pattern of regularly-spaced R8 cells that gives rise to the crystalline lattice structure of the adult eye [31]. The process dynamically transforms a disordered collection of cells into an ordered template for subsequent stages of patterning. In other words, it refines the spatial resolution of the developing eye field.

Experiments suggest cell decisions to commit to an R8 fate are non-deterministic. Differentiation is driven by the expression of Atonal, a transcription factor initially induced in all cells along the leading edge of the furrow [30, 32, 33]. Stochastic differences in cells reception of signaling cues yields variation in Atonal levels and, consequently, cells propensities to adopt an R8 fate [34, 35]. The eventual R8 cell therefore does not appear to be predetermined. Equivalent mechanisms have been shown to control many other cell fate decisions, including sensory bristle specification and the choice between epidermal and neural fates in the neuro-ectoderm of *Drosophila* [36, 37].

These examples expose another important feature of GRNs: their output is non-deterministic. Dual-reporter experiments in *E. coli* have elegantly shown that variation may be attributed to intrinsic thermal fluctuations at the molecular scale, as well as extrinsic variation in cell state [38]. Both types of noise have since been shown to manifest in the population-wide penetrance of complex organismal phenotypes [39–42]. Some developmental systems appear to leverage noise, using GRNs to amplify and reinforce stochastic fluctuations in order to limit signal responses to a randomly chosen subset of cells [34, 36, 37]. Indeed, Pel’aez et al. recently speculated that a similar mechanism may explain why subsequent R cell

fate transitions coincide with rapid increases in transcription factor expression heterogeneity [43].

These experiments and others like them form our contemporary systems-level view of development, in which complex networks of regulatory interactions guide cells toward the appropriate fates by refining their stochastic behavior over time. However, it is unknown precisely how cell fates are resolved from the dynamic activities of GRNs. It is also unclear how cells integrate stochastic inputs to execute reliable cell fate decisions. These ambiguities persist despite the wealth of experimental data generated throughout the past century. New approaches are therefore needed to unravel the complexities of GRNs and their execution of cell fate decisions.

## 1.2. The power of quantitative analysis

Experimental perturbations continue to supply some of the most potent tools in the arsenal available to researchers, but it has become clear that quantitative and systematic analysis frameworks are needed to tease apart the complex interactions that govern cell fate decisions [44, 45]. Moreover, attempts to either rectify faulty decision mechanisms or engineer new ones, such as for cancer treatment or the generation of induced pluripotent stem cells, demand predictive models backed by quantitative data [4].

The focus has therefore gradually shifted from break something and *see* what happens to break something and *measure* what happens. The transition is supported by simultaneous advances in the resolution with which we can quantify GRN activity during development. Early approaches assayed the aggregate transcript or protein content of entire tissues. Strategies have since diverged in three separate directions, each of which sharpens the focus to

single-cell resolution while prioritizing a different dimension of measurement. First, high-throughput transcriptomics strategies emphasize the breadth of genes surveyed. Second, flow cytometric approaches stress measurement precision. In both cases, cells identities may be crudely inferred from lineage-specific barcodes or biomarkers [46], but the precise spatial identity of each cell is discarded. In contrast, methods based on quantitative microscopy prioritize spatial resolution because cells are measured in their native context. These techniques generally entail imaging gene-specific fluorescent reporters before using software to detect individual cells and quantify their fluorescence levels. They are particularly well-suited to probing the intricacies of GRNs subject to signaling from adjacent cells, and have consequently become a preferred option for studying tissue-scale patterning. In terms of microscopy, the state of the art has steadily progressed from static ex vivo imaging of fixed tissues to in situ recording of single-cell expression dynamics [47]. Computational methods to support quantitative image analysis have enjoyed similar progress [48]. Combined, these advances allow researchers to collect the data needed to conduct rigorous analyses and make testable predictions [49]. That is, they promote quantification.

Quantification has completely redefined the precision of GRN analysis. The enhanced resolution has inspired exciting new research into the specific mechanisms of information transmission and processing that underlie cell fate decisions in a broad variety of model systems [43, 50–52]. These pursuits have also emboldened researchers to pause and reconsider canonical wisdom. Petkova et al. recently combined quantitative dynamic measurements and information theory to show that gap gene expression provides sufficient spatial context to uniquely define the positions of all cells [52]. The authors argued that downstream GRNs likely decode these signals with near-optimal efficiency, implying that many cell fates could, at least in principle, be determined during the earliest stages of *Drosophila* embryogenesis. The findings imply that the resolution of positional information does not necessarily need

to be refined as development proceeds. Similarly, cell fates need not be spatially resolved over time, prompting closer scrutiny of the extent to which cell fate decisions during later stages of development are truly non-deterministic. In causing the field to revisit these tenets of established dogma, the study exemplifies the power of quantification to produce novel insight.

Despite its revelatory potential, biologists have not universally embraced quantification, instead continuing to rely on visual analysis of imaging data. Resistance may partially come down to the scarcity of computational proficiency in experimental labs. Collaborations help alleviate this bottleneck, but they are often plagued by long turnaround times as researchers attempt to juggle many unrelated commitments. User-friendly analysis software would benefit experimentalists caught in this predicament by eliminating the need for computational proficiency altogether. Automated analysis frameworks have already played a similar role in many other subdisciplines of biology [53–65]. Indeed, without the support of automated alignment software, next-generation sequencing would be inaccessible to all but a few labs with extensive programming and statistical modeling experience. Similar platforms are available to support quantification of microscopy data [48, 66], but comparatively few are tailored to address the intricacies of specific model systems and experimental pipelines [67]. Further development of context-specific quantification platforms is therefore needed in order to lower the barrier to adoption of data-driven analysis.

### 1.3. Mathematical modeling of cell fate decisions

Mathematical models have reinvigorated the study of cell fate decisions in a broad variety of contexts. Most modeling efforts fall into one of two categories; those that use a

data-driven approach to recapitulate molecular mechanism, and those that provide a sparse representation of systems-level behavior.

The first class of models strive to parameterize specific biomolecular interactions by fitting a model directly to data. They typically describe the time-evolution of transcripts and proteins using systems of coupled ordinary differential equations (ODEs) reminiscent of those familiar to chemical engineers and ecologists. Despite the illusion of mechanistic detail, these models still deploy a healthy dose of abstraction. None of the commonly used rate represent true elementary reactions, instead opting for empirical representations such as linear degradation and cooperative binding kinetics. Nevertheless, many novel GRN behaviors and functions have been elegantly proposed and tested in this manner [40, 68, 69].

The second class of models forego molecular detail in favor of a coarse-grained representation of a particular phenomenon. These approaches provide a powerful means to identify, characterize, and predict behaviors that span a broad variety of model systems and developmental contexts. Among the common modeling frameworks, control theory has proven particularly fertile for generating and testing hypotheses related to GRN dynamics. Bacterial chemotaxis offers a compelling example in which molecular models were supplanted by a simple integral control framework [68, 70–72]. Analogous strategies have drawn inspiration from several disciplines to discover numerous novel functions of GRNs [42, 73–77].

Coarse-grained models are particularly well suited to studying how cell fates are resolved from spatiotemporal signaling cues. These problems would otherwise be intractable due to the many complex transport processes that shuttle signaling molecules between neighboring cells. Lubensky et al. developed a relatively simple reaction-diffusion approach to model pattern formation in the larval eye. They showed that inductive signaling cues could drive cell-autonomous positive feedback to account for the emergence of a hexagonal lattice

of R8 cells, as well as an otherwise inexplicable striped pattern observed in some mutants [78]. Gavish et al. used an even simpler model to show that an additional inhibitory signal is required to stabilize retinal patterning against minor fluctuations in cells spatial arrangement. The authors then used an experimental technique known as quantitative mosaic analysis to identify the unknown diffusible inhibitor [35].

Mathematical models have also shone light on the regulatory interactions that implement cell fate decisions within individual cells. One study explored how cells generate all-or-none responses to morphogen gradients in the *Drosophila* ventral ectoderm [79]. An ultrasensitive response mechanism was proposed to dictate the expression of Yan, a transcriptional repressor known to impede cell fate transitions [80–82]. A later study proposed that Yan plays a different role in the larval eye, instead forming a bi-stable switch through reciprocal antagonism with a transcriptional activator named Pointed (Pnt) [83]. The authors used a psuedo-molecular model to demonstrate that the decision to adopt an R cell fate could be triggered by an irreversible transition between two stable states; one characterized by high Yan and low Pnt expression, and the other by low Yan and high Pnt expression. Both transcription factors are known to be subject to several seemingly redundant sources of negative feedback. The model rationalized the purpose of these inhibitors by postulating that they enforce bi-stability of the two distinct states. Transitions between the states are triggered by an increase in Pnt levels. Shwartz et al. extended the model to include autoregulatory interactions that help flip the switch by converting transient inductive signals into sustained Pnt expression [84].

Soon thereafter, the advent of a Pnt-specific fluorescent reporter curiously revealed that Yan and Pnt are co-expressed in several developmental contexts [85]. Peláez et al. then published quantitative measurements indicating Yan exhibits mono-stable expression

dynamics in the larval eye [43]. These two studies fundamentally contradict the existing model, prompting renewed interest in both Yan and Pnt. Namely, how do they cooperate to mediate cell fate decisions? How do they do so reliably? And why are they subject to so much negative feedback?

#### 1.4. Roles for negative feedback in GRNs

Theory supports many potential uses for negative feedback in developmental GRNs [86]. Perhaps the most obvious application is the maintenance of cell states — that is, rejecting exogenous disturbances and driving cells toward a desired set point [71, 87, 88]. Negative feedback has also been shown to improve information transmission by linearizing input-output relationships and expanding dynamic range in several developmental signaling cascades [40, 69, 89–91]. Yi et al. used a FRET-based reporter of the G $\alpha$ -subunit to demonstrate dose-response alignment of G-protein signaling activity at the highest level of the yeast mating response pathway [91]. Yu et al. later quantified pFus3 activity to show that negative feedback further increases signal fidelity throughout much of the downstream pathway [69]. These functions mimic critical roles for negative feedback in human-engineered systems [92].

Negative feedback is also vital to the reliable execution of cell fate decisions. Rahimi et al. showed that Wnt-mediated negative feedback enhances the precision of morphogen gradients in the ventral domain of the *Drosophila* embryo [74]. Paulsen et al. showed that the fidelity of BMP4 signaling in the *Xenopus* embryo is improved by concomitant expression of a repressor that suppresses transduction of extrinsic noise. Eliminating the repressor leads to increased variation of pathway outputs, as well as the cell fate decisions they control

[40]. These studies provide direct experimental evidence that negative feedback can suppress phenotypic variation.

Experiments in *Drosophila* indicate that short non-coding transcripts, called microRNAs, confer a similar function by buffering developmental processes against both environmental and genetic variability [93–96]. Li et al. studied the effect of perturbing miR-7 activity during sensory organ development, and found that the microRNA stabilizes both gene expression and fate commitment decisions against fluctuating environmental conditions [95]. Cassidy et al. used directional selection to quantify the heritability of bristle formation defects in miR-9a mutants, revealing that microRNAs can also suppress the influence of intrinsic genetic variation [94]. The same group later showed that miR-9 suppresses the penetrance of genetic variants deleterious to viability at elevated temperatures [93]. Negative feedback mediated by microRNAs has therefore been shown to directly affect the evolutionary fitness of adult organisms by promoting robustness against varying environmental conditions. Precisely how this effect is achieved remains unknown.

### 1.5. Evolutionary drivers of robust cell fate decisions

Studies of developmental robustness are often traced back to C.H. Waddingtons claim that there is scarcely a mutant that is comparable in constancy with the wild type [97]. His comment reflects the consistency of adult phenotypes amidst modest levels of genetic and environmental variation, which gives way to variation when severe perturbations are introduced [98–101]. Robustness has since been attributed to individual genes and their products [102–104], as well as the systems-level architectures that they comprise [40, 93–95, 104–106]. A handful of conserved regulatory motifs, and the strategies they implement, are now understood to pervade most developmental processes [86, 87, 107, 108]. Robustness

has thus come to be accepted as a fundamental organizational principle underlying the evolution of all biological systems [109, 110]. That is, selection for genotypes that confer robustness is widely assumed to shape the evolution of gene regulatory network topologies.

The specific evolutionary drivers for increased robustness of GRNs remain unclear [111]. Citing phenomenological examples, Waddington attributed robustness to evolutionary selection for optimal traits. More recently, Siegal et al. argued that developmental robustness can arise without stabilizing selection for any particular phenotype [112]. They used a gene interaction network model to demonstrate that robustness is an emergent property of complex networks subject only to selection on the basis of their own functional stability. The same authors used *in silico* evolution to show that most genes in developmental networks buffer phenotypic variation, indicating that robustness may be an emergent byproduct of GRN architecture [113]. Computational studies have also questioned the notion that global GRN properties are a consequence of adaptive evolution, instead suggesting that local properties of genetic circuits could drive the emergence of macroscopic features, such as those that ensure robust development [114, 115].

When a gene or its products are absent or fail to perform their functions, the simplest contingency is to have others ready to compensate. It consequently seems intuitive that cells could incorporate functional redundancy to guarantee reliable cell fate decisions [107, 116]. One computational study showed that genetic redundancy is particularly evolutionarily stable in developmental systems. By simulating functionally overlapping pairs of genes, the authors demonstrated that increasing the probability with which each gene independently fails to carry out its function places increasing selection pressure upon the pair as a whole [117]. They reasoned that the resultant increase in evolutionary stability explains the prevalence of redundancy in developmental systems, where failure rates are high and erroneous cell

fate decisions yield deleterious phenotypes. Their suggestion is consistent with extensive documentation of functional redundancy in developmental systems [109].

Redundancy could arise through gene duplication or convergent evolution. A. Wagner argued against the former. He analyzed gene sequence and expression data in yeast to show that the phenotypic severity of loss-of-function mutations does not correlate with the availability of functionally equivalent genes. He concluded that functional redundancy is not the primary source of robustness against genetic variation, instead favoring epistatic interactions between otherwise unrelated gene products [118]. Combined, these studies suggest the prevalence of functional redundancy in GRNs may be driven by as-of-yet unknown forces.

## CHAPTER 2

## Automated analysis of mosaic eye imaginal discs in *Drosophila*

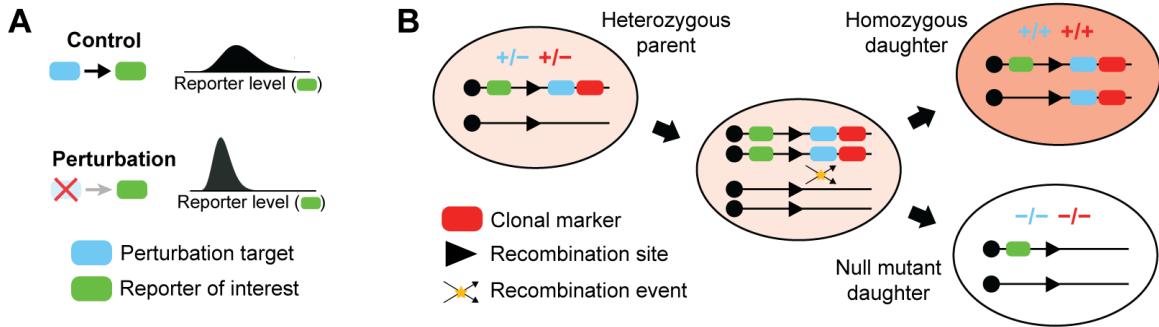
A manuscript resembling this chapter was coauthored with Neda Bagheri and Luís Amaral, using experimental data collected by Nicolás Peláez. These data were borrowed from the study presented in Chapter 3.

### 2.1. Background on quantitative mosaic analysis

Quantification will be essential as biologists study increasingly complex facets of organismal development [45]. Unfortunately, qualitative analysis remains common because it is often difficult to measure cellular processes in their native context. Modern fluorescent probes and microscopy techniques make such measurements possible [119–121], but the ensuing image analysis demands specialized skills that fall beyond the expertise of most experimentalists. Automated analysis strategies have addressed similar challenges in cytometry [53–55], genomics and transcriptomics [56–59], and other subdisciplines of biology [60, 61]. Image analysis has proven particularly amenable to automation, with several computer vision tools having gained traction among biologists [62–65]. These platforms are popular because they increase productivity, improve the consistency and sensitivity of measurements, and obviate the need for specialized computational proficiency [48, 66, 67]. Designing similar tools to help biologists probe and measure developmental processes *in vivo* will further transform studies of embryogenesis and development into quantitative endeavors.

Developmental biologists study how the expression and function of individual genes coordinate the emergence of adult phenotypes. They often ask how cells respond when a specific gene is perturbed during a particular stage of development. Cell response may be characterized by changes in morphology, or by changes in the expression of other genes (Fig. 2.1A). Experimental efforts to answer this question were historically stifled by the difficulty of isolating perturbations to a single developmental context, as the most interesting perturbation targets often confer pleiotropic function across several stages of development [122–124].

Mosaic analysis addressed this challenge in *Drosophila* by limiting perturbations to a subset of cells within the larval eye [125, 126], a model system with enduring relevance [17]. The technique yields a heterogeneous tissue comprised of genetically distinct patches of cells, or clones. Clone formation may be restricted to specific developmental contexts by using endogenously-activated promoters to drive trans-chromosomal recombination events [127, 128]. The timing of these events determines the number and size of the resultant clones [129]. Cells within each clone are genetically identical. Perturbations are applied by engineering the dosage of a target gene to differ across clones (Fig. 2.1B), resulting in clones whose cells are either mutant (−/−), heterozygous (+/−), or homozygous (+/+) for the perturbed gene. Labeling these clones with fluorescent markers enables direct comparison of cells subject to control or perturbation conditions (Fig. 2.2A). Additional reporters may be used to monitor differences in gene expression or morphology across clones (Fig. 2.2B). Variants of this strategy led to seminal discoveries in neural patterning [130–132] and morphogenesis [133, 134], and remain popular today [18, 19, 135].



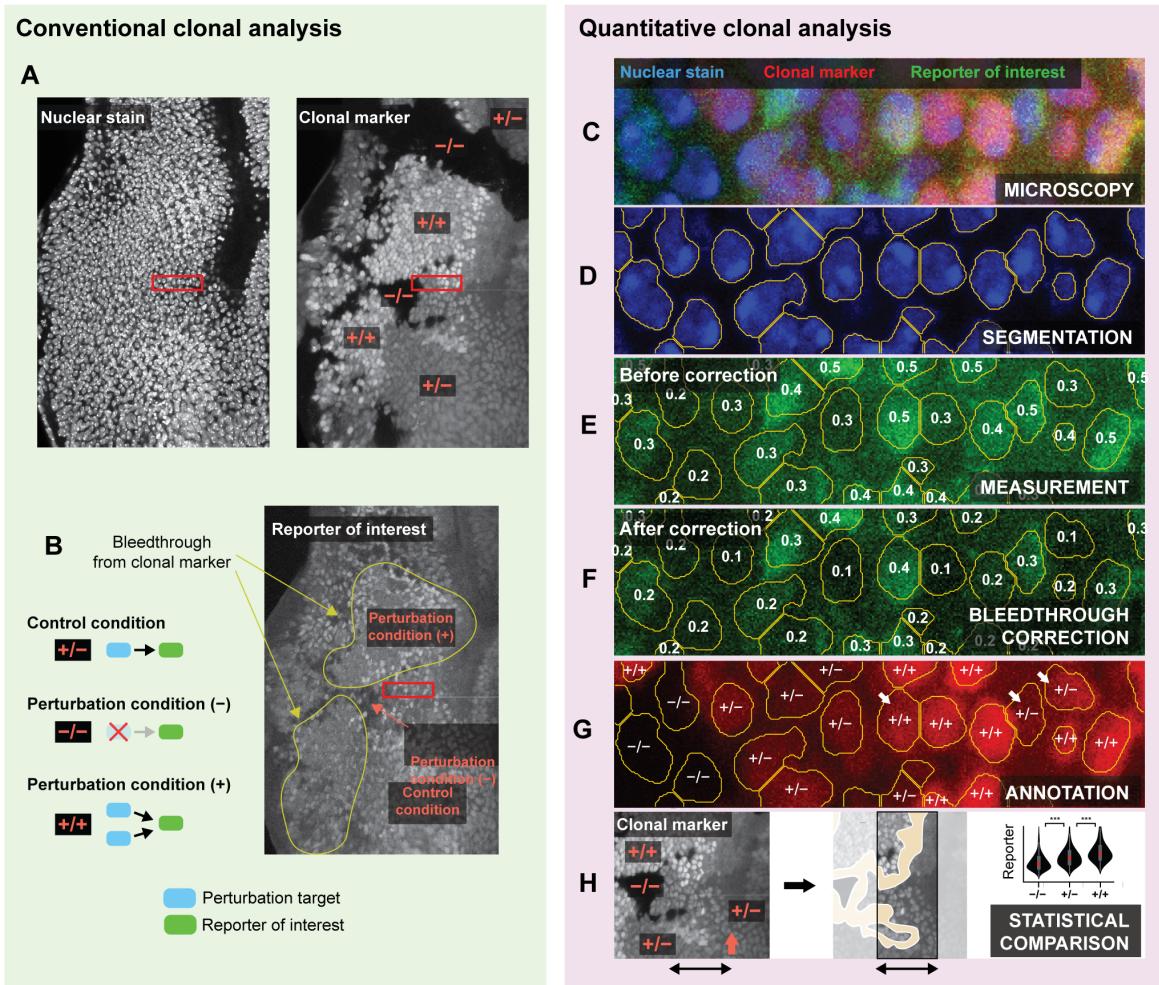
**Figure 2.1. Perturbing gene expression via mitotic recombination.** Experimental framework using mitotic clones to test whether or not regulatory interactions occur between a perturbation target and reporter of interest. Blue and green ovals represent the respective genes encoding the perturbation target and the reporter. (A) A perturbation-induced decrease in reporter levels would confirm that regulation occurs. (B) Mitotic recombination generates clonal subpopulations carrying zero, one, or two copies of the gene encoding a perturbation target. Black lines depict a genetic locus. Only genes downstream of the recombination site are subject to recombination. Red ovals represent a gene encoding a clonal marker used to identify the resultant clones.

Quantitative microscopy techniques are well suited to measuring differences in cell behavior across clones. One reporter (a clonal marker) labels the clones, while others quantitatively report properties of their constituent cells, such as the expression level of a gene product of interest (Fig. 2.2C). The former then defines the stratification under which the latter are compared. We call this strategy quantitative mosaic analysis because it replaces subjective visual comparison with a rigorous statistical alternative. Although many recent studies have deployed this approach [136–139], qualitative visual comparison remains pervasive in the literature.

We suspect the adoption of quantitative mosaic analysis is hindered by an overwhelming demand for specialized computational skills or, in their stead, extensive manual labor. Researchers must first draw or detect boundaries around individual nuclei in a procedure commonly known as segmentation (Fig. 2.2D). Averaging the pixel intensities within each boundary then yields a fluorescence intensity measurement for each reporter in each identified nucleus (Fig. 2.2E). The measurements should then be corrected to account for any fluorescence bleedthrough between reporter channels (Fig. 2.2F). Correction often requires

single-reporter calibration experiments to quantify the crosstalk, followed by complex calculations to remedy the data [140, 141]. Researchers must then label, or annotate, each identified nucleus as mutant, heterozygous, or homozygous for the clonal marker. Annotation is typically achieved through visual inspection (Fig. 2.2G). Cells carrying zero, one, or two copies of the clonal marker should exhibit low, medium, or high average levels of fluorescence, respectively. However, both measurement and biological noise introduce the possibility that some cells measured fluorescence levels may not reliably reflect their genetic identity. Annotation must therefore also consider the spatial context surrounding each nucleus. For instance, a nucleus whose neighbors express high levels of the clonal marker is likely to be homozygous for the clonal marker, even if its individual fluorescence level is comparable to that of heterozygous cells (Fig. 2.2G, white arrows). With many biological replicates containing thousands of cells each, annotation can quickly become insurmountably tedious. The corrected and labeled measurements are then curated for statistical comparison by excluding those on the border of each clone, and limiting their scope to particular regions of the image field (Fig. 2.2H). Combined, all of these tasks ultimately burden researchers and raise the barrier for adoption of quantitative analysis strategies.

Automation promises to alleviate this bottleneck, yet the literature bears surprisingly few computational resources designed to support quantitative mosaic analysis. The Clonal-Tools plugin for ImageJ deploys an image-based approach to measure macroscopic features of clone morphology, but is limited to binary classification of mutant versus non-mutant tissue and offers no functionality for comparing reporter expression across clones [142]. Alternatively, the MosaicSuite plugin for ImageJ deploys an array of image processing, segmentation, and analysis capabilities to automatically detect spatial interactions between objects found in separate fluorescence channels [143, 144]. While useful in many other settings, neither of these tools support automated labeling of individual cells or explicit comparison of clones with



**Figure 2.2. Conventional versus quantitative mosaic analysis.** (A,B) Conventional clonal analysis in the larval fruit fly eye. (A) Clones are identified by visual comparison of clonal marker fluorescence among nuclei. (B) Regions labeled mutant ( $-/-$ ) or homozygous ( $+/+$ ) for the clonal marker are compared with those labeled heterozygous ( $+/-$ ) to assess whether reporter expression differs across clones. Fluorescence bleed-through is arbitrarily diagnosed. (C-H) Quantitative clonal analysis. Panels depict a magnified view of the region enclosed by red rectangles in panels A and B. (C) Raw confocal image of the nuclear stain, clonal marker, and reporter of interest. (D) Segmentation identifies distinct nuclei. (E) Reporter expression is quantified by averaging the pixel intensities within each segment. Numbers reflect measured values. (F) Measurements may be corrected to mitigate fluorescence bleedthrough. (G) Individual nuclei are labeled mutant, heterozygous, or homozygous for the clonal marker. White arrows mark nuclei with ambiguous fluorescence levels. (H) Reporter levels are compared across clones to determine whether the perturbation affects reporter expression. Yellow region marks excluded clone borders. Comparison may exclude clone borders (yellow regions) and focus on a particular region of the image field (black arrows). In the larval eye, comparison is often limited to a narrow window near the MF (orange arrow).

single-cell resolution. Most modern studies employing a quantitative mosaic analysis instead report using some form of ad hoc semi-automated pipeline built upon ImageJ [136, 138, 139].

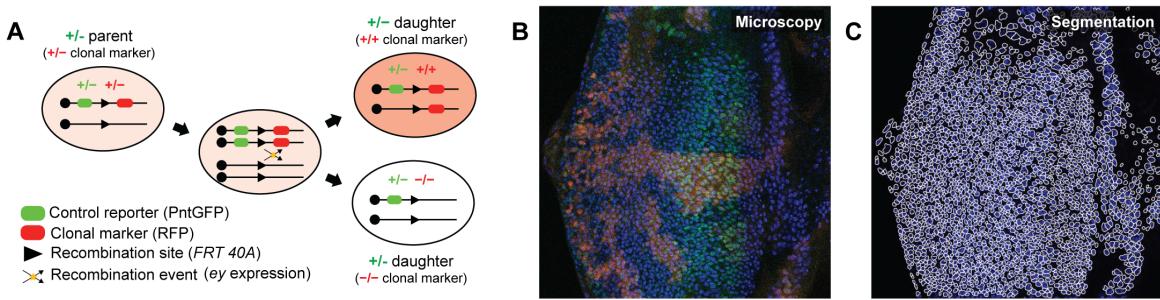
We are therefore unaware of any platforms that offer comprehensive support for an automated quantitative mosaic analysis workflow.

We previously published a framework for automated segmentation of cell nuclei in the larval eye [43], in addition to a collection of tools for analyzing reporter expression in the segmented nuclei [137]. Here, we leverage this experience to create an open-source framework for automated quantitative mosaic analysis of eye imaginal discs. The framework supports segmentation, bleedthrough correction, and annotation of confocal microscopy data (Fig. 2.2D-H). We demonstrate each of these functions by applying them to real confocal images of clones in the eye, and find that our automated approach yields results consistent with manual analysis by a human expert. We then generate and use synthetic data to survey the performance of our framework across a broad range of biologically plausible conditions.

## 2.2. Experimental data

We borrowed an experimental control dataset from a separate study of neuronal fate commitment during retinal patterning that is later discussed in Chapter 3. The data consist of six eye imaginal discs dissected and fixed during the third larval instar of *Drosophila* development. Within each disc, *ey>FLP* and *FRT40A* were used to generate mitotic clones. The chromosome targeted for recombination was marked with a *Ubi-mRFPnls* transgene (Fig. 2.3A), enabling automated detection of subpopulations marked by distinct levels of mRFP fluorescence (Fig. 2.3B). The discs also carried a *pnt-GFP* transgene located on a different chromosome that was not subject to mitotic recombination. Discs were dissected, fixed, and co-stained with a DAPI nuclear marker prior to confocal imaging. Please refer to A.1 for additional details regarding genetics and experimental conditions.

As shown in Chapter 3, the PntGFP reporter is predominantly expressed in two narrow stripes of progenitor cells during eye development. The first stripe occurs immediately posterior to a wave of developmental signaling that traverses the eye disc. Progenitor cells located in this region are suitable for comparison because they are of approximately equivalent developmental age. We applied our automated analysis framework to a total of nine images of these cells.



**Figure 2.3. Example clones in the larval fly eye.** (A) Genetic schema for a bleedthrough control experiment. Red and green ovals represent genes encoding a RFP-tagged clonal marker and a GFP-tagged control reporter, respectively. Black lines depict a genomic locus. Recombination does not affect gene dosage of the control reporter, so GFP variation across clones is attributed to fluorescence bleedthrough. (B) Confocal image of an eye imaginal disc. Red, green, and blue reflect clonal marker, control reporter, and nuclear stain fluorescence, respectively. (C) Segmentation of the DAPI nuclear stain. White lines show individual segments.

### 2.3. Image segmentation and quantification of nuclear fluorescence levels

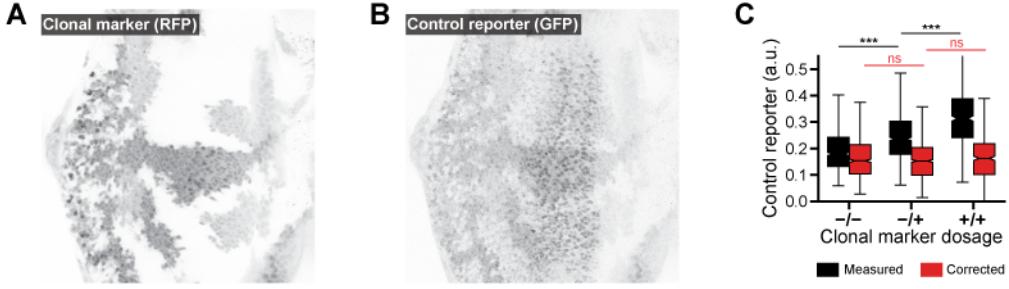
We implemented a segmentation strategy based upon a standard watershed approach [145]. Briefly, we construct a foreground mask by Otsu thresholding the nuclear stain image following a series of smoothing and contrast-limited adaptive histogram equalization operations [145, 146]. We then apply a Euclidean distance transform to the foreground mask, identify the local maxima, and use them as seeds for watershed segmentation. When applied to the microscopy data, few visible spots in the nuclear stain were neglected, and the vast majority of segments outlined individual nuclei (Fig. 2.3C).

This approach is flexible and should perform adequately in many scenarios. However, we acknowledge that no individual strategy can address all microscopy data because segmentation is strongly context dependent. All subsequent stages of analysis were therefore designed to be compatible with any data that conform to our standardized file structure. This modular arrangement grants users the freedom to use one of the many other available segmentation platforms [147], including FlyEye Silhouette [43], before applying the remaining functionalities of our framework. Regardless of how nuclear contours are identified, averaging the pixel intensities within them yields fluorescence intensity measurements for each reporter in each identified nucleus. We next sought to ensure that these measurements were suitable for comparison across clones.

#### **2.4. Bleedthrough correction**

Despite efforts to select non-overlapping reporter bandwidths and excite them sequentially, it is not uncommon for reporters excited at one wavelength to emit some fluorescence in another channel (Fig. 2.2B, yellow lines) [140, 148]. The end result is a positive correlation, or crosstalk, between the measured fluorescence intensities of two or more reporters. Exogenous correlations are problematic given that the purpose of the experiment is to detect changes in reporter levels with respect to the clonal marker.

In our microscopy data, individual clones were distinguished by their low, medium, or high expression levels of an RFP-tagged clonal marker (Fig. 2.4A). These images should not have shown any detectable difference in GFP levels across clones because all cells carried an equivalent dosage of the control reporter (Fig. 2.3A). However, the images visibly suffered from bleedthrough between the RFP and GFP channels (Fig. 2.4A,B). Bleedthrough was similarly evident when we compared measured GFP levels across labeled clones. Nuclei



**Figure 2.4. Automated correction of fluorescence bleedthrough in the larval eye.** (A) Low, medium, and high expression levels of the RFP-tagged clonal marker. (B) GFP-tagged control reporter expression. RFP fluorescence bleedthrough is visually apparent upon comparison with A. (C) Comparison of control reporter expression between clones. Includes data aggregated across nine images taken from six separate eye discs. Data were limited to cells within the region of elevated GFP expression that were of approximately comparable developmental age (see Fig. 2.5E-G). Measurements are stratified by their assigned labels. Before correction, expression differs between clones (black boxes,  $p < 10^{-5}$ ). No difference is detected after correction (red boxes,  $p > 0.05$ ).

labeled mutant, heterozygous, or homozygous for the clonal marker had low, medium, and high expression levels of the control reporter, respectively (Fig. 2.4C, black boxes). The data were therefore ripe for systematic correction.

Spectral bleedthrough correction is common practice in other forms of cross-correlation and co-localization microscopy [140, 148]. These methods typically entail characterizing the extent of crosstalk between fluorophores globally [149, 150], on a pixel-by-pixel basis [141], or by experimental calibration [140], then detrending all images or measurements prior to subsequent analysis. Our framework adopts the global approach, using the background pixels in each image to infer the extent of fluorescence bleedthrough across spectral channels.

#### 2.4.1. Statistical basis of bleedthrough correction

Specifically, we assume the fluorescence intensity  $F_{ij}$  for channel  $i$  at pixel  $j$  is a superposition of a background intensity  $B_{ij}$  and some function of the expression level  $E_{ij}$  that we seek to compare across cells [151]:

$$F_{ij} = B_{ij} + f(E_{ij}) \quad (2.1)$$

We further assume that the background intensity of a channel includes linear contributions from the fluorescence intensity of each of the other channels:

$$B_{ij} = \sum_{k \neq i} \alpha_k F_{kj} + \beta \quad (2.2)$$

where  $k$  is indexed over  $K$  anticipated sources of bleedthrough. Given estimates for each  $\{\alpha_1, \alpha_2, \dots, \alpha_K\}$  and  $\beta$  we can then estimate the background intensity of each measurement:

$$\langle B_{ij} \rangle = \sum_{k \neq i} \alpha_k \langle F_{kj} \rangle + \beta \quad (2.3)$$

where the braces denote the average across all pixels within a single nucleus. The corrected signal value is obtained by subtracting the background intensity from the measured fluorescence level:

$$\langle f(E_{ij}) \rangle = \langle F_{ij} \rangle - \langle B_{ij} \rangle \quad (2.4)$$

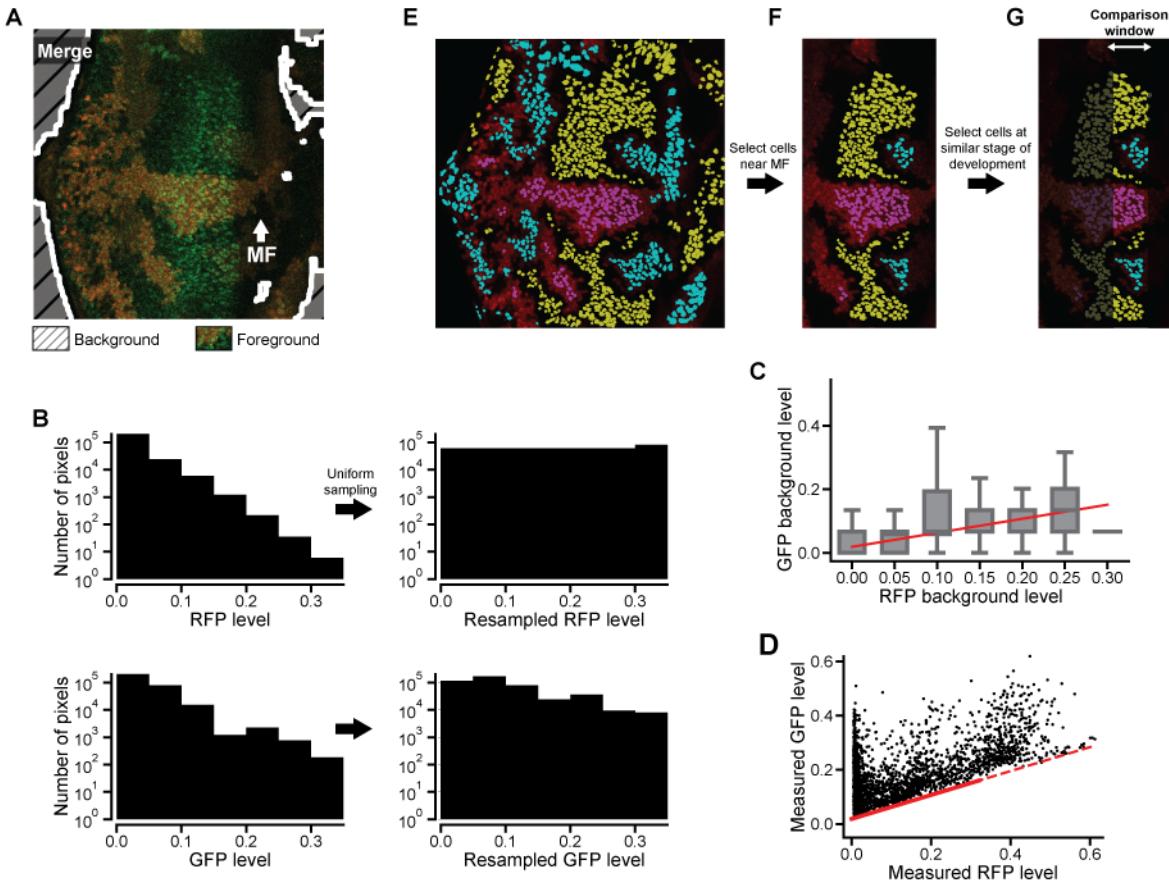
Repeating this procedure for each nucleus facilitates comparison of relative expression levels across nuclei in the absence of bleedthrough effects. Bleedthrough correction performance is therefore strongly dependent upon accurate estimation of the bleedthrough contribution strengths,  $\{\alpha_1, \alpha_2, \dots, \alpha_K\}$ .

#### 2.4.2. Characterization of fluorescence bleedthrough coefficients

We estimate these parameters by characterizing their impact on background pixels. For each image, we morphologically dilate the foreground until no features remain visible (Fig. 2.5A). We then extract the background pixels and resample them such that the distribution of pixel intensities is approximately uniform (Fig. 2.5B). Resampling helps mitigate the skewed distribution of pixel intensities found in the background. We then estimate values for each

$\{\alpha_1, \alpha_2, \dots, \alpha_K\}$  and  $\beta$  by fitting a generalized linear model to the fluorescence intensities of the resampled pixels (Fig. 2.5C). Each model is a variant of Equation 2.3 in which angled braces instead denote averages across all background pixels. We formulate these models with identity link functions under the assumption that residuals are gamma distributed. Their coefficients provide an estimate of the bleedthrough contribution strengths that may then be used to estimate the background fluorescence intensity of each nucleus in the corresponding image (Fig. 2.5D). The measurements may then be corrected through application of Equation 2.4.

We applied this procedure to the microscopy data, then compared measured control reporter expression across labeled clones. To mitigate edge effects, cells residing on the periphery of each clone were excluded from all comparisons (Fig. 2.5E). Border cells were identified by using a Delaunay triangulation to find all cells connected to a neighbor within a different clone. Our framework includes a simple graphical user interface that permits manual curation of which regions of the image field are included in subsequent analyses. We used this tool to limit our analysis to the region of elevated GFP expression near the morphogenetic furrow (Fig. 2.5F). Comparisons were further restricted to cells undergoing similar stages of development (Fig. 2.5G). These restrictions served to buffer against differences in developmental context and ensured that all compared cells were of similar developmental age. The remaining fluorescence measurements were then aggregated across all eye discs and compared between pairs of clones, confirming that bleedthrough correction successfully eliminated any detectable difference in GFP expression (Fig. 2.4C, red boxes,  $p > 0.05$  two-sided Mann-Whitney  $U$  test).



**Figure 2.5. Using background pixels to characterize bleedthrough contributions in the foreground.** (A) Extraction of background pixels (striped region). Foreground includes the merged RFP and GFP images, surrounded by a white line. White arrow marks the morphogenetic furrow (MF). (B) Background pixel values are resampled such that RFP intensities are uniformly distributed. (C) A generalized linear model characterizes the contribution of RFP bleedthrough to GFP fluorescence. Boxes reflect windowed distributions of resampled background pixel intensities. Red line shows the model fit. (D) Measured GFP levels before bleedthrough correction. Markers represent individual nuclei. Red line shows the inferred contributions of RFP fluorescence bleedthrough. Dashed portion is extrapolated. (E-G) Data curation prior to statistical comparison of GFP levels. (E) Cells on the periphery of each clone are excluded. (F) The selection is limited to the region of elevated GFP expression near the MF. (G) It is further limited to cells of the same developmental age, defined by their relative positions along the x-axis.

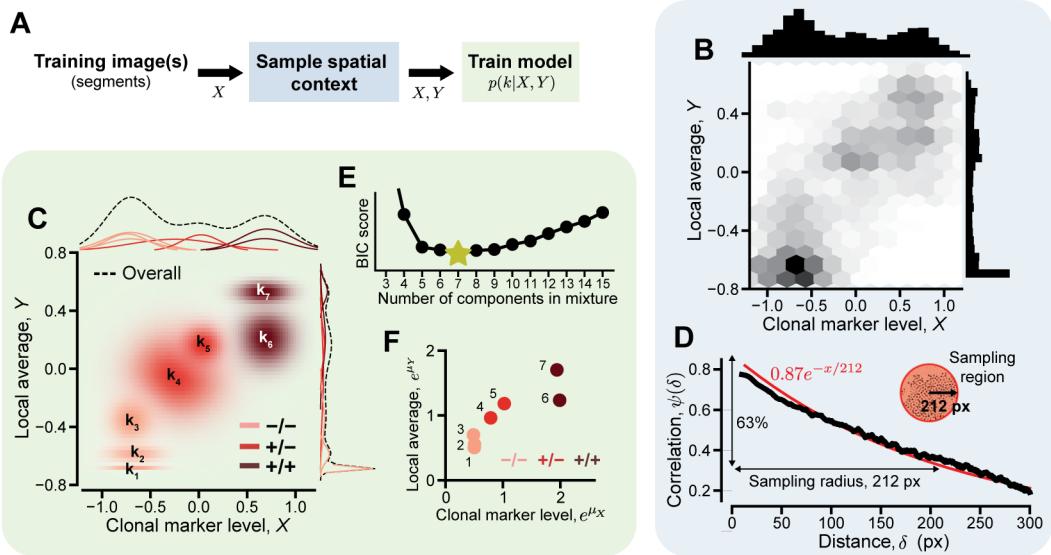
## 2.5. Automated annotation of clones

Our annotation strategy seeks to label each identified cell as mutant, heterozygous, or homozygous for the clonal marker. Variation within each clone precludes accurate classification of a cell's genotype solely on the basis of its individual expression level. However, clonal

lineages are unlikely to exist in isolation because recombination events are typically timed to generate large clones. Our strategy therefore integrates both clonal marker expression and spatial context to identify clusters of cells with locally homogeneous expression behavior, then maps each cluster to one of the possible labels. This unsupervised approach lends itself to automated annotation because the clusters are inferred directly from the data without any guidance from the user.

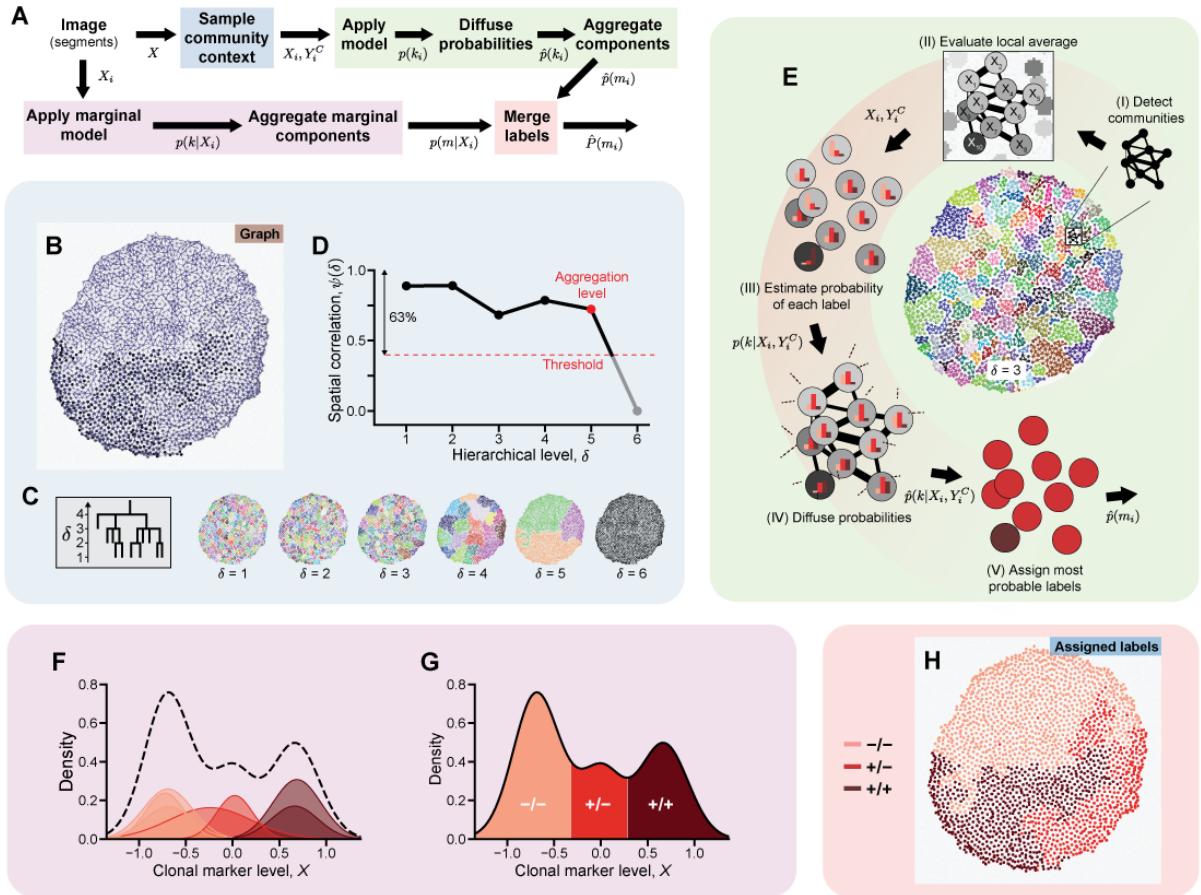
### 2.5.1. Qualitative overview of clone annotation algorithm

We first train a statistical model to estimate the probability that a given measurement came from a cell carrying zero, one, or two copies of the clonal marker (Fig. 2.6A). This entails fitting a weighted mixture of three or more bivariate lognormal distributions (components) to a two dimensional set of observations (Fig. 2.6B,C). The first dimension corresponds to the clonal marker fluorescence level measured within each cell. The second dimension describes the local average expression level within the region surrounding each cell. We evaluate the latter by estimating a neighborhood radius from the decay of the radial correlation of the expression levels, then averaging the expression levels of all cells within that radius (Fig. 2.6D). The second dimension therefore measures the spatial context in which a cell resides. We balance model fidelity against overfitting by using the Bayesian information criterion to determine the optimal number of model components (Fig. 2.6E). We then cluster the components into three groups on the basis of their mean values (Fig. 2.6F), effectively mapping each component to one of the three possible gene dosages. The model may be trained using observations derived from a single image, or with a collection of observations derived from multiple images. Once trained, the model is able to predict the conditional probability that an individual observation belongs to one of the model's components, given its measured expression level.



**Figure 2.6. Training a clone annotation model.** (A) One or more images are segmented, yielding a set of fluorescence measurements  $X$ . These are used to sample the spatial context  $Y$  of the neighborhood surrounding each cell. Both sets of values are used to train a mixture model. Subsequent panels demonstrate these procedures using the example shown in Figure 2.6C. (B) Expression levels are jointly distributed with the local average among neighboring cells. Center panel shows the joint distribution. Top and right bar plots show marginal distributions. (C) Mixture model identifies seven distinct components  $k_i$ . Center panel shows position and spread of each component. Top and right panels show marginal components scaled by their respective weights. Red shading denotes the label  $m_i$  assigned to each component. The model predicts the posterior probabilities that a given sample ( $X, Y$ ) belongs to each component. (D) Neighborhood size is estimated by computing the decay constant of the spatial correlation function,  $\psi(\delta)$ . Black line shows the moving average of  $\psi(\delta)$ , red line shows an exponential fit. Inset shows the resultant sampling region. (E) The optimal number of mixture components is determined by minimizing BIC score. (F) Mixture components are labeled by k-means clustering their mean values. Markers reflect the component means, colors denote the assigned label.

We then use the learned conditional probabilities to detect entire clones, thus assigning a label to each cell. Rather than using the trained model to classify each observation, we compile a new set of observations by limiting each estimate of spatial context to spatially collocated communities with similar expression behavior (Fig. 2.7A). We identify these communities by applying a community detection algorithm to an undirected graph connecting adjacent cells (Fig. 2.7B). Edges in this graph are weighted by the similarity of clonal marker expression between neighbors, resulting in communities with similar expression levels (Fig. 2.7E, Steps I and II). The graph-based approach increases spatial resolution by limiting the information shared by dissimilar neighbors. Applying the mixture model yields an initial



**Figure 2.7.** Label assignment using a trained clone annotation model.

estimate of the probability that an observation belongs to one of the model's components (Fig. 2.7E, Step III). We further refine these estimates by allowing the probabilities estimated for each cell to diffuse throughout the graph (Fig. 2.7E, Step IV). The rate of diffusion between neighbors is determined by the weight of the edge that connects them, with more similar neighbors exerting stronger influence on each other. We then use the diffused probabilities to identify the most probable source component and label each observation (Fig. 2.7E, Step V). These probabilities also provide a measure of confidence in the assigned labels. We replace any low-confidence labels with alternate labels assigned using a marginal classifier that neglects spatial context (Fig. 2.7F,G), resulting in a fully labeled image (Fig. 2.7H).

**Figure 2.7.** (A) The measurements  $X$  from a segmented image are used to sample the spatial context  $Y^C$  of the community surrounding each cell before the mixture model is applied (blue and green path). They are simultaneously labeled using a marginal projection of the trained model (magenta path). The two sets of labels are then merged (red path). Subsequent panels demonstrate these procedures using the example shown in Figure 2.6C. (B-D) Spatial context sampling. (B) Weighted undirected graph connecting adjacent cells. Line thickness reflects the expression similarity between neighbors. (C) Community resolution is defined by aggregating clusters that fall below a cut level  $\delta$  in the hierarchy. Images show potential levels of aggregation. Colors denote distinct communities. (D) Cut level is chosen by finding the maximum level (red dot) that remains lower than the decay constant of the spatial correlation function,  $\psi(\delta)$  (black line). In this example, clusters are aggregated below the fifth level. Panel E instead depicts aggregation below the third level for ease of visualization. (E) Application of the mixture model. (I) The graph connecting adjacent cells contains distinct communities of locally similar expression. (II) Mean expression level within each community serves as the local average for each cell. (III) Mixture model estimates the probability that each cell belongs to each of its component. Bar plots within each cell illustrate the cumulative probability of each label. (IV) Posterior probabilities are diffused across the entire graph. (V) Each cell is assigned its most probable label. (F,G) Application of a marginal mixture model that neglects spatial context. (F) Marginal mixture model components obtained by summing across the spatial context dimension of the full mixture model. Red shading denotes the label assigned to each component. Dashed black line is the overall marginal density. (G) Marginal classifier that labels cells strictly on the basis of their individual fluorescence level. Red shading denotes the most probable label for each expression level. (H) Annotated measurements. Red shading denotes the assigned label. Labels with low confidence  $\hat{P}(m_i) < 0.8$  are replaced by their marginal counterparts.

The algorithm leverages the collective wisdom of neighboring measurements to override spatially isolated fluctuations in clonal marker expression, and thereby enforces consistent annotation within contiguous regions of the image field. The size of these regions depends upon the granularity of estimates for the spatial context surrounding each cell. We used an unsupervised approach to choose an appropriate spatial resolution in a principled manner. In short, the resolution is matched to the approximate length scale over which expression levels remain correlated among cells. Both the training and application stages of our annotation algorithm use this automated approach (Figs. 2.6D and 2.7D), thus averting any need for user input.

### 2.5.2. Statistical basis of clone annotation

We assume the measured fluorescence level  $x_i$  for cell  $i$  is sampled from an underlying distribution  $p_m(x)$  for cells carrying  $m$  copies of the gene encoding the clonal marker:

$$x_i \sim p_m(x) \quad (2.5)$$

We further assume that  $p_m(x)$  is comprised of a mixture of one or more lognormal distributions:

$$p_m(\ln x) = \sum_{n=1}^N \lambda_n \mathcal{N}(\ln x | \theta_n) \quad (2.6)$$

$$\sum_{n=1}^N \lambda_n = 1 \quad (2.7)$$

where  $0 \leq \lambda \leq 1$  are the mixing proportions,  $\theta_n = (\mu_n, \sigma_n^2)$  are the mean and variance of the  $n$ th distribution. This assumption is supported by both empirical observations and theoretical insights [152, 153]. By superposition, the global distribution of measured fluorescence levels  $p(\ln x)$  for all values of  $m$  are also sampled from a mixture of  $K$  components:

$$p(\ln x) = \sum_{m=0}^2 \alpha_m p_m(\ln x) = \sum_{m=0}^2 \alpha_m \sum_{n=1}^N \lambda_n \mathcal{N}(\ln x | \theta_n) = \sum_{k=1}^K \lambda_k \mathcal{N}(\ln x | \theta_k) \quad (2.8)$$

$$\sum_{k=1}^K \lambda_k = 1 \quad (2.9)$$

where  $\alpha_m$  denotes the overall fraction of cells with  $m$  copies of the gene encoding the clonal marker. For brevity, we substitute  $X = \ln x$  yielding:

$$p(X) = \sum_{k=1}^K \lambda_k \mathcal{N}(X | \theta_k) \quad (2.10)$$

Given a collection of sampled fluorescence levels,  $\{X_i\}_{i=1\dots N}$ , we use expectation maximization to find values of  $\theta_k$  and  $\lambda_k$  for each of the model's  $K$  components that maximize the log-likelihood of the observed sample. We repeat this procedure for a range of sequential values of  $K$ , resulting in multiple models of increasing size. We then balance model resolution against overfitting by selecting the model that yields the smallest value of the Bayesian Information Criterion (BIC):

$$BIC(K) = \ln(N)q_K - 2\ln(\hat{L}_K) \quad (2.11)$$

$$q_K = K - 1 + 2^K \quad (2.12)$$

where  $N$  is the sample size,  $\ln(\hat{L})_K$  is the maximum value of the log-likelihood, the subscript  $K$  denotes the number of mixture components in the model, and  $q_K$  is the total number of parameters (i.e.  $K - 1$  values of  $\lambda_k$  and  $2^K$  values of  $\mu_k$  and  $\sigma_k^2$ ).

Applying Bayes' rule to the selected model infers the posterior probabilities that each sample  $X_i$  belongs to the  $k$ th component:

$$p(k|X_i) = \frac{p(X_i|k)p(k)}{p(X_i)} = \frac{p(X_i|k)\lambda_k}{p(X_i)} \quad (2.13)$$

where  $p(X_i | k)$  is evaluated using the model's likelihood function and  $p(X_i)$  is evaluated by marginalizing across each of the model's  $K$  components. The end result is a mixture model that allows us to predict the probability that a given measurement of clonal marker expression belongs to a particular one of its component distributions.

We then define a many-to-one mapping,  $f$ , from each of the  $K$  components of the mixture to each of the three possible values of  $m$ :

$$f : \{0, 1, \dots, K\} \rightarrow \{0, 1, 2\} \quad (2.14)$$

We determine the mapping by k-means clustering the  $K$  component distributions into three groups on the basis of their mean values,  $e^{\mu_k}$ . We may then assign a genotype label  $m$  to each measurement  $X_i$  by predicting the component  $k$  from which it was sampled. The accuracy of these labels depends upon how closely the fitted mixture model reflects the true partitioning of gene copies among clones. While finite mixtures are always identifiable given a sufficiently large sample [154], the algorithm used to fit the mixture tends toward local maxima of the likelihood function when the true components are similar (Wu, 1983). An approach based on a univariate mixture is thus inherently prone to failure when expression levels extensively overlap across clones, as variation within each clone precludes accurate classification of a cell's genotype solely on the basis of its individual expression level. However, clonal lineages are unlikely to exist in isolation because recombination events are usually timed to generate large clones. Our strategy therefore integrates both clonal marker expression and spatial context to identify clusters of cells with locally homogeneous expression behavior.

We incorporate spatial context by introducing a second jointly-distributed variable  $Y_i$ :

$$Y_i = \frac{1}{M_i} \sum_{j=0}^{M_i} X_j \quad (2.15)$$

where the subscript  $j$  indexes all  $M_i$  neighbors of cell  $i$ . The new variable reflects the average expression level among the neighbors surrounding each cell. We define neighbors as pairs of cells located within a critical distance of each other. This distance, or sampling radius, is derived from the approximate length scale over which cells retain approximately similar clonal marker expression levels. Specifically, we determine the exponential decay constant of the spatial correlation function,  $\psi(\delta)$ :

$$\psi(\delta) = \frac{<(X_i - \mu_X)(X_j - \mu_X)>_{i,j \in \delta}}{\sigma_X^2} \quad (2.16)$$

where  $\mu_X$  and  $\sigma_X^2$  are the global mean and standard deviation, and angled brackets denote the mean across all pairs of cells separated by distance  $\delta$ . We efficiently implement this procedure by fitting an exponential decay function to the down-sampled moving average of  $\psi(\delta)$  as a function of increasing separation distance.

Following the introduction of spatial context, the mixture model becomes:

$$p(X, Y) = \sum_{k=1}^K \lambda_k \mathcal{N}(X, Y | \theta_k) \quad (2.17)$$

where  $\theta_k = (\vec{\mu}_k, \vec{\sigma}_k^2)$  contains the mean and variance of each component given by vectors of length two. This formulation constrains each component's covariance matrix to be diagonal.

The posterior is now:

$$p(k|X_i, Y_i) = \frac{p(X_i, Y_i|k)\lambda_k}{p(X_i, Y_i)} \quad (2.18)$$

We can recover the univariate model by marginalizing the posterior over all values of  $Y$ :

$$p(k|X_i) = \sum_j p(k|X_i, Y_j) \quad (2.19)$$

When neglecting spatial context, we use this expression to classify each sample by applying the mapping  $f$  to the value of  $k$  that maximizes  $p(k | X_i)$ :

$$f(\operatorname{argmax}_k p(k|X_i)) \quad (2.20)$$

In all other cases, we deploy a graph-based approach to refine the estimate of  $p(k | X_i, Y_i)$ . This first entails constructing an undirected graph connecting adjacent cells within each image. We obtain the graph's edges through Delaunay triangulation of the measured cell positions, then exclude distant neighbors by thresholding the edge lengths. Each edge is assigned a weight  $w_{ij}$  reflecting the similarity of clonal marker expression between adjacent

cells  $i$  and  $j$ :

$$w_{ij} = \exp\left(\frac{-E_{ij}}{\langle E \rangle}\right) \quad (2.21)$$

$$E_{ij} = |X_i - X_j| \quad (2.22)$$

where  $E_{ij}$  is the absolute log fold-change in measured expression level and angled brackets denote the mean across all edges. We chose an exponential formulation because it yields an approximately uniform distribution of edge weights. We then detect communities within the graph using the Infomap algorithm [155]. The algorithm provides a hierarchical partitioning of nodes into non-overlapping clusters. We aggregate all clusters below a critical level that is again chosen by estimating the spatial correlation decay constant. We then enumerate  $p(k | X_i, Y_i^c)$  where  $Y_i^c$  is the spatial context obtained by averaging expression levels among all neighbors in the same community as cell  $i$ .

We further incorporate spatial context by allowing the posterior probabilities  $p(k | X_i, Y_i^c)$  to diffuse among adjacent cells. We define the modified posterior probability  $\hat{p}(k | X_i, Y_i^c)$  through a recursive relation analogous to the Katz centrality [156], initialized by  $p(k | X_i, Y_i^c)$ :

$$\hat{p}(k | X_i, Y_i^c) = \alpha \sum_j w_{ij} \hat{p}(k | X_j, Y_j^c) + \beta \quad (2.23)$$

$$\beta = (1 - \alpha)p(k | X_i, Y_i^c) \quad (2.24)$$

where  $\alpha$  is the attenuation factor and  $w_{ij}$  are the edge weights. Expressed in matrix form, the solution for  $\hat{p}(k | X, Y^c)$  is given by:

$$\hat{p}(k | X, Y^c) = (I - \alpha W)^{-1} (1 - \alpha)p(k | X, Y^c) \quad (2.25)$$

where  $I$  denotes the identity matrix and  $W$  is the matrix of edge weights  $w_{ij}$ . We then assign a label to each measurement  $X_i$  by applying  $f$  to the value of  $k$  that maximizes  $\hat{p}(k | X_i, Y_i^c)$ :

$$f(\operatorname{argmax}_k \hat{p}(k | X_i, Y_i^c)) \quad (2.26)$$

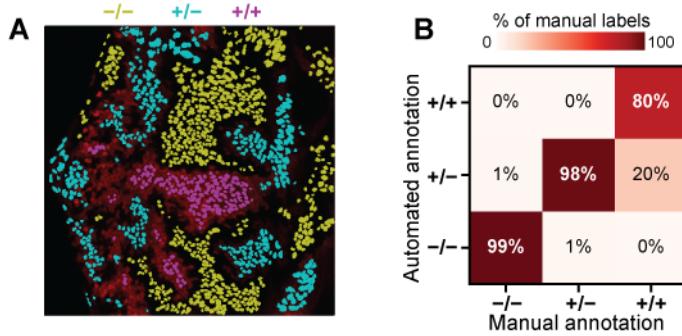
Finally, we assess the total posterior probability of each assigned label,  $\hat{P}(m_i)$ :

$$\hat{P}(m_i) = \sum_{\{k | f(k) = m_i\}} \hat{p}(k | X_i, Y_i^c) \quad (2.27)$$

This measure reflects the overall confidence that  $m_i$  is the appropriate label. Labels whose confidence falls below 80% are replaced by their counterparts estimated using the marginal classifier. This substitution helps preserve classification accuracy in situations where spatial context is not informative, and is particularly useful when the annotated clones are relatively small.

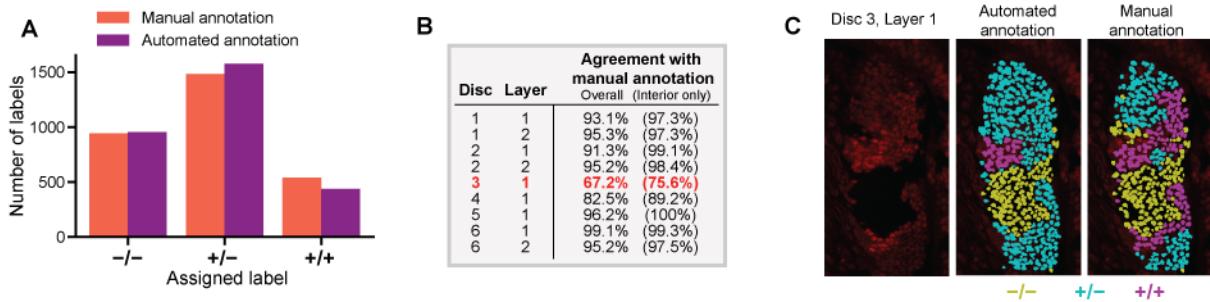
## 2.6. Manual assessment of annotation performance

We sought to validate the performance of the annotation algorithm by assessing its ability to accurately label a dataset in which the true labels are known. While it is common practice to use human-labeled data as the gold standard, manually assigned labels do not represent a reliable and reproducible ground truth. Furthermore, we contend that validation with manually-labeled data entrains implicit human biases in the selection of performant algorithms. These biases are particularly pronounced in biological image data where intrinsic variation, measurement noise, and transient processes can make cell-type annotation a highly subjective, and thus irreproducible, task.



**Figure 2.8. Automated unsupervised annotation of clones in the larval eye.** (A) Labels assigned by automated annotation. Yellow, cyan, and magenta denote the label assigned to each contour. Labels are overlaid on the RFP channel of the image shown in 2.3B. Cells on the periphery of each clone are excluded. (B) Comparison of automated annotation with manually-assigned labels. Confusion matrix includes data aggregated across nine images taken from six separate eye discs. Cells on the periphery of each clone are included. Columns sum to one.

Nevertheless, we begrudgingly labeled nuclei in each eye imaginal disc as mutant, heterozygous, or homozygous for the clonal marker, then automatically labeled the same cells (Fig. 2.8A). The two sets of labels showed strong overall agreement (Figs. 2.8B and 2.9A). Excluding cells on the border of each clone revealed greater than 97% agreement in seven of the nine annotated images (Fig. 2.9B). Upon secondary inspection of the sole instance of substantial disagreement (Fig. 2.9C), we are unable to confidently discern which set of labels are more accurate.



**Figure 2.9. Comparison of automated annotation with manually assigned labels.** (A) Distribution of labels among each possible value. (B) Agreement between automated and manual annotation. Values in parentheses exclude cells on the periphery of each clone. The sole instance of low agreement is marked in red. (C) Visual comparison of the instance in which automated and manual annotation differ. Image shows clonal marker fluorescence, overlaid colors denote the assigned label.

Synthetic benchmarking provides a powerful alternative to validation against manually labeled data. The idea is simple; measure how accurately an algorithm is able to label synthetic data for which the labels are known. The synthetic data generation procedure may be modeled after the process underlying formation of the real data, providing a means to assess the performance of an algorithm across the range of conditions that it is likely to encounter. The strategy therefore provides a means to survey the breadth of biologically plausible conditions under which the algorithm provides adequate performance. Synthetic benchmarking also facilitates unbiased comparison of competing algorithms, resulting in a reliable standard that may be called upon at any time. We therefore developed a platform to generate synthetic microscopy data that could be used to benchmark the performance of our annotation algorithm.

## 2.7. Generation of synthetic microscopy data

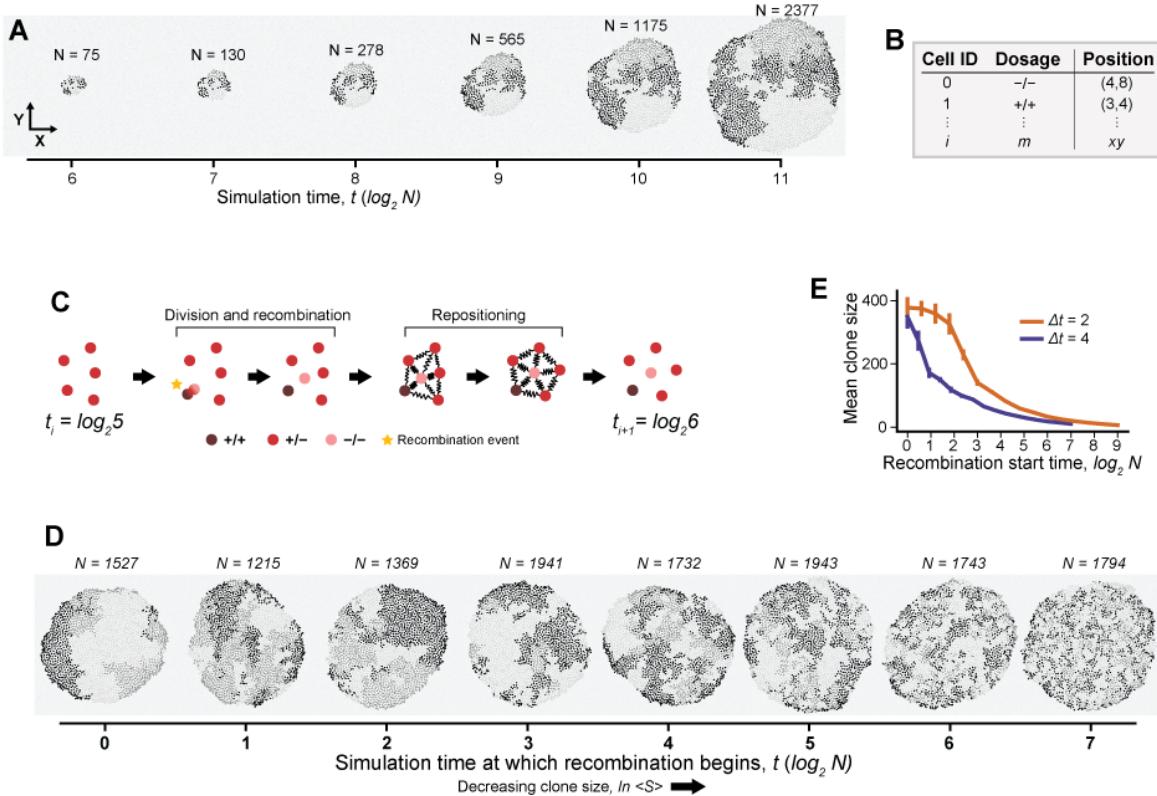
Each synthetic dataset depicts a culture of cells distributed roughly uniformly in space. To generate each culture, we simulated the two dimensional growth of a population seeded with a single cell. Growth proceeds through sequential division of cells (Fig. 2.10A). Not all cells divide at each time-step because cell division is a stochastic process. Instead, each cell divides stochastically with a rate controlled by a global growth rate parameter.

Cells in this culture carry a gene encoding a clonal marker (Fig. 2.10B). During growth, the gene is subject to mitotic recombination (Fig. 2.10C). Each time a cell divides, its genes are duplicated and equally partitioned between the two daughter cells. However, in some instances a heterozygous parent may instead partition its two duplicate genes unequally, with one daughter receiving both and the other receiving none. These mitotic recombination events occur stochastically with a frequency defined by a global recombination rate parameter.

After each round of cell division, all cells are repositioned in order to preserve approximately uniform spatial density (Fig. 2.10C). Repositioning is achieved by equilibrating a network of springs connecting each cell with its neighbors. This undirected network is constructed through Delaunay triangulation of all cells spatial positions. Edges on the periphery of the culture are systematically excluded by establishing a maximum polar angle between neighbors. This filtration removes spurious edges between distant pairs of cells. Edges connecting pairs of cells with the same clonal marker dosage are assigned a 10% higher spring constant than edges that connect dissimilar cells. This modest bias ensures that cells tend to remain proximal to their clonal lineages. Cell positions are then updated using a force-directed graph drawing algorithm [157]. Alternating cell division and repositioning steps are then repeated until a predefined population size is reached.

The timing and duration of recombination events affects the number and size of the resultant clones. In real experiments, recombination events are restricted to a particular stage of the developmental program through localized exogenous expression of the recombination machinery. We incorporated this feature into our cell growth simulations via two adjustable parameters. The first determines the minimum population size at which recombination may begin, while the second determines the number of generations over which recombination may continue to occur. These two parameters provide a means to tune the average number and size of clonal subpopulations in the synthetic data (Fig. 2.10D). Early recombination events generally entail larger clones, while shorter recombination periods limit the extent of clone formation (Fig. 2.10E).

Each simulation yields a list of spatial coordinates and gene dosages for each nucleus (Fig. 2.10B). We generated a synthetic measurement for each nucleus by randomly sampling its fluorescence level in a dosage-depend manner. Specifically, the measured fluorescence



**Figure 2.10.** Simulated growth of a synthetic cell culture. (A) Partial simulation time course. Each marker depicts a cell. Greyscale intensity reflects clonal marker gene dosage. Simulation time reflects the approximate number of cell divisions since the initial seed. (B) Simulations yield gene dosages and spatial coordinates for each cell. (C) Single iteration of an example simulation. Circles represent individual cells, red shading denotes clonal marker dosage. Cycles of cell division, recombination, and repositioning are repeated until the simulation reaches a specified end time ( $t > 11$  in panel A). (D) Cultures simulated with varying recombination start times. All cultures were subject to four generations of recombination ( $\delta t = 4$ ). Recombination start time increases from left to right. Later recombination events generally yield smaller clones. (E) Mean clone size (cells per clone) as a function of the recombination start time. Colors denote recombination period duration. Error bars reflect standard error of the mean across 50 replicates. Clone size generally decreases as recombination is limited to later times.

levels  $\{x_1, x_2, \dots, x_{i=N}\}$  were sampled from a lognormal distribution conditioned upon the corresponding gene dosage (Fig. 2.11A-C):

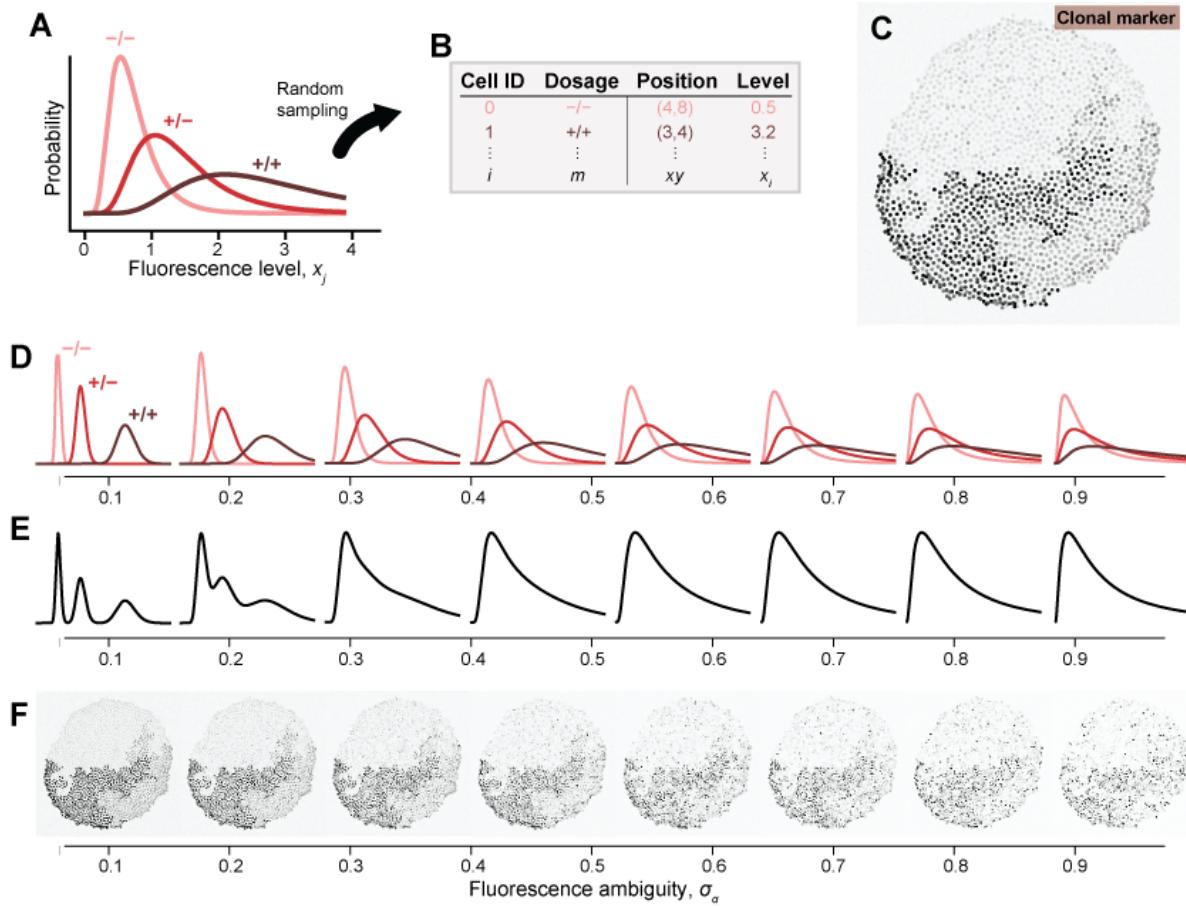
$$\ln x \sim \mathcal{N}_n(\theta_n) \quad (2.28)$$

where the subscript  $n$  denotes the gene copy number and  $\theta_n = (\mu_n, \sigma_n^2)$  are the mean and variance of the corresponding distribution. We define  $\mu_n$  such that the mean fluorescence

level doubles for each additional copy of the gene:

$$\mu_n = \ln(2^{n-1}) \quad (2.29)$$

We refer to  $\sigma_\alpha$  as the *fluorescence ambiguity* because it modulates the similarity of fluorescence levels across gene dosages. Increasing  $\sigma_\alpha$  increases the overlap among  $\mathcal{N}_0$ ,  $\mathcal{N}_1$ , and  $\mathcal{N}_2$  (Fig. 2.11D,E), and consequently increases the difficulty of the annotation task (Fig. 2.11F).



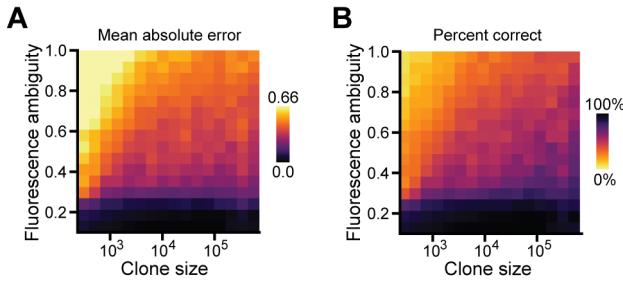
**Figure 2.11. Tunable generation of synthetic microscopy data.** (A) Fluorescence levels are sampled from lognormal distributions conditioned upon gene dosage. (B) Synthetic data include a measured fluorescence level for each reporter in each cell. Text color reflects the generative distribution in A. (C) Synthetic image of clonal marker fluorescence when  $\sigma_\alpha = 0.25$ . Each nucleus is shaded in accordance with its sampled fluorescence intensity. (D-F) Left to right, increasing the fluorescence ambiguity parameter broadens the overlap in fluorescence levels across gene dosages. (D) Distributions used to generate clonal marker fluorescence levels. Red shading denotes gene dosage. (E) Evenly weighted sum of the generative distributions. (F) Example images of clonal marker fluorescence.

## 2.8. Synthetic benchmarking of annotation performance

We generated a large synthetic dataset spanning a broad range of sixteen different clone sizes and fluorescence ambiguities (Figs. 2.10D and 2.11F, only half are shown for each). We performed 50 replicate simulations for each condition. All simulations were terminated when the total population exceeded 2048 cells. We assigned each cell a 20% probability of division upon each iteration, and each cell division event was accompanied by a 20% chance of mitotic recombination. Parent cells containing zero or two copies of the recombined genes were ineligible for recombination, effectively sealing the genetic fates of their respective lineages.

We then annotated each set of measurements, and compared the assigned labels with their true values. The mixture model given by Equation 2.17 was independently trained and applied to each replicate. Training a single model on all replicates yields modestly stronger performance on average (not shown), but also yields more variable variable results across the parameter space because all labels are dependent upon the outcome of a single expectation maximization routine. We used the mean absolute error (MAE) as a comparison metric because it provides a stable measure of accuracy for multiclass classification problems in which the labels are intrinsically ordered [158]. In other words, it penalizes egregious misclassifications more severely than mild ones.

We observed very strong annotation performance for all cases in which  $\sigma_\alpha \leq 0.3$  (Fig. 2.12A). Unsurprisingly, performance suffers as the difficulty of the classification problem is increased. As cells on the periphery of each clone were not excluded from these analyses, the observed metrics provide a lower bound on the performance that may be anticipated in practice. The same trends are evident when performance is graded strictly on accuracy (Fig. 2.12B).

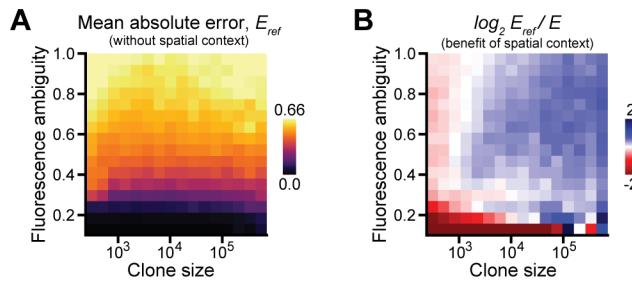


**Figure 2.12. Synthetic benchmarking of automated annotation performance.** (A) MAE and (B) accuracy of the assigned labels. Each pixel reflects the average across 50 replicates. Clone size reflects the mean number of cells per clone. Accuracy denotes the fraction of cells that were correctly labeled. Performance improves with increasing clone size and worsens with increasing fluorescence ambiguity.

Performance improved with increasing clone size. We suspected this was caused by larger clones offering additional spatial context to inform the identify of each cell. We verified our assertion by re-evaluating performance relative to a variant of our annotation algorithm that neglects spatial context (Fig. 2.7G). As expected, the variant's performance exhibited no dependence on clone size (Fig. 2.13A). Comparing the two strategies confirmed that spatial context confers the most benefit when clones are large (Fig. 2.13B). Inclusion of spatial context also becomes increasingly advantageous as the fluorescence ambiguity is increased, even for smaller clones. Thus, spatial context adds progressively more value as the classification task becomes more difficult.

This observation may be rationalized from a statistical perspective. Each cell is classified by maximizing the probability that the assigned label is correct. We compute these probabilities using the estimated expression level of each cell. Neglecting spatial context, this estimate is limited to a single sample and is therefore highly sensitive to both measurement and biological noise. Incorporating spatial context expands the sample size and thereby reduces the standard error of the estimated fluorescence level. The strategy is thus generally well suited to scenarios in which fluorescence intensities correlate across large clones, and closely parallels computer vision methods that exploit spatial contiguity to segment image

features with ill-defined borders [159]. Because increased measurement precision comes at the expense of spatial resolution, we expect strong performance when measurements are aggregated across relatively large clones, but failure to detect small, heterogeneous clones. These expectations are consistent with the observed results. They are also conveniently aligned with the anticipated properties of real data, as experiments typically attempt to mitigate edge effects by driving early recombination events to generate large clones.



**Figure 2.13. Spatial context is most informative for large clones with ambiguous fluorescence.** (A) MAE of labels assigned using a marginal classifier that neglects spatial context. Performance worsens with increasing fluorescence ambiguity but does not depend upon clone size. (B) Annotation performance relative to the marginal classifier. Color scale reflects the  $\log_2$  fold-change in MAE when spatial context is neglected. Blue indicates that spatial context improves performance.

## 2.9. Potential applications and continued development

We used synthetic data to survey the performance of our annotation strategy across a much broader range of conditions than would have otherwise been possible with manually labeled data. This included conditions well beyond those of practical use. In particular, experiments designed to compare gene expression levels across clones would likely seek to avoid generating small clones with ambiguous clonal marker expression. Synthetic data provided a means to survey these edge cases and establish a lower bound on annotation performance. The strong performance observed across the remaining conditions bolsters our confidence that our annotation strategy is well suited to the images it is likely to encounter.

In each of our examples, clones were distinguished by ternary segregation of clonal marker fluorescence levels. Modern mosaic analysis techniques continue to deploy ternary labeling [160, 161], but also frequently opt for binary labeling of mutant versus non-mutant clones [162–164] and dichromic labeling of twin-spots [165, 166]. Our annotation scheme readily adapts to each of these scenarios provided that the number of anticipated labels is adjusted accordingly. In the case of dichromic labeling, binary classification would be performed separately for each color channel before merging the assigned labels. Extending the same logic to combinatorial pairs of colors suggests that our framework may also be compatible with multicolor labeling schemes used to simultaneously trace many clonal lineages over time [167–169]. Our framework is thus well suited to many different mosaic analysis platforms deployed in the larval eye.

In principle, the framework described here should also be applicable to a wide variety of other tissues [170, 171] and model organisms [172–174] in which mosaics are studied. In practice, application to alternate contexts would require modifying some stages of the analysis. Most notably, image segmentation is strongly context dependent and any attempts to develop a universally successful strategy are likely to prove futile [175]. For this reason, we implemented a modular design in which each stage of analysis may be applied separately. For example, a user could perform their own segmentation before using our bleedthrough correction and clone annotation tools. By offering modular functionalities we hope to extend the utility of our software to the wider community of developmental biologists. Furthermore, the open-source nature of our framework supports continued development of more advanced features as various demands arise. Our synthetic benchmarking platform could then be used to objectively confirm the benefit conferred by any future developments.

## CHAPTER 3

### Ratiometric control of a transit to differentiation

A manuscript resembling this chapter was coauthored with Jean-Francois Boisclair Lachance, Nicolás Peláez, Rachael Bakker, Heliodoro Navarro, Luís Amaral, Neda Bagheri, Ilaria Rebay, and Richard Carthew. A preprint is publicly available at <https://doi.org/10.1101/430744>. All of the experiments were conceived, designed, and executed by my colleagues. My contributions include all of the computational analysis and modeling and figures. I also wrote most of the the text under the guidance of Professor Richard Carthew.

#### 3.1. Background on the coordination of cell fate decisions

Organismal development proceeds through a sequence of transitions that yield increasingly specific cellular states. Developmental programs employ a variety of strategies to coordinate transitions, ultimately ensuring that cells adopt the correct state in space and time. Elucidating these decision mechanisms is crucial to understanding the processes that guide development, as well as the diseases that arise when they fail.

How cells dynamically integrate the activities of one or more transcription factors to reliably execute state transitions is a long-standing question. Transcription factors initiate and enforce transitions by differentially regulating the expression of specific target genes [176–178]. Coordination among transcription factors can thus broaden the spectrum of possible cell responses to developmental cues [179, 180]. Positive and negative feedback can create bi-stable patterns of mutually exclusive transcription factor expression or activity that

render state transitions irreversible [79, 181–183]. Cells can also regulate state transitions by differentially partitioning transcription factors between daughter cells [51]. In all of these models, mRNA and protein expression are believed to depend upon the absolute concentrations of the transcription factors involved [39, 184–186].

Studies of intercellular signaling suggest cells are capable of sensing relative levels of signaling molecules. For example, the TGF- $\beta$  pathway elicits a cell response following changes in input signaling relative to the preceding background [50]. Fold-change, rather than absolute levels of  $\beta$ -catenin, dictate Wnt signal transduction in eukaryotic cells [187]. Likewise, aggregation of the social amoeba *Dyctiostellium* depends on fold-change detection of extracellular cAMP concentrations [188]. Relative measurements can also involve multiple molecular components. In the BMP pathway, cells interpret multi-ligand inputs based on the relative levels of each ligand pair, with additive, differential, and ratiometric response types arising directly from the relative competition between different ligand-receptor pairs [189]. In yeast, pheromone response is insensitive to the absolute abundance of the Gprotein-coupled receptor Ste2. Instead, cells respond to the fractional occupancy of the signal receptor by forming a ratiometric sensor between Ste2 and its regulatory inhibitor Sst2 [190]. Topological features of molecular interaction networks, such as feed-forward loops, can also sense relative changes in molecule abundance [76, 191]. Combining such circuits with precise coordination of signaling inputs would yield a molecular decision mechanism that is robust to fluctuations in the abundance of participating molecules. Notably, while relevant transcription factors are precisely controlled during development [192, 193], it remains unknown whether cells sense the relative concentrations or activities of different transcription factors when executing cell state transitions.

Two ETS-domain transcription factors, the activator Pointed (Pnt) and the repressor Yan (also known as Anterior open, Aop), are essential regulators of cell fate transitions at numerous stages of *Drosophila* development [194–198]. Consistent with their opposing regulatory effects, genetic studies have shown that Pnt and Yan can act antagonistically in guiding numerous cell fate transitions [194, 195, 199, 200]. The *pnt* locus encodes two protein isoforms, PntP1 and PntP2, that differ in their N-terminal transactivation domains but share the same DNA-binding domain (Fig. 3.1A) [201, 202]. Specifically, PntP1 is constitutively active whereas PntP2 requires phosphorylation via the RTK signaling pathway to become a potent activator [199, 200]. Because both Pnt isoforms and Yan bind to a common ETS-binding DNA sequence motif GGA(A/T) [203], competition for occupancy and regulation of common target genes must be precisely orchestrated [195, 197, 198, 200, 204–207].

Pnt and Yan display mutually exclusive expression patterns in several developing tissues [85]. The embryonic ventral ectoderm provides a classic example in which cells unambiguously reside in one of two stable states [79, 194]. These and similar observations inspired a bistable switch model of cell fate specification in which the multipotent state is defined by high absolute Yan levels and the differentiated state is defined by high absolute Pnt levels [83]. The model posits that RTK signaling triggers a transition from target gene repression to activation by degrading Yan and activating PntP2 [82, 199]. A positive feedback loop in which transient phosphorylation of PntP2 activates expression of *pntP1* is thought to stabilize the transition by enabling prolonged signaling-independent stimulation of target genes [84]. Sustained PntP1 expression in cells devoid of Yan thereby recapitulates a complementary expression pattern under the control of RTK signaling [85]. However, Pnt and Yan are also co-expressed during cell fate commitment in several developmental contexts [85]. For example, the two proteins are co-expressed in posterior follicle cells of the early egg chamber and throughout the embryonic mesoderm [85], where they are required for

specification of cell fates [195, 196]. Co-expression also occurs in the larval eye [85], despite the repeated use of RTK signaling to designate cell fates [208]. Eye development thus prompts exploration of how state transitions are resolved from concurrent Pnt and Yan expression in response to signaling cues.

Eye development is divided into two distinct phases, growth and differentiation. In the first phase, multipotent progenitor cells in the eye-field asynchronously proliferate from the earliest larval stage until the third instar stage of larval life [209]. The differentiation phase of eye development begins in the early third instar larva, when cells situated at the posterior margin of the eye disc start to differentiate into photoreceptor (R) cells, followed by progressively more anterior cells (Fig. 3.1B). This wave of differentiation is initiated and coordinated by a morphogenetic furrow (MF), which traverses the eye disc from posterior to anterior for the remainder of the third instar stage up to the early pupal stage [29]. Progenitor cells located immediately anterior to the MF arrest in G1 of the cell cycle and express a transcription factor called Atonal. Refinement of Atonal expression within this field establishes the differentiation program by specifying individual R8-type photoreceptors in a periodic pattern immediately posterior and parallel to the MF [30, 210]. Each R8 cell then locally secretes an RTK ligand that induces R8's multipotent neighbors to differentiate into other photoreceptor types (Fig. 3.1C) [29, 208]. Transitions sequentially occur approximately every two hours to form R2/R5, R3/R4, R1/R6, and R7 photoreceptors [209]. Many cells remain as multipotent progenitors even after all R cell fates have been adopted. These cells will adopt other fates at later stages of eye development, with any surplus eliminated by apoptosis [211].

R cell fate specification violates three central tenets of the bistable switch model. First, Pnt and Yan do not exhibit a mutually exclusive expression pattern during eye development.

The two proteins are extensively co-expressed in both progenitor and differentiating cells within and posterior to the MF [85]. Second, transitioning cells do not originate in a stable high Yan state. We recently quantified Yan protein dynamics during larval and early pupal eye development [43]. In progenitor cells, Yan displays pulsatile dynamics in which protein levels rapidly increase as the MF passes and then gradually decay back to low initial values. Third, transitioning cells do not adopt a stable high Pnt state. Visual inspection of eye discs carrying a fluorescent reporter for Pnt expression suggest that Pnt levels decay on a timescale comparable to Yan in transitioning cells [85]. Despite their similar expression patterns, the two proteins still exhibit antagonistic effects on R cell fate determination. Pnt stimulates progenitor cells to transit to an R cell fate while Yan inhibits these transitions [82, 200].

Here, we explore this apparent paradox to understand how two coexpressed transcription factors with comparable DNA binding specificity but opposing transcriptional functions elicit precisely controlled cell state transitions in the developing eye. We find that cell states are regulated by the relative abundance of these two proteins, rather than by their absolute concentrations. Progenitor cells dynamically maintain an approximately constant ratio of Pnt-to-Yan protein despite their absolute concentrations varying over time. Cells that transition to R cell fates rapidly increase their Pnt-to-Yan ratio, which remains elevated over time. We show that a ratio control strategy buffers this ratio against variable abundance of either protein. We also find that the signaling inputs of the Yan-Pnt network regulate the dynamics of the transcription factor ratio. Although Notch and Ras signals can both promote and inhibit differentiation in the developing eye [208, 212], we find that Notch signaling predominantly decreases the Pnt-to-Yan ratio in progenitor cells while Ras signaling mainly increases the ratio. Notch and Ras signals thus inhibit and promote cell state transitions in the eye by dynamically tuning the ratio of the two transcription factors. We conclude that

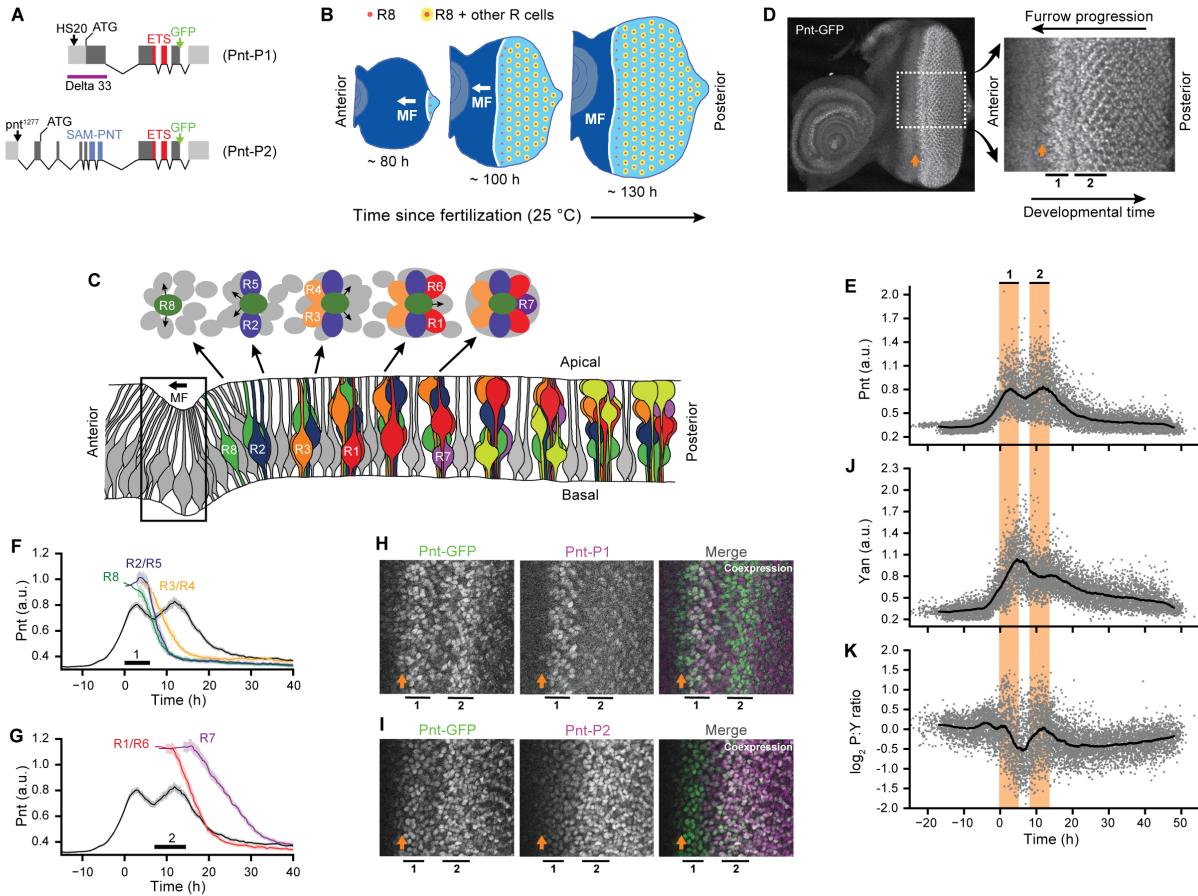
progenitor cells interpret an increase in the abundance of Pnt relative to Yan as a cue to differentiate.

### 3.2. PntGFP expression dynamics in the developing eye

Although Pnt expression has been qualitatively studied in the eye [85], its dynamics have not been quantified. Therefore, we took advantage of a fully functional genomic transgene in which both *pntP1* and *pntP2* are C-terminally tagged with GFP. As previously reported [85], *pnt-gfp* rescues *pnt* null mutants to viable, fertile adults with wild type external morphology (Fig. A.1A,B). Qualitative examination of Pnt-GFP expression in 100h eye-antennal imaginal discs revealed a region of very low expression in cells anterior to the MF, followed by strong expression in two parallel stripes of cells immediately posterior to the MF (Fig. 3.1D, regions 1 and 2).

### 3.3. The Pnt-to-Yan ratio varies between cells in different states

We used Histone-RFP fluorescence from a His2Av-mRFP transgene to label all eye cell nuclei for automated segmentation following direct fluorescence microscopy of fixed specimens [43, 213]. Average Pnt-GFP fluorescence levels and exact 3D positions were then calculated for all nuclei in each developing eye disc. Pnt-GFP fluorescence levels were normalized to Histone-RFP, which provided some control over measurement noise and nuclear constriction that occurs at the MF [43, 213]. We mapped each cell's position along the anterior-posterior coordinate of the eye disc to a point in developmental time. This linear approximation is sufficiently accurate because the MF moves across the eye field with approximately constant velocity, forming one column of R8 cells every two hours [214, 215]. The distance between a cell and the MF is therefore proportional to the time elapsed since the MF passed. We



**Figure 3.1.** (Continued on the next page.)

also manually assigned a state value to each cell. Cell state classification is possible because nuclei can be unambiguously identified without cell-specific markers by their morphology, apical-basal position, and relative distance to the furrow [27, 28, 43, 209, 213, 216]. Combined, these data allowed us to infer a macroscopic view of cell state transition dynamics from the spatial arrangement of cells relative to each other and the MF. Although our approach cannot measure the developmental progression of an individual cell, it provides a dynamic view of thousands of cells across a developing eye. From this information, average cell behaviors can be reconstructed and modeled.

**Figure 3.1. Dynamics of Pnt and Yan expression during eye development.** (A) Diagram of the *pnt* locus encoding the Pnt-P1 and Pnt-P2 protein isoforms. The isoforms share an ETS DNA-binding domain (red) but are distinguished by the presence of a SAM domain (blue) within Pnt-P2. Green arrows at the C-termini depict insertion sites of GFP in the *pnt-gfp* allele, black arrows at the N-termini depict insertion sites of the *pnt<sup>HS20</sup>* and *pnt<sup>1277</sup>* enhancer-trap alleles. Adapted from Shwartz et al. (2013). (B) Differentiation is initiated in the developing eye by the MF, which moves across the eye epithelium from posterior to anterior (white arrow). On the furrow's posterior side, G1-arrested progenitor cells differentiate (light blue). Formation of regularly spaced R8 photoreceptors (red dots) precedes recruitment of additional R cell types (yellow dots). On the anterior side, progenitor cells are still proliferating (dark blue). Axis refers to time elapsed since fertilization. Adapted from Peláez et al. (2015). (C) Top, cartoon of an apical view of the sequential differentiation of eight R cell types from multipotent progenitor cells (grey) and their relative positions within a single ommatidium. Arrows denote signals transmitted from the R8 to nearby cells. Bottom, a cross-section view through an eye disc, showing the epithelial constriction that marks the MF (boxed region) and then the relative nuclear positions of progenitors (grey) and specified R cells (various colors). Together, the stereotyped features depicted in this schematic enable unambiguous identification of each cell type as ommatidial assembly proceeds. Adapted from Peláez et al. (2015). (D) Maximum intensity projection of Pnt-GFP fluorescence in an imaginal disc fixed ~ 100 h after fertilization. Right panel corresponds to the region enclosed by dashed white lines in the left panel. Morphogenetic furrow (orange arrow) precedes the first and second stripes of strong Pnt-GFP expression (black lines, labeled 1 and 2). (E) Pnt-GFP expression in progenitor cells. Grey points are individual cells, solid line is the smoothed moving average. Orange shading indicates the first and second stripes of Pnt-GFP expression. (F, G) R cell recruitment from the (F) first and (G) second pulses of Pnt-GFP expression. Solid lines and shaded regions denote moving averages and their 95% confidence intervals. (H, I) Confocal images of (H) Pnt-P1 and (I) Pnt-P2 enhancer-trap co-expression with Pnt-GFP. Orientation is consistent with panel D. Morphogenetic furrow (orange arrow) precedes first and second stripes of Pnt-GFP induction (black lines, labeled 1 and 2). In merged images, Pnt-GFP is green and enhancer-trap expression is magenta. (J, K) Measured (J) Yan expression and (K) log<sub>2</sub>-transformed Pnt-GFP to Yan ratios in progenitor cells. Grey points are individual cells, solid line is a smoothed moving average.

Progenitor cells anterior to the MF expressed a basal level of Pnt-GFP, but expression dramatically increased in cells immediately anterior to the MF (Fig. 3.1E). This was followed by two successive pulses of Pnt-GFP expression, marked by peaks where protein expression reached maximal amplitude. The pulses matched the visual stripes seen in regions 1 and 2 (Fig. 3.1D). Thereafter, Pnt-GFP decayed to a low basal level. The two pulses of Pnt-GFP in progenitor cells coincided with the two periods of transition to R cell states (Figs. 3.1F,G and A.1C). Transitions to R8, R2/R5, and R3/R4 states occurred during the first pulse, while transitions to R1/R6, and R7 states occurred during the second pulse.

Based on prior description of the distinct temporal expression patterns of isoform-specific *pnt* transcriptional reporters [84], we suspected that each pulse of Pnt-GFP corresponded to the induction of either PntP1 or PntP2. Using flies that carried the *pnt-gfp* transgene and either a *pntP1*- or *pntP2*-specific reporter, we found that region 1 overlapped with the domain of strongest *pntP1* reporter expression, and region 2 corresponded to the domain of strongest *pntP2* reporter expression (Fig. 3.1H,I). Low levels of *pntP2* reporter expression were detected in region 1 and low levels of *pntP1* reporter were detected in region 2. Therefore, Pnt expression appears as a PntP1-PntP2 pulse sequence. The two groups of differentiating R cells predominantly expressed PntP1 or PntP2 respectively (data not shown, see [213]), suggesting that specific Pnt isoforms are used to specify distinct cell fates.

All cell state transitions coincided with a rapid increase in Pnt-GFP (Fig. 3.1F,G). The earliest identified R cells had, on average, 25-50% higher levels of Pnt-GFP than progenitor cells at comparable times. Pnt-GFP then rapidly decayed in all differentiating R cells, with all but the R7 cell type exhibiting faster decay kinetics than progenitor cells. Thus, transitioning R cells did not adopt stable high Pnt levels as predicted by a bi-stable model of R cell fate specification. Rather, the measured Pnt-GFP expression dynamics were similar to those previously reported for Yan [43]. Average levels of both proteins increased as the MF passed, then decayed during cell state transitions. These similar population-wide dynamics led us to ask how cell states are resolved from the co-expression of two transcription factors with opposing transcriptional functions at the single cell level.

We explored this question by simultaneously measuring Pnt-GFP and Yan in each nucleus using an anti-Yan monoclonal antibody. Previously, we had shown that Yan dynamics measured with the antibody were almost identical to those measured by a YFP tagged version of Yan [43], validating our approach. Pnt-GFP and Yan were induced at the same time in

progenitor cells (Figs. 3.1E,J and A.1D). Yan levels reached a maximum amplitude between the two pulses of Pnt-GFP. Yan then decayed back to a basal steady-state level, interrupted by a transient plateau during the second pulse of Pnt-GFP. Despite their alternating maxima, the overall induction and decay of Pnt-GFP and Yan were concurrent in progenitor cells.

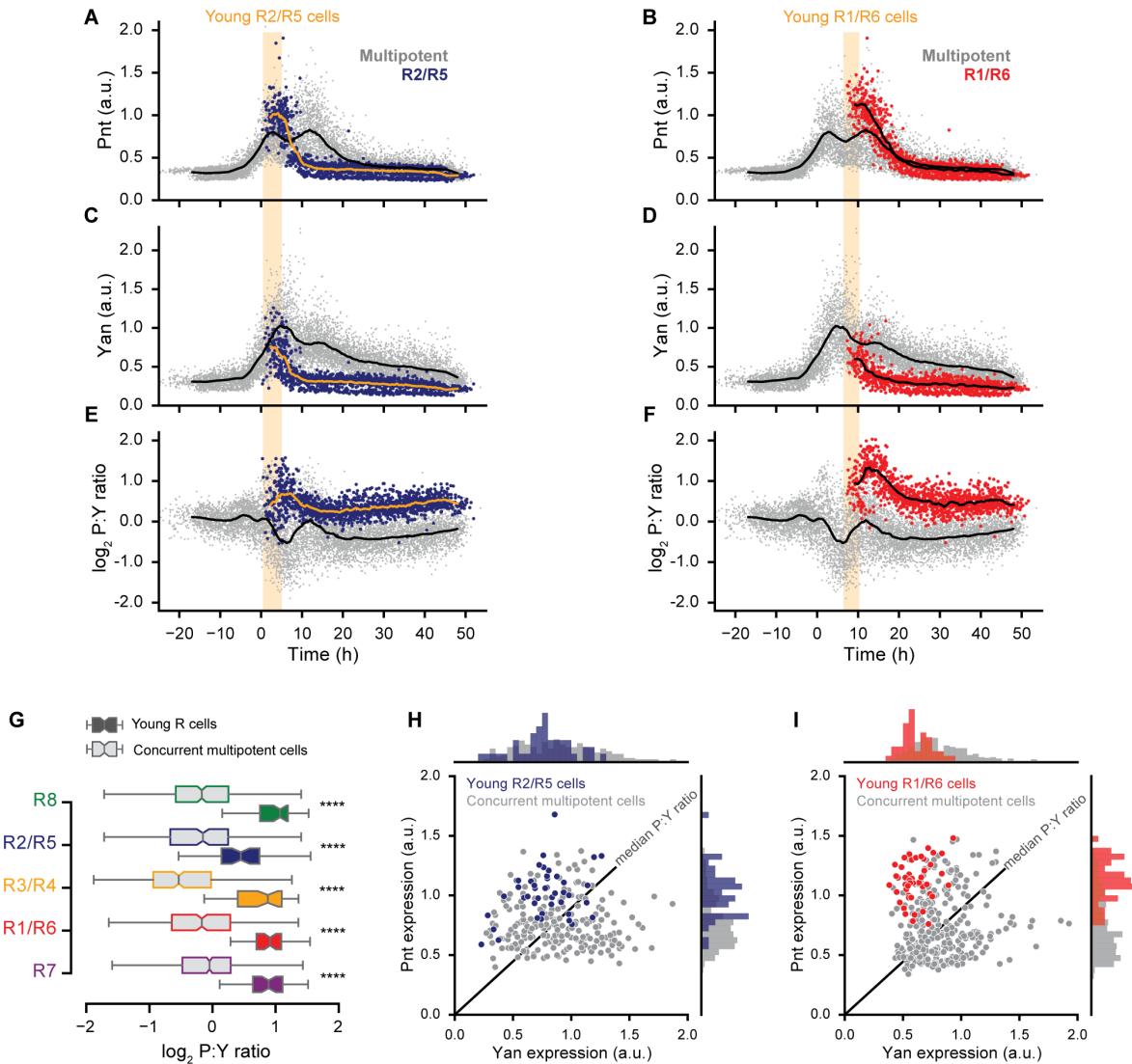
The similar dynamics prompted us to consider whether relative levels of Pnt and Yan dictate cell state transitions in the eye. To explore this possibility, we measured the ratio of Pnt-to-Yan in each progenitor cell. Strikingly, the average Pnt-to-Yan ratio remained dynamically stable about a constant value over time (Fig. 3.1K). However, from 0 to 15 h, there was considerable cell-to-cell heterogeneity in the ratio. Some cells had above-average ratios when they were expressing peak levels of Pnt-GFP, and many cells had below-average ratios when they were between the Pnt-GFP pulses. After the second pulse of Pnt-GFP expression, cells acquired a slight bias towards Yan. These are the progenitor cells that remain multipotent and are used to differentiate into other cell types later in development [211]. As the two positive spikes in the ratio coincided with the two periods of cell state transition, we reasoned that dynamic changes in the ratio might control the state of cells and regulate their transit to differentiation.

As a first test of this idea, we quantified the levels of Pnt-GFP and Yan in cells that had undergone R cell state transitions. We focused on R2/R5 and R1/R6 cells, since they are representative of transitions of cells derived from groups 1 and 2, respectively. As previously noted, Pnt-GFP levels were elevated in both sets of “young” R cells as soon as we could confidently identify them (Fig. 3.2A,B). In contrast, Yan levels were lower in these young R cells than in progenitor cells of comparable age (Fig. 3.2C,D). This meant that the average ratio of Pnt-to-Yan was elevated 1.5- to 2-fold in young R cells (Fig. 3.2E,F). The ratio elevation was more modest in young R2/R5 cells than in young R1/R6 cells, but all increases

were significant (KS test,  $p < 0.001$ ). The elevated ratio persisted for all times thereafter as differentiation proceeded. Analogous ratio trends were evident with the other R cell types as well (Figs. 3.2G and A.2), suggesting that different state transitions share a common requirement for sustained change in the Pnt-to-Yan ratio relative to progenitor cells (Fig. A.2C-G).

We next asked whether elevated Pnt-to-Yan ratios precede the onset of R cell state transitions. If R cells are recruited from a subpopulation of progenitors with relatively high ratios, some progenitors should exhibit ratios comparable to those of early R cells. We compared the distribution of Pnt-to-Yan ratios in young R cells versus concurrent progenitor cells (Fig. 3.2H,I). The extensive overlap between the two populations implies that some cells we morphologically classified as progenitors were actually transitioning to R cell fates. Additionally, ratios may have increased among a subset of progenitors that did not ultimately transition to an R cell state. Many progenitors also adopted low Pnt-to-Yan ratios during this time period that did not overlap with the early R cell population. We reasoned that these cells were not viable candidates for recruitment, suggesting that the extent of variation in the ratio among progenitors constrains their competence for differentiation.

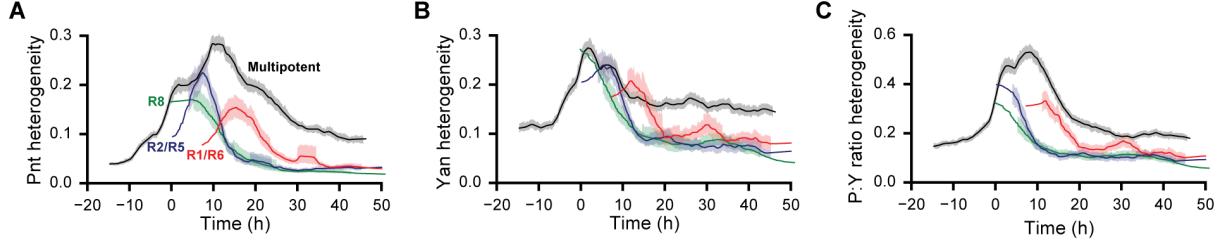
We then asked whether variation in the Pnt-to-Yan ratio strictly coincides with R cell fate transitions. We anticipated that variability should arise within the pool of progenitors from which R cells are recruited, then subside as fates are resolved. We previously reported methods to quantify the dynamic cell-to-cell heterogeneity of Yan concentration [43]. We applied similar analysis to simultaneously quantify the heterogeneity of Pnt, Yan, and the Pnt-to-Yan ratio among both progenitors and early R cells (Fig. 3.3). A broad increase in variation of the ratio among progenitors coincided with the time periods in which state transitions occurred. Ratio variation was predominantly attributed to Yan and Pnt variability



**Figure 3.2. Pnt-to-Yan ratios differ between cellular states.** (A-F) Measured (A, B) Pnt-GFP expression, (C, D) Yan expression, and (E, F) Pnt-GFP to Yan ratio dynamics in R2/R5 (blue) and R1/R6 (red) cells. Progenitors are grey. Solid lines are smoothed moving averages across 250 and 50 samples for progenitor and R cells, respectively. Yellow shading indicates time spanned by young R cells. (G) Comparison of Pnt-to-Yan ratio levels between young R cells (color filled boxes) and their concurrent progenitors (grey filled boxes). Colors denote R cell types. For each R cell type, the ten earliest identifiable R cells in each disc were designated as young R cells. Progenitor cells that fall within the time window spanned by these young R cells were designated as concurrent progenitors. Asterisks denote significance (KS 2-sample test,  $p < 0.001$ ). (H, I) Joint distributions of Pnt-GFP and Yan protein levels for young (H) R2/R5 and (I) R1/R6 cells. Progenitor cells concurrent with the corresponding young R cells are shown in grey. Black line denotes the median Pnt-to-Yan ratio among the progenitor cells shown.

during the first and second groups of R cell state transitions, respectively. Heterogeneity among progenitors then decreased back to basal levels after all R cell fate transitions were

complete. Similar trends were evident among transitioning R cells, but with a more rapid approach toward a consensus ratio following fate specification.



**Figure 3.3. Dynamics of expression variability in progenitor and R8, R2/R5 and R1/R6 cells.** Heterogeneities of (A) Pnt expression, (B) Yan expression, and (C) the log<sub>2</sub>-transformed ratio are estimated by de-trending fluctuations about a moving average of 250 sequential cells. Lines are moving averages of 250 sequential fluctuations, shaded regions are bootstrapped 95% confidence intervals for the moving average. Colors denote cell type.

### 3.4. Cooperative DNA-binding sensitizes promoters to changes in Pnt-to-Yan ratio

How could cells reprogram transcription in response to a change in the Pnt-to-Yan ratio across a wide range of absolute protein concentrations? Since both transcription factors have overlapping sequence specificity for DNA binding [195, 197, 198, 203–205, 217], the underlying mechanism may be a natural consequence of competition for binding sites in target genes. To further explore this idea, we composed a simple equilibrium model in which two species, Yan ( $Y$ ) and Pnt ( $P$ ), competing for a finite pool of shared binding sites,  $S$ :



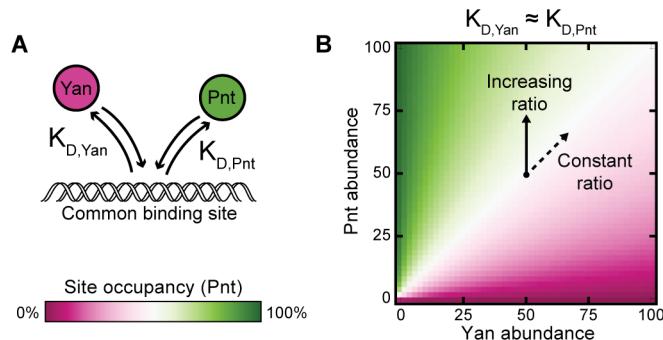
where  $K_{D,Yan}$  and  $K_{D,Pnt}$  are equilibrium association constants and  $SY$  and  $SP$  denote the bound species. Applying a mass balance to the total protein and binding site ( $S_0$ ) abundances:

$$Y_0 = Y + SY$$

$$P_0 = P + SP \quad (3.2)$$

$$S_0 = S + SY + SP$$

yields an analytically tractable system of nonlinear equations [218]. Given a pair of absolute protein abundances ( $Y_0, P_0$ ), the Pnt binding site occupancy is simply  $SP/S_0$ . The phase diagram offered by this simple model suggests that if the sites are saturated, equilibrium occupancy by either factor is more sensitive to the relative concentration of the two factors than to the absolute concentration of both species (Fig. 3.4A,B).



**Figure 3.4. A simple two-species competitive binding model** (A) Model schematic. (B) Theoretical Pnt site occupancy as a function of transcription factor abundance. Equivalent binding affinities are used for illustrative purposes. Simultaneous proportional increases in absolute abundance of both species have minimal impact on Pnt occupancy, while varying ratio confers maximal change.

However, the situation is more complex for Pnt and Yan. While there are several well-documented target genes that contain common binding sites for Pnt and Yan [195, 197, 198, 206], Yan binds these enhancers with higher affinity than Pnt [197]. Moreover, recent experiments suggest a scenario in which Pnt and Yan differentially interpret the structural syntax of *cis*-regulatory modules [206]. This complex phenomenon is a consequence of

cooperative recruitment between adjacent chromatin-bound Yan molecules. Yan monomers are able to polymerize via their sterile alpha motif (SAM) binding domains, enabling tightly-bound Yan monomers at strong ETS sites to stabilize the recruitment of additional Yan monomers to adjacent, weaker ETS sites or non-ETS sites [206, 219]. These cooperative effects could conceivably bias the competition between Pnt and Yan, which led us to consider a more complex model.

Hope, Rebay, and Reinitz recently introduced a modeling framework in order to probe the effects of *cis*-regulatory syntax on Yan binding site occupancy [220]. The model considers an ensemble of microstates, each defined by a unique configuration of vacant or Yan-bound sites. Each microstate is assigned a thermodynamic potential based on the cumulative influence of strong ETS-binding, weak non-ETS binding, and polymerization. We augmented this model by incorporating Pnt as a second transcription factor that competes for occupancy of the same binding sites (Fig. 3.5A).

Our formulation of the model is based on a single *cis*-regulatory element consisting of  $n$  adjacent binding sites, each of which may be designated as ETS or non-ETS. Each binding site may only exist in one of three binding states; bound by a single copy of Yan, bound by a single copy of Pnt, or unbound. Thermodynamic potentials are assigned to each binding state using two parameters for each transcription factor. The parameter  $\alpha_X$  defines the free energy of transcription factor  $X$  binding to an ETS site, while  $\beta_X$  defines the free energy of binding to a non-ETS site (Fig. 3.5A). A unique configuration of binding states for all  $n$  binding sites constitutes a single microstate,  $k$ . The thermodynamic potential of each microstate is given by the superposition of thermodynamic potentials for each of its constituent binding sites. For each microstate, the stabilizing effect of polymerization is incorporated via a third parameter,  $\gamma_X$ , that defines the free energy of SAM-SAM binding

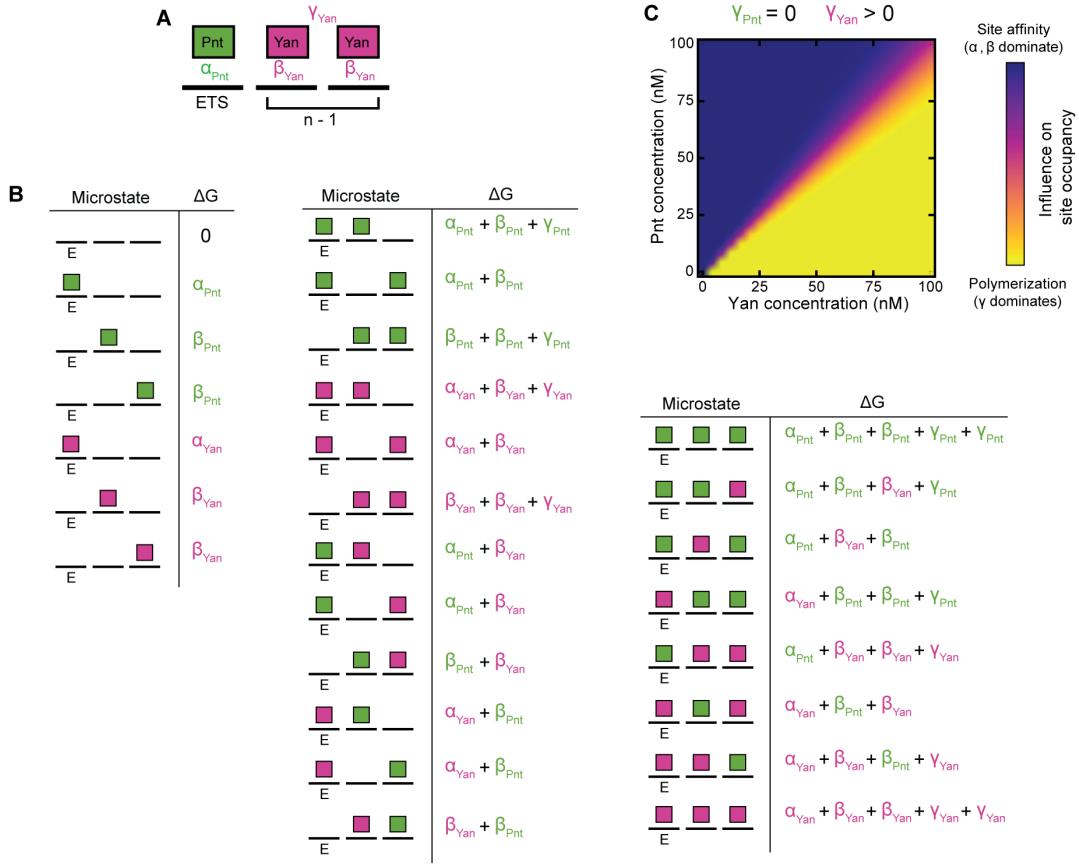
between a pair of similar transcription factors bound to adjacent sites. The net result is a total thermodynamic potential,  $\Delta G_k$ , for each microstate. An example enumeration of all possible microstates for an element consisting of one ETS site preceding two non-ETS sites is provided in Figure 3.5B. The statistical frequencies of each microstate are obtained via the canonical ensemble:

$$p_k = \frac{\exp\left(\frac{-\Delta G_k}{RT}\right)[P]^{a_P(k)}[Y]^{a_Y(k)}}{\sum_k \exp\left(\frac{-\Delta G_k}{RT}\right)[P]^{a_P(k)}[Y]^{a_Y(k)}} \quad (3.3)$$

in which  $p_k$  is the frequency of microstate  $k$ ,  $[P]$  and  $[Y]$  are the Pnt and Yan concentrations,  $a_P(k)$  and  $a_Y(k)$  are functions representing the number of bound molecules of  $P$  and  $Y$  within microstate  $k$ ,  $T$  is a fixed temperature set to 300 K, and  $R$  is the gas constant. Fractional occupancies for each binding site correspond to the cumulative frequency of all microstates in which the site is occupied by a given transcription factor. Overall fractional occupancies may then be computed by summing across all sites within the element.

Using this model, we sought to characterize the sensitivity of Pnt binding site occupancy to changes in the Pnt-to-Yan ratio without neglecting cooperativity derived from *cis*-regulatory syntax. We first considered a scenario in which Yan and Pnt did not exhibit cooperativity (Fig. 3.6A). In the absence of stabilizing SAM-SAM interactions, the landscape of overall binding site occupancy is identical to that obtained with the simple binding model described above (Figs. 3.6B and 3.4B). Increasing the Pnt-to-Yan ratio revealed a gradual increase in Pnt occupancy for all individual binding sites (Fig. 3.6C). This titration contour closely resembles a Langmuir isotherm or Michaelis-Menten saturation curve [221].

We then introduced a stabilizing SAM-interaction for Yan (Fig. 3.6D). The resultant landscape of overall Pnt binding site occupancy is clearly distinguished from the simple



**Figure 3.5. Thermodynamic model of transcription factor DNA binding.** (A) Summary of thermodynamic interactions within one microstate of a cis-regulatory element containing one ETS site and two non-ETS sites. Solid black lines represent individual binding sites. Green and magenta rectangles denote Pnt and Yan molecules. Example thermodynamic potentials of strong ETS-binding, weak non-ETS binding, and polymerization interactions are denoted by  $\alpha_{Pnt}$ ,  $\beta_{Yan}$ , and  $\gamma_{Yan}$ , respectively. For this microstate,  $a_P(k) = 1$  and  $a_Y(k) = 2$ . (B) Enumeration of all possible microstates for a cis-regulatory element of length 3 in which only the first site carries the ETS designation. Solid black lines denote binding sites, green and magenta rectangles denote bound Pnt and Yan molecules. The cumulative thermodynamic potentials of each microstate,  $\Delta G_k$ , are listed beside each graphical depiction. Visual representation is adapted from [220]. (C) Relative thermodynamic contributions of binding site affinity versus polymerization to microstate statistical frequencies as a function of Pnt and Yan concentration. For each point in the plane, influence of site affinity was calculated by weighting the sum of all ETS and non-ETS thermodynamic potentials for each microstate by the statistical frequency of the corresponding microstate. The influence of polymerization was analogously determined. The shown color scale reflects the relative magnitude of these two summations, normalized by limits of zero and complete polymerization.

binding model by a sharpening of the transition from Yan to Pnt dominance in occupancy (Fig. 3.6E). Weighting the energetic contributions of binding strength and polymerization by the statistical frequency of each microstate revealed that the transition is driven by an

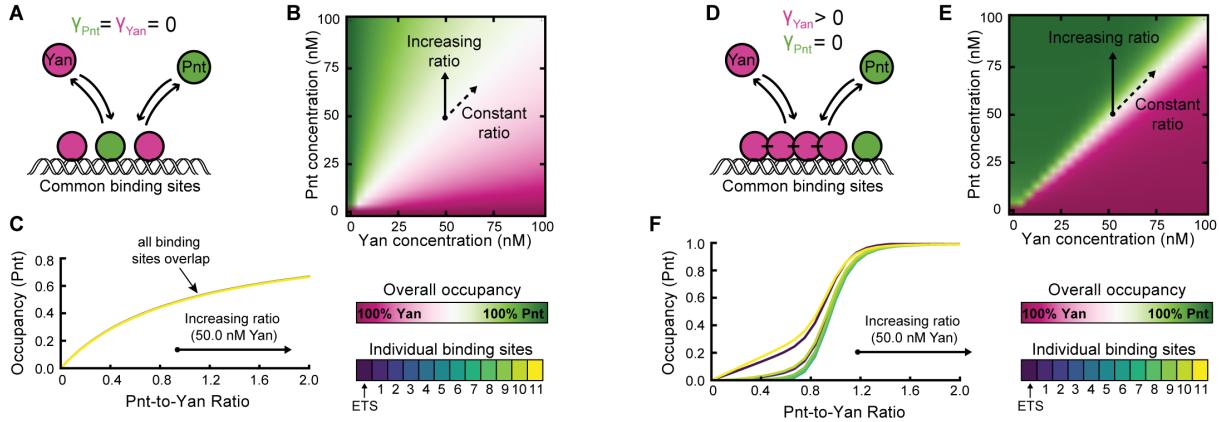
abrupt change in the dominant binding mechanism. Polymerization effects dominate binding site occupancy when the Pnt-to-Yan ratio is low, while binding strength dominates when the ratio is high (Fig. 3.5C).

Increasing the Pnt-to-Yan ratio revealed nonlinear transitions from low to high Pnt occupancy for each individual binding site (Fig. 3.6F). These transitions resemble Hill functional forms [221], indicating the emergence of sharp thresholds that delimit distinct regimes of transcriptional output. At low Pnt-to-Yan ratios, Yan is able to polymerize and occupies all binding sites. At some critical Pnt-to-Yan ratio, Pnt-bound sites intersperse Yan-bound sites such that Yan is no longer able to polymerize. Pnt then out-competes Yan as the ratio increases further. These results recapitulate the long-standing notion that cooperative DNA-binding sensitizes transcriptional output to changes in transcription factor activity.

Assuming binding sites are saturated, then relative occupancy by Pnt and Yan is agnostic to changes in the absolute abundance of either factor, as long as the Pnt-to-Yan ratio remains constant. This mechanism, coupled with cooperativity, would enable modest changes in the Pnt-to-Yan ratio to elicit large changes in DNA binding site occupancy by either factor, and presumably large changes in mRNA synthesis given the opposing transcriptional effects of Yan and Pnt.

### **3.5. Regulation stabilizes the ratio against varying Pnt and Yan concentrations**

Equilibrium modeling suggests cell state transitions can proceed normally amidst individual or cell-to-cell fluctuations in the absolute concentrations of Pnt and Yan. We tested this idea by varying the genetic dosage of the *pnt* gene from one to two copies. Protein output in *Drosophila* is approximately proportional to the number of copies of any given



**Figure 3.6. Cooperative DNA-binding sensitizes transcriptional output to the Pnt-to-Yan ratio.** (A) Cartoon of competition between Pnt and Yan for occupancy of mutual binding sites in the absence of Yan polymerization. (B) Overall binding site occupancy as a function of transcription factor abundance in the absence of Yan polymerization. Color scale reflects overall Pnt site occupancy. A diverging scale was used because all sites are fully saturated at total transcription factor concentrations above 1 nM. Under the range of conditions shown, this implies that Yan occupies all sites left vacant by Pnt. Simultaneous proportional increases in absolute abundance of both species have minimal impact on Pnt occupancy (dashed arrow), while varying ratio confers gradual change (solid arrow). (C) Pnt occupancy of individual binding sites as a function of Pnt-to-Yan ratio in the absence of Yan polymerization. Contours correspond to a vertical path traversed across panel B at a fixed Yan concentration of 50 nM. All binding sites behave similarly. (D) Cartoon of competition between Pnt and Yan for occupancy of mutual binding sites when Yan polymerizes via its SAM domain. (E) Overall binding site occupancy as a function of transcription factor abundance when Yan polymerizes via its SAM domain. Color scale and arrows retain their meaning from panel B. (F) Pnt occupancy of individual binding sites as a function of Pnt-to-Yan ratio when Yan polymerizes via its SAM domain. Contours correspond to a vertical path traversed across panel E at a fixed Yan concentration of 50 nM. Line colors reflect binding site positions within the *cis*-regulatory element. Sites at intermediate distances from the strong ETS site (green lines) transition at higher ratios than those nearest and furthest from the strong ETS site (blue and yellow lines).

gene [222], validating our strategy. We found that the eyes of adult flies were morphologically indistinguishable across this *pnt* dosage range (Fig. A.1A,B). A similar lack of dosage sensitivity had been previously observed with the *yan* gene [43]. Because both sets of genetic manipulations should in theory change the Pnt-to-Yan ratio, the absence of overt phenotypes suggested that either cell state transitions are not sensitive to this ratio or that there are active feedback mechanisms that drive cells back to the ideal ratio.

To distinguish between these possibilities we asked whether the ratio of Pnt-GFP to Yan protein is sensitive to the abundance of Pnt-GFP protein. We quantified Pnt-GFP levels

in eye cells containing either one or two copies of the *pnt-gfp* transgene in a *pnt* mutant background. As expected, Pnt-GFP protein concentration varied proportionally to *pnt-gfp* gene copy number (Fig. 3.7A,B). Interestingly, average Yan protein concentration also scaled with *pnt-gfp* gene copy number (Fig. 3.7C,D) resulting in an essentially identical Pnt-to-Yan protein ratio (Fig. 3.7E). The dependence of Yan protein output on *pnt-gfp* gene copy number parallels our previous finding that *pnt* mutant cells had lower Yan protein concentrations [43]. We conclude that the network regulating Yan protein output compensates for variation in the abundance of Pnt to maintain a constant Pnt-to-Yan ratio.

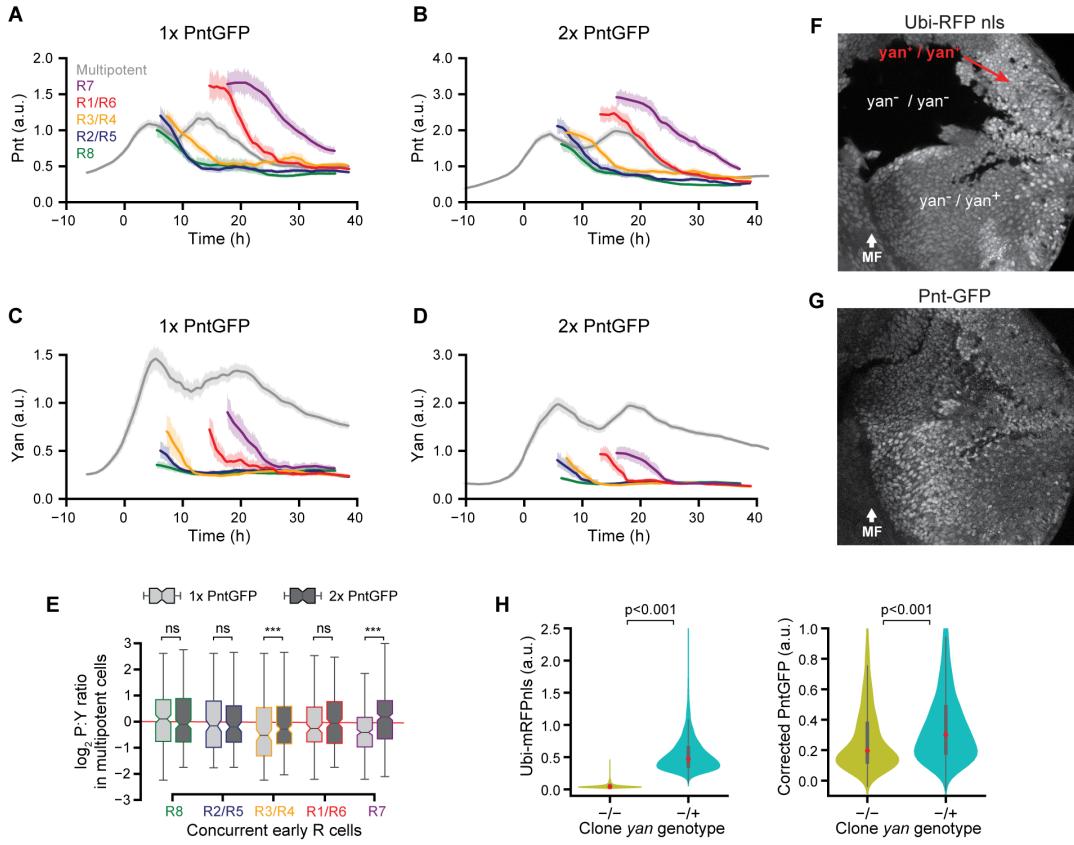
Effective control of the Pnt-to-Yan ratio would require a similar dependence of Pnt protein output on the concentration of Yan. We assessed this prediction by quantifying Pnt-GFP protein in cells with different copy numbers of the *yan* gene. Due to the embryonic lethality of *yan*-null mutations, we conducted this experiment by inducing *yan*-null clones in the developing eye, and using an Ubi-mRFPnls marker to identify genotypes of cells in the clones. Cells with different *yan* gene dosages exhibited different Pnt-GFP protein concentration (Figs. 3.7F,G and A.3A,B). Notably, Pnt-GFP levels were correlated with *yan* gene dosage in progenitor cells immediately posterior to the MF. We quantified this effect by measuring Pnt-GFP protein concentration in cells with or without a copy of the wildtype *yan* gene (Figs. 3.7H). Pnt-GFP expression was higher in cells with one or more copies of *yan* than in cells with zero copies (Mann-Whitney *U* test,  $p < 0.001$ ), suggesting that Pnt protein output is dependent on the abundance of Yan protein. Overall, these data suggest cells compensate for fluctuations in Pnt or Yan protein abundance by adjusting the level of the opposing protein.

Yan expression was not quantified in this experiment due to limiting availability of fluorescence reporters with non-overlapping emission spectra. We were therefore unable

to determine whether the Pnt-to-Yan ratio is robust to changes in *yan* dosage. However, the data suggest that cells respond to an increase in *yan* gene dosage by increasing their Pnt protein levels. Our wildtype data indicate that the Pnt-to-Yan ratio remains constant over time (Fig. 3.1K). It is therefore possible that mutual compensation between Pnt and Yan help preserve the ratio by buffering fluctuations in the abundance of either protein. If this mutual compensation occurs on a sufficiently fast timescale, it could account for the constant Pnt-to-Yan ratio we observed in progenitor cells, despite large changes in Pnt and Yan concentration over time (Fig. 3.1E,J).

### 3.6. Notch signaling lowers the Pnt-to-Yan ratio in progenitor cells

Progenitor cells are dynamically stable about a constant Pnt-to-Yan ratio, but this fixed ratio changes as cells transition to an R cell state. We next asked how the ratio is set in individual cells targeted for differentiation. Notch and RTK signaling provide both transcriptional and post-transcriptional inputs to the Yan-Pnt network [83], and therefore present prime candidates to establish and modulate the Pnt-to-Yan ratio. The two pathways generally exert opposing influence on R cell state transitions. Notch signaling is required to maintain progenitor cells in a multipotent state [212], and ensures proper patterning of the first group of R cells by constraining the proximity of adjacent R8 cells [35, 78]. Conversely, RTK signaling is required to initiate each of the subsequent R cell state transitions [208]. Both signaling pathways influence the Pnt-Yan network in multipotent eye cells [81, 82, 199]. Notch stimulates Yan expression [223], while RTK activates PntP2 and stimulates PntP1 expression [84, 199] while attenuating the pulse of Yan expression [43, 82]. The precise influence of Notch and RTK signaling on the Pnt-to-Yan ratio are difficult to predict as the two pathways are coupled by feedback within and beyond the Pnt-Yan network [29, 223, 224].



**Figure 3.7. The Pnt-to-Yan ratio is stabilized against varying Pnt and Yan concentrations in progenitor cells.** (A-D) Moving averages of (A, B) Pnt-GFP and (C, D) Yan levels in progenitor and differentiating cells with one (A, C) versus two (B, D) copies of the *pnt-gfp* gene. Measurements used DAPI to mark nuclei. Colors denote cell type. Shaded regions are bootstrapped 95% confidence intervals for the moving average. (E) Comparison of Pnt-to-Yan ratios between progenitor cells with one versus two copies of *pnt-gfp* during cell fate transitions. Colors denote cell fate transition periods for each R cell type. These time periods are defined in each disc by the times spanned by the first ten identifiable R cells. The concurrent progenitor cell populations are selected from these time windows. Light grey filled boxes denote 1x *pnt-gfp*, dark grey filled boxes denote 2x *pnt-gfp*. Pnt-to-Yan ratios in progenitor cells are indistinguishable between gene dosages during R8, R2/R5, and R1/R6 cell fate transitions (KS 2-sample test,  $p < 0.001$ ). (F-G) Confocal image slice of progenitor nuclei in a disc containing loss-of-function *yan* clones. RFP fluorescence marks wildtype *yan*. (H) Quantitative comparison of Pnt-GFP expression between *yan* genotypes. Progenitor cells were assigned *yan* genotypes based on measured RFP level, and Pnt-GFP levels were corrected to account for fluorescence bleed-through (see 2.4). Pnt-GFP levels decrease when no gene copies of *yan* are present (Mann-Whitney *U* tests). Red dots denote the median of each distribution, thick grey lines denote the interquartile range.

We used a temperature-sensitive *Notch* mutant [225] to measure the impact of Notch-mediated signaling on Pnt-GFP and Yan. We divided the eye field into two regions for analysis purposes (Fig. A.4A, dashed yellow line). The first region starts anterior to the MF

and ends in the region of R8 cell specification. The second region extends  $\sim 10$  columns of R8 cells posterior to the first region. These regions contain the first and second pulses of Pnt-GFP, respectively (Fig. A.4A, thick black lines).

At the non-permissive temperature, progenitor cells had visibly reduced Yan levels (Fig. A.4A,B), consistent with previous reports of Yan expression's dependence on Notch signaling in the developing eye disc [223]. Pnt-GFP levels were also reduced in region 1, but appeared close to normal at later times (Fig. A.4A,B). Attempts to study ratio dynamics in mutant eye discs were challenging. Notch is essential for proper patterning of R8 cells, so the mutant eye discs had distorted spacing of R8 cells at the non-permissive temperature (Fig. A.4C). The irregularity of the intervals between adjacent columns of R8 cells precluded the conversion of spatial position along the anterior-posterior axis to developmental time. As an alternative to our standard quantitative analysis, we visualized the effects of Notch by mapping the pixel-wise difference between Pnt-GFP and Yan to a diverging color scale (Fig. 3.8A,B). Direct visualization of the ratio yielded a very similar view of differential Pnt-GFP/Yan expression, but was prone to computational errors imparted by zero-valued pixels. This qualitative analysis was limited to optical sections specifically spanning the progenitor cells.

At the non-permissive temperature, the *Notch* mutant showed a consistently higher Pnt-to-Yan ratio in progenitor cells in region 2 (Fig. 3.8A). This observation suggests that Notch signaling maintains a low Pnt-to-Yan ratio in these undifferentiated cells. The known role of Notch in maintaining multipotency in region 2 [212] is consistent with our hypothesis that cell state transitions are mediated by the transcription factor ratio.

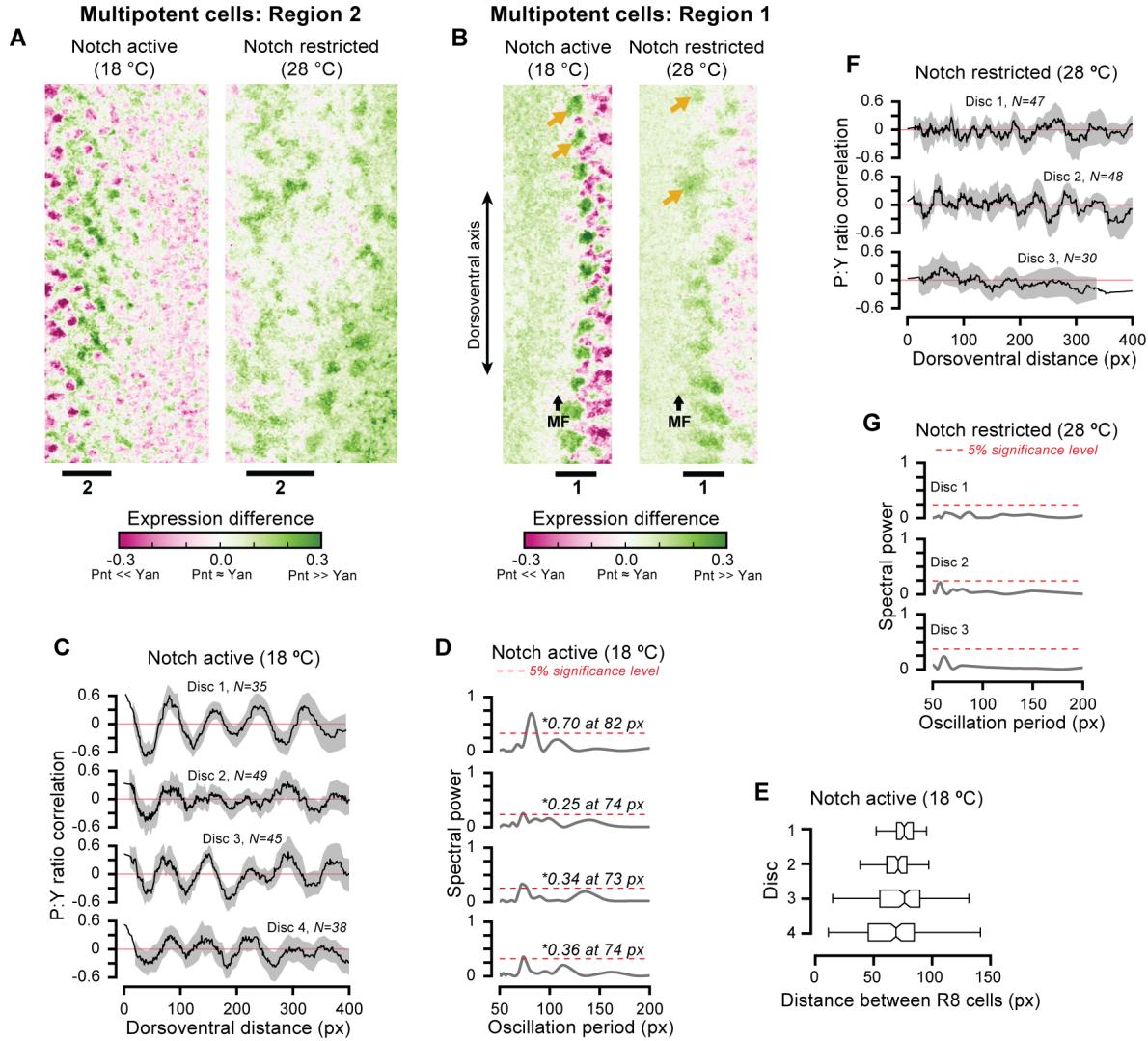
The Notch mutant revealed more complex behavior in region 1, where the first group of R cell transitions occur. In this region Notch is required for patterning of the R8 lattice

[78]. At the permissive temperature, there was a periodic pattern in the Pnt-to-Yan ratio of progenitor cells (Fig. 3.8B). Clusters of cells with higher ratio alternated with clusters with lower ratio. We quantified the periodicity of this pattern by evaluating the similarity of ratios between cells as a function of their separation distance (Fig. 3.8C). We detected periodic spatial patterns with a constant period of oscillation that was approximately equivalent to the length scale separating adjacent R8 cells (Fig. 3.8D,E). Since young R8 cells have elevated Pnt-to-Yan ratios, we infer that periodic clusters of high-ratio progenitor cells give rise to the R8, R2/R5, and R3/R4 cells. At the non-permissive temperature, the Pnt-GFP/Yan pattern was strongly impaired (Fig. 3.8B). The ratio was more uniform along the dorso-ventral axis, and while there were modestly detectable oscillations in the ratio, their period was not stable (Fig. 3.8F,G).

These results are consistent with the consensus understanding that Notch signaling serves dual roles in R8 fate determination. Initially, Notch pushes clusters of progenitor cells towards an R8 cell state. Later, Notch restricts differentiation to ensure only one cell per cluster adopts an R8 state [32, 78, 226]. If Notch signaling is inhibited, very few high Pnt-to-Yan ratio clusters and R8 cells are formed because the first step is blocked [32]. If cell state transitions are coupled to the Pnt-to-Yan ratio, then the *Notch* mutant would not be expected to have high-ratio clusters in region 1. This is precisely what we observed.

### 3.7. Ras signaling elevates the Pnt-to-Yan ratio in progenitor cells

RTK signals received by progenitor cells trigger their transition to R cell states [208]. We quantitatively probed the effect of RTK signaling on Pnt-GFP dynamics and the Pnt-to-Yan ratio by using a temperature-sensitive *EGFR<sup>ts</sup>* allele that restricts RTK signaling [227]. At high temperatures, the mutant blocks RTK signal transduction, which triggers



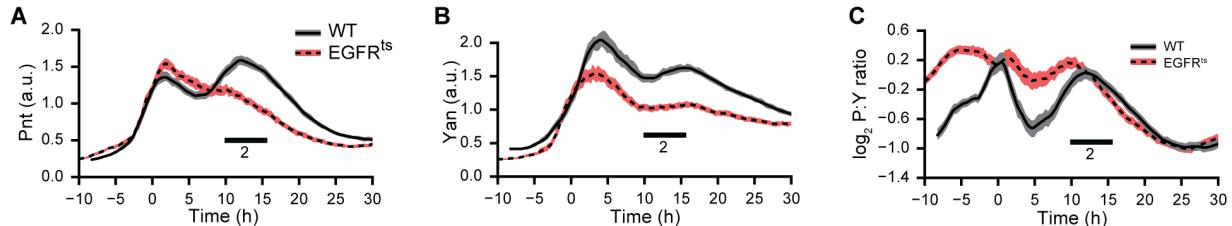
**Figure 3.8.** (Continued on the next page.)

cell death and allows only R8 neuron patterning. However, animals raised at intermediate temperatures achieve normal recruitment of R8, R2/R5, and R3/R4 neurons, but fail to recruit most R1/R6 and R7 cells [43].

At intermediate temperatures, the entire second pulse of Pnt-GFP expression disappeared upon restriction of EGFR activity (Fig. 3.9A). Because the second pulse can be predominantly ascribed to PntP2 expression (Fig. 3.1I), this result conflicts with a report that PntP2 expression is not dependent upon RTK signaling [84]. At intermediate signaling,

**Figure 3.8. Notch signaling lowers the Pnt-to-Yan ratio in progenitor cells.** (A) Visualization of relative Pnt and Yan expression in progenitor cells in region 2 when Notch signaling is active (left panel) and restricted (right panel). Color scale reflects the difference between Pnt-GFP and Yan fluorescence. Black lines denote periods of elevated Pnt-GFP expression. See methods for details on post-processing of images. (B) Visualization of relative Pnt and Yan expression in progenitor cells during the first wave of cell state transitions. Black arrow marks the morphogenetic furrow. Gold arrows annotate clusters of elevated ratio. (C-E) Quantification of spatial periodicity in the Pnt-to-Yan ratio among progenitor cells immediately posterior to the MF when Notch signaling is active. (C) Spatial correlation functions for progenitor cells in four eye discs. Black lines show the moving average pairwise correlation of Pnt-to-Yan ratios between cells as a function of their separation distance along the dorso-ventral axis. Oscillatory forms indicate alternating regions of similar and dissimilar behavior relative to the population-wide mean. Lines are obtained via first-order Savitzky-Golay filtration with a window size of 50. Shaded region shows a bootstrapped 95% confidence interval for the moving average. Cell counts are annotated above each correlation function. Red lines are the expected outcome for random expression (no pattern). (D) Normalized Lomb-Scargle periodograms for each disc. Spectra are constructed from individual progenitor cell measurements for periods ranging 50 to 200 px. Grey lines denote spectral power attributed to each oscillation period. Dashed red significance thresholds are obtained by bootstrap resampling the ratio intensities. Asterisks denote signal frequencies exceeding the confidence threshold. In all discs, a pattern in Pnt-to-Yan ratios repeats on a length scale of 73-82 px when Notch signaling is active. (E) Distribution of dorso-ventral separation distances between adjacent R8 neurons within a single column of ommatidia within each disc. Mean values are comparable to the detected oscillation periods. (F,G) No periodicity is detected above the significance threshold when Notch signaling is restricted.

Yan levels were also reduced, resulting in Pnt-to-Yan ratios that were indistinguishable from wildtype during the second wave of state transitions (Fig. 3.9B,C).



**Figure 3.9. Pnt and Yan expression dynamics in  $EGFR^{ts}$  eye discs.** (A-C) Effects of  $EGFR^{ts}$  on (A) Pnt-GFP, (B) Yan, and (C) Pnt-to-Yan ratio dynamics in progenitor cells. Lines are moving averages across 250 sequential cells. Shaded regions are bootstrapped 95% confidence intervals for the mean. Solid lines and grey shading denote wildtype controls. Dashed lines and red shading denote restricted EGFR signaling. Black bars denote second period of elevated Pnt-GFP expression.

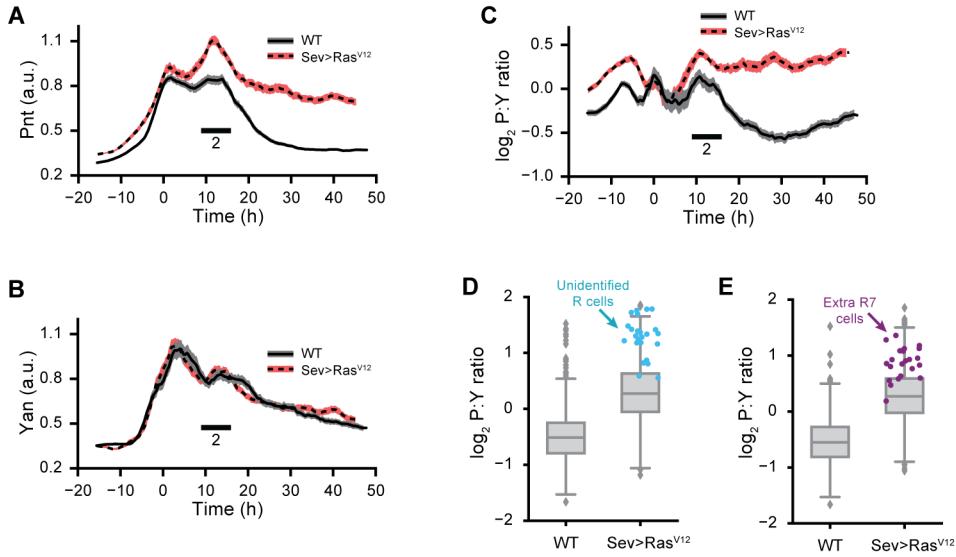
We next asked whether RTK signaling is sufficient to induce an increase in the Pnt-to-Yan ratio of progenitor cells by expressing a constitutively-active form of Ras [228, 229]. This construct uses a *sev* promoter to drive Ras expression, limiting its effects to the second region

of Pnt-GFP expression. Constitutive Ras activity dramatically increased the amplitude and duration of the second pulse of Pnt-GFP expression in progenitor cells (Fig. 3.10A) but did not significantly alter Yan expression dynamics (Fig. 3.10B). This yielded a sustained increase in Pnt-to-Yan ratio in progenitors during the second wave of state transitions (Fig. 3.10C), as well as the ectopic differentiation of R cells. As previously reported [229], supernumerary cells included extra R7 cells, as well as additional R cells whose identities were not discerned due to their aberrant positioning.

We sought to determine whether these ectopic R cells emerged from a pool of progenitors with abnormally high Pnt-to-Yan ratios. Focusing our analysis on progenitor cells concurrent with the ectopic induction of R cells revealed that their ratios were higher than those of wildtype progenitor cells at comparable times (Fig. 3.10D,E grey boxes). Nearly all young supernumerary R cells had ratios that were within the range of mutant progenitor cells, and were above the range observed in wildtype cells (Fig. 3.10D,E blue and purple markers). These observations suggest that abnormally high Pnt-to-Yan ratios in progenitor cells accompany ectopic R cell state transitions.

### 3.8. Ratiometric control as a model for cell fate commitment

Successful cell state transitions usually require changes in mRNA and protein expression, which are often dictated by transcription factors. It is widely believed that expression depends on the absolute concentration of these factors [230]. We have identified a scenario in which cells respond to the ratio in abundance of two transcription factors rather than to the absolute concentration of either protein. This novel mechanism is made possible by several characteristics of the system. First, target genes of these factors are induced when cells transit to differentiated states, and expression of these targets is necessary for differentiation



**Figure 3.10. Ras signaling elevates the Pnt-to-Yan ratio in progenitor cells.** (A-C) Effects of constitutive Ras signaling on (A) Pnt-GFP, (B) Yan, and (C) Pnt-to-Yan ratio dynamics in progenitor cells. Lines are moving averages across 250 sequential cells. Shaded regions are bootstrapped 95% confidence intervals for the mean. Solid lines and grey shading denote wildtype controls. Dashed lines and red shading denote constitutive Ras signaling by *Sev > Ras<sup>V12</sup>*. Black bars denote periods of elevated Pnt-GFP expression. We previously reported a modest increase in the duration of Yan-YFP expression in *Sev > Ras<sup>V12</sup>* mutant progenitor cells [43], but this difference was not detected using the Yan antibody. (D, E) Comparison of Pnt-to-Yan ratios between wildtype and *Sev > Ras<sup>V12</sup>* progenitor cells concurrent with the ectopic differentiation of (D) unidentified R cells and (E) R7 cells in *Sev > Ras<sup>V12</sup>* discs. Markers denote the first 25 supernumerary R cells.

[197, 231]. Second, both factors bind to the same DNA sites in their target genes. Third, they exert opposing effects on transcription.

A theoretical analysis reveals that, under these conditions, relative site occupancy by either factor determines whether or not a target gene is transcribed. When binding sites are saturated, the probability a site is occupied by one of the factors is controlled by the ratio of factor concentrations and not by the absolute concentration of the factors. This mechanism allows the transcription factors to have pulsatile expression dynamics and still consistently regulate transcription. Regulation of genes through pulsatile dynamics of competing transcription factors with opposing effects has been reported in yeast [232]. In the developing eye, however, this competition relies on the ratio of the two factors to differentially

regulate genes. Perhaps the pulsatile dynamics of Pnt and Yan allow R cell state transitions to be restricted to a specific period of developmental time.

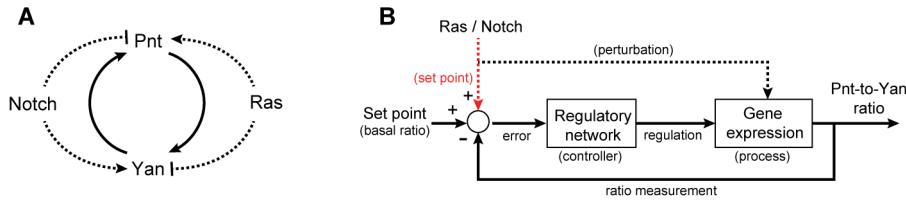
We developed a model to demonstrate that stabilizing interactions between adjacently bound monomers can sensitize enhancer occupancy to the relative abundance of two competing transcription factors. The model does not offer a rigorous characterization of the competition between Pnt and Yan for a specific enhancer, as the data required to construct such a model are not available. For example, it is unknown whether nuclear concentrations of Pnt and Yan protein are of an equivalent order of magnitude in differentiating eye cells. Similarly, we do not know the strength, number, or arrangement of binding sites in the enhancers targeted by Pnt and Yan. However, the conclusions we have drawn from our model do not rely upon specific values for these variables, and are consequently not anchored to the specific details of any one system.

In general, the proposed ratiometric sensing mechanism is well suited when either or both transcription factors bind their target sites in a manner stabilized by cooperative interactions. Like its human ortholog TEL1 [233], DNA-bound Yan monomers enhance recruitment of Yan to adjacent binding sites through stabilizing SAM-SAM polymerization [234]. We show theoretically that such cooperativity could generate threshold-like behavior and cause ultrasensitive switching in site occupancy between Yan and Pnt. In this model, the switching is agnostic to the absolute abundance of either transcription factor as long as their relative ratio can be precisely controlled. This mechanism would enable state transitions to proceed despite variation in protein concentrations driven by fluctuations in metabolism, cell volume, expression noise, environmental conditions, genetic polymorphisms, or gene copy number.

The average Pnt-to-Yan ratio is dynamically stable about an approximately constant value within each cellular state, and only exhibits transient fluctuations when cells switch states. These dynamics reflect the capacity of the system to coordinate the relative expression of the two transcription factors. We have found that the abundance of each transcription factor depends upon the expression of the other factor. Dependencies of this type are often depicted as positive or negative regulation in cartoons of qualitative regulatory interactions (Fig. 3.11A). We believe that as biology becomes increasingly quantitative it will be more fruitful to emphasize an empirical description of system dynamics based on control theory (Fig. 3.11B).

From a control perspective, a system of cellular components monitors the relative abundance of Pnt and Yan and takes corrective action when the ratio deviates from a specified reference value. The particular components responsible for implementing control may remain unspecified. This perspective eschews molecular events in favor of minimizing complexity, but preserves the salient features of a detailed molecular mechanism and can enable quantitative predictions.

Fluctuations in the absolute abundance of one factor are mitigated by compensatory adjustment of the other. Notch or RTK activity could modulate Pnt or Yan protein levels to transiently perturb the Pnt-to-Yan ratio (Fig. 3.11B, dashed black arrow). These signals could permanently set the ratio by adjusting the reference value (Fig. 3.11B, dashed red arrow). We advocate this control theoretic perspective because it more accurately conveys the fundamental strategy underlying system behavior. Furthermore, accurate model predictions would only require the evaluation of a small number of parameters that characterize Pnt-to-Yan ratio dynamics, obviating the need for experimental measurement of reaction rates during R cell specification in *Drosophila*.



**Figure 3.11. Conceptual models for regulation of the Pnt-to-Yan ratio.** (A) Cartoon of qualitative regulatory interactions suggests Pnt and Yan protein levels are coupled by reciprocal positive feedback (solid lines), while Notch and Ras signaling adjust the Pnt-to-Yan ratio by modulating the levels of each protein (dashed lines). (B) Block diagram of ratio control in the Pnt-Yan network. Lines represent values, rectangles indicate functions, and the circle is a comparison point. The Pnt-to-Yan ratio is compared against a basal reference value, and the difference is fed into a regulatory network that acts to drive the ratio back toward the reference value. Extracellular signals transiently perturb the ratio by modulating Pnt or Yan protein levels (dashed black line), or set the ratio by adjusting the reference value (dashed red line).

We have shown that dynamic changes in the Pnt-to-Yan ratio are coupled to cell state. Our favored interpretation is that the ratio determines which state a cell is in. Direct testing of this hypothesis remains difficult because the observed ratio control strategy precludes gene-level manipulation of the ratio. Instead, we acknowledge that our observations are correlative. We cannot discard the possibility that Notch and RTK signaling regulate state transitions in a manner that is not only mediated by the Pnt-to-Yan ratio but by other mechanisms as well. We emphasize, however, that ratio control enforces stability when cells are in one state, which implicates the ratio as an active cell state determinant. Moreover, qualitatively different approaches to experimentally manipulate the ratio affected cell state transitions in a consistent manner. Both Notch inhibition and Ras activation increase the ratio and cause abnormal R cell state transitions [229, 235]. Single-cell dynamical measurements of Yan and specific Pnt isoforms during isoform-specific perturbations may ultimately prove necessary to determine definitively whether ratios directly mediate transitions. The regulatory mechanism described here provides insight into how the relative dynamics of competing transcription factors can be used to pattern complex epithelia, and may also aid design of synthetic regulatory systems based on ratiometric sensing.

## CHAPTER 4

## Layered repression synchronizes development with cellular metabolism

A manuscript resembling this chapter was coauthored with Justin Cassidy, Rachael Bakker, Ritika Giri, Nicolás Peláez, Bryan Eder, Anna Bobrowska, Neda Bagheri, Luís Amaral, and Richard Carthew. The preprint is available at <https://doi.org/10.1101/548032>. All of the experiments were conceived, designed, and executed by my colleagues. In particular, Justin Cassidy obtained the wealth of data that ultimately made this work possible. My contributions include all of the computational modeling and simulations. Much of the text was either written by or under the guidance of Professor Richard Carthew.

### 4.1. Background on the environmental dependence of developmental tempo

Animal development occurs over a defined timescale, which requires control of the rates of developmental processes. Developmental timescales are an intrinsic feature of a species, and are not necessarily determined by external clocks [236]. Rather, the pace of development is encoded in the genome. Development occurs via a stereotypic sequence of events involving cell division, growth, movement, apoptosis, polarization, and differentiation. Correct assembly of functional structures depends upon synchronization of cell division and differentiation events [237, 238]. Small variation in timing produces variation in structure that is observed between individuals [239, 240]. Abnormal timing can result in structural defects that lead to compromised survival [241].

While developmental tempo is a fundamental property of a species, it can vary under different conditions. For example, temperature affects the pace of development in many ectotherms, such as arthropods, nematodes, fish, and reptiles [242–245]. Diet and food intake also affect organismal growth rate and the pace of development for many species, including humans [246–249]. Finally, cellular metabolism can alter the pace of development. For example, the evolutionarily conserved *Clk1* gene encodes a mitochondrial enzyme necessary for normal cellular respiration [250], and loss of the *clk1* gene in nematodes and mice results in developmental delays [251–253]. In *Drosophila*, restricting glucose consumption by cells slows development [254–257]. West and colleagues formulated a general quantitative model that relates developmental tempo to both cellular metabolic rate and temperature [258]. Strikingly, the model fits meta-data spanning several kingdoms, suggesting a universal relationship between metabolism and developmental tempo.

Many developmental processes involve specification of different cell types in a stereotyped sequence. All of these differentiated cell types originate from progenitor cells. The sequence of cell differentiation is driven by changes in the gene expression program within progenitors. Gene regulators, typically transcription factors, are sequentially activated and repressed, resulting in transient periods of increased activity. During these periods, they change gene expression in the progenitors. This coincides with and causes a temporal series of cell fate decisions. Since these regulators frequently interact with one another, the entire cascade constitutes a gene regulatory network (GRN). Such GRNs have been characterized for embryogenesis [259, 260], development of the central nervous system [261], and development of the sensory nervous system [262]. Because the tempo of development can vary, GRN dynamics must be able to reliably adjust to a variable timing mechanism. Therefore, understanding how these GRNs adapt to a variable timescale is crucial for understanding the mechanisms of animal development.

Phenomenological observations suggest that there are limits to the timescales to which development may adapt. While broiler chickens have been successfully bred for rapid growth, frequent abnormalities in musculoskeletal development are evident in such breeds [263, 264]. Animals (and humans) experience hyper-normal growth rates if they initially experience delayed growth [246]. Such compensatory growth is linked to a variety of developmental and physiological defects [248]. Conversely, slowing growth can alleviate defects caused by mutations that impair development. As first noted by T.H. Morgan, morphological phenotypes can be suppressed by limiting the nutrition of mutant animals [265–268]. Likewise, raising animals under lowered temperatures can sometimes suppress the phenotypes of mutations that are not classical *ts* alleles [269–272]. Collectively, these observations suggest an unknown mechanism ensures successful developmental outcomes amidst variability in developmental tempo.

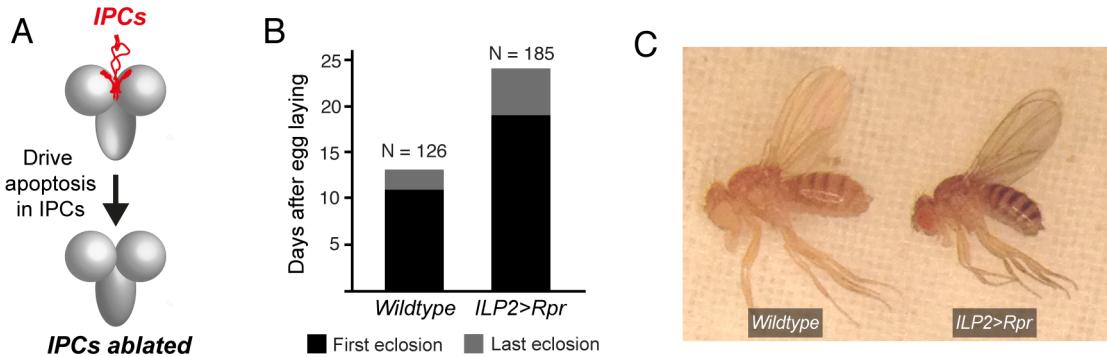
Here, we have explored this mechanism. We find that impairing gene repression in GRNs causes developmental errors but only when cell metabolism and growth rate are normal. When either energy metabolism or protein anabolism are reduced, developmental errors are reduced or even suppressed. We find that this relationship between metabolism and repression is so prevalent that the entire microRNA family becomes unnecessary when metabolism is slowed. Using a general quantitative modeling framework for regulated gene expression, we show that multiple layers of weak repression render gene expression dynamics independent of variable biochemical rates. When rates are modestly reduced, fewer repressors are needed to ensure normal expression dynamics. We experimentally validate this model prediction by following GRN dynamics in *Drosophila*. Our findings support a new mechanism whereby layers of gene repression allow development to occur over a wider range of time scales, enabling development to proceed faster if metabolic conditions allow for it. The need

for flexible error frequency suppression could provide an evolutionary impetus for the high prevalence of genetic redundancy.

#### **4.2. Repressors are less impactful when metabolism is reduced**

Developmental patterns arise from directed dynamics of cell-cell signaling and gene regulation. The sensory organs of *Drosophila* are a classic system with which to study these phenomena [273]. A broad collection of gene mutations has specific effects on the formation of various sensory organs, and these mutations have been instrumental in uncovering the molecular mechanisms of sensory organ development. The affected genes encode transcription factors, microRNAs, signaling factors, and other gene regulators. We used such gene mutations to readdress the relationship between reduced metabolism and phenotype suppression that was first observed by Morgan [266, 267]. We did so by scoring *Drosophila* sensory mutant phenotypes under conditions of reduced energy metabolism. We generated animals that had reduced metabolism by genetic ablation of their insulin producing cells (IPCs) in the brain (Fig. 4.1A). This ablation reduces the amount of glucose that cells consume [256], resulting in 70% slower development (Fig. 4.1B), and small but normally proportioned adults (Fig. 4.1C) [256].

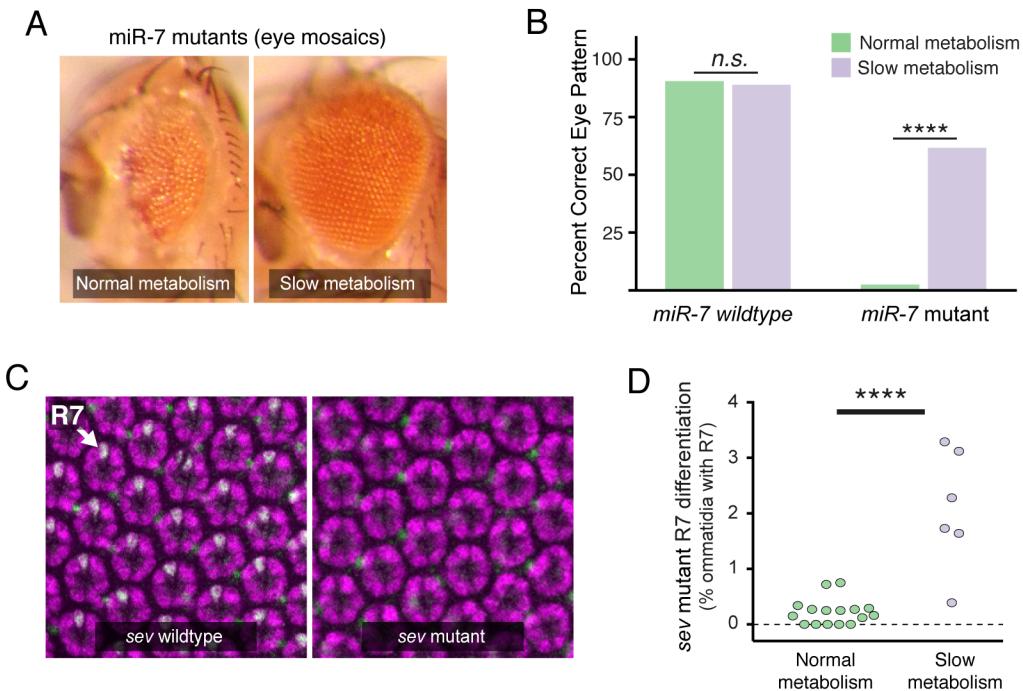
We first examined mutations affecting formation of the compound eye. The microRNA miR-7 represses expression of the Yan transcription factor in the developing eye [16]. Yan protein is transiently expressed in the eye [43], and is cleared from differentiating photoreceptor (R) cells by multiple repressors acting on its transcription, mRNA stability, and protein stability [83]. When the *miR-7* gene was specifically ablated in the compound eye of an otherwise wildtype animal, it resulted in small malformed adult eyes due to errors in R cell differentiation (Fig. 4.2A). This phenotype was highly penetrant in genetically mosaic



**Figure 4.1. IPC ablation slows development by restricting energy metabolism.** (A) Strategy to ablate IPCs (red) in the young fly brain. Gal4 expressed under control of the promoter for the *Insulin-Like Peptide 2* (*ILP2*) gene drives production of the pro-apoptotic protein Reaper (Rpr) specifically in IPCs of the brain. (B) The number of days after egg laying (AEL) at which the first individual in either wildtype or *ILP2 > Rpr* populations eclosed (hatched from pupa into adult) is shown, as is the time at which the last individual in each population eclosed. Population sizes for wildtype and *ILP2 > Rpr* were 126 and 185, respectively. (C) Adult body size is affected by IPC ablation. Two females that were raised at the same time and temperature. The left *yw* animal has normal metabolism, whereas the right animal has slowed metabolism due to ablation of its IPCs.

animals (Fig. 4.2B). However, when energy metabolism was slowed by IPC ablation, loss of *miR-7* was much less important for the formation of correctly patterned eyes (Fig. 4.2B). We also examined mutations affecting cell-cell signaling. The Sevenless (Sev) receptor tyrosine kinase hyper-activates MAP kinase in certain eye progenitor cells, leading to enhanced turnover of the Yan protein [82]. This enables cells to differentiate into R7 photoreceptors [29]. When *sev* is mutated, cells completely fail to differentiate as R7 photoreceptors. This effect was readily apparent by staining for an R7-specific marker protein (Fig. 4.2C). However, slowing metabolism allowed a small but significant number of *sev* mutant cells to become R7 photoreceptors (Fig. 4.2D). Importantly, since the *sev* mutant makes no protein products [274], rescue of the mutant phenotype was not simply due to more functional Sev protein molecules being present in slowly metabolizing cells.

We also examined formation of other sensory organs for evidence of metabolic interactions. Large sensory bristles develop in a highly stereotypic pattern over the *Drosophila* body.

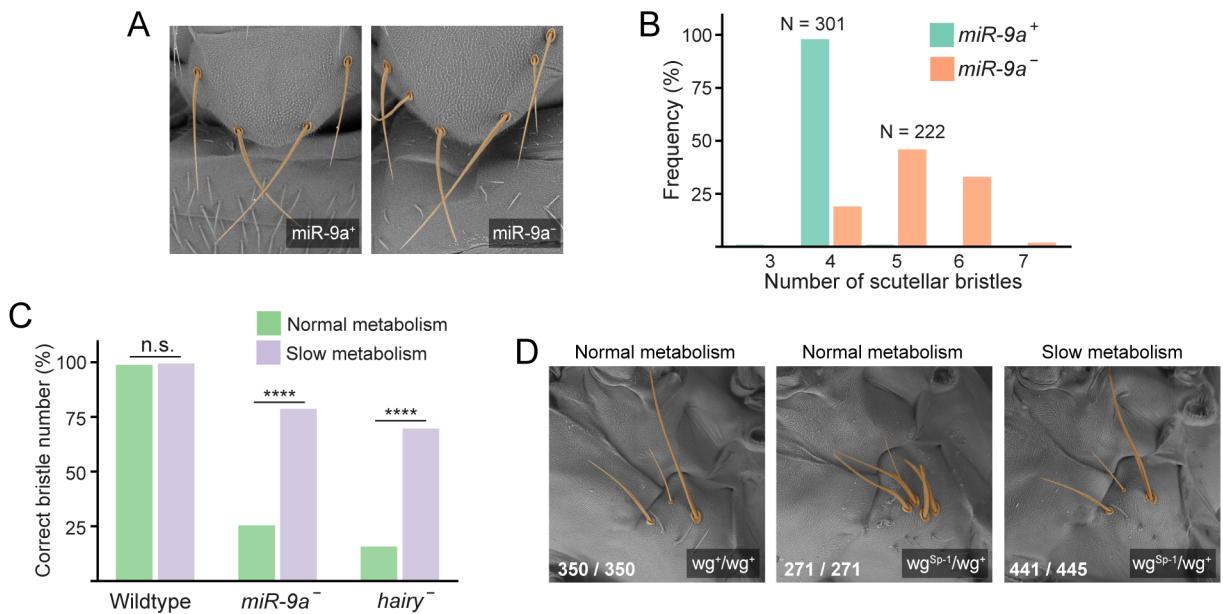


**Figure 4.2. Eye developmental defects are rescued by slower energy metabolism.** (A) Genetically mosaic individuals with a *miR-7*<sup>+</sup> body and a *miR-7* mutant eye. Left individual with mispatterned eye has its IPCs intact while the right individual with a normally patterned eye has had its IPCs ablated by *ILP2 > Rpr*. (B) Eye patterning is more normal if mosaic individuals slowly metabolize due to IPC ablation. Sample population sizes were between 264 and 467 individuals. P values from Chi-square test with Yates correction. (C) Eye cells stained for specific protein markers such that R7 cells (white) can be distinguished from other R cells (purple) and bristle cells (green). Each ring-like cluster of R cells is an ommatidium. Null mutation of *sev* results in no R7 cells (right). (D) Slow metabolism due to IPC ablation increases the fraction of ommatidia that contain an R7 cell in *sev* mutants. Each data point represents one eye sample; between 481 and 837 ommatidia were scored for R7 cells within each eye sample. P value is from a one-way ANOVA with Bonferroni correction.

The protein Senseless (Sens) transiently appears in a cluster of proneural cells before one cell is chosen to differentiate into a sensory bristle [275]. MicroRNA miR-9a represses Sens protein expression, and *miR-9a* mutants frequently develop ectopic sensory bristles because this repression is missing (Fig. 4.3A,B) [94, 276]. However, when *miR-9a* mutants had their IPCs ablated, errors in bristle number were greatly reduced (Fig. 4.3C).

The protein Hairy directly represses transcription of the proneural genes *achaete* and *scute* during selection of cells for bristle fates [277]. Mutation of *hairy* causes some

individuals to develop ectopic large bristles. However, this effect of *hairy* mutation was strongly suppressed when energy metabolism was slowed (Fig. 4.3C). We saw a similar effect on a *cis*-regulatory module (CRM) that represses gene transcription. The *Sternopleural* (*Sp-1*) mutation is present in a CRM located on the 3' side of the *wingless* (*wg*) gene [278], causing Wg misexpression and development of ectopic bristles [278]. However, the ectopic bristle phenotype of the *wg<sup>Sp-1</sup>* mutant was completely reversed under conditions of slowed energy metabolism (Fig. 4.3D).



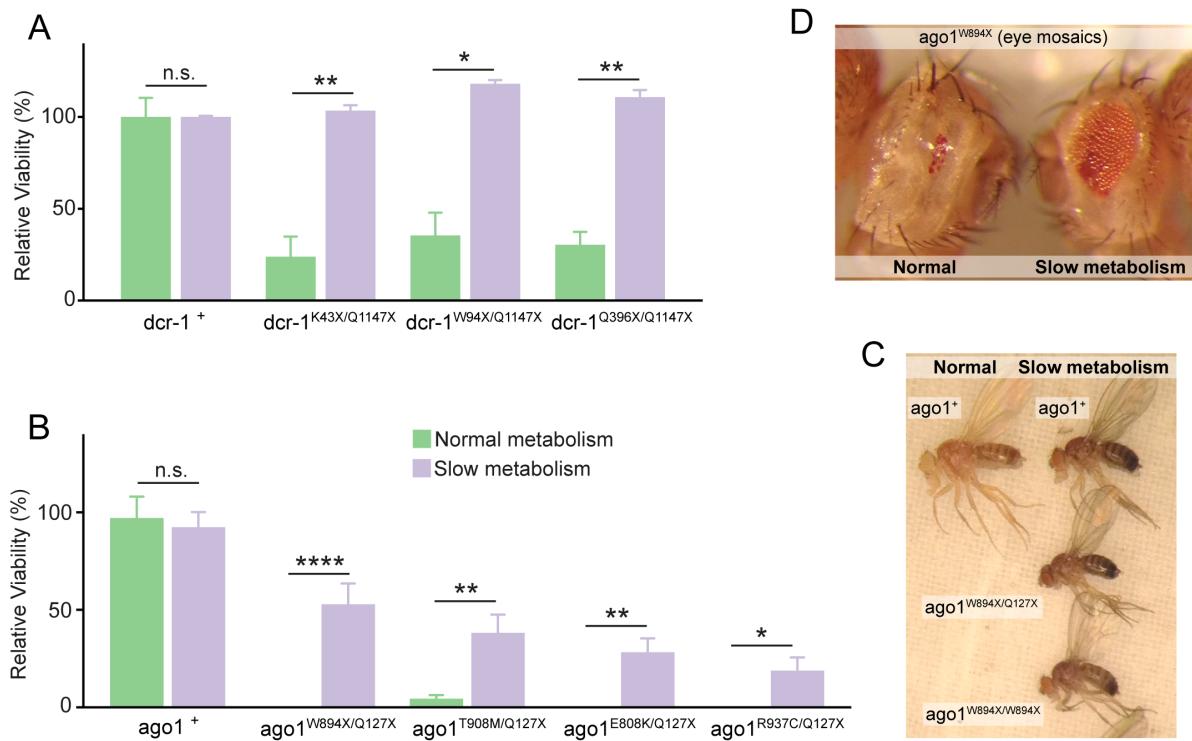
**Figure 4.3. Sensory bristle developmental defects are rescued by slower energy metabolism.** (A) Number of scutellar bristles is frequently greater than four in a *miR-9a* mutant whereas it is almost invariably four in wildtype. (B) Distribution of scutellar bristle numbers in wildtype and *miR-9a* mutant populations. Population sizes for wildtype and *miR-9a* were 301 and 222, respectively. The cumulative frequency distributions between wildtype and mutant were significantly different ( $p < 0.0001$ , KS test). (C) IPC ablation increases the proportion of *miR-9a* mutants and *hairy* mutants that have the wildtype number of scutellar bristles. \*\*\*\*,  $p < 0.0001$ ; n.s.,  $p > 0.05$  (D) Under normal metabolic conditions, *wg<sup>Sp-1</sup>* displays an increased number of sternopleural bristles. IPC ablation dramatically increases the number of mutant individuals with the wildtype number of three sternopleural bristles. Shown in each panel is the number of individuals with bristle number of three versus the total number of individuals scored. IPC ablation significantly suppresses the *wg<sup>Sp-1</sup>* mutant phenotype ( $p < 0.0001$ , Fishers exact test).

#### 4.3. MicroRNAs are dispensable when metabolism is reduced

The mutations thus far examined affect diverse types of regulators, including microRNAs, transcription factors, and signaling molecules. Despite this diversity, all of the mutations have something in common: they affect repressive interactions between genes. To explore the prevalence of this relationship between gene repression and metabolism, we eliminated an entire family of regulatory repressors that control all stages of *Drosophila* development. The microRNA family is composed of 466 distinct microRNAs in *Drosophila melanogaster* [279]. Virtually all microRNAs require Dicer-1 (Dcr-1) protein for their proper biosynthesis, and Ago1 protein as a partner to repress target gene expression [280]. Protein-null mutations in either *dcr-1* or *ago1* genes are lethal [281]. We raised different null *dcr-1* mutants under conditions of slower energy metabolism, and found that many more animals survived development (Fig. 4.4A). *Ago1* null mutants are normally 100% embryonic lethal, but mutant lethality was broadly suppressed when animals slowly metabolize due to IPC ablation (Fig. 4.4B). The mutants survived to adulthood, and most survivors had normal eye and bristle patterns as well as other body structures, indicating the rescue of a massive array of developmental defects (Fig. 4.4C). Rescue could also be seen when Ago1 was specifically ablated in cells of the compound eye; eye development was strongly rescued by slower energy metabolism (Fig. 4.4D). Therefore, a major class of regulatory repressors is rendered non-essential when energy metabolism is slowed.

#### 4.4. Modeling the emergence of developmental errors in variable environments

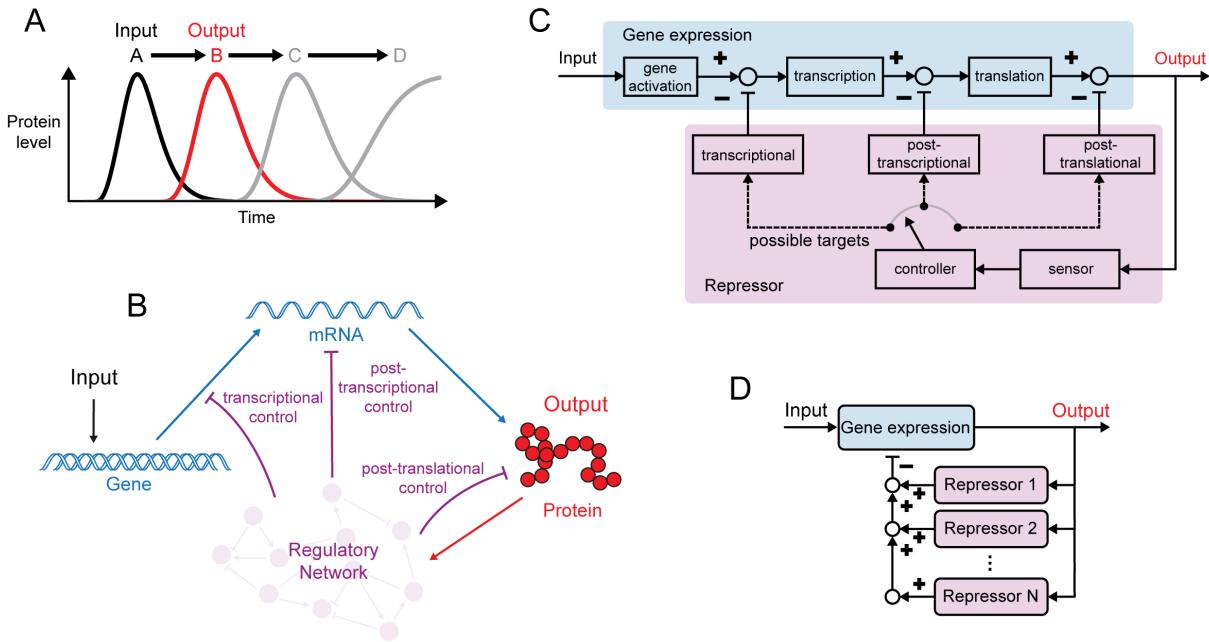
We turned to computational modeling in order to elucidate the biochemical mechanism linking gene repression, developmental phenotypes, and metabolism. Because this relationship appeared relevant to many GRNs during many stages of *Drosophila* development, we sought



**Figure 4.4. The microRNA family is dispensable when energy metabolism is slowed.** (A) The pupal viability of various *dcr-1* nonsense mutants is fully rescued when IPCs are ablated in the mutants. (B) Adult viability of various *ago1* missense and nonsense mutants is rescued when IPCs are ablated in the mutants. (C) Representative *ago1* adults with normal or slowed metabolism. (D) Genetically mosaic individuals with *ago1*<sup>+</sup> bodies and *ago1*<sup>W894X</sup> mutant eyes. Left, representative individual with normal metabolism has almost no eye tissue (24/24 animals had this phenotype). Right, representative individual with slowed metabolism has rescued eye tissue. Of 70 such animals analyzed, 46 had this phenotype, 20 had normal eyes, and 4 had eyes that resembled the left animal. This is a statistically significant difference ( $p < 0.0001$ ; Chi square with Yates correction). Error bars, s.d. \*\*\*\*,  $p < 0.0001$ ; \*\*,  $p < 0.01$ ; \*,  $p < 0.05$ ; n.s.,  $p > 0.05$ .

to directly model the emergent properties of these systems rather than the specific regulatory mechanisms behind them. We therefore explored the mechanism using a general modeling framework premised on the progressive restriction of cell potential over time. Each step of restriction corresponds to a change in gene expression; gene products are synthesized, act, and then are eliminated until they are again needed in other cells (Fig. 4.5A). Temporally localized expression allows signaling molecules and transcription factors to be repeatedly used to build different body structures at different times. On the other hand, transient expression dynamics require GRN components to remain synchronized with the processes they govern.

If they fall out of rhythm, developmental errors occur. We therefore modeled the transient expression and regulation of a single gene within an intracellular cascade of developmental gene expression (Fig. 4.5A, red line).



**Figure 4.5. A dynamic model of gene regulation during cell fate determination.** (A) A program of gene expression occurs as a single cell passes through a series of developmental states. The model focuses on transient expression of a single gene within a cascade of gene expression. A state change is defined as the induction of gene expression by upstream gene products (input) and the action of the gene product (output). (B) Schematic representation of the response to a transient input, which can be either an extracellular or intracellular signal. Gene expression output is subject to layers of negative regulation acting at the gene, transcript, and protein levels. (C) Control representation of a single feedback loop as depicted in B. Boxes contain transfer functions, open circles indicate summation points, and closed circles indicate exclusive switches for each repressor. (D) Protein expression may be subject to layers multiple repressors acting in parallel.

#### 4.4.1. Mathematical model of gene expression

To model the dynamics of gene expression, we defined a linear time invariant system that describes the time evolution of activated DNA ( $D$ ), mRNA ( $R$ ), and protein ( $P$ ) in response to an inductive stimulus ( $I$ ). These state variables describe the extent of gene expression at

any point in time. Transitions between each of the variables' discrete states are governed by the set of psuedo-elementary reactions listed in Table 4.1.

**Table 4.1. Elementary steps of gene expression**

Reaction	Transition	Propensity	Param. [min <sup>-1</sup> ]
Gene activation	$\Delta D \rightarrow \Delta D + 1$	$k_1 \Delta I$	1
Transcription	$\Delta R \rightarrow \Delta R + 1$	$k_2 \Delta D$	1
Translation	$\Delta P \rightarrow \Delta P + 1$	$k_3 \Delta R$	1
Gene deactivation	$\Delta D \rightarrow \Delta D - 1$	$\gamma_1 \Delta D$	1
Transcript decay	$\Delta R \rightarrow \Delta R - 1$	$\gamma_2 \Delta R$	$10^{-2}$
Protein decay	$\Delta P \rightarrow \Delta P - 1$	$\gamma_3 \Delta P$	$10^{-3}$

GRNs use layers of negative regulation to attenuate expression of target genes (Fig. 4.5B). In our model, a system of regulatory components monitors the relative abundance of a target protein that dictates a cell fate transition. When the abundance of this protein increases, the regulatory components sense the increase in protein levels and act to down-regulate it, either at the level of gene transcription, mRNA stability, or protein stability (Fig. 4.5C). The particular mechanisms responsible for implementing regulation may remain unspecified. Instead, we abstract all modes of regulation using the independent and linear elementary reactions listed in Table 4.2. Multiple repressors often act in parallel to attenuate the expression of target genes. These control elements can be thought of as independent repressors working in parallel to bring the protein level back to a basal steady-state (Fig. 4.5D).

**Table 4.2. Elementary steps of gene regulation**

Reaction	Transition	Propensity	Param. [min <sup>-1</sup> ]
Transcriptional feedback	$\Delta D \rightarrow \Delta D - 1$	$\eta_1 \Delta P$	$5.0 \times 10^{-4}$
Feedback on mRNA	$\Delta R \rightarrow \Delta R - 1$	$\eta_2 \Delta P$	$1.0 \times 10^{-4}$
Feedback on protein	$\Delta P \rightarrow \Delta P - 1$	$\eta_3 \Delta P$	$5.0 \times 10^{-4}$

In the continuum limit, this model yields a deterministic system of ordinary first-order differential equations:

$$\begin{aligned}\frac{dD}{dt} &= k_1 I - \gamma_1 D - \sum_{i=1}^N \eta_i P \\ \frac{dR}{dt} &= k_2 D - \gamma_2 R - \sum_{i=1}^N \eta_i P \\ \frac{dP}{dt} &= k_3 R - \gamma_3 P - \sum_{i=1}^N \eta_i P\end{aligned}\tag{4.1}$$

where  $k_i$  are activation, transcription, or translation rate constants,  $\gamma_i$  are degradation constants,  $\eta_i$  are feedback strengths, and each species may be subject to  $N$  independent repressors. Equation 4.1 describes the time evolution of transcription rates, transcript abundance, and protein abundance. When these targets are induced by exogenous stimuli, their timely attenuation ensures that protein expression remains transient. The resultant dynamics resemble a simple pulse response. Our model therefore eschews molecular detail while preserving the salient features of gene expression dynamics that are relevant to developmental outcomes. This coarse-grained modeling strategy emphasizes an empirical view of GRN dynamics, and facilitates quantitative predictions at the organismal scale.

The model was not designed to capture the specific details of the various GRNs probed by our experiments. Instead, it provides a platform to survey the general principles that govern the dynamics of developmental processes. Namely, we developed a model that allowed us to ask how protein expression dynamics change when repressors are removed. In each of our model systems, protein expression is transient. Beginning at a basal steady state, expression is driven by upstream components of the developmental program. The ensuing expression dynamics can therefore be considered a response to perturbation.

#### 4.4.2. Model relation to control theory

Control theory provides a theoretical foundation underpinning the response of systems displaced from steady state. One of its core principles is the notion of local stability; that is, systems deviate linearly about a fixed point. Models based on this principle seek to describe how system output deviates from its steady state value in time. Describing our model with control terminology, protein level remains fixed about a basal steady state. Deviations from the basal level are driven by a transient disturbance. The disturbance induces activation of a gene, which induces transcription of mRNAs, which proceed to induce translation of protein. These three state variables are linearized about their steady state values:

$$\begin{aligned}\Delta D &= D - \lim_{t \rightarrow \infty} D(t) \\ \Delta R &= R - \lim_{t \rightarrow \infty} R(t) \\ \Delta P &= P - \lim_{t \rightarrow \infty} P(t)\end{aligned}\tag{4.2}$$

where the prefix  $\Delta$  signifies a deviation variable and the limit denotes the steady state concentration for a fixed level of the input signal.

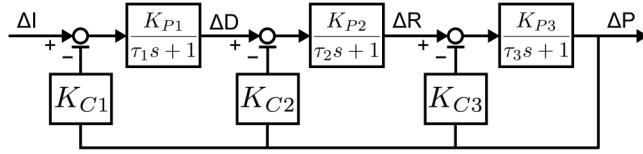
Protein levels relax back to steady state as the stimulus subsides. Control theory offers further insight when the relaxation process is mediated by one or more regulatory actors. In each of the experimentally surveyed GRNs, regulatory species such as microRNAs detected an increase in protein levels and acted to attenuate protein expression. These actors implement feedback control; they sense deviations in system output and exert an opposing response to drive the system back toward steady state. Neglecting their precise mechanisms of action, we can capture the influence of these *controllers* on system output with a single parameter, the feedback strength  $\eta_i$  for controller  $i$ . For simplicity we assume these regulatory mechanisms provide *proportional control*, meaning they modulate the deviations defined by

Equation 4.2 with a strength proportional to the output protein level. This *proportional only* scheme is incapable of input tracking and could not reject a sustained disturbance [71]. However, proportional only control provides an adequate representation of system dynamics because our model depicts an intermediate step in a cascade of developmental processes whose inputs and outputs are inherently localized in space and time (Fig. 4.5A).

When expressed in the Laplace frequency domain (see [282]), the system is readily described by three sequential first-order transfer functions:

$$\begin{aligned}\Delta D(s) &= \left( \frac{\frac{k_1}{\gamma_1}}{\frac{1}{\gamma_1}s + 1} \right) \left[ \Delta I(s) - \sum_{i=1}^N \frac{\eta_i}{k_i} \Delta P(s) \right] \\ \Delta R(s) &= \left( \frac{\frac{k_2}{\gamma_2}}{\frac{1}{\gamma_2}s + 1} \right) \left[ \Delta D(s) - \sum_{i=1}^N \frac{\eta_i}{k_i} \Delta P(s) \right] \\ \Delta P(s) &= \left( \frac{\frac{k_3}{\gamma_3}}{\frac{1}{\gamma_3}s + 1} \right) \left[ \Delta P(s) - \sum_{i=1}^N \frac{\eta_i}{k_i} \Delta P(s) \right]\end{aligned}\quad (4.3)$$

where the argument  $s$  is the complex frequency. Given this formulation, control theory provides a wealth of insight into the stability and dynamic character of pulsatile protein expression when the feedback strengths are varied. The transfer functions propagate deviations in the stimulus level (the input) to deviations in protein level (the output). A signal block diagram representation reveals that each regulatory process listed in Table 4.2 confers closed-loop control of protein levels following their induction (Fig. 4.6). The internal dynamics of this system dictate its sensitivity to perturbation. Restated symbolically, the magnitude and duration of protein expression are governed by the relative influence of the process and controller gains,  $K_{Pi} = k_i/\gamma_i$  and  $K_{Ci} = \eta_i/k_i$ . The contributions of each regulatory element  $\eta_i$  are additive, and may therefore be combined into a single controller gain for each point of actuation  $i$ . This formulation allows us to emulate the loss of a repressor simply by lowering the relevant controller gain.



**Figure 4.6. Block diagram depiction of the mathematical model.** Boxes contain transfer functions relating upstream and downstream variables, open circles indicate summation points. Transfer functions are expressed in the Laplace frequency domain.

#### 4.4.3. Simulation of developmental errors due to repressor loss

Protein expression follows a biphasic trajectory after reception of a transient stimulus (Fig. 4.7A,B, left panel). If there were no noise or variability, the protein level would be deterministic over time. However, protein dynamics vary because gene expression is noisy [283], something that can be captured in the model simulations by incorporating intrinsic noise.

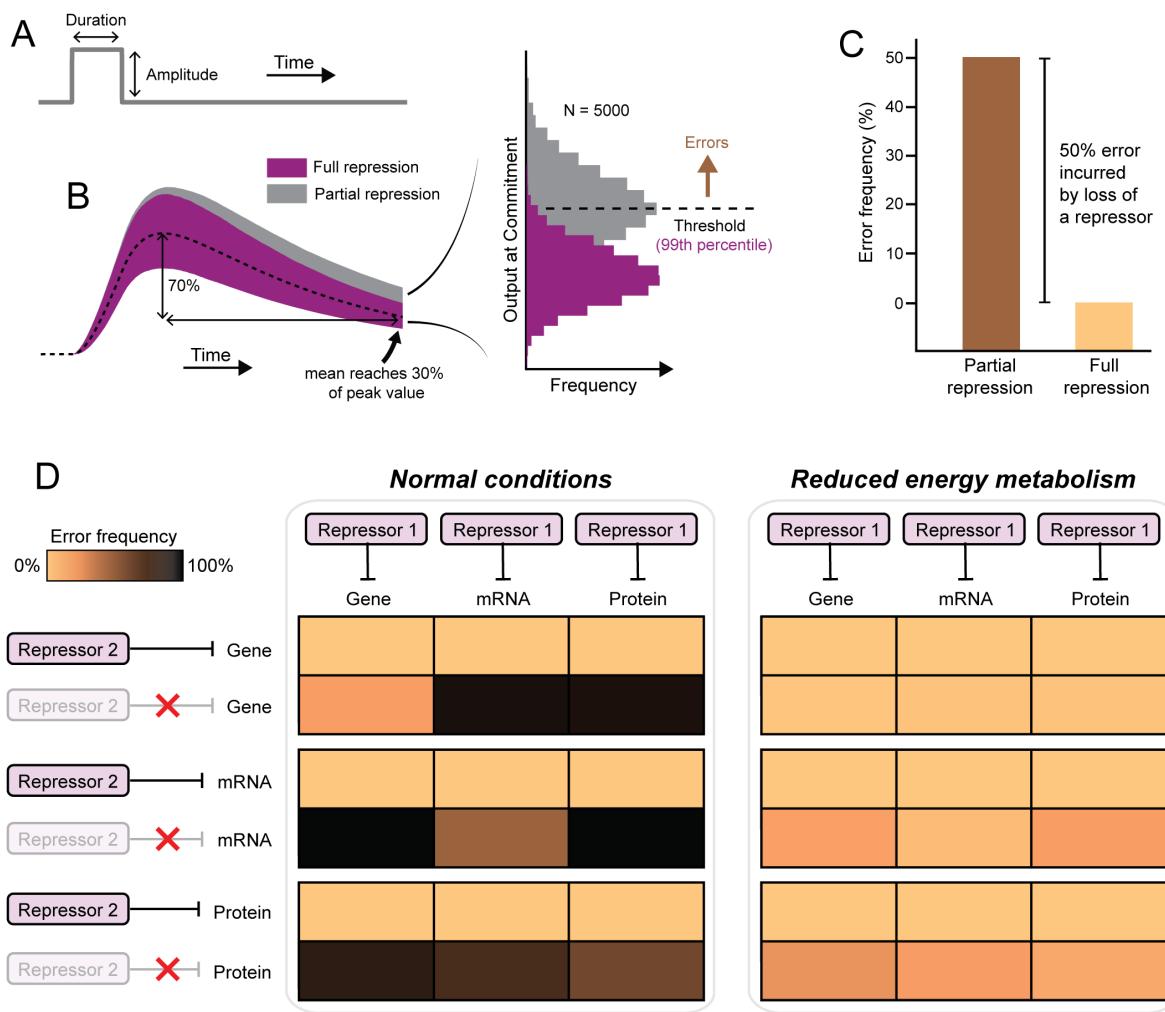
We then devised a scheme to relate protein expression dynamics to the likelihood of a successful developmental outcome. We define success as the ability of a GRN to attenuate protein expression in a timely manner, thus keeping pace with parallel components of the developmental program by triggering subsequent developmental events. We quantified errors in developmental outcome by defining a threshold that the output protein level must cross before a subsequent event can be triggered. Protein levels exceeding the threshold constitute errors in developmental outcome (Fig. 4.7B, right panel). Notably, such errors become more frequent when one repressor is removed (Fig. 4.7C). This property is observed over a broad range of parameter values, regardless of the manner in which repressors act, or the value at which the threshold is established (see Section 4.9).

The modeling framework allowed us to ask whether multiple layers of repression are less important for developmental outcome when energy metabolism is reduced. To answer this question, we halved the rate parameters of each ATP-utilizing reaction to reflect conditions of reduced energy metabolism (see Section B.7). Although ATP content remains fairly constant

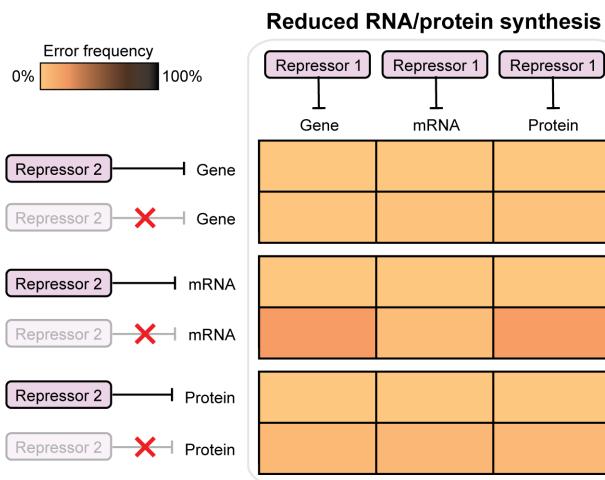
in cells facing limited respiration, the fluxes of ATP synthesis and turnover are affected, manifesting in altered ratios of ATP to ADP and free phosphate [284]. Anabolic processes such as protein synthesis are highly dependent on the ATP/ADP ratio [285]. When we halved ATP-dependent rate parameters and compared model results from full versus partial repression, we observed that error frequency in developmental outcome did not increase when a repressor was lost (Fig. 4.7D). This insensitivity to loss of a repressor persisted whether repression was transcriptional, post-transcriptional, or post-translational. The effect was observed across a wide range of model parameter values, irrespective of where the threshold was set, and regardless of whether a basal stimulus was present (see Section 4.9). In many cases the effect remained modestly apparent when the stimulus duration was extended to maintain comparable protein levels under conditions of reduced energy metabolism. In general, our modeling framework suggests that the frequency of developmental errors is less sensitive to changes in repression when energy metabolism is reduced.

Reduced glucose consumption by cells might not only limit ATP fluxes, but also hinder the synthesis of nucleotide and amino acid precursors required for RNA and protein synthesis. To simulate this scenario, we specifically reduced the rate parameters for RNA and protein production. Constraining these synthesis rates also suppressed the rise in error frequency when a repressor was lost (Fig. 4.8).

Overall, our modeling framework is fully consistent with our experimental observations that multiple layers of repression cease to be important for developmental success under conditions of reduced carbon and energy metabolism. Furthermore, the modeling framework suggests that the phenotype suppression phenomenon may be driven by differences in protein expression dynamics that are dependent on metabolic conditions.



**Figure 4.7. Simulated developmental errors are less frequent when metabolism is reduced.** (A-C) Simulated emergence of developmental errors. (A) A transient input signal drives (B) pulsatile protein expression dynamics. Simulations may be performed with two post-translational repressors in place (full repression), or with only one repressor in place (partial repression). Shaded regions correspond to the 98% confidence band of simulated protein trajectories. We define the commitment time as the time needed for 99% of simulations with full repression to cross a predefined threshold. With partial repression, the protein levels take longer to decay, so fewer simulations cross the threshold within the defined commitment time. We interpret each failure of a simulated protein level to decay below the threshold in time as a developmental error. (C) Error frequency is greater with partial repression. (D) The model suggests that errors will occur more frequently under partial repression regardless of how repressors act on gene expression (left panel). However, partial repression imparts fewer developmental errors when ATP-dependent parameter values are reduced by 50% (right panel).



**Figure 4.8. Partial repression imparts few developmental errors when biosynthesis is reduced.** Construction is equivalent to the right panel of Figure 4.7D. However, rather than fully reduced energy metabolism, only the RNA and protein synthesis rate parameter values were reduced by 50%.

#### 4.5. Protein expression dynamics after partial repressor loss

We quantified the extent to which the expression dynamics underlying developmental events were affected in our model. We first constructed a 98% confidence band around the set of trajectories simulated with the full complement of repressors (see Section B.6.2). This confidence band provides lower and upper bounds for the expected protein level. We then evaluated the fraction of trajectories,  $E(t)$ , simulated with one repressor missing that fell above or below the confidence band at each point in time. Averaging these values across the time course yields a single metric that reflects the extent to which protein dynamics are affected by the loss of a repressor:

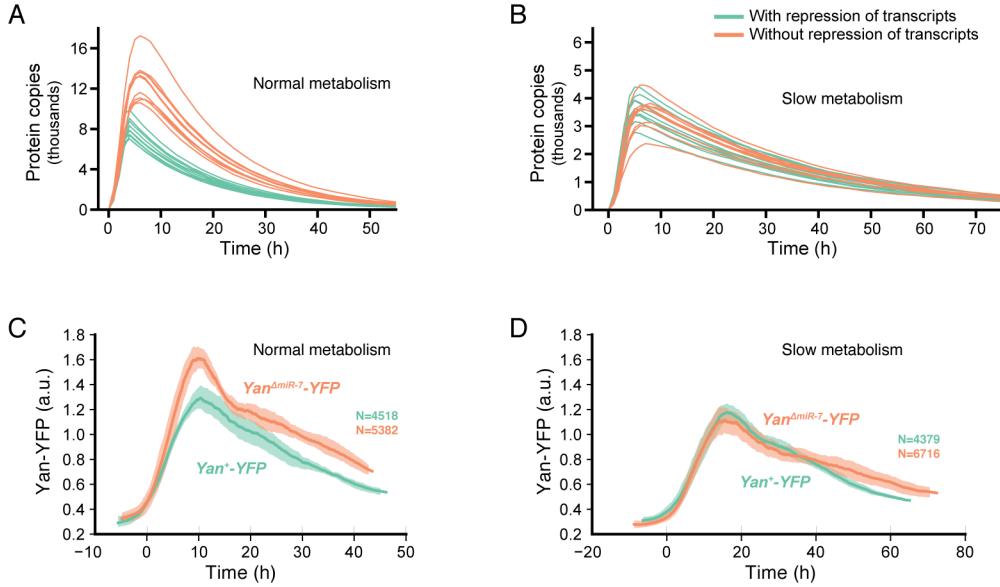
$$\text{Protein overexpression} \equiv \frac{1}{\tau} \int_0^\tau E(t) dt \quad (4.4)$$

We evaluated this metric for a scenario in which an auxiliary post-transcriptional repressor, akin to a microRNA, is lost. Using typical metabolic parameters, 78% of trajectories simulated without the post-transcriptional repressor exceed the confidence band generated

under full repression (Fig. 4.9A). This overexpression effect is highly robust to parameter variation in the model (see Section 4.9). When ATP-dependent parameters were halved, only 16% of trajectories exceeded the confidence band (Fig. 4.9B). The strong diminishment of overexpression under low metabolic conditions was also robust to extensive parameter variation. These results led us to predict that protein expression dynamics would be much less sensitive to repressor loss if we reduced metabolic rate.

We experimentally tested these predictions by measuring the expression dynamics of a key developmental regulatory protein, Yan. As shown in Chapter 3, Yan exhibits pulsatile dynamics in the larval eye disc, where its expression is induced by a morphogenetic furrow that traverses the eye disc. Eye disc cells located in the morphogenetic furrow rapidly upregulate Yan protein abundance, as quantified by a YFP-tagged version [43]. Yan levels then gradually decay back to initial conditions within these cells, thus exhibiting pulsatile dynamics. We compared Yan-YFP dynamics in eye disc cells from normally metabolizing larvae and larvae with ablated IPCs (Fig. 4.9C,D). The same pulsatile dynamics were observed in both, but the amplitude of the pulse was slightly reduced and the duration was extended when metabolism was slower. Similar trends were predicted *in silico* (Fig. 4.9A,B).

In the eye disc, Yan expression is repressed by the microRNA miR-7 [16]. There are four binding sites for miR-7 in the 3'UTR of *yan* mRNA, and their mutation causes de-repression of Yan output. We eliminated miR-7 repression of *Yan-YFP* by mutating the four binding sites in the 3'UTR of *Yan-YFP* mRNA to make *Yan<sup>ΔmiR-7</sup>-YFP*. In normally metabolizing eye discs, Yan-YFP protein made from the mutated gene pulsed with greater amplitude and showed impaired decay when compared to Yan-YFP made from the wildtype gene (Fig. 4.9C). These dynamics recapitulate the effect of repressor loss predicted by our model (Fig. 4.9A). In contrast, Yan-YFP made from the mutated gene showed similar



**Figure 4.9. Expression dynamics are resistant to repressor loss when energy metabolism is reduced.** (A,B) Simulated expression of target protein output when it is under control of an auxiliary post-transcriptional repressor (green) or not under control of the repressor (orange). All simulations (green and orange) are also under control of a constitutive repressor. Shown are ten randomly-chosen samples from a total population of 5000 trajectories for each condition. (A) Simulations performed with normal ATP-dependent reaction rates. (B) Simulations performed following a 50% reduction in the rate of ATP-dependent reactions. (C,D) Yan-YFP protein dynamics in eye disc progenitor cells. Time 0 marks the time at which Yan-YFP induction occurs. Solid lines are moving line averages. Shaded regions denote 95% confidence intervals. Each line average is calculated from a composite of measurements of between 4,379 and 6,716 cells. (C) Yan-YFP dynamics for wildtype *Yan-YFP* and mutant *Yan<sup>ΔmiR-7</sup>-YFP* eyes under normal metabolic conditions. (D) Yan-YFP dynamics for wildtype and mutant genes when the IPCs have been ablated.

dynamics to protein made from the wildtype gene when metabolism was slowed (Fig. 4.9D). This behavior clearly resembled the simulated dynamics under conditions of reduced energy metabolism (Fig. 4.9B).

These measurements demonstrate that miR-7 has little to no impact on Yan expression dynamics when metabolism is slowed, and are consistent with the observed suppression of developmental errors when the same repressor is lost in the eye. The breadth of our model predictions further suggests that these effects are generalizable to other genes and repressors.

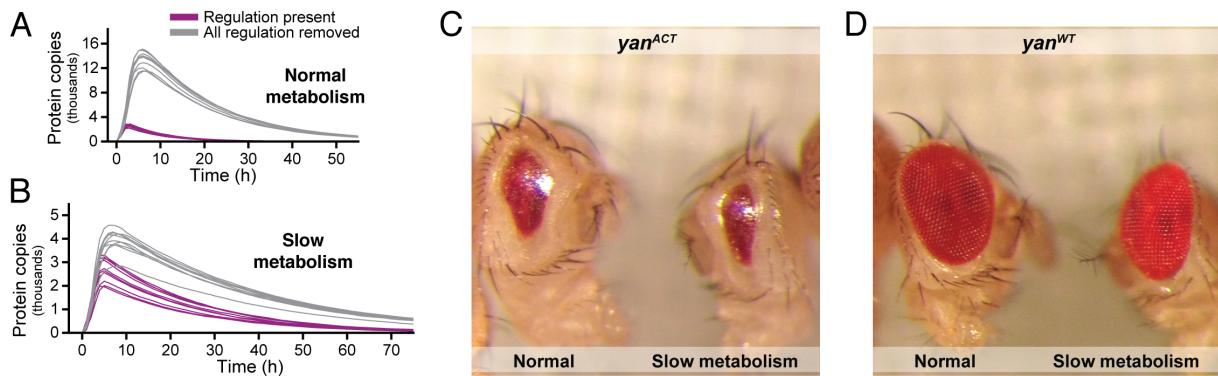
#### 4.6. Effect of full repression loss

Our modeling framework is consistent with the hypothesis that multiple weak repressors allow GRN dynamics to faithfully couple to variable energy metabolism, with fewer repressors required when metabolic conditions are reduced. We then asked whether repression is needed at all under such conditions. We studied a model with a full complement of negative control elements and compared the results to a scenario in which all control elements were removed (Fig. 4.10A). Error frequencies approached 100% under normal growth conditions. While expression dynamics were visibly less affected by repressor loss when ATP-dependent parameters were reduced, the error frequency remained very high (Fig. 4.10B). These results suggest that there are limits to the severity of perturbations for which reductions in energy metabolism can compensate, and reducing energy metabolism does not eliminate the need for gene repression altogether.

To test this prediction, we expressed in the eye a *yan* mutant transgene that is insensitive to all known repression of *yan* transcription, mRNA stability and protein stability [82]. The *Yan<sup>ACT</sup>* mutant adults had severely disrupted compound eye patterning (Fig. 4.10C). This mutant eye phenotype was not suppressed by ablation of the animals' IPC cells. Wildtype *yan* transgenic adults with normal eye patterning were also unaffected by IPC ablation (Fig. 4.10D).

#### 4.7. Limiting protein synthesis reduces the need for repressors

The coupling of developmental dynamics to time can be explored with other aspects of metabolism. In particular, protein synthesis is an important determinant of rates of growth and development [286]. We used our modeling framework to investigate the impact of a twofold reduction in overall protein synthesis rate on GRN dynamics. The model suggests

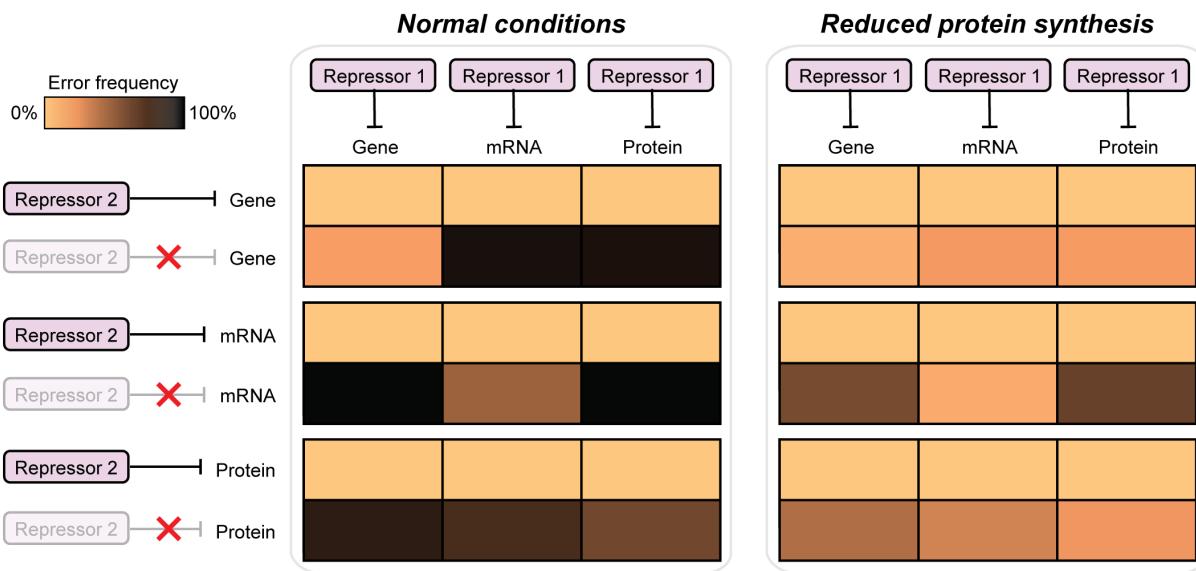


**Figure 4.10. Reduced energy metabolism cannot compensate for complete loss of repression.** (A,B) Simulated expression of protein output both with (purple) and without (grey) any repression of the target gene. Shown are ten randomly chosen samples from a total population of 5000 trajectories for each condition. Error frequencies exceed 99% irrespective of metabolic conditions. (A) Simulations performed under normal conditions. (B) Simulations performed following a 50% reduction in the rate of ATP-dependent reactions. (C) Loss of eye tissue in a *yan<sup>ACT</sup>* mutant is not suppressed by slower metabolism. Representative individuals were taken from  $N > 100$  individuals for each condition. (D) Eye patterning in a *yan<sup>WT</sup>* control is not affected by slower metabolism. Representative individuals were taken from  $N > 100$  individuals for each condition.

that expression dynamics are less affected and fewer developmental errors are incurred by loss of a repressor when protein synthesis rates are reduced (Fig. 4.11). We again found this effect is robust to a wide range of parameter values and model assumptions (see Section 4.9).

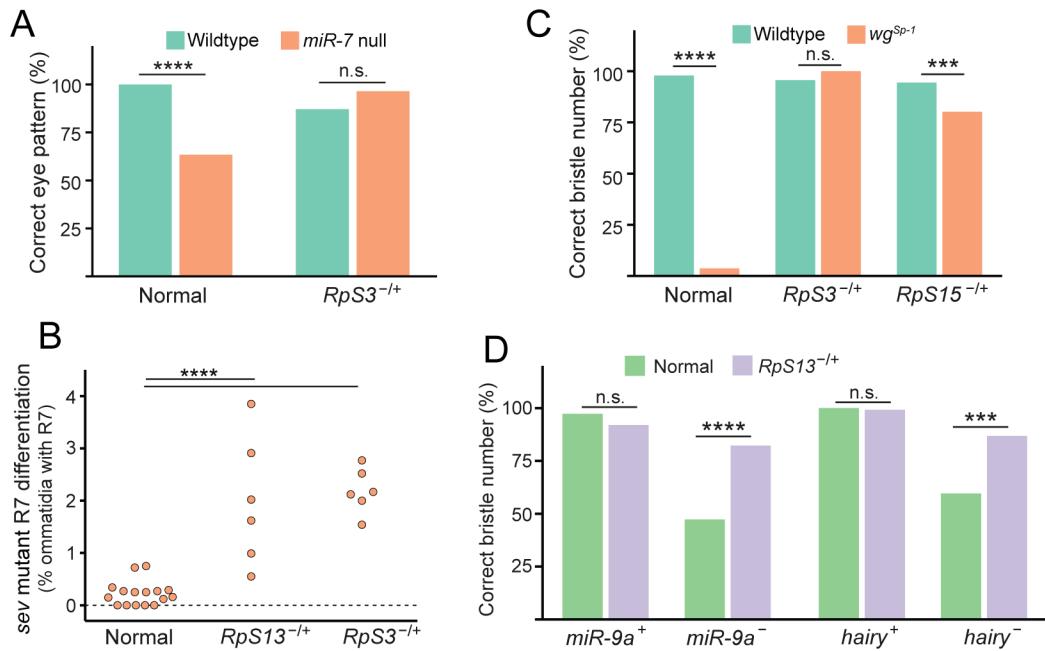
We tested this model prediction by genetically reducing the abundance of cytobiosomes in all cells in *Drosophila*. We made use of loss-of-function mutations in genes encoding various ribosomal proteins (RPs), which cause the “Minute” syndrome of dominant, haploinsufficient phenotypes, including slower growth and development [287, 288]. Heterozygous *RP* mutants reduce the number of ribosomes per cell by approximately 50%, and a total of 64 *RP* genes exhibit a Minute syndrome when mutated. We selected a subset of these genes to reduce ribosome number.

We combined heterozygous *RP* mutants with the repressor mutations we had previously studied. In all cases, the *RP* mutants suppressed the developmental phenotypes of mutations in *wg*, *miR-7*, *sev*, *hairy*, and *miR-9a* (Fig. 4.12). This error frequency suppression was precisely the result predicted by our modeling.



**Figure 4.11. Simulated developmental errors are less frequent when protein synthesis is reduced.** The model predicts increased frequency of error with partial repression regardless of how auxiliary repressors act on gene expression (left panel is copied from 4.7D). However, partial repression induces fewer errors when protein synthesis-dependent parameter values are reduced by 50% (right panel).

We also tested whether expression dynamics are affected by repressor loss under limiting translation conditions. The Sens protein is transiently expressed in proneural cells during selection of sensory bristle fates in the imaginal wing disc [289]. Bordering the presumptive wing margin, stripes of proneural cells express Sens protein over a spectrum of levels, reflecting heterogeneity in Wg and Notch regulation of its expression [273, 290]. Moreover, miR-9a weakly represses *sens* expression in these cells [276]. We recombined a 19 kb *sens* transgene, amino-terminally tagged with superfold GFP (sfGFP), that functionally replaced the endogenous *sens* gene [94, 291]. Quantitative measurement of sfGFP fluorescence in individual proneural cells yielded the expected distribution of *sens* expression (Fig. 4.13A). We compared this distribution to one derived from individuals expressing a mutated sfGFP-*sens* transgene in which its miR-9a binding sites had been mutated [94]. Mutation of the miR-9a binding sites in *sfGFP-sens* shifted the fluorescence distribution, and resulted in an

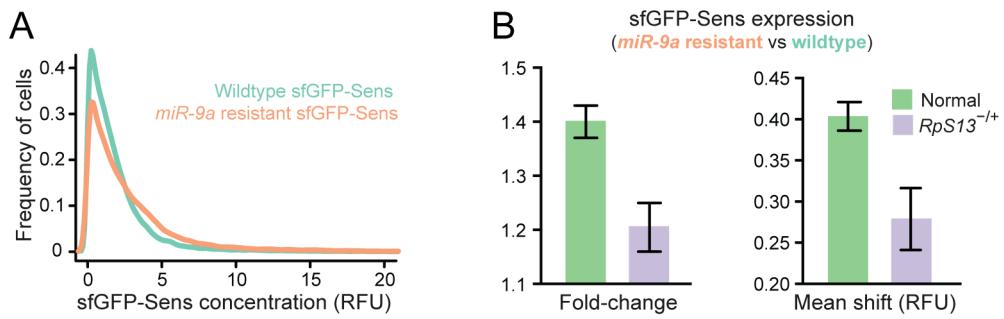


**Figure 4.12. Reducing ribosome number rescues sensory organ development when repressors are lost.** (A) Loss of *miR-7* does not cause adult eye mispatterning when *RpS3* is heterozygous mutant. (B) *sev* mutants have more R7-positive ommatidia when either *RpS3* or *RpS13* are heterozygous mutant. Each datapoint represents one eye sample, and between 481 and 837 ommatidia were scored for R7 cells within each eye sample. (C) *wg<sup>Sp1</sup>* heterozygous individuals that are also heterozygous mutant for different *RpS* genes have sternopleural bristle numbers more similar to wildtype. (D) Developmental accuracy is recovered for both *miR-9a* and *hairy* mutants that are also heterozygous mutant for *RpS13*. For all panels in B-E, error bars, s.d. \*\*\*\*,  $p < 0.0001$ ; \*\*\*,  $p < 0.001$ ; n.s.,  $p > 0.05$ .

average 1.45-fold increase in sfGFP-Sens levels (Fig. 4.13B). We then tested the effects of miR-9a on *sfGFP-sens* expression when the *RpS13* gene was heterozygous mutant. Strikingly, loss of miR-9a regulation had less effect on sfGFP-Sens protein levels when ribosome numbers were reduced (Fig. 4.13B). This behavior clearly resembled the effect predicted by the modeling framework.

#### 4.8. Implications for the evolution of GRNs

Growth and development are fueled by metabolism. This means that the tempo of development depends on metabolic rate. Thus, the dynamics of developmental gene expression



**Figure 4.13. Reducing ribosome number diminishes *sfGFP-sens* overexpression when miR-9a repression is lost.** (A) Frequency distribution of sfGFP-Sens protein level in cells bordering the wing margin of white prepupal wing discs. Shown are distributions of cells expressing either wildtype *sfGFP-sens* or *sfGFP-sens* in which miR-9a binding sites have been mutated. Each group represents > 15,000 cells. (B) *sfGFP-sens* overexpression caused by miR-9a binding site mutations in *RpS13* wildtype (green) and heterozygous mutant backgrounds (violet). Left panel shows median fold-change. Right panel shows the shift in the fluorescence distribution of sfGFP-Sens-positive cells as determined by a Mann-Whitney-Wilcoxon test. Error bars denote 95% confidence intervals. Overexpression is attenuated in the *RpS13* heterozygous mutant background.

must faithfully adjust to a variable time scale. We have shown that multi-layered weak repression within GRNs plays an unexpected function in synchronizing gene expression dynamics with the variable pace of the developmental program. Multiple repressors are required for accelerated development when metabolism is high, and they become functionally redundant when metabolism is low. Multiple repressors therefore allow for reliable development across a broader range of metabolic conditions than would otherwise be tolerated.

Our model explains long-standing observations linking nutrient limitation to suppression of mutant phenotypes [266, 267]. Presumably, such mutations cripple regulatory genes acting on developmental GRNs. Our model might also offer an explanation as to why animals that undergo above-normal growth exhibit compromised development [246, 248]. Wildtype GRNs might function across a limited range of metabolism, with functionality breaking down when metabolism exceeds that range.

Our varied analyses suggest that this relationship between metabolism and repression is ubiquitous. We found that the entire family of 466 microRNAs in *Drosophila melanogaster*

can become functionally dispensable when energy metabolism is slowed. The extensive literature on microRNA function in *Drosophila* implicates them in practically all facets of the fruit fly's life [292, 293]. Various explanations have been provided for why this family of weak repressors has flourished in the animal kingdom, chief among them the idea that they act as buffers for gene expression [96]. We now posit that microRNAs also provide broad and flexible coupling of many developmental processes to variable timescales resulting from fluctuations in metabolism.

There is an alternative mechanism to explain phenotype suppression by reduced metabolism. This mechanism relies on a steady-state and not dynamical perspective of gene expression. Genome-wide gene expression patterns could conceivably change with organismal growth rate. This is the case for chemostat-grown yeast cells, where the expression of 27% of all genes correlates with growth rate [294]. Most genes associated with stress response are overexpressed when cells grow at a slow rate [294, 295]. Such differential gene expression could globally modulate dynamical processes such as protein folding and turnover, among others, and thereby attenuate phenotypes of genetic mutations. Abundance of molecular chaperones has been found to affect the penetrance of diverse gene mutations in *C. elegans* and *Drosophila* [104, 296]. However, these global effects do not explain why gene expression dynamics are conditionally dependent upon mutations in regulatory genes. We found that repression of Yan and Sens dynamics by microRNAs become more redundant when metabolic rates are slowed.

Metabolic rate increases exponentially with temperature as described by the Arrhenius equation [245], resulting in an indirect temperature dependence of developmental tempo [258]. Temperature also directly affects the rates of reactions within developmental GRNs [245], yet developmental outcomes are generally robust to fluctuations in temperature across

a limited range. Various molecular mechanisms have been invoked to explain this robustness. These include chaperones that create large protein-folding reservoirs [104, 297], and regulatory circuits within interaction networks [95]. Our model suggests a complementary mechanism for developmental robustness against temperature variation. By coupling gene expression dynamics with metabolism, weak repressors might neutralize the metabolic effects of temperature on developmental tempo. Indeed, loss of miR-9a regulation is less impactful on sensory organ development if the growth temperature is lowered [94]. Likewise, raising animals under lowered temperatures can suppress the phenotypes of mutations that are not classical *ts* alleles [269–272].

Metabolic conditions drive variation about the intrinsic developmental tempo of each species. We have shown that layered weak repression within GRNs enables these fluctuations to occur without causing developmental errors. Metabolic conditions change in both space and time. Perhaps the selective advantage of a reliable developmental outcome amidst variable environmental conditions is a driving force in the evolution of gene regulatory networks.

#### **4.9. Robustness of modeling results**

We conducted numerous parameter sweeps to confirm the robustness of each result presented in this chapter. In each sweep, all model parameters were varied across a ten-fold range ( $\pm \sim$ three-fold). To efficiently sample the parameter space, we quasi-randomly sampled 2500 parameter sets from the multi-dimensional hyperspace defined by one order of magnitude variation in each of the model parameters. For each parameter set we independently ran six sets of five thousand simulations:

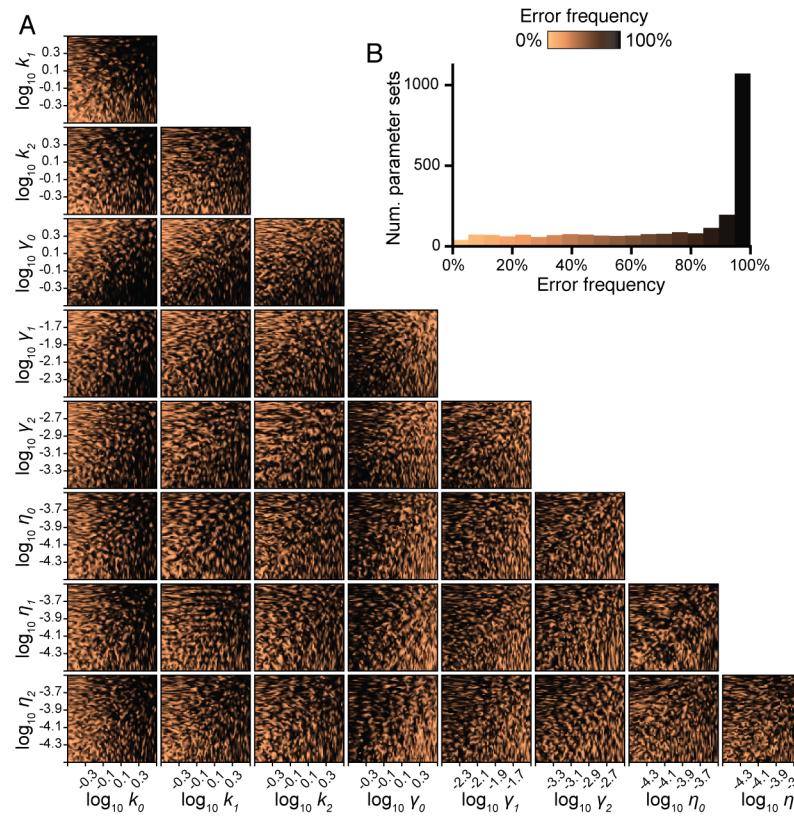
- Full feedback with normal metabolism and translation
- Partial feedback with normal metabolism and translation

- Full feedback with reduced energy metabolism
- Partial feedback with reduced energy metabolism
- Full feedback with reduced protein synthesis
- Partial feedback with reduced protein synthesis

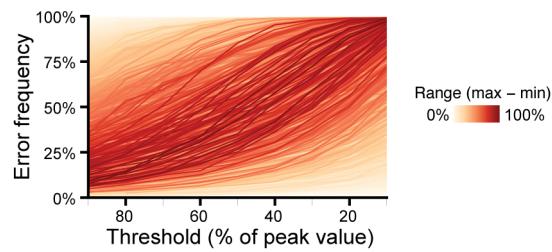
Full-repression systems were assigned two copies of each feedback element present in the corresponding partial-repression system. Error frequencies were evaluated as described previously. In total, this procedure constitutes one parameter sweep.

Our parameter sweeps sampled an  $N$ -dimensional space, prompting us to seek a lower dimensional representation in order to visualize and discern any characteristic trends. One option is to project the results of all simulations onto each of the  $N(N - 1)/2$  orthogonal 2-D planes. Figure 4.14A demonstrates this approach for error frequencies simulated with partial feedback and normal metabolism. While it clearly shows that error frequency is greater than 1% for most combinations of parameter values (lots of dark spots), we find that the complexity of the 2-D visualization offers little additional insight. Instead, for all remaining parameter sweeps we opt for the simpler one dimensional projection of parameter sweep results, as illustrated by Figure 4.14B. The histogram simply conveys the global trend in simulated error frequency, which in this case is skewed heavily toward large error frequencies, indicating that partial loss of repression induces an increase in error frequency across a broad parameter range.

We also varied the level of the success threshold, and recalculated all error frequencies accordingly. Error frequency is greater than 1% for almost all definitions of the success threshold, indicating that loss of a repressor increases developmental error irrespective of where the success threshold is set (Fig. 4.15).



**Figure 4.14. Error frequencies are broadly increased when an auxiliary repressor is lost.** (A) Simulated error frequencies are projected onto two dimensional planes and then linearly interpolated onto  $(100 \text{ px})^2$  grids. (B) Binned counts of the 2500 simulated error frequencies depicted in A. The majority of parameter sets yield high error frequencies.



**Figure 4.15. Error frequencies increase when a repressor is lost irrespective of where the success threshold is set.** Error frequencies for parameter sets from Fig. 4.14 were re-calculated across a range of different success thresholds. Thresholds are defined by the 99<sup>th</sup> percentile of protein levels simulated with all repressors, and are evaluated at the time when the mean protein level with all repressors reaches the indicated fraction of its maximum value. Each line represents one of the parameter sets from Figure 4.14. Color scale reflects maximum difference in error frequency across the range of thresholds tested.

We also surveyed how error frequency changes when metabolism is reduced. For the control theoretic model, error frequencies simulated across the parameter space were

heavily skewed toward lower values (Fig. 4.16A). For each parameter set, we then computed the difference in simulated error frequency between conditions of normal and reduced energy metabolism. The one-dimensional representation clearly shows that simulated error frequency decreases when metabolism is reduced for the vast majority of the sampled parameter space (Fig. 4.16B). Similarly, simulated error frequency decreases when metabolism is reduced regardless of where the threshold is set (Fig. 4.16C).

Our conclusion also persists when a nonzero basal stimulus is introduced. We conducted an additional parameter sweep in which the stimulus consists of a transient step change between input values of  $\Delta I = 0.1$  and  $\Delta I = 1.0$ . Simulations were carried out on an absolute basis, and were allowed sufficient time to reach a non-zero steady state before and after the stimulus was applied. The resultant protein level trajectories for each of the six sets of simulations were converted to deviation form by subtracting the respective population-wide mean final value. Error frequencies were then evaluated as previously described. Despite the inclusion of a nonzero basal stimulus, error frequencies remained broadly suppressed under conditions of reduced energy metabolism (Fig. 4.16C).

All preceding simulations assume the stimulus (input) is a unit step that persists for three hours regardless of metabolic conditions. Alternatively, metabolic conditions might affect stimulus (input) duration, particularly if the upstream processes responsible for the input are also governed by metabolically delayed processes. We find that the general prediction made by our model – that reduced energy metabolism and reduced protein synthesis limit sensitivity to loss of regulation – persists in roughly half of cases if we apply a four-fold extension of input duration when energy metabolism is reduced (Fig. 4.16E). Notably, in many cases scaling the input duration with metabolic condition yields the opposite effect. However, these instances correspond to simulations in which the extended stimulus yields

output protein levels greater than those observed under normal metabolic conditions, suggesting that a four-fold increase in stimulus duration may be excessive. Nevertheless, we take this to be an upper bound on the observed phenotype suppression phenomenon.

Our control theoretic modeling framework suffers two notable limitations. First, the number of transcriptionally active sites within a cell is limited by gene copy number, but the activated-DNA state in our initial linear model was unbounded. To test whether error frequency suppression persists when an upper bound on gene activity is introduced, we considered a simple two-state transcription model:

$$\begin{aligned} \frac{dG_{on}}{dt} &= k_G G_{off} I - \gamma_G G_{on} - \sum^{N_g} \eta_G G_{on} P \\ \frac{dG_{off}}{dt} &= -\frac{dG_{on}}{dt} \\ \frac{dR}{dt} &= k_R G_{on} - \gamma_R R - \sum^{N_r} \eta_R P \\ \frac{dP}{dt} &= k_P R - \gamma_P P - \sum^{N_p} \eta_P P \end{aligned} \tag{4.5}$$

where  $G_{on}$  and  $G_{off}$  are the on- and off- states of a gene;  $I$ ,  $R$ , and  $P$  are the input, transcript, and protein levels;  $k_i$ ,  $\gamma_i$ , and  $\eta_i$  are the synthesis, decay, and feedback rate constants for species  $i$ ; and  $N_g$ ,  $N_r$ , and  $N_p$  are the number of transcriptional, post-transcriptional, and post-translational repressors, respectively. We performed another parameter sweep varying each of the model's nine parameters across one order of magnitude. All simulations were initialized as diploid ( $G_{off} = 2$ ) then subject to a constant 3 h stimulus before reverting to a basal level of zero gene expression. Despite the limitation placed on gene activity, error frequency remained elevated under normal growth conditions and broadly suppressed when ATP-dependent rate parameters were reduced (Fig. 4.16F).

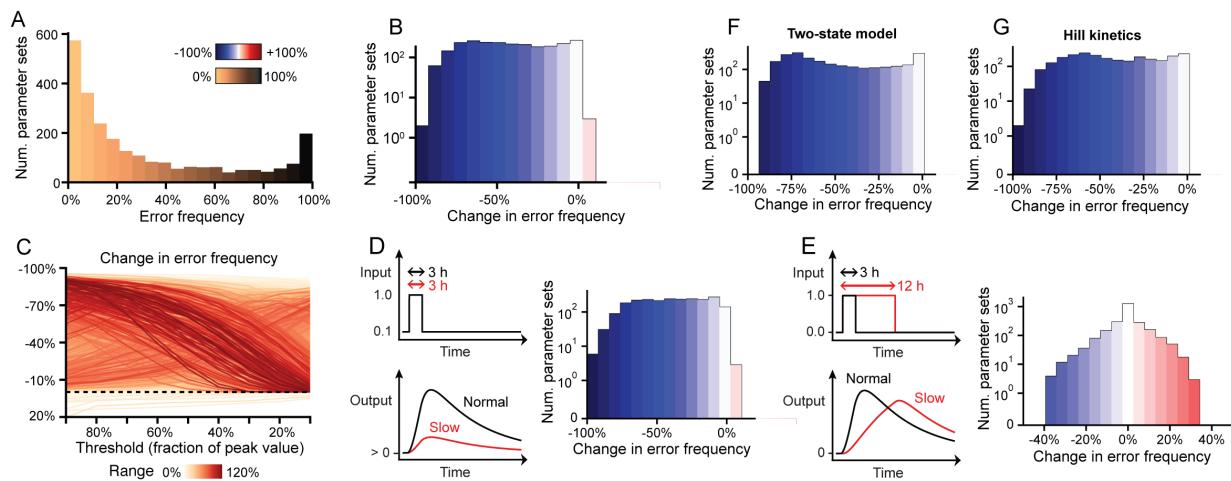
Second, gene expression models frequently utilize cooperative kinetics in order to reproduce the nonlinearities and thresholds encountered in transcriptional regulation. We captured these dynamics by reformulating our model in terms of Hill kinetics:

$$\begin{aligned} \frac{dR}{dt} &= \frac{k_R}{1 + (\frac{1}{2I})^H} \prod_{i=1}^{N_g} \left[ \frac{1}{1 + (\frac{P}{K_r})^{H_r}} \right] - \gamma_R R - \sum_{i=1}^{N_r} \eta_i P \\ \frac{dP}{dt} &= k_P R - \gamma_P P - \sum_{i=1}^{N_p} \eta_i P \end{aligned} \quad (4.6)$$

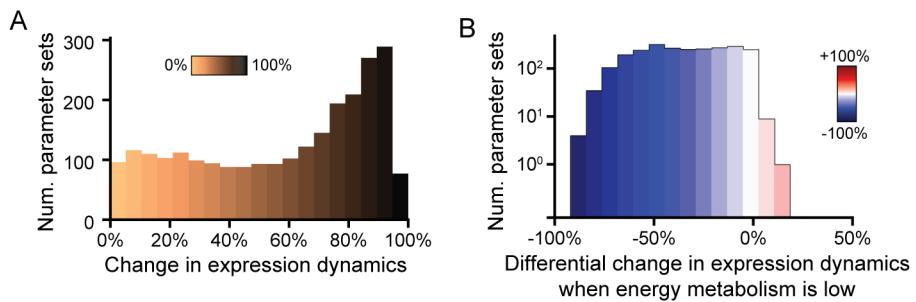
where  $I$ ,  $R$ , and  $P$  are the input, transcript, and protein levels;  $k_i$ ,  $\gamma_i$ , and  $\eta_i$  are the synthesis, decay, and linear feedback rate constants for species  $i$ ;  $N_r$  and  $N_p$  are the number of post-transcriptional, and post-translational linear repressors;  $H$  is a transcriptional Hill coefficient; and  $K_r$  and  $H_r$  are the half-maximal occupancy level and Hill coefficient of each of the  $N_g$  transcriptional repressors. The stimulus level corresponding to half-maximal transcription rate was fixed at 0.5 because we only consider a binary input signal. Despite the incorporation of cooperative kinetics, error frequencies remained elevated under normal conditions and broadly suppressed when ATP-dependent rate parameters are reduced (Fig. 4.16G).

Our simulations revealed that protein expression is less sensitive to repressor loss when metabolic conditions are reduced (Fig. B.2). We computed the same protein overexpression metric for each set of simulations in our parameter sweep (Fig. B.2). Under normal metabolic conditions, we find that protein levels generally increase when a repressor is lost (Fig. 4.17A), but this effect is broadly diminished when metabolism is reduced (Fig. 4.17B).

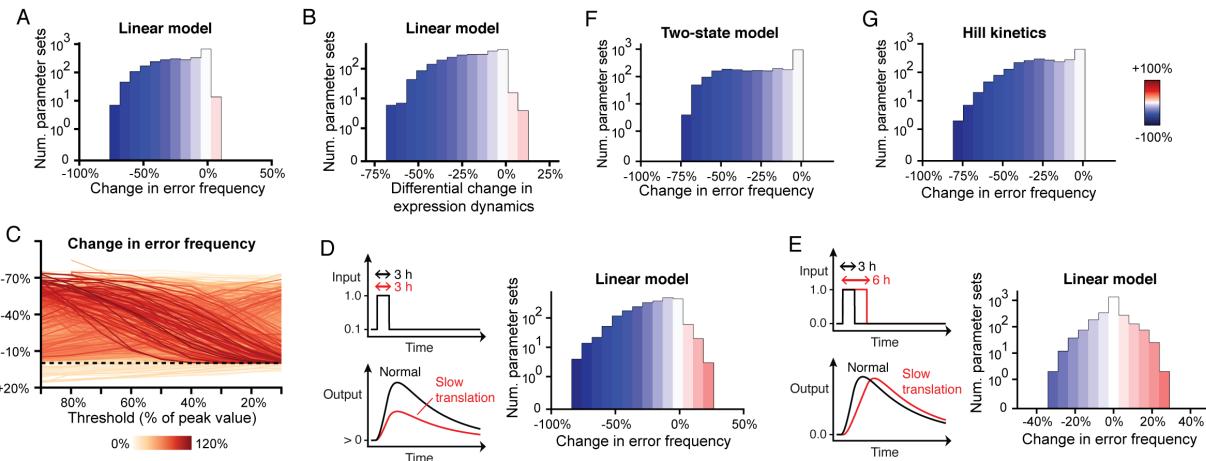
Finally, we repeated all of the above robustness checks for the scenario in which protein synthesis rates are reduced (Fig. 4.18C). The results persist in all cases, leading us to conclude that there is a general trend of decreased error frequency with partial feedback under both reduced energy metabolism and reduced protein synthesis conditions.



**Figure 4.16. Reduced energy metabolism diminishes the importance of auxiliary repressors over a wide range of model conditions.** (A-G) For each panel, 2500 simulations were performed with parameter sets quasi-randomly sampled from the nine-dimensional hyperspace defined by one order of magnitude variation in each of the respective model parameters, as done for Fig. 4.14. For each parameter set, error frequencies pertain to 50% loss of repression mimicking partial repressor loss. (A) One-dimensional representation of the error frequencies for all parameter sets under conditions of normal or diminished energy metabolism. (B) Change in error frequencies with diminished metabolism relative to normal metabolic conditions for all parameter sets. Blue-Red color scale corresponds to the difference in error frequency between low-metabolic and normal conditions, e.g. blue indicates error suppression by reduced energy metabolism. (C) Results from (B) re-calculated across a range of different success thresholds. Each line corresponds to a single parameter set. Color scale reflects maximum change in error frequency across the threshold range. Black dashed line corresponds to unchanged error frequency by reduced energy metabolism. The vast majority of simulations exhibit some reduction in error frequency across all thresholds. (D-G) Systematic modification of model conditions showing the change in error frequencies with diminished metabolism relative to normal metabolic conditions for all parameter sets. Blue-Red color scale corresponds to the difference in error frequency between low-metabolic and normal conditions, e.g. blue indicates error suppression by reduced energy metabolism. (D) Simulations where a nonzero basal stimulus is applied. (E) Simulations where input duration is increased four-fold by a reduction in energy metabolism. (F) Simulations when an upper bound is placed on the number of sites firing transcription. (G) Simulations when cooperative transcription kinetics are considered.



**Figure 4.17. Reductions in energy metabolism limit the extent to which protein expression dynamics are affected by loss of a repressor.** (A) Percent overexpression caused by loss of a repressor for model simulations performed with 2500 independent parameter sets. Color scale reflects the strength of overexpression. Overexpression is large for most parameter sets. (B) Percent overexpression caused by loss of a repressor was calculated for simulations implementing normal energy metabolism and reduced energy metabolism. The difference in percent overexpression between the two metabolic conditions is shown for model simulations performed with 2500 independent parameter sets. Color scale reflects the difference. The majority of simulations are blue, indicating that expression dynamics are less affected by repressor loss when energy metabolism is low.



**Figure 4.18. Reduced protein synthesis capacity diminishes the importance of auxiliary repression over a wide range of model conditions.** Each panel depicts a parameter sweep of the nine-dimensional hyperspace defined by one order of magnitude variation in each of the respective model parameters. For each parameter set, error frequency and percent overexpression were calculated as previously described. They pertain to 50% reduced repression mimicking auxiliary repressor loss. Error frequency and percent overexpression were calculated independently for conditions of normal and reduced protein synthesis. The difference in error frequency or overexpression between metabolic conditions are shown color-coded, e.g. blue indicates error suppression by reduced protein synthesis. (A-C) Simulations in which the duration of the stimulus input is constant. Shown are the (A) differential error frequencies and (B) differential changes in expression dynamics relative to normal protein synthesis conditions. (C) Differential error frequencies for varying definitions of the success threshold. Each line represents one parameter set, colored by the corresponding range of differential error frequencies. (D) Simulations where a nonzero basal stimulus is applied. (E) Simulations where input duration is increased two-fold by a reduction in protein synthesis capacity. (F) Simulations when an upper bound is placed on the number of sites firing transcription. (G) Simulations when cooperative transcription kinetics are considered.

## CHAPTER 5

### Conclusions

The preceding chapters are unified by their pursuit of a common question: how do the structure and function of gene regulatory networks control cell fate decisions and yield emergent properties at the organismal scale? Each chapter combined quantitative data and mathematical analysis to address subtle aspects of this question in a unique way. The following sections provide an introspective summary of how each chapters findings, and the lessons learned therein, impact the present understanding and future research directions of quantitative, developmental, and systems biology.

#### 5.1. Implications for developmental and systems biology

Chapter 2 introduced a computational framework for automated analysis of genetic mosaics; a class of experiments commonly used to probe cell fate decisions *in situ* [19, 135]. The framework combines computer vision and statistics to automate the labor-intensive steps of a quantitative work-flow, enabling automated and systematic comparison of cells subject to control and perturbation conditions in an otherwise equivalent background.

Quantitative mosaic analysis is not new, nor is it uncommon in high profile publications [35, 136, 139]. Yet, these studies deploy an irreproducible mix of ad hoc implementations and costly commercial software. Worse still, qualitative analysis pervades less prominent corners of the literature. Contributing a fully automated framework to the open-source ecosystem will make quantitative mosaic analysis equally accessible to the entire community. A unified

framework will also dramatically simplify the reproduction of existing analyses. Chapter 2 therefore advances the quantification of developmental biology by adding potential for rigor and reproducibility where they are currently lacking.

Chapter 3 explored a novel cell fate decision mechanism underlying photoreceptor specification in the *Drosophila* larval eye. Computer vision techniques were used to extract quantitative measurements of Pnt and Yan dynamics from a wealth of confocal microscopy data. Statistical analysis revealed that differentiation is driven by dynamic changes in the ratio between Pnt and Yan, and is agnostic to changes in their absolute concentrations as long as the ratio remains constant. The data therefore provide the first direct evidence that cell fate decisions can be triggered by changes in the relative abundance of separate transcription factors. This finding rebukes the canonical model of photoreceptor specification [83]. More importantly, it adds a new dimension to our understanding of how multiple transcription factors cooperatively control cell fate decisions, with broad implications for many developmental contexts within and beyond *Drosophila*.

The ratiometric sensing mechanism identified in Chapter 3 adds to a growing body of evidence that cells are able to sense relative changes in the abundance of GRN components [50, 187]. These discoveries are exciting because relative sensing could potentially isolate cell fate decisions from environmental sources of variation. This is because environmental fluctuations would likely manifest as correlated extrinsic noise that affects all GRN components in a similar manner, causing absolute but not relative concentrations to vary between cells. Relative sensing might then increase fitness in variable environments.

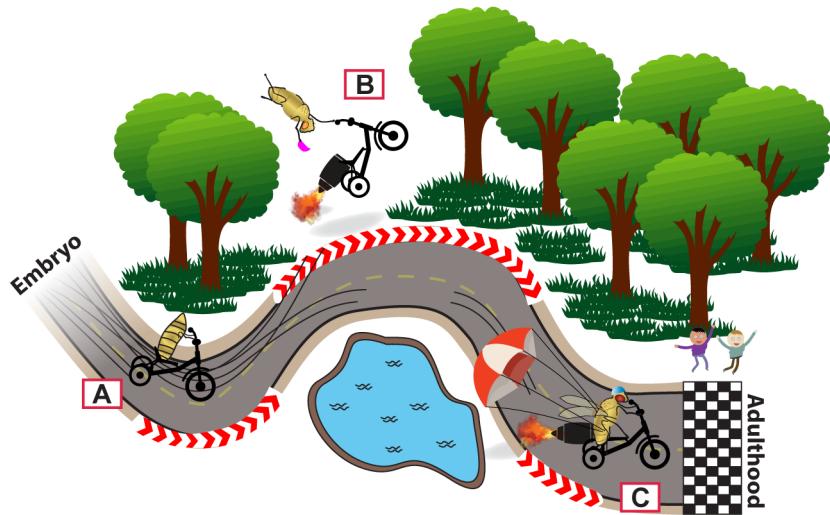
Regulatory interactions may provide additional layers of stability. Dual-reporter experiments have shown that regulation buffers cell-by-cell differences in yeast gene expression to enhance the precision of decisions to commit to a mating response phenotype [42]. Indeed,

Chapter 4 also showed that the microRNA miR-7 buffers Yan expression levels, and by extension R cell fate decisions, against varying biosynthesis capacity. Perhaps future experiments could address how Pnt levels are affected before and after IPC ablation in *yan<sup>ΔmiR-7</sup>* mutants.

These observations reflect a central theme of this dissertation; the structure and function of developmental GRNs dictate the robustness of cell fate decisions to environmental variation. Chapter 4 directly embraced this sentiment. It proposed a new theory to explain why the regulatory networks that coordinate cell fate decisions often contain many seemingly redundant repressors acting upon the same target genes. The theory posits that auxiliary negative regulators mitigate erroneous cell fate decisions when cells are rapidly metabolizing, and implies that auxiliary repressors may help GRNs adapt their behavior to environmentally driven variation in cell metabolism. The theory is supported by a diverse collection of experiments in which repressor loss-of-function phenotypes were reversed when biosynthesis rates were slowed. A quantitative modeling framework was used to explore the mechanistic origin of this effect, and theoretically demonstrated that auxiliary repressors could avert erroneous decisions by expanding cells capacity to buffer excess protein expression. Quantitative measurements of transcription factor activity confirmed the hypothesized mechanism *in vivo*.

Chapter 1 introduced the idea that developmental GRNs guide organisms through a tortuous journey from embryo to adulthood. The journey is not a sprint. Rather, cells must carefully navigate the many twists and turns of developmental programs; rapidly synthesizing GRN components when and where they are needed, then degrading them with equal urgency. Protein synthesis and degradation machineries supply the engine and brakes needed to negotiate these obstacles (Fig. 5.1A). Individuals stand to benefit from completing the journey quickly because they are generally vulnerable until adulthood, with little means to defend themselves against predation and other dangers. They could naively swap out the

engine for something more potent, but adding power escalates risk, particularly around sharp corners (Fig. 5.1B). Simultaneously upgrading the brakes is a winning strategy (Fig. 5.1C). Analogously, evolution could accelerate development by expanding biosynthesis capacity. The resultant boost in protein expression would strain each cell's ability to localize distinct GRN activities in time, causing increased risk of erroneous fate decisions that lead to the emergence of deleterious phenotypes. However, by selecting for auxiliary repressors, cells could tolerate faster biosynthesis rates without compromising the quality of fate decisions. The data presented in Chapter 4 confirm this hypothesis by illustrating the inverse perspective. Repressors were shown to be dispensable when metabolism was slow, much in the way that brakes aren't needed to round corners at low speed.



**Figure 5.1. The race from embryo to adulthood.** Individuals must navigate the many twists and turns of the developmental program to reach the finish line. (A) Slow and steady loses the race. (B) Adding power improves acceleration, but increases the risk of failure. (C) Adding power along with better brakes is a winning strategy. The analogous experimental conditions are (A) slow metabolism without auxiliary repression, (B) fast metabolism without auxiliary repression, and (C) fast metabolism with auxiliary repression. Evolution suggests this wildtype configuration consistently wins the race.

This line of reasoning implies that a novel evolutionary driving force may shape the structure and function of developmental gene regulatory networks. Shorter generational times confer a selective advantage beyond reducing individuals exposure during infancy. They

facilitate rapid exploration of the phenotypic landscape, enabling fast adaptation to variable environments. GRNs should therefore be expected to incorporate any topological features that allow development to proceed more quickly. The abundance of seemingly redundant regulation found in developmental GRNs certainly appears to support this hypothesis, and thus reinforce our contemporary understanding that robustness is a fundamental organizational principle underlying the evolution of biological systems [109, 110].

The findings also contribute to an emerging view that cells capacity to rapidly adjust protein homeostasis directly affects organismal fitness and health [298–300]. The assertion is backed by convincing experimental evidence. Burnaevskiy et al. used a dual-reporter scheme in *C. elegans* to show that cellular differences in the abundance of protein synthesis machinery manifest in the population-wide penetrance of adult phenotypes. The authors went on to speculate that longevity might be similarly be affected [300]. Tollerson et al. showed that Elongation factor P alleviates a translational bottleneck caused by ribosomal queuing in *E. coli*, facilitating adaption to environmentally-driven increases in cell metabolism [299]. Chapter 4 contributes unique evidence that the accuracy of cell fate decisions depends upon cells ability to dynamically balance the push and pull of protein synthesis and degradation.

## 5.2. Avenues for further exploration

This dissertation prompts several exciting new directions for future research. This section begins with a survey of those that merit further attention, before elaborating on some preliminary analyses to guide prospective efforts.

Chapter 3 proposed that R cell fate commitment in the larval eye is driven by dynamic changes in the relative abundance of Pnt and Yan. Experimental evidence was limited to

correlative observations because intrinsic regulation precluded direct manipulation of the Pnt-to-Yan ratio by varying gene dosage. Future studies could rigorously confirm the hypothesis by applying the same gene dosage perturbations in a genetic background that lacks the complete set of regulatory interactions needed to stabilize the Pnt-to-Yan ratio. Disrupting the ratio control mechanism would first require characterizing its biomolecular implementation. Computational simulations could explore the space of plausible GRN topologies, then leverage insight derived from existing experimental data to distill a manageable number of options for experimental validation. Establishing an unambiguous picture of R cell fate commitment might then allow for a complete model of retinal patterning dynamics, as is an ongoing mission among computational biologists [35, 78].

Chapter 3 used an equilibrium modeling framework to explore the effect of *cis*-regulatory interactions between Yan monomers on the equilibrium binding occupancy of promoters regulated by Pnt and Yan. The approach was inspired by the work of Hope et al., who used an equivalent model, limited to a single binding component, to explore how the *cis*-regulatory syntax of target genes modulates transcriptional output [220]. The multi-species implementation introduced in Chapter 3 was comparatively underutilized. Future studies could ask many questions related to how *cis*-regulatory syntax modulates the transcriptional output of genes regulated by two or more polymerizing transcription factors. How do the number and arrangement of individual binding sites modulate transcriptional output? What about the spacing or distribution of high and low affinity sites? What if anti-cooperative or steric interactions are included? Does the number of unique binding species matter? All of these questions could readily be explored using the open-source platform developed to support this dissertation (see Appendix C).

The theory developed in Chapter 4 posits that auxiliary repressors stabilize cell fate decisions against environmental variation in cells capacity to synthesize and degrade proteins. This assertion was backed by both experiments and computational analysis showing that repressors were dispensable when either ATP turnover or translation capacity were reduced. It is well known that many mutant phenotypes are also suppressed in animals raised at low temperatures. Future studies could ask whether reduced temperature imparts similar effects on the GRN dynamics that govern cell fate decisions. From a modeling perspective, the primary challenge would be deciding precisely how to incorporate the relative effects of temperature on the protein synthesis and degradation machineries. The analogous decisions were comparatively obvious for *ILP2-GAL4 UAS-Rpr* and *RP* mutants, in which protein synthesis is disproportionately affected. In principle, experimental efforts to quantify the dependence of protein synthesis and degradation rates on temperature could prove fruitful.

Alternatively, researchers could computationally survey the landscape of plausible relationships between proteostasis and the environment to identify conditions under which auxiliary repressors are dispensable. For instance, consider the simplest possible model of protein expression dynamics (Fig. 5.2A):

$$\frac{dP}{dt} = kI - \gamma P \quad (5.1)$$

where  $P$  and  $I$  are the protein and stimulus levels, and  $k$  and  $\gamma$  are the synthesis and degradation rate parameters. Repressors may be implemented as proportional feedback, as they were in Chapter 4:

$$\frac{dP}{dt} = kI - \gamma P - \eta P \quad (5.2)$$

where  $\eta$  is the feedback strength. The frequency of developmental errors induced by losing the repressor is readily evaluated using the same procedure described in Chapter 4 (Fig. 5.2B).

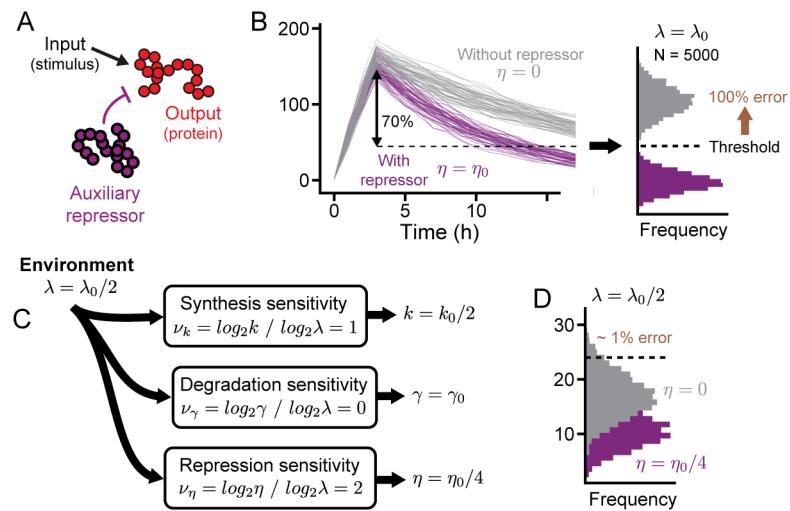
Rather than hard-coding an explicit dependence of  $k$ ,  $\gamma$ , and  $\eta$  on environmental conditions, each parameter can be scaled by a latent dimension  $\lambda$  that reflects the environmental state of the cell:

$$\begin{aligned} k &\propto \lambda^{\nu_k} \\ \gamma &\propto \lambda^{\nu_\gamma} \\ \eta &\propto \lambda^{\nu_\eta} \end{aligned} \tag{5.3}$$

where  $\nu_k$ ,  $\nu_\gamma$ , and  $\nu_\eta$  define the respective sensitivities of synthesis, degradation, and feedback strength to environmental conditions. Consider an example in which the environmental state of the cell is halved relative to some reference condition, i.e.  $\lambda = \lambda_0/2$ . For  $\nu_{i \in k, \gamma, \eta} = 1$ , parameter  $i$  is also halved. If  $0 < \nu_i < 1$  or  $\nu_i > 1$ , parameter  $i$  exhibits sub- or super-linear dependence on the environment, respectively. Finally,  $\nu_i < 0$  implies that parameter  $i$  should actually increase when  $\lambda$  is halved.

The robustness checks presented in Section 4.9 surveyed a handful of discrete values for the model parameters analogous to  $\nu_k$ ,  $\nu_\gamma$ , and  $\nu_\eta$ . The scaling assumptions used to represent conditions of reduced energy metabolism throughout Chapter 4 were loosely equivalent to  $(\nu_k, \nu_\gamma, \nu_\eta) = (1, 0, 2)$ . Applying the same assumptions to the simplified model recapitulates a core result of Chapter 4; error frequency is dependent upon the environmental state of the cell (Fig. 5.2C,D).

Combined, equations 5.2 and 5.3 facilitate continuous enumeration of the phase diagram spanned by  $\nu_k$ ,  $\nu_\gamma$ , and  $\nu_\eta$  in order to identify the range of scaling assumptions under which auxiliary repressors may be dispensable. Performing this exercise revealed that induced error frequencies are highest when the strength of the removed repressor is more sensitive to the environment than the intrinsic rate of protein turnover (Fig. 5.3A, region IV). The same is true of error frequency suppression, indicating that highly sensitive repressors have

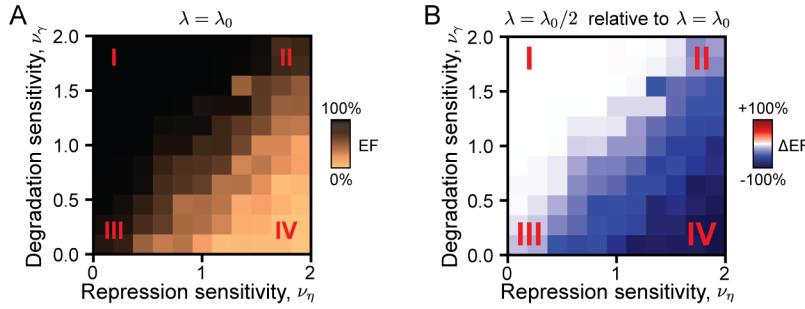


**Figure 5.2. Simplified framework for modeling the loss of an auxiliary repressor.** (A) Schematic representation of protein expression in response to a transient input. Output is subject to regulation by a single auxiliary repressor. (B) Simulated emergence of developmental errors. Simulations may be performed with (purple) or without (grey) the auxiliary repressor. Lines reflect a random sample of 5000 trajectories. The two sets of trajectories are compared when 99% of trajectories simulated with the repressor cross a threshold value (dashed line). Without the auxiliary repressor, very few trajectories successfully cross the threshold. (C) Graphic representation of the relation between environmental conditions and the rate parameters that dictate protein synthesis, degradation, and repressor strength. (D) Error frequency is dramatically suppressed by a change in environmental conditions.

the highest propensity to become dispensable (Fig. 5.3B, region IV). This observation is consistent with intuition, as environmental conditions that limit the influence of a repressor would also be expected to mitigate the impact of its removal.

This preliminary analysis could guide the design of future experiments that survey the effects of temperature on mutations that compromise repressor function. For example, experiments could quantify  $\nu_k$ ,  $\nu_\gamma$ , and  $\nu_\eta$  for a particular cell-fate determinant by measuring steady-state protein levels in both wildtype and repressor loss-of-function mutants across a range of temperatures. They could then use the model to generate testable predictions for alternate temperatures.

Chapter 4 also raises the question of whether any other features of GRNs are dispensable under certain environmental conditions. For instance, how about promoters? Preliminary



**Figure 5.3. Suppression depends on parameter sensitivity to the environment.** Phase diagrams span a range of protein degradation and repressor strength sensitivities to environmental conditions. Heatmaps show (A) error frequency induced by the loss of an auxiliary repressor when  $\lambda = \lambda_0$  and (B) change in error frequency when  $\lambda = \lambda_0/2$ . All simulations use  $k = 1$ ,  $\gamma = 0.001$ ,  $\eta = 0.001$ , and  $\lambda_0 = 1$ . Dark blue indicates strong error frequency suppression. Red numerals label quadrants. The example shown in Fig. 5.2 is taken from quadrant IV.

analysis may again provide some insight to guide future work. Consider another scenario in which an auxiliary promoter is added to the simple model given by Equation 5.1:

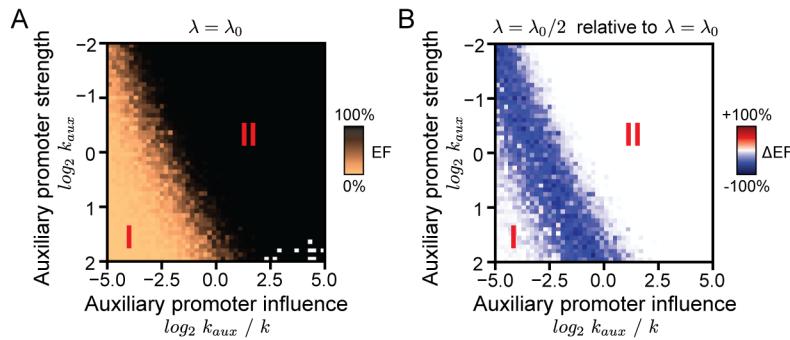
$$\frac{dP}{dt} = kI + k_{aux}I - \gamma P \quad (5.4)$$

where  $k_{aux}$  is the rate constant for synthesis driven by the auxiliary promoter. The relative influence of the auxiliary promoter is given by its strength relative to the primary promoter, i.e.  $\log_2(k_{aux}/k)$ . The respective sensitivities of both promoters and degradation to environmental conditions are again parameterized in terms of  $\lambda$ :

$$\begin{aligned} k &\propto \lambda^{\nu_k} \\ k_{aux} &\propto \lambda^{\nu_{k-aux}} \\ \gamma &\propto \lambda^{\nu_\gamma} \end{aligned} \quad (5.5)$$

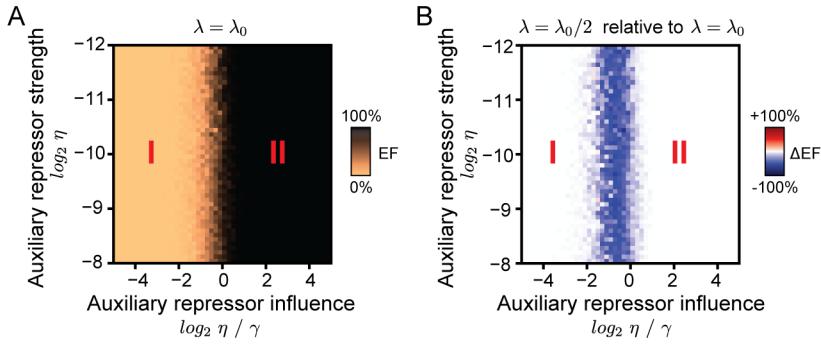
The simulation procedure described in 4 is readily modified to evaluate the frequency of developmental errors induced by removing the auxiliary promoter, i.e. by setting  $k_{aux} = 0$ . Intuition suggests promoter loss should cause a decrease in output protein levels. Error frequency is therefore redefined to reflect the extent to which protein is *under-expressed* when

the auxiliary promoter is removed. The metric is evaluated by computing the fraction of trajectories simulated with a single promoter that fail to reach the lower bound of trajectories simulated with both promoters. Surveying a range of promoter strengths and relative influences reveals that error frequencies are most severe when strong and influential auxiliary promoters are removed (Fig. 5.4A). Figure 5.4B shows the subsequent change in error frequencies when  $(\nu_k, \nu_{k-aux}, \nu_\gamma) = (1, 1, 1)$  and  $\lambda = \lambda_0/2$ . The phase diagram is punctuated by a diagonal band in which error frequencies are suppressed. Below this band, suppression is minimal because the auxiliary promoter is inconsequential to normal protein expression dynamics (Fig. 5.4A,B, region I). Above it, removing the auxiliary promoter imparts a severe perturbation that cannot be recovered by the proposed change in environmental conditions (Fig. 5.4A,B, region II). Similar zones arise when the equivalent procedure is performed using the model of auxiliary repressor loss defined by equation 5.2 (Fig. 5.5). Here, the influence of the auxiliary repressor is defined relative to the intrinsic degradation rate, i.e.  $\log_2(\eta/\gamma)$ .



**Figure 5.4. Phase diagram of auxiliary promoter loss.** Phase diagrams span a range of auxiliary promoter strengths and influences. Heatmaps show (A) error frequency induced by the loss of the auxiliary promoter when  $\lambda = \lambda_0$  and (B) change in error frequency when  $\lambda = \lambda_0/2$ . All simulations use  $\gamma = 0.001$  and  $\lambda_0 = 1$ . Dark blue indicates strong error frequency suppression. Perturbations targeting promoters in region I are inconsequential. Those targeting promoters in region II are too severe to be recovered. Blue band is the Goldilocks zone.

The bands observed in Figures 5.4B and 5.5B indicate the existence of a Goldilocks zone in which perturbations are strong enough to be felt but weak enough to be recoverable. In



**Figure 5.5. Phase diagram of auxiliary repressor loss.** Phase diagrams span a range of auxiliary repressor strengths and influences. Heatmaps show (A) error frequency induced by the loss of the auxiliary repressor when  $\lambda = \lambda_0$  and (B) change in error frequency when  $\lambda = \lambda_0/2$ . All simulations use  $k = 1$  and  $\lambda_0 = 1$ . Dark blue indicates strong error frequency suppression. Perturbations targeting repressors in region I are inconsequential. Those targeting repressors in region II are too severe to be recovered. Blue band is the Goldilocks zone.

other words, there exists a finite and continuous range of conditions in which error frequencies can be suppressed by slowing biosynthesis. Given appropriate knowledge of promoter and repressor strengths, a model could inform the selection of experimental perturbation targets that fall within the range. Such an approach would also require quantification of  $\nu_i$  and  $\lambda$  for a given model system.

The value of the models defined by equations 5.2 and 5.4 lies in their simplicity. Because each strictly employs linear kinetics, the stochastic dynamics are analytically tractable via a closed system of moment equations [301]. Even the more complex models presented in Chapter 4 are approximately solvable, courtesy of modern moment closure and finite state projection techniques [302, 303]. Future theoretical efforts could be directed toward the development of a complete analytical framework for studying the interplay of GRN dynamics, cell fate decisions, and the environment. An analytical platform would vastly reduce the computational burden imposed by the numerical methods used to conduct all of the analysis presented in this section and throughout Chapter 4. It would also expedite fitting the model to experimental data in order to generate testable predictions.

### 5.3. Perspective on quantitative biology

This dissertation surveyed developmental cell fate decisions through a quantitative lens. It used numbers and equations to derive nuanced understanding of processes that are notoriously difficult to characterize. In many cases, doing so required forcible reconciliation with time-honored traditions in biological analysis. Two major obstacles encountered along the way merit further discussion.

First, models should fit the data; not the other way around. While it may be possible to coax data into a conventional framework, tailoring one appropriate for the task at hand will generally return more meaningful insight. Mathematical flexibility was a key strength of this dissertation. Each chapter leveraged a customized modeling framework to elicit deeper meaning out of experimental data than would have been possible using conventional techniques.

Chapter 2 deployed a Bayesian statistical framework to infer cell genotypes from an image of clonal marker expression. Biological intuition suggests it should have enforced detection of three distinct components strictly delimited by clonal marker level. Instead, the model was designed to tolerate an arbitrary number of components that also considered spatial context, which were later aggregated into the three anticipated genotypes. This empirical formulation buffered uncertainties imparted by expression heterogeneity and imaging noise to improve annotation performance overall.

Chapter 3 used an equilibrium binding model to explore how *cis*-regulatory interactions between Yan monomers bias the transcriptional output of genes simultaneously regulated by Pnt. Conventional wisdom suggests a competitive binding model using the Hill-Langmuir equation would be appropriate [304]. Instead, a statistical mechanical approach was adapted

from earlier work by Hope et al. [220]. Contrary to the empirical strategies used in other chapters, this model substantially *increased* the resolution of analysis. Doing so provided detailed mechanistic insight into the intricacies of polymerizing transcription factors, which would have otherwise been inaccessible (see Fig. 3.5C).

Chapter 4 sought to model the dynamic output of GRNs controlling a broad spectrum of cell fate decisions. The conventional first step would have been to identify and consolidate all known regulatory interactions in each system. Such an endeavor was certainly possible for Yan-mediated control of retinal patterning, as the pertinent interactions have been studied for decades [27, 82, 85, 95, 223] and in some cases quantified [43]. Chapter 4 instead drew inspiration from control theory to develop a model strictly concerned with the two salient features of pulsatile dynamics; the magnitude of induction and timescale of decay. The model jettisoned the molecular details of protein expression and regulation in each system, instead favoring a coarse-grained depiction of output dynamics. The reductionist approach was vital to simultaneously depicting the behavior of a diverse assortment of cell fate decisions.

Chapter 4 was also forced to circumvent the second major obstacle: The resolution of analysis should match the resolution of the data. Biological networks are complex, often far more complex than we can intuitively comprehend. Their emergent behavior arises from the collective interactions of numerous components, many of which are often unknown. These uncertainties make it particularly dangerous to think of a GRN as the sum of its parts. Similarly, predictions are all but meaningless when made by aggregating disparate interactions that were characterized in isolation.

Despite slowly coming under scrutiny, these practices remain tragically common. They are perhaps most strongly embodied in the abundance of cartoons that purport to depict systems level-behavior with a compendium of qualitative regulatory interactions. Figure

3.11A provides a modest example. This cartoon is harmless in isolation, but problems arise when unsuspecting viewers ascribe physical meaning to the various interactions. After all, there is no unified standard to define what an arrow, or “edge,” means. A single edge might actually represent an entire sub-network of complex nonlinear processes. Intermediate steps within an edge might even interact with other intermediates in other edges drawn elsewhere in the diagram. Cartoons are thus rife with ambiguity that hinders the communication of otherwise outstanding research. For instance, a basic attempt to simulate the system shown in Figure 3.11A will reveal that the illustrated “regulatory loop” cannot maintain a constant Pnt-to-Yan ratio in response to varying *pnt* or *yan* gene dosage (data not shown). Yet, Figure 3.11A was necessary because these depictions are so deeply ingrained in the biological literature that the representation in Figure 3.11B would likely be misconstrued without it.

This ambiguity starkly contrasts the standardized descriptive languages commonly used in engineering [44]. To be fair, most human-engineered systems do not suffer the same extent of uncertainty as their biological counterparts. Notable exceptions arise in electrical, chemical, and biomolecular engineering. Autonomous vehicles, advanced robotics, process plants, and synthetic biological circuits must all contend with disturbances whose origin and character may be unpredictable or unknown. Fortunately, a viable solution to this challenge already exists. Control theory emphasizes an empirical representation of systems-level dynamics that is easily matched to the resolution of available information [282]. In many cases, the specific details of a system’s internal dynamics may remain unknown. Sensors need only monitor the minimum set of variables necessary to ensure that a given process is controllable. Often, this simply entails monitoring systems-level output and taking corrective action when it deviates from a desired set point.

Control theory offers a particularly compelling perspective for rationalizing the behavior of developmental GRNs. Most are inherently dynamic, contain numerous unknown components and interactions, and ultimately coordinate a manageable number of outputs. This rationale inspired the conceptual model of ratiometric control used to interpret the R cell fate decision analyzed in Chapter 3 (see Fig. 3.11B). After all, Pnt and Yan are transiently expressed, subject to extensive uncharacterized regulation, and appear to mediate R cell fate transitions through a single observable output; the Pnt-to-Yan ratio. Similar reasoning inspired the modeling framework used throughout Chapter 4. The breadth of experiments pointed toward a dynamic phenomenon agnostic to the molecular detail of repressors and their targets. Furthermore, among all systems surveyed, only Yan and Sens expression were measurable. The resolution of analysis was therefore matched to the resolution of the data by modeling the representative dynamics of a generic protein, subject to feedback at each of the three levels that were experimentally surveyed (see Fig. 4.6). Combined, these approaches reflect a growing trend of interpreting biological robustness from a control perspective [92].

As biology and engineering converge on common interests, it is important to reconcile the strengths of both disciplines. Biology contributes a wealth of prior knowledge and experimental techniques that frequently bewilder even the most seasoned engineers. Engineering contributes quantitatively rigorous frameworks to systematically analyze, interpret, predict, and design complex systems. The promise of successful integration prompts continued dialogue to resolve any growing pains and advance the field of quantitative biology as a whole.

## References

- [1] Adam S Wilkins. *Genetic analysis of animal development*. Wiley-Liss Inc, 1993.
- [2] Editorial. Elements of a good decision. *Nat. Immunol.*, 11(8):645–645, aug 2010. doi: 10.1038/ni0810-645.
- [3] Zhiwei Wang, Yiwei Li, Sanjeev Banerjee, and Fazlul H. Sarkar. Emerging role of Notch in stem cells and cancer, jun 2009.
- [4] Jorrit J Hornberg, Frank J Bruggeman, Hans V Westerhoff, and Jan Lankelma. Cancer: A Systems Biology disease. In *BioSystems*, volume 83, pages 81–90, 2006. doi: 10.1016/j.biosystems.2005.05.014.
- [5] Thomas Hunt Morgan. *The mechanism of Mendelian heredity*. Holt, 1915.
- [6] G W Beadle and B Ephrussi. The Differentiation of Eye Pigments in Drosophila as Studied by Transplantation. *Genetics*, 21(3):225–47, may 1936.
- [7] E. B. Lewis. A gene complex controlling segmentation in Drosophila. *Nature*, 276 (5688):565–570, dec 1978. doi: 10.1038/276565a0.
- [8] Christiane Nüsslein-volhard and Eric Wieschaus. Mutations affecting segment number and polarity in drosophila. *Nature*, 287(5785):795–801, oct 1980. doi: 10.1038/287795a0.
- [9] G Halder, P Callaerts, and W. Gehring. Induction of ectopic eyes by targeted expression of the eyeless gene in Drosophila. *Science (80-)*, 267(5205):1788–1792, mar 2006. doi: 10.1126/science.7892602.
- [10] Norbert Perrimon. The genetic basis of patterned baldness in Drosophila, mar 1994.
- [11] Hugo J Bellen, Chao Tong, and Hiroshi Tsuda. 100 years of Drosophila research and its impact on vertebrate neuroscience: A history lesson for the future, jul 2010.
- [12] T R Parody and M. A T Muskavitch. The pleiotropic function of Delta during postembryonic development of *Drosophila melanogaster*. *Genetics*, 135(2):527–539, 1993.
- [13] T. Ian Simpson and David J. Price. Pax6; a pleiotropic player in development, nov 2002.

- [14] P Bhanot, M Fish, J A Jemison, R Nusse, J Nathans, and K M Cadigan. Frizzled and Dfrizzled-2 function as redundant receptors for Wingless during *Drosophila* embryonic development. *Development*, 1999.
- [15] Kartik S Pappu. Dual regulation and redundant function of two eye-specific enhancers of the *Drosophila* retinal determination gene *dachshund*. *Development*, 132(12):2895–2905, jun 2005. doi: 10.1242/dev.01869.
- [16] Xin Li and Richard W Carthew. A microRNA mediates EGF receptor signaling and promotes photoreceptor differentiation in the *Drosophila* eye. *Cell*, 123(7):1267–1277, 2005. doi: 10.1016/j.cell.2005.10.040.
- [17] Jorge V. Beira and Renato Paro. The legacy of *Drosophila* imaginal discs, sep 2016.
- [18] Masato Enomoto, Carmen Siow, and Tatsushi Igaki. *Drosophila* as a cancer model. In *Adv. Exp. Med. Biol.*, volume 1076, pages 173–194. Springer, Singapore, 2018. doi: 10.1007/978-981-13-0529-0\_10.
- [19] Federico Germani, Cora Bergantinos, and Laura A Johnston. Mosaic analysis in *Drosophila*. *Genetics*, 208(2):473–490, feb 2018. doi: 10.1534/genetics.117.300256.
- [20] M. Levine and E. H. Davidson. Gene regulatory networks for development. *Proc. Natl. Acad. Sci.*, 2005. doi: 10.1073/pnas.0408031102.
- [21] Angelike Stathopoulos and Michael Levine. Genomic regulatory networks and animal development, oct 2005.
- [22] Martin Hüskamp, Christine Pfeifle, and Diethard Tautz. A morphogenetic gradient of hunchback protein organizes the expression of the gap genes Krüppel and knirps in the early *Drosophila* embryo. *Nature*, 346(6284):577–580, aug 1990. doi: 10.1038/346577a0.
- [23] Martin Hüskamp and Diethard Tautz. Gap genes and gradients The logic behind the gaps, jun 1991.
- [24] Daniel St Johnston and Christiane Nüsslein-Volhard. The origin of pattern and polarity in the *Drosophila* embryo, jan 1992.
- [25] M. W. Perry, A. N. Boettiger, and M. Levine. Multiple enhancers ensure precision of gap gene-expression patterns in the *Drosophila* embryo. *Proc. Natl. Acad. Sci.*, 2011. doi: 10.1073/pnas.1109873108.
- [26] Gary Struhl and Konrad Basler. Organizing activity of wingless protein in *Drosophila*. *Cell*, 72(4):527–540, feb 1993. doi: 10.1016/0092-8674(93)90072-X.
- [27] Donald F. Ready, Thomas E. Hanson, and Seymour Benzer. Development of the *Drosophila* retina, a neurocrystalline lattice. *Dev. Biol.*, 53(2):217–240, oct 1976. doi: 10.1016/0012-1606(76)90225-6.

- [28] Andrew Tomlinson and Donald F. Ready. Neuronal differentiation in the Drosophila ommatidium. *Dev. Biol.*, 120(2):366–376, apr 1987. doi: 10.1016/0012-1606(87)90239-9.
- [29] Matthew G. Voas and Ilaria Rebay. Signal integration during development: insights from the Drosophila eye. *Dev Dyn*, 229(1):162–175, jan 2004. doi: 10.1002/dvdy.10449.
- [30] Andrew P. Jarman, Ellsworth H. Grell, Larry Ackerman, Lily Y. Jan, and Yuh Nung Jan. *atonal* is the proneural gene for Drosophila photoreceptors. *Nature*, 369(6479):398–400, 1994.
- [31] Benjamin Jay Frankfort and Graeme Mardon. R8 development in the Drosophila eye: a paradigm for neural selection and differentiation. *Development*, 129(6):1295–306, 2002.
- [32] Nicholas E. Baker and Sung-Yun Yu. Proneural function of neurogenic genes in the developing Drosophila eye. *Curr. Biol.*, 7:122–132, feb 1997. doi: 10.1016/S0960-9822(06)00056-X.
- [33] F. Hsiung. Retinal development in Drosophila: specifying the first neuron. *Hum. Mol. Genet.*, 11(10):1207–1214, may 2002. doi: 10.1093/hmg/11.10.1207.
- [34] Nicholas E Baker, Marek Mlodzik, and Gerald M Rubin. Spacing differentiation in the developing Drosophila eye: A fibrinogen-related lateral inhibitor encoded by scabrous. *Science (80-. ).*, 250(4986):1370–1377, dec 1990. doi: 10.1126/science.2175046.
- [35] Avishai Gavish, Arkadi Shwartz, Abraham Weizman, Eyal Schejter, Ben-Zion Shilo, and Naama Barkai. Periodic patterning of the Drosophila eye is stabilized by the diffusible activator Scabrous. *Nat. Commun.*, 7:10461, 2016. doi: 10.1038/ncomms10461.
- [36] Alain Ghysen and Christine DamblyChaudière. The specification of sensory neuron identity in Drosophila, may 1993.
- [37] P Simpson. Notch signalling in development: on equivalence groups and asymmetric developmental potential. *Curr. Opin. Genet. Dev.*, 7(4):537–42, aug 1997.
- [38] Michael B Elowitz, Arnold J Levine, Eric D Siggia, and Peter S Swain. Stochastic gene expression in a single cell. *Science (80-. ).*, 297(5584):1183–1186, aug 2002. doi: 10.1126/science.1070919.
- [39] Arjun Raj, Scott A. Rifkin, Erik Andersen, and Alexander van Oudenaarden. Variability in gene expression underlies incomplete penetrance. *Nature*, 463(7283):913–918, 2010. doi: 10.1038/nature08781.
- [40] Malte Paulsen, Stefan Legewie, Roland Eils, Emil Karaulanov, and Christof Niehrs. Negative feedback in the bone morphogenetic protein 4 (BMP4) synexpression group governs its dynamic signaling range and canalizes development. *Proc. Natl. Acad. Sci.*, 108(25):10202–10207, jun 2011. doi: 10.1073/pnas.1100179108.

- [41] Alejandro Burga, M Olivia Casanueva, and Ben Lehner. Predicting mutation outcome from early stochastic variation in genetic interaction partners. *Nature*, 480(7376):250–253, 2011. doi: 10.1038/nature10665.
- [42] Alejandro Colman-Lerner, Andrew Gordon, Eduard Serra, Tina Chin, Orna Resnekov, Drew Endy, C. Gustavo Pesce, and Roger Brent. Regulated cell-to-cell variation in a cell-fate decision system. *Nature*, 437(7059):699–706, 2005. doi: 10.1038/nature03998.
- [43] Nicolás Peláez, Arnau Gavalda-Miralles, Bao Wang, Heliodoro Tejedor Navarro, Herman Gudjonson, Ilaria Rebay, Aaron R. Dinner, Aggelos K. Katsaggelos, Luís A. Nunes Amaral, and Richard W. Carthew. Dynamics and heterogeneity of a fate determinant during transition towards cell differentiation. *eLife*, 4, nov 2015. doi: 10.7554/eLife.08924.
- [44] Y. Lazebnik. Can a biologist fix a radio? - Or, what I learned while studying apoptosis, 2004.
- [45] Andrew C. Oates, Nicole Gorfinkiel, Marcos González-Gaitán, and Carl Philipp Heisenberg. Quantitative approaches in developmental biology, aug 2009.
- [46] Charles A. Herring, Bob Chen, Eliot T. McKinley, and Ken S. Lau. Single-Cell Computational Strategies for Lineage Reconstruction in Tissue Systems, jan 2018.
- [47] Philipp J Keller. Imaging morphogenesis: Technological advances and biological insights, jun 2013.
- [48] Ivo F. Sbalzarini. Seeing is believing: Quantifying is convincing: Computational image analysis in biology. *Adv. Anatomy, Embryol. Cell Biol.*, 219:1–39, 2016. doi: 10.1007/978-3-319-28549-8\_1.
- [49] Brian Munsky, Karen Tkach Tuzman, Dirk Fey, Maciej Dobrzynski, Boris N. Khodenko, Sarah Olson, Jianjun Huang, Zachary Fox, Abhyudai Singh, Ramon Grima, Roberto Bertolusso, Marek Kimmel, Margaritis Voliotis, Philipp Thomas, Clive G. Bowsher, Thomas R. Sokolowski, Pieter Rein ten Wolde, Petr Sulc, Jonathan P. Doye, Ard A. Louis, Joe Natale, David Hofmann, Damian Hernandez, Ilya Nemenman, Heeseob Lee, Jason G. Lomnitz, Michael A. Savageau, Bryan C. Daniels, Karol Nienaltowski, Tomasz Jetka, Michal Komorowski, Jr. Bernie J. Daigle, Dongya Jia, Mohit Kumar Jolly, Herbert Levine, Martin Lopez-Garcia, Maria Nowicka, Gareth W. Fearnley, Sreenivasan Ponnambalam, Grant Lythe, Carmen Molina-Paris, Herbert M. Sauro, Paolo Cazzaniga, Marco S. Nobile, Andrea Tangherloni, Daniela Besozzi, Michael L. Blinov, Dan Vasilescu, Ion I. Moraru, James C. Schaff, Boris M. Slepchenko, Leslie M. Loew, Brian Drawert, Kevin R. Sanft, John H. Abel, Stefan Hellander, Arya Pourzanjani, Linda R. Petzold, Sanjana Gupta, Jacob Czech, Robert Kuczewski, Thomas M. Bartol, Terrence J. Sejnowski, Robin E. C. Lee, James R. Faeder, Fiona A. Chandra, Ganhui Lan, Alan S. Perelson, Ruy M. Ribeiro, Andrew Mugler, Sean Fancher, Lisa Weber, William Raymond, Brian Munsky, William S. Hlavacek, and Lev S. Tsimring. *Quantitative Biology: Theory, Computational Methods, and Models*. The MIT Press,

2018.

- [50] Christopher L. Frick, Clare Yarka, Harry Nunns, and Lea Goentoro. Sensing relative signal in the Tgf- $\beta$ /Smad pathway. *Proc. Natl. Acad. Sci.*, 114(14):E2975–E2982, mar 2017. doi: 10.1073/pnas.1611428114.
- [51] Samuel C. Wolff, Kasia M. Kedziora, Raluca Dumitru, Cierra D. Dungee, Tarek M. Zikry, Adriana S. Beltran, Rachel A. Haggerty, JrGang Cheng, Margaret A. Redick, and Jeremy E. Purvis. Inheritance of OCT4 predetermines fate choice in human embryonic stem cells. *Mol. Syst. Biol.*, 14(9):e8140, jan 2018. doi: 10.1101/137299.
- [52] Mariela D. Petkova, Gašper Tkačik, William Bialek, Eric F. Wieschaus, and Thomas Gregor. Optimal Decoding of Cellular Identities in a Genetic Network. *Cell*, 176(4):844–855.e15, feb 2019. doi: 10.1016/j.cell.2019.01.007.
- [53] Nima Aghaeepour, Greg Finak, Holger Hoos, Tim R Mosmann, Ryan Brinkman, Raphael Gottardo, Richard H Scheuermann, David Dougall, Alireza Hadj Khodabakhshi, Phillip Mah, Gerlinde Obermoser, Josef Spidlen, Ian Taylor, Sherry A. Wuensch, Jonathan Bramson, Connie Eaves, Andrew P. Weng, Edgardo S. Fortuno, Kevin Ho, Tobias R. Kollmann, Wade Rogers, Stephen De Rosa, Bakul Dalai, Ariful Azad, Alex Pothen, Aaron Brandes, Hannes Bretschneider, Robert Bruggner, Rachel Finck, Robin Jia, Noah Zimmerman, Michael Linderman, David Dill, Gary Nolan, Cliburn Chan, Faysal El Khettabi, Kieran O’Neill, Maria Chikina, Yongchao Ge, Stuart Sealfon, István Sugár, Arvind Gupta, Parisa Shooshtari, Habil Zare, Philip L. De Jager, Mike Jiang, Jens Keilwagen, Jose M. Maisog, George Luta, Andrea A. Barbo, Peter Májek, Jozef Vilček, Tapio Manninen, Heikki Huttunen, Pekka Ruusuvuori, Matti Nykter, Geoffrey J. McLachlan, Kui Wang, Iftekhar Naim, Gaurav Sharma, Radina Nikolic, Saumyadipta Pyne, Yu Qian, Peng Qiu, John Quinn, Andrew Roth, Pablo Meyer, Gustavo Stolovitzky, Julio Saez-Rodriguez, Raquel Norel, Madhuchhanda Bhattacharjee, Michael Biehl, Philipp Bucher, Kerstin Bunte, Barbara Di Camillo, Francesco Sambo, Tiziana Sanavia, Emanuele Trifoglio, Gianna Toffolo, S. D. Slavica Dimitrieva, Rene Dreos, Giovanna Ambrosini, Jan Grau, Ivo Grosse, Stefan Posch, Nicolas Guex, Miron Kursa, Witold Rudnicki, Bo Liu, Mark Maienschein-Cline, Petra Schneider, Michael Seifert, Marc Strickert, and Jose M.G. Vilar. Critical assessment of automated flow cytometry data analysis techniques. *Nat. Methods*, 10(3):228–238, mar 2013. doi: 10.1038/NMETH.2365.
- [54] Xiaoyi Chen, Milena Hasan, Valentina Libri, Alejandra Urrutia, Benoît Beitz, Vincent Rouilly, Darragh Duffy, Étienne Patin, Bernard Chalmond, Lars Rogge, Lluis Quintana-Murci, Matthew L. Albert, and Benno Schwikowski. Automated flow cytometric analysis across large numbers of samples and cell types. *Clin. Immunol.*, 157(2):249–260, apr 2015. doi: 10.1016/j.clim.2014.12.009.
- [55] Saumyadipta Pyne, Lisa M Maier, T.-I Lin, Kui Wang, Elizabeth Rossin, Xinli Hu, Pablo Tamayo, David A Hafler, Philip L De Jager, Geoffrey J McLachlan, Clare Baecher-Allan, and Jill P Mesirov. Automated high-dimensional flow cytometric data analysis. *Proc. Natl. Acad. Sci.*, 106(21):8519–8524, may 2009. doi: 10.1073/pnas.0903028106.

- [56] Bradley E Bernstein, Myles Brown, David S Johnson, X Shirley Liu, Chad Nussbaum, Richard M Myers, Tao Liu, Wei Li, Clifford A Meyer, Jérôme Eeckhoute, and Yong Zhang. Model-based Analysis of ChIP-Seq (MACS). *Genome Biol.*, 9(9):R137, sep 2008. doi: 10.1186/gb-2008-9-9-r137.
- [57] Jan Hellemans, Geert Mortier, Anne De Paepe, Frank Speleman, and Jo Vandesompele. qBase relative quantification framework and software for management and automated analysis of real-time quantitative PCR data. *Genome Biol.*, 8(2), feb 2007. doi: 10.1186/gb-2007-8-2-r19.
- [58] Ben Langmead and Steven L Salzberg. Fast gapped-read alignment with Bowtie 2. *Nat. Methods*, 9(4):357–9, apr 2012. doi: 10.1038/nmeth.1923.
- [59] Cole Trapnell, Lior Pachter, and Steven L. Salzberg. TopHat: Discovering splice junctions with RNA-Seq. *Bioinformatics*, 25(9):1105–1111, may 2009. doi: 10.1093/bioinformatics/btp120.
- [60] Sylvain V. Costes, Dirk Daelemans, Edward H. Cho, Zachary Dobbin, George Pavlakis, and Stephen Lockett. Automatic and quantitative measurement of protein-protein colocalization in live cells. *Biophys. J.*, 86(6):3993–4003, jun 2004. doi: 10.1529/biophysj.103.038422.
- [61] Lawrence A Kelley, Stefans Mezulis, Christopher M Yates, Mark N Wass, and Michael J.E. Sternberg. The Phyre2 web portal for protein modeling, prediction and analysis. *Nat. Protoc.*, 10(6):845–858, jun 2015. doi: 10.1038/nprot.2015.053.
- [62] Anne E Carpenter, Thouis R Jones, Michael R Lamprecht, Colin Clarke, In Han Kang, Ola Friman, David A Guertin, Joo Han Chang, Robert A Lindquist, Jason Moffat, Polina Golland, and David M Sabatini. CellProfiler: Image analysis software for identifying and quantifying cell phenotypes. *Genome Biol.*, 7(10):R100, oct 2006. doi: 10.1186/gb-2006-7-10-r100.
- [63] Ahmad Paintdakhi, Bradley Parry, Manuel Campos, Irnov Irnov, Johan Elf, Ivan Surovtsev, and Christine Jacobs-Wagner. Oufti: An integrated software package for high-accuracy, high-throughput quantitative microscopy analysis. *Mol. Microbiol.*, 99(4):767–777, feb 2016. doi: 10.1111/mmi.13264.
- [64] Johannes Schindelin, Ignacio Arganda-Carreras, Erwin Frise, Verena Kaynig, Mark Longair, Tobias Pietzsch, Stephan Preibisch, Curtis Rueden, Stephan Saalfeld, Benjamin Schmid, Jean Yves Tinevez, Daniel James White, Volker Hartenstein, Kevin Eliceiri, Pavel Tomancak, and Albert Cardona. Fiji: An open-source platform for biological-image analysis, jul 2012.
- [65] Christoph Sommer, Christoph Straehle, Ullrich Kothe, and Fred A. Hamprecht. Ilastik: Interactive learning and segmentation toolkit. In *Proc. - Int. Symp. Biomed. Imaging*, pages 230–233. IEEE, mar 2011. doi: 10.1109/ISBI.2011.5872394.

- [66] Johannes Schindelin, Curtis T. Rueden, Mark C. Hiner, and Kevin W. Eliceiri. The ImageJ ecosystem: An open platform for biomedical image analysis, jul 2015.
- [67] Florian Jug, Tobias Pietzsch, Stephan Preibisch, and Pavel Tomancak. Bioimage Informatics in the context of Drosophila research. *Methods*, 68(1):60–73, jun 2014. doi: 10.1016/j.ymeth.2014.04.004.
- [68] N. Barkai and S. Leibler. Robustness in simple biochemical networks. *Nature*, 387 (6636):913–917, jun 1997. doi: 10.1038/43199.
- [69] Richard C Yu, C Gustavo Pesce, Alejandro Colman-Lerner, Larry Lok, David Pincus, Eduard Serra, Mark Holl, Kirsten Benjamin, Andrew Gordon, and Roger Brent. Negative feedback that improves information transmission in yeast signalling. *Nature*, 456 (7223):755–61, 2008. doi: 10.1038/nature07513.
- [70] U. Alon, M. G. Surette, N. Barkai, and S. Leibler. Robustness in bacterial chemotaxis. *Nature*, 1999. doi: 10.1038/16483.
- [71] T.-M. M Yi, Yun Huang, Melvin I Simon, and John Doyle. Robust perfect adaptation in bacterial chemotaxis through integral feedback control. *Proc. Natl. Acad. Sci.*, 97 (9):4649–53, 2000. doi: 10.1073/pnas.97.9.4649.
- [72] Dale Muzzey, Carlos A. Gómez-Uribe, Jerome T Mettetal, and Alexander van Oude-naarden. A Systems-Level Analysis of Perfect Adaptation in Yeast Osmoregulation. *Cell*, 138(1):160–171, 2009. doi: 10.1016/j.cell.2009.04.047.
- [73] Wenzhe Ma, Ala Trusina, Hana El-Samad, Wendell A Lim, and Chao Tang. Defining Network Topologies that Can Achieve Biochemical Adaptation. *Cell*, 138(4):760–773, 2009. doi: 10.1016/j.cell.2009.06.013.
- [74] Neta Rahimi, Inna Averbukh, Michal Haskel-Ittah, Neta Degani, Eyal D Schejter, Naama Barkai, and Ben Zion Shilo. A WntD-Dependent Integral Feedback Loop Attenuates Variability in Drosophila Toll Signaling. *Dev. Cell*, 36(4):401–414, 2016. doi: 10.1016/j.devcel.2016.01.023.
- [75] Dirk Benzinger and Mustafa Khammash. Pulsatile inputs achieve tunable attenuation of gene expression variability and graded multi-gene regulation. *Nat. Commun.*, 9(1): 3521, dec 2018. doi: 10.1038/s41467-018-05882-2.
- [76] Miri Adler and Uri Alon. Fold-change detection in biological systems. *Curr. Opin. Syst. Biol.*, 8:81–89, apr 2018. doi: 10.1016/j.coisb.2017.12.005.
- [77] Pencho Yordanov and Jörg Stelling. Steady-State Differential Dose Response in Biological Systems. *Biophys. J.*, 114(3):723–736, feb 2018. doi: 10.1016/j.bpj.2017.11.3780.
- [78] David K. Lubensky, Matthew W. Pennington, Boris I. Shraiman, and Nicholas E. Baker. A dynamical model of ommatidial crystal formation. *Proc. Natl. Acad. Sci.*, 108(27):11145–11150, 2011. doi: 10.1073/pnas.1015302108.

- [79] Gustavo J. Melen, Sagi Levy, Naama Barkai, and Ben-Zion Shilo. Threshold responses to morphogen gradients by zero-order ultrasensitivity. *Mol. Syst. Biol.*, pages 1–11, 2005. doi: 10.1038/msb4100036.
- [80] Zhi-Chun Lai and Gerald M Rubin. Negative Control of Photoreceptor Development in Drosophila by the Product of the yan Gene, an ETS Domain Protein. *Cell*, 70(609420), 1992.
- [81] R. Rogge, P. J. Green, J. Urano, S. Horn-Saban, M. Mlodzik, B Z Shilo, V. Hartenstein, and U. Banerjee. The role of yan in mediating the choice between cell division and differentiation. *Development*, 121(12):3947–3958, 1995.
- [82] Ilaria Rebay and Gerald M. Rubin. Yan functions as a general inhibitor of differentiation and is negatively regulated by activation of the Ras1/MAPK pathway. *Cell*, 81:857–866, 1995.
- [83] Thomas G. W. Graham, S. M. Ali Tabei, Aaron R. Dinner, and Ilaria Rebay. Modeling bistable cell-fate choices in the Drosophila eye: qualitative and quantitative perspectives. *Development*, 137(14):2265–78, jul 2010. doi: 10.1242/dev.044826.
- [84] Arkadi Shwartz, Shaul Yogev, Eyal D. Schejter, and Ben-Zion Shilo. Sequential activation of ETS proteins provides a sustained transcriptional response to EGFR signaling. *Development*, 140(13):2746–2754, 2013. doi: 10.1242/dev.093138.
- [85] Jean-François Boisclair Lachance, Nicolás Peláez, Justin J. Cassidy, Jemma L. Webber, Ilaria Rebay, and Richard W. Carthew. A comparative study of Pointed and Yan expression reveals new complexity to the transcriptional networks downstream of receptor tyrosine kinase signaling. *Dev. Biol.*, 385:263–278, jan 2014. doi: 10.1016/j.ydbio.2013.11.002.
- [86] Matthew Freeman. Feedback control of intercellular signalling in development. *Nature*, 408(November):313–319, 2000. doi: 10.1038/35042500.
- [87] Uri Alon. Network motifs: theory and experimental approaches. *Nat. Rev. Genet.*, 8(6):450–61, 2007. doi: 10.1038/nrg2102.
- [88] Marcelo Behar, Nan Hao, Henrik G Dohlman, and Timothy C Elston. Mathematical and computational analysis of adaptation via feedback inhibition in signal transduction pathways. *Biophys. J.*, 93(3):806–21, 2007. doi: 10.1529/biophysj.107.107516.
- [89] Upinder S. Bhalla, Prahlad T. Ram, and Ravi Iyengar. MAP Kinase Phosphatase As a Locus of Flexibility in a Mitogen-Activated Protein Kinase Signaling Network. *Science* (80-. ), 297(5583):1018–1023, 2002. doi: 10.1126/science.1068873.
- [90] Raymond Cheong, Alex Rhee, C. J. Wang, Ilya Nemenman, and Andre Levchenko. Information Transduction Capacity of Noisy Biochemical Signaling Networks. *Science* (80-. ), 334(6054):354–358, 2011. doi: 10.1126/science.1204553.

- [91] Tau-Mu Yi, Hiroaki Kitano, and Melvin I Simon. A quantitative characterization of the yeast heterotrimeric G protein cycle. *Proc. Natl. Acad. Sci. U. S. A.*, 100(19):10764–9, 2003. doi: 10.1073/pnas.1834247100.
- [92] Mustafa Khammash. An engineering viewpoint on biological robustness, mar 2016.
- [93] Justin J. Cassidy, Alexander J. Straughan, and Richard W. Carthew. Differential masking of natural genetic variation by miR-9a in Drosophila. *Genetics*, 202(2):675–687, 2016. doi: 10.1534/genetics.115.183822.
- [94] Justin J Cassidy, Aashish R Jha, Diana M Posadas, Ritika Giri, Koen J T Venken, Jingran Ji, Hongmei Jiang, Hugo J Bellen, Kevin P White, and Richard W Carthew. MiR-9a minimizes the phenotypic impact of genomic diversity by buffering a transcription factor. *Cell*, 155(7):1556–1567, 2013. doi: 10.1016/j.cell.2013.10.057.
- [95] Xin Li, Justin J Cassidy, Catherine A Reinke, Stephen Fischboeck, and Richard W Carthew. A MicroRNA Imparts Robustness against Environmental Fluctuation during Development. *Cell*, 137(2):273–282, 2009. doi: 10.1016/j.cell.2009.01.058.
- [96] Margaret S. Ebert and Phillip A. Sharp. Roles for microRNAs in conferring robustness to biological processes. *Cell*, 149(3):515–524, 2012. doi: S0092-8674(12)00464-3[pii]10.1016/j.cell.2012.04.005.
- [97] CH H Waddington. Canalization of Development and the Inheritance of Acquired Characters. *Nature*, 150(3811):563–565, 1942. doi: 10.1038/150563a0.
- [98] K G Bateman. The genetic assimilation of four venation phenocopies. *J. Genet.*, 56(3):443–474, 1959. doi: 10.1007/BF02984796.
- [99] J M Rendel. Canalisation of the scute phenotype of Drosophila. *Evolution (N. Y.)*, 13:425–439, 1959. doi: 10.1111/j.1558-5646.1959.tb03033.x.
- [100] J M Rendel, B L Sheldon, and D E Finlay. Selection for Canalization of the Scute Phenotype. II. *Am. Nat.*, 100(910):13–31, 1966. doi: 10.2307/2459376.
- [101] Willem Scharloo. Canalization: genetic and developmental aspects. *Annu. Rev. Ecol. Syst.*, 22(1991):65–93, 1991. doi: 10.1146/annurev.es.22.110191.000433.
- [102] R.B. Dun and A.S. Fraser. Selection for an Invariant Character - 'Vibrissa Number' - in the House Mouse. *Nature*, 182:1638 – 1640, 1958.
- [103] G Gibson and David S. Hogness. Effect of Polymorphism in the Drosophila Regulatory Gene Ultrabithorax on Homeotic Stability. *Science (80-. ).*, 271(5246):200–203, 1996. doi: 10.1126/science.271.5246.200.
- [104] Suzanne L Rutherford and Susan Lindquist. Hsp90 as a capacitor for morphological evolution. *Nature*, 396(6709):336–342, 1998. doi: 10.1038/24550.

- [105] Avigdor Eldar, Ruslan Dorfman, Daniel Weiss, Hilary Ashe, Ben-Zion Shilo, and Naama Barkai. Robustness of the BMP morphogen gradient in Drosophila embryonic patterning. *Nature*, 419(September):304–308, 2002. doi: 10.1038/nature01061.
- [106] Charles M Denby, Joo Hyun Im, Richard C Yu, C. Gustavo Pesce, and Rachel B Brem. Negative feedback confers mutational robustness in yeast transcription factor regulation. *Proc. Natl. Acad. Sci.*, 109(10):3874–3878, 2012. doi: 10.1073/pnas.1116360109.
- [107] John L I V Hartman, Barbara Garvik, and Lee Hartwell. Principles for the Buffering of Genetic Variation. *Science (80-. ).*, 1001(2001):1–5, 2001. doi: 10.1126/science.1056072.
- [108] David C Marciano, Rhonald C Lua, Panagiotis Katsonis, Shivas R Amin, Christophe Herman, and Olivier Lichtarge. Negative Feedback in Genetic Circuits Confers Evolutionary Resilience and Capacitance. *Cell Rep.*, 7(6):1789–1795, 2014. doi: 10.1016/j.celrep.2014.05.018.
- [109] Hiroaki Kitano. Biological Robustness. *Nat. Rev. Genet.*, 837(November):826–837, 2004. doi: 10.1038/nrg1471.
- [110] Jörg Stelling, Uwe Sauer, Zoltan Szallasi, Francis J. Doyle, and John Doyle. Robustness of cellular functions, sep 2004.
- [111] Mark L Siegal and Jun-Yi Leu. On the Nature and Evolutionary Impact of Phenotypic Robustness Mechanisms. *Annu. Rev. Ecol. Evol. Syst.*, 45:495–517, 2014. doi: 10.1146/annurev-ecolsys-120213-091705.
- [112] Mark L Siegal and Aviv Bergman. Waddington’s canalization revisited: developmental stability and evolution. *Proc. Natl. Acad. Sci. U. S. A.*, 99:10528–10532, 2002. doi: 10.1073/pnas.102303999.
- [113] Aviv Bergman and Mark L. Siegal. Evolutionary capacitance as a general feature of complex gene networks. *Nature*, 424(6948):549–552, 2003. doi: 10.1038/nature01765.
- [114] Michael Lynch. The evolution of genetic networks by non-adaptive processes. *Nat. Rev. Genet.*, 8(10):803–13, 2007. doi: 10.1038/nrg2192.
- [115] Andreas Wagner. How the global structure of protein interaction networks evolves. *Proc. Biol. Sci.*, 270(1514):457–66, 2003. doi: 10.1098/rspb.2002.2269.
- [116] Harley H. McAdams and Adam Arkin. It’s a noisy business! Genetic regulation at the nanomolar scale, 1999.
- [117] M A Nowak, M C Boerlijst, J Cooke, and John Maynard Smith. Evolution of genetic redundancy. *Nature*, 388(6638):167–71, 1997. doi: 10.1038/40618.
- [118] Andreas Wagner. Robustness against mutations in genetic networks of yeast. *Nat. Genet.*, 24(4):355–61, 2000. doi: 10.1038/74174.

- [119] Dale Muzzey and Alexander van Oudenaarden. Quantitative Time-Lapse Fluorescence Microscopy in Single Cells. *Annu. Rev. Cell Dev. Biol.*, 25(1):301–327, nov 2009. doi: 10.1146/annurev.cellbio.042308.113408.
- [120] Ernst H K Stelzer. Light-sheet fluorescence microscopy for quantitative biology. *Nat. Methods*, 12(1):23–26, dec 2014. doi: 10.1038/nmeth.3219.
- [121] Thai V. Truong and Willy Supatto. Toward high-content/high-throughput imaging and analysis of embryonic morphogenesis. *Genesis*, 49(7):555–569, jul 2011. doi: 10.1002/dvg.20760.
- [122] Ian T. Simpson and David J. Price. Pax6; a pleiotropic player in development, nov 2002.
- [123] T R Parody and M. A T Muskavitch. The pleiotropic function of Delta during postembryonic development of *Drosophila melanogaster*. *Genetics*, 135(2):527–539, 1993.
- [124] Ben Zion Shilo and Erez Raz. Developmental control by the *Drosophila* EGF receptor homolog DER, nov 1991.
- [125] T Xu and G M Rubin. Analysis of genetic mosaics in developing and adult *Drosophila* tissues. *Development*, 117(4):1223–37, apr 1993. doi: 8404527.
- [126] Tian Xu and Gerald M Rubin. The effort to make mosaic analysis a household tool. *Development*, 139(24):4501–4503, dec 2012. doi: 10.1242/dev.085183.
- [127] T P Newsome, B Asling, and B J Dickson. Analysis of *Drosophila* photoreceptor axon guidance in eye-specific mosaics. *Development*, 127(4):851–60, feb 2000.
- [128] Nicole A. Theodosiou and Tian Xu. Use of FLP/FRT system to study *Drosophila* development. *Methods A Companion to Methods Enzymol.*, 14(4):355–365, apr 1998. doi: 10.1006/meth.1998.0591.
- [129] Gary Struhl and Konrad Basler. Organizing activity of wingless protein in *Drosophila*. *Cell*, 1993. doi: 10.1016/0092-8674(93)90072-X.
- [130] K Halfar, C Rommel, H Stocker, and E Hafen. Ras controls growth, survival and differentiation in the *Drosophila* eye by different thresholds of MAP kinase activity. *Development*, 128(9):1687–96, may 2001.
- [131] A Tomlinson and G Struhl. Delta/Notch and Boss/Sevenless signals act combinatorially to specify the *Drosophila* R7 photoreceptor. *Mol. Cell*, 7(3):487–95, mar 2001.
- [132] L Yang and N E Baker. Role of the EGFR/Ras/Raf pathway in specification of photoreceptor cells in the *Drosophila* retina. *Development*, 128(7):1183–91, apr 2001.
- [133] Jianbin Huang, Shian Wu, Jose Barrera, Krista Matthews, and Duoqia Pan. The Hippo signaling pathway coordinately regulates cell proliferation and apoptosis by

- inactivating Yorkie, the Drosophila homolog of YAP. *Cell*, 122(3):421–434, aug 2005. doi: 10.1016/j.cell.2005.06.007.
- [134] Barry J. Thompson and Stephen M. Cohen. The Hippo Pathway Regulates the bantam microRNA to Control Cell Proliferation and Apoptosis in Drosophila. *Cell*, 126(4):767–774, aug 2006. doi: 10.1016/j.cell.2006.07.013.
- [135] Mardelle Atkins. Drosophila Genetics: The Power of Genetic Mosaic Approaches. In *Methods Mol. Biol.*, volume 1893, pages 27–42. Humana Press, New York, NY, 2019. doi: 10.1007/978-1-4939-8910-2\_2.
- [136] Wei Dai, Amy Peterson, Thomas Kenney, Haley Burrous, and Denise J. Montell. Quantitative microscopy of the Drosophila ovary shows multiple niche signals specify progenitor cell fate. *Nat. Commun.*, 8(1):1244, dec 2017. doi: 10.1038/s41467-017-01322-9.
- [137] Sebastian M. Bernasek, Jean-François Boisclair Lachance, Nicolás Peláez, Rachael Bakker, Heliodoro Tejedor Navarro, Luis A. N. Amaral, Neda Bagheri, Ilaria Rebay, and Richard W. Carthew. Ratio-based sensing of two transcription factors regulates the transit to differentiation. *bioRxiv*, page 430744, sep 2018. doi: <http://dx.doi.org/10.1101/430744>.
- [138] Christian Ghiglione, Patrick Jouandin, Delphine Cérézo, and Stéphane Noselli. The Drosophila insulin pathway controls Profilin expression and dynamic actin-rich protrusions during collective cell migration. *Development*, 145(14):dev161117, jul 2018. doi: 10.1242/dev.161117.
- [139] Ke Li and Nicholas E Baker. Regulation of the Drosophila ID protein Extra macrochaetae by proneural dimerization partners. *Elife*, 7, 2018. doi: 10.7554/elife.33967.
- [140] Kirsten Bacia, Zdeněk Petrášek, and Petra Schwille. Correcting for spectral cross-talk in dual-color fluorescence cross-correlation spectroscopy. *ChemPhysChem*, 13(5):1221–1231, apr 2012. doi: 10.1002/cphc.201100801.
- [141] Masilamani Elangovan, Horst Wallrabe, Ye Chen, Richard N. Day, Margarida Barroso, and Ammasi Periasamy. Characterization of one- and two-photon excitation fluorescence resonance energy transfer microscopy. *Methods*, 29(1):58–73, jan 2003. doi: 10.1016/S1046-2023(02)00283-9.
- [142] Richard L. Mort. Quantitative analysis of patch patterns in mosaic tissues with ClonalTools software. *J. Anat.*, 215(6):698–704, dec 2009. doi: 10.1111/j.1469-7580.2009.01150.x.
- [143] Jo A Helmuth, Grégory Paul, and Ivo F Sbalzarini. Beyond co-localization: Inferring spatial interactions between sub-cellular structures from microscopy images. *BMC Bioinformatics*, 11(1):372, jul 2010. doi: 10.1186/1471-2105-11-372.

- [144] Arun Shivanandan, Aleksandra Radenovic, and Ivo F Sbalzarini. MosaicIA: An ImageJ/Fiji plugin for spatial pattern and interaction analysis. *BMC Bioinformatics*, 14(1):349, dec 2013. doi: 10.1186/1471-2105-14-349.
- [145] Stéfan van der Walt, Johannes L. Schönberger, Juan Nunez-Iglesias, François Boulogne, Joshua D. Warner, Neil Yager, Emmanuelle Gouillart, and Tony Yu. scikit-image: image processing in Python. *PeerJ*, 2014. doi: 10.7717/peerj.453.
- [146] Nobuyuki Otsu. A Threshold Selection Method from Gray-Level Histograms. *IEEE Trans. Syst. Man. Cybern.*, 1979.
- [147] Milica Bugarski, Maysam Mansouri, Axel Niemann, Aurélien Rizk, Philipp Berger, Urs Ziegler, Ivo F Sbalzarini, Pietro Incardona, and Grégory Paul. Segmentation and quantification of subcellular structures in fluorescence microscopy images using Squash. *Nat. Protoc.*, 9(3):586–596, mar 2014. doi: 10.1038/nprot.2014.037.
- [148] Vadim Zinchuk, Olga Zinchuk, and Teruhiko Okada. Quantitative Colocalization Analysis of Multicolor Confocal Immunofluorescence Microscopy Images: Pushing Pixels to Explore Biological Phenomena. *Acta Histochem. Cytochem.*, 40(4):101–111, 2007. doi: 10.1267/ahc.07002.
- [149] Paul T. Arsenovic, Carl R. Mayer, and Daniel E. Conway. SensorFRET: A Standardless Approach to Measuring Pixel-based Spectral Bleed-through and FRET Efficiency using Spectral Imaging. *Sci. Rep.*, 7(1), dec 2017. doi: 10.1038/s41598-017-15411-8.
- [150] Dahan Kim, Nikki M Curthoys, Matthew T Parent, and Samuel T Hess. Bleed-through correction for rendering and correlation analysis in multi-colour localization microscopy. *J. Opt.*, 15(9), sep 2013. doi: 10.1088/2040-8978/15/9/094011.
- [151] Patrick D. McMullen, Richard I. Morimoto, and Luís A. Nunes Amaral. Physically grounded approach for estimating gene expression from microarray data. *Proc. Natl. Acad. Sci.*, 107(31):13690–13695, aug 2010. doi: 10.1073/pnas.1000938107.
- [152] Chikara Furusawa, Takao Suzuki, Akiko Kashiwagi, Tetsuya Yomo, and Kunihiko Kaneko. Ubiquity of Log-normal Distributions in Intra-cellular Reaction Dynamics. *Biophysics (Oxf.)*, 1:25–31, 2005. doi: 10.2142/biophysics.1.25.
- [153] Jacob Beal. Biochemical complexity drives log-normal variation in genetic expression. *Eng. Biol.*, 1(1):55–60, jun 2017. doi: 10.1049/enb.2017.0004.
- [154] Henry Teicher. Identifiability of Finite Mixtures. *Ann. Math. Stat.*, 34(4):1265–1269, dec 1963. doi: 10.1214/aoms/1177703862.
- [155] M. Rosvall, D. Axelsson, and C. T. Bergstrom. The map equation. *Eur. Phys. J. Spec. Top.*, 2009. doi: 10.1140/epjst/e2010-01179-1.
- [156] Leo Katz. A new status index derived from sociometric analysis. *Psychometrika*, 1953. doi: 10.1007/BF02289026.

- [157] Tomihisa Kamada and Satoru Kawai. An algorithm for drawing general undirected graphs. *Inf. Process. Lett.*, 31(1):7–15, apr 1989. doi: 10.1016/0020-0190(89)90102-6.
- [158] Lisa Gaudette and Nathalie Japkowicz. Evaluation methods for ordinal classification. In *Lect. Notes Comput. Sci.*, volume 5549 LNAI, pages 207–210. Springer, Berlin, Heidelberg, 2009. doi: 10.1007/978-3-642-01818-3\_25.
- [159] Thanh Minh Nguyen and Q. M.Jonathan Wu. Gaussian mixture-model-based spatial neighborhood relationships for pixel labeling problems. *IEEE Trans. Syst. Man. Cybern.*, 42(1):193–202, feb 2012. doi: 10.1109/TSMCB.2011.2161284.
- [160] Alexis Gambis, Pierre Dourlen, Hermann Steller, and Bertrand Mollereau. Two-color *in vivo* imaging of photoreceptor apoptosis and development in *Drosophila*. *Dev. Biol.*, 351(1):128–134, mar 2011. doi: 10.1016/j.ydbio.2010.12.040.
- [161] Pierre Dourlen, Clemence Levet, Alexandre Mejat, Alexis Gambis, and Bertrand Mollereau. The Tomato/GFP-FLP/FRT Method for Live Imaging of Mosaic Adult *Drosophila* Photoreceptor Cells. *J. Vis. Exp.*, 79, sep 2013. doi: 10.3791/50610.
- [162] Yvette E Fisher, Helen H Yang, Jesse Isaacman-Beck, Marjorie Xie, Daryl M Gohl, and Thomas R Clandinin. FlpStop, a tool for conditional gene control in *Drosophila*. *Elife*, 6, 2017. doi: 10.7554/eLife.22279.
- [163] Joy S Wu and Liqun Luo. A protocol for mosaic analysis with a repressible cell marker (MARCM) in *Drosophila*. *Nat. Protoc.*, 1(6):2583–2589, jan 2007. doi: 10.1038/nprot.2006.320.
- [164] Qingxiang Zhou, Scott J. Neal, and Francesca Pignoni. Mutant analysis by rescue gene excision: New tools for mosaic studies in *Drosophila*. *Genesis*, 54(11):589–592, nov 2016. doi: 10.1002/dvg.22984.
- [165] Elleard Heffern, Norbert Perrimon, Amber M Hohl, Alberto del Valle Rodriguez, Chris Bakal, Marianne Bonvin, Anne Sustar, Didier Grunwald, Ruth Griffin, Christians Villalta, C-ting Wu, Jack R Bateman, Claude Desplan, Richard Binari, and Gerold Schubiger. The twin spot generator for differential *Drosophila* lineage analysis. *Nat. Methods*, 6(8):600–602, aug 2009. doi: 10.1038/nmeth.1349.
- [166] Hung Hsiang Yu, Chih Fei Kao, Yisheng He, Peng Ding, Jui Chun Kao, and Tzumin Lee. A complete developmental sequence of a *Drosophila* neuronal lineage as revealed by twin-spot MARCM. *PLoS Biol.*, 8(8):39–40, aug 2010. doi: 10.1371/journal.pbio.1000461.
- [167] Alexandru S. Denes, Emmanuel Caussinus, Markus Affolter, Oguz Kanca, and Anthony Percival-Smith. Raeppli: a whole-tissue labeling tool for live imaging of *Drosophila* development. *Development*, 141(2):472–480, jan 2013. doi: 10.1242/dev.102913.
- [168] Dafni Hadjieconomou, Shay Rotkopf, Cyrille Alexandre, Donald M Bell, Barry J Dickson, and Iris Salecker. Flybow: Genetic multicolor cell labeling for neural circuit analysis in *Drosophila melanogaster*. *Nat. Methods*, 8(3):260–266, mar 2011. doi:

10.1038/nmeth.1567.

- [169] Stefanie Hampel, Phuong Chung, Claire E McKellar, Donald Hall, Loren L Looger, and Julie H Simpson. Drosophila Brainbow: a recombinase-based fluorescence labeling technique to subdivide neural expression patterns. *Nat. Methods*, 8(3):253–259, mar 2011. doi: 10.1038/nmeth.1566.
- [170] Thomas P Neufeld, Aida Flor A. De La Cruz, Laura A Johnston, and Bruce A Edgar. Coordination of growth and cell division in the Drosophila wing. *Cell*, 93(7):1183–1193, jun 1998. doi: 10.1016/S0092-8674(00)81462-2.
- [171] Michael Tworoger, Michele Keller Larkin, Zev Bryant, and Hannele Ruohola-Baker. Mosaic analysis in the Drosophila ovary reveals a common Hedgehog- inducible precursor stage for stalk and polar cells. *Genetics*, 1999.
- [172] Russell T Collins, Claudia Linker, and Julian Lewis. MAZe: A tool for mosaic analysis of gene function in zebrafish. *Nat. Methods*, 7(3):219–223, mar 2010. doi: 10.1038/nmeth.1423.
- [173] Celia Muñoz-Jiménez, Cristina Ayuso, Agnieszka Dobrzynska, Antonio Torres-Mendéz, Patricia de la Cruz Ruiz, and Peter Askjaer. An efficient FLP-based toolkit for spatiotemporal control of gene expression in *Caenorhabditis elegans*. *Genetics*, 206(4):1763–1778, aug 2017. doi: 10.1534/genetics.117.201012.
- [174] Wei Wang, Madhuri Warren, and Allan Bradley. Induced mitotic recombination of p53 in vivo. *Proc. Natl. Acad. Sci.*, 104(11):4501–4505, mar 2007. doi: 10.1073/pnas.0607953104.
- [175] Erik Meijering. Cell Segmentation: 50 Years Down the Road. *IEEE Signal Process. Mag.*, 29(5):140–145, sep 2012. doi: 10.1109/msp.2012.2204190.
- [176] Wei-ping Zheng and Richard A. Flavell. The transcription factor GATA-3 is necessary and sufficient for Th2 cytokine gene expression in CD4 T cells. *Cell*, 89(4):587–596, may 1997. doi: 10.1016/S0092-8674(00)80240-8.
- [177] Patricia Ducy, Rui Zhang, Valérie Geoffroy, Amy L. Ridall, and Gérard Karsenty. Osf2/Cbfa1: a transcriptional activator of osteoblast differentiation. *Cell*, 89(5):747–754, may 1997. doi: 10.1016/S0092-8674(00)80257-3.
- [178] James D. McGhee, Tetsunari Fukushige, Michael W. Krause, Stephanie E. Minnema, Barbara Goszczyński, Jeb Gaudet, Yuji Kohara, Olaf Bossinger, Yongjun Zhao, Jaswinder Khattra, Martin Hirst, Steven J.M. Jones, Marco A. Marra, Peter Ruzanov, Adam Warner, Richard Zapf, Donald G. Moerman, and John M. Kalb. ELT-2 is the predominant transcription factor controlling differentiation and function of the *C. elegans* intestine, from embryo to adult. *Dev. Biol.*, 327(2):551–565, 2009. doi: 10.1016/j.ydbio.2008.11.034.

- [179] Mary O'Riordan and Rudolf Grosschedl. Coordinate regulation of B cell differentiation by the transcription factors EBF and E2A. *Immunity*, 11:21–31, 1999. doi: 10.1016/S1074-7613(00)80078-3.
- [180] Rachel A. Evans, Y. Chung Tian, Robert Steadman, and Aled O. Phillips. TGF- $\beta$ 1-mediated fibroblast-myofibroblast terminal differentiation - The role of Smad proteins. *Exp. Cell Res.*, 282(2):90–100, jan 2003. doi: 10.1016/S0014-4827(02)00015-0.
- [181] Hao Y. Kueh, Ameya Champhekar, Stephen L. Nutt, Michael B. Elowitz, and Ellen V. Rothenberg. Positive feedback between PU.1 and the cell cycle controls myeloid differentiation. *Science (80-. ).*, 341:670–673, aug 2013. doi: 10.1126/science.1240831.
- [182] Guang Yao, Tae Jun Lee, Seiichi Mori, Joseph R. Nevins, and Lingchong You. A bistable Rb-E2F switch underlies the restriction point. *Nat. Cell Biol.*, 10(4):476–482, 2008. doi: 10.1038/ncb1711.
- [183] Byung Ouk Park, Robert Ahrends, and Mary N. Teruel. Consecutive Positive Feedback Loops Create a Bistable Switch that Controls Preadipocyte-to-Adipocyte Conversion. *Cell Rep.*, 2(4):976–990, oct 2012. doi: 10.1016/j.celrep.2012.08.038.
- [184] Peter Tontonoz, Erding Hu, and Bruce M Spiegelman. Stimulation of adipogenesis in fibroblasts by PPAR $\gamma$ 2, a lipid-activated transcription factor. *Cell*, 79(7):1147–1156, 1994. doi: 10.1016/0092-8674(94)90006-X.
- [185] Peter Laslo, Chauncey J. Spooner, Aryeh Warmflash, David W. Lancki, Hyun-Jun Lee, Roger Sciammas, Benjamin N. Gantner, Aaron R. Dinner, and Harinder Singh. Multilineage transcriptional priming and determination of alternate hematopoietic cell fates. *Cell*, 126(4):755–766, 2006. doi: 10.1016/j.cell.2006.06.052.
- [186] Hitoshi Niwa, Jun-Ichi Miyazaki, and Austin G. Smith. Quantitative expression of Oct-3/4 defines differentiation, dedifferentiation or self-renewal of ES cells. *Nat. Genet.*, 24(April):372–376, 2000.
- [187] Lea Goentoro and Marc W. Kirschner. Evidence that Fold-Change, and Not Absolute Level, of  $\beta$ -Catenin Dictates Wnt Signaling. *Mol. Cell*, 36(5):872–884, dec 2009. doi: 10.1016/j.molcel.2009.11.017.
- [188] Keita Kamino, Yohei Kondo, Akihiko Nakajima, Mai Honda-Kitahara, Kunihiko Kaneko, and Satoshi Sawai. Fold-change detection and scale invariance of cellcell signaling in social amoeba. *Proc. Natl. Acad. Sci.*, 114(21):E4149–E4157, may 2017. doi: 10.1073/pnas.1702181114.
- [189] Yaron E. Antebi, James M. Linton, Heidi Klumpe, Bogdan Bintu, Mengsha Gong, Christina Su, Reed McCardell, and Michael B. Elowitz. Combinatorial Signal Perception in the BMP Pathway. *Cell*, 170:1184–1196, sep 2017. doi: 10.1016/j.cell.2017.08.015.

- [190] Alan Bush, Gustavo Vasen, Andreas Constantinou, Paula Dunayevich, Inés Lucía Patop, Matías Blaustein, and Alejandro ColmanLerner. Yeast GPCR signaling reflects the fraction of occupied receptors, not the number. *Mol. Syst. Biol.*, 12:898, dec 2016. doi: 10.15252/msb.20166910.
- [191] Lea Goentoro, Oren Shoval, Marc W. Kirschner, and Uri Alon. The incoherent feedforward loop can provide fold-change detection in gene regulation. *Mol. Cell*, 36: 894–899, dec 2009. doi: 10.1016/j.molcel.2009.11.018.
- [192] Chris Q. Doe. Temporal patterning in the Drosophila CNS. *Annu. Rev. Cell Dev. Biol.*, 33(1):219–240, oct 2017. doi: 10.1146/annurev-cellbio-111315-125210.
- [193] Ted Erclik, Xin Li, Maximilien Courgeon, Claire Bertet, Zhenqing Chen, Ryan Baumert, June Ng, Clara Koo, Urfa Arain, Rudy Behnia, Alberto Del Valle Rodriguez, Lionel Senderowicz, Nicolas Negre, Kevin P. White, and Claude Desplan. Integration of temporal and spatial patterning generates neural diversity. *Nature*, 541(7637):365–370, jan 2017. doi: 10.1038/nature20794.
- [194] Limor Gabay, Henrike Scholz, Myriam Golembio, Andrea Klaes, Ben-Zion Shilo, and Christian Klambt. EGF receptor signaling induces pointed P1 transcription and inactivates Yan protein in the Drosophila embryonic ventral ectoderm. *Development*, 122:3355–3362, 1996.
- [195] Marc S. Halfon, Ana Carmena, Stephen Gisselbrecht, Charles M. Sackerson, Fernando Jiménez, Mary K. Baylies, and Alan M. Michelson. Ras pathway specificity is determined by the integration of multiple signal-activated and tissue-restricted transcription factors. *Cell*, 103(1):63–74, sep 2000. doi: 10.1016/S0092-8674(00)00105-7.
- [196] Alyssa M. Morimoto, Katherine C. Jordan, Kyria Tietze, Jessica S. Britton, Elizabeth M. O'Neill, and Hannele Ruohola-Baker. Pointed, an ETS domain transcription factor, negatively regulates the EGF receptor pathway in Drosophila oogenesis. *Development*, 122:3745–3754, 1996.
- [197] Chunyan Xu, Rachele C. Kauffmann, Jianjun Zhang, Susan Kladny, and Richard W. Carthew. Overlapping activators and repressors delimit transcriptional response to receptor tyrosine kinase signals in the Drosophila eye. *Cell*, 103:87–97, 2000. doi: 10.1016/S0092-8674(00)00107-0.
- [198] Gail V Flores, Hong Duan, Huajun Yan, Raghavendra Nagaraj, Weimin Fu, Yu Zou, Markus Noll, and Utpal Banerjee. Combinatorial signaling in the specification of unique cell fates. *Cell*, 103:75–85, sep 2000. doi: 10.1016/S0092-8674(00)00106-9.
- [199] D Brunner, K Dücker, N Oellers, E Hafen, H Scholz, and C Klämbt. The ETS domain protein pointed-P2 is a target of MAP kinase in the sevenless signal transduction pathway. *Nature*, 370(6488):386–9, aug 1994. doi: 10.1038/370386a0.
- [200] Elizabeth M. O'Neill, Ilaria Rebay, Robert Tjian, and Gerald Rubin. The activities of two ETS-related transcription factors required for Drosophila eye development are

- modulated by the Ras/MAPK pathway. *Cell*, 78:137–147, 1994.
- [201] C Klämbt. The *Drosophila* gene *pointed* encodes two ETS-like proteins which are involved in the development of the midline glial cells. *Development*, 117(1):163–176, 1993.
- [202] H. Scholz, J Deatrick, A Klaes, and C Klämbt. Genetic dissection of *pointed*, a *Drosophila* gene encoding two ETS-related proteins. *Genetics*, 135(2):455–468, 1993.
- [203] Gong Hong Wei, Gwenael Badis, Michael F. Berger, Teemu Kivioja, Kimmo Palin, Martin Enge, Martin Bonke, Arttu Jolma, Markku Varjosalo, Andrew R. Gehrke, Jian Yan, Shaheynoor Talukder, Mikko Turunen, Mikko Taipale, Hendrik G. Stunnenberg, Esko Ukkonen, Timothy R. Hughes, Martha L. Bulyk, and Jussi Taipale. Genome-wide analysis of ETS-family DNA-binding in vitro and in vivo. *EMBO J.*, 29(13):2147–2160, jul 2010. doi: 10.1038/emboj.2010.106.
- [204] Jemma L. Webber, Jie Zhang, Lauren Cote, Pavithra Vivekanand, Xiaochun Ni, Jie Zhou, Nicolas Negre, Richard W. Carthew, Kevin P. White, and Ilaria Rebay. The relationship between long-range chromatin occupancy and polymerization of the *drosophila* ETS family transcriptional repressor Yan. *Genetics*, 193(2):633–649, 2013. doi: 10.1534/genetics.112.146647.
- [205] Jemma L. Webber, Jie Zhang, Aaron Mitchell-Dick, and Ilaria Rebay. 3D chromatin interactions organize Yan chromatin occupancy and repression at the even-skipped locus. *Genes Dev.*, 27(21):2293–2298, nov 2013. doi: 10.1101/gad.225789.113.
- [206] Jean-François Boisclair Lachance, Jemma L Webber, Lu Hong, Aaron R Dinner, and Ilaria Rebay. Cooperative recruitment of Yan via a high-affinity ETS supersite organizes repression to confer specificity and robustness to cardiac cell fate specification. *Genes Dev.*, 32:389–401, mar 2018. doi: 10.1101/gad.307132.117.
- [207] Jemma L Webber, Jie Zhang, Alex Massey, Nicelio Sanchez-Luege, and Ilaria Rebay. Collaborative repressive action of the antagonistic ETS transcription factors Pointed and Yan fine-tunes gene expression to confer robustness in *Drosophila*. *Development*, 145(13):dev165985, may 2018. doi: 10.1242/dev.165985.
- [208] Matthew Freeman. Reiterative use of the EGF receptor triggers differentiation of all cell types in the *Drosophila* eye. *Cell*, 87:651–660, nov 1996. doi: 10.1016/S0092-8674(00)81385-9.
- [209] Tanya Wolff and Donald F. Ready. Pattern formation in the *Drosophila* retina. In M. Bate and A. Martinez-Arias, editors, *Dev. Drosoph. Melanogaster Vol. 2*, pages 1277–1325. Cold Spring Harbor Press, Plainview, NY, 1993.
- [210] Tianyi Zhang, Swati Ranade, Chuan Qi Cai, Christopher Clouser, and Francesca Pignoni. Direct control of neurogenesis by selector factors in the fly eye: regulation of *ataonal* by *Ey* and *So*. *Development*, 133(24):4881–4889, dec 2006. doi: 10.1242/dev.02669.

- [211] T. Wolff and Donalf F. Ready. Cell death in normal and rough eye mutants of *Drosophila*. *Development*, 113(3):825–839, 1991. doi: 8223268.
- [212] Mark E. Fortini, Ilaria Rebay, Laurent A. Caron, and Spyros Artavanis-Tsakonas. An activated Notch receptor blocks cell-fate commitment in the developing *Drosophila* eye. *Nature*, 365:555–557, oct 1993. doi: 10.1038/365555a0.
- [213] Nicolás Peláez. *Dynamics and heterogeneity of fate determinants during transition towards cell differentiation*. PhD thesis, Northwestern University, 2016.
- [214] Konrad Basler and Ernst Hafen. Dynamics of *Drosophila* eye development and temporal requirements of sevenless expression. *Development*, 107:723–731, 1989.
- [215] J. A. Campos-Ortega and A. Hofbauer. Cell Clones and Pattern Formation: On the Lineage of Photoreceptor Cells in the Compound Eye of *Drosophila*. *Roux's Arch. Dev. Biol.*, 181:227–245, 1977.
- [216] Andrew Tomlinson. The cellular dynamics of pattern formation in the eye of *Drosophila*. *Development*, 89:313–331, 1985.
- [217] Kazuhiro R. Nitta, Artuu Jolma, Yimeng Yin, Ekaterina Morgunova, Teemu Kivioja, Junaid Akhtar, Korneel Hens, Jarkko Toivonen, Bart Deplancke, Eileen E.M. Furlong, and Jussi Taipale. Conservation of transcription factor binding specificities across 600 million years of bilateria evolution. *eLife*, 4, 2015. doi: 10.7554/eLife.04837.
- [218] Zhi-Xin Wang. An exact mathematical expression for describing competitive binding of two different ligands to a protein molecule. *FEBS Lett.*, 360:111–114, 1995. doi: 10.1016/0014-5793(95)00062-E.
- [219] Feng Qiao, Haiyun Song, Chongwoo A. Kim, Michael R. Sawaya, Jacob B. Hunter, Mari Gingery, Ilaria Rebay, Albert J. Courey, and James U. Bowie. Derepression by depolymerization: Structural insights into the regulation of Yan by Mae. *Cell*, 118: 163–173, jul 2004. doi: 10.1016/j.cell.2004.07.010.
- [220] Matthew C. Hope, Ilaria Rebay, and John Reinitz. DNA Occupancy of polymerizing transcription factors: A chemical model of the ETS family factor Yan. *Biophys. J.*, 112(1):180–192, jan 2017. doi: 10.1016/j.bpj.2016.11.901.
- [221] H. S. Fogler. *Elements of chemical reaction engineering*. Prentice Hall, 1987. doi: 10.1016/0009-2509(87)80130-6.
- [222] John C. Lucchesi and John M. Rawls. Regulation of gene function: A comparison of enzyme activity levels in relation to gene dosage in diploids and triploids of *Drosophila melanogaster*. *Biochem. Genet.*, 9(1):41–51, may 1973. doi: 10.1007/BF00485589.
- [223] Margaret Rohrbaugh, Edward Ramos, Duc Nguyen, Mitch Price, Yu Wen, and Zhi-Chun Lai. Notch activation of yan expression is antagonized by RTK/Pointed signaling in the *Drosophila* eye. *Curr. Biol.*, 12:576–581, 2002.

- [224] Justin P. Kumar, Frank Hsiung, Maureen A. Powers, and Kevin Moses. Nuclear translocation of activated MAP kinase is developmentally regulated in the developing Drosophila eye. *Development*, 130(16):3703–14, apr 2003. doi: 10.1242/dev.00556.
- [225] Ross L. Cagan and Donald F. Ready. Notch is required for successive cell decisions in the developing Drosophila retina. *Genes Dev.*, 3:1099–1112, 1989. doi: 10.1101/gad.3.8.1099.
- [226] Yanxia Li and Nicholas E. Baker. Proneural enhancement by Notch overcomes suppressor-of-hairless repressor function in the developing Drosophila eye. *Curr. Biol.*, 11(5):330–338, mar 2001. doi: 10.1016/S0960-9822(01)00093-8.
- [227] Justin P. Kumar, Murni Tio, Frank Hsiung, Sevak Akopyan, Limor Gabay, Rony Seger, Ben-Zion Shilo, and Kevin Moses. Dissecting the roles of the Drosophila EGF receptor in eye development and MAP kinase activation. *Development*, 125(19):3875–3885, 1998.
- [228] Michael A. Simon, David D.L. Bowtell, G. Steven Dodson, Todd R. Laverty, and Gerald M. Rubin. Ras1 and a putative guanine nucleotide exchange factor perform crucial steps in signaling by the sevenless protein tyrosine kinase. *Cell*, 67(4):701–716, nov 1991. doi: 10.1016/0092-8674(91)90065-7.
- [229] Mark E. Fortini, Michael A. Simon, and Gerald M. Rubin. Signalling by the sevenless protein tyrosine kinase is mimicked by Ras1 activation. *Nature*, 355:559–61, 1992. doi: 10.1038/355559a0.
- [230] François Spitz and Eileen E. M. Furlong. Transcription factors: From enhancer binding to developmental control. *Nat. Rev. Genet.*, 13(9):613–626, sep 2012. doi: 10.1038/nrg3207.
- [231] Raghavendra Nagaraj, Jude Canon, and Utpal Banerjee. Cell fate specification in the Drosophila eye. In *Results Probl. Cell Differ. Vol. 37*, pages 73–88. Springer, Berlin, Heidelberg, 2002. doi: 10.1007/978-3-540-45398-7\_6.
- [232] Yihan Lin, Chang Ho Sohn, Chiraj K. Dalal, Long Cai, and Michael B. Elowitz. Combinatorial gene regulation by modulation of relative pulse timing. *Nature*, 527: 54–58, nov 2015. doi: 10.1038/nature15710.
- [233] Chongwoo A. Kim, Martin L. Phillips, Woojae Kim, Mari Gingery, Hoang H. Tran, Michael A. Robinson, Salem Faham, and James U. Bowie. Polymerization of the SAM domain of TEL in leukemogenesis and transcriptional repression. *EMBO J.*, 20(15): 4173–4182, 2001. doi: 10.1093/emboj/20.15.4173.
- [234] Jie Zhang, Thomas G.W. Graham, Pavithra Vivekanand, Lauren Cote, Maureen Cetera, and Ilaria Rebay. Sterile alpha motif domain-mediated self-association plays an essential role in modulating the activity of the Drosophila ETS family transcriptional repressor Yan. *Mol. Cell. Biol.*, 30(5):1158–1170, mar 2010. doi: 10.1128/MCB.01225-09.

- [235] Lihui Yang and Nicholas E. Baker. Notch activity opposes ras-induced differentiation during the second mitotic wave of the developing Drosophila eye. *BMC Dev. Biol.*, 6(8):1–10, feb 2006. doi: 10.1186/1471-213X-6-8.
- [236] Miki Ebisuya and James Briscoe. What does time mean in development? *Development*, 145:1–6, 2018. doi: 10.1242/dev.164368.
- [237] V E Foe. Mitotic domains reveal early commitment of cells in Drosophila embryos. *Development*, 107(1):1–22, 1989.
- [238] J E Sulston, E Schierenberg, J G White, and J N Thomson. The embryonic cell lineage of the nematode *Caenorhabditis elegans*. *Dev Biol*, 100(1):64–119, 1983. doi: 0012-1606(83)90201-4[pii].
- [239] M Francesconi and B Lehner. The effects of genetic variation on gene expression dynamics during development. *Nature*, 505(7482):208–211, 2014. doi: 10.1038/nature12772[nature12772][pii].
- [240] N Poulet, A Vielle, C Gimond, S Carvalho, H Teotonio, and C Braendle. Complex heterochrony underlies the evolution of *Caenorhabditis elegans* hermaphrodite sex allocation. *Evolution (N. Y.)*, 70(10):2357–2369, 2016. doi: 10.1111/evo.13032.
- [241] Eric G Moss. Heterochronic genes and the nature of developmental time. *Curr. Biol.*, 17(11):R425–R434, 2007.
- [242] M Atlas. The effect of temperature on the development of *Rana pipiens*. *Physiol. Zool.*, 8(3):290–310, 1935.
- [243] J Davidson. On the relationship between temperature and rate of development of insects at constant temperatures. *J. Anim. Ecol.*, pages 26–38, 1944.
- [244] S G Kuntz and M B Eisen. Drosophila embryogenesis scales uniformly across temperature in developmentally diverse species. *PLoS Genet*, 10(4):e1004293, 2014. doi: 10.1371/journal.pgen.1004293PGGENETICS-D-13-02464[pii].
- [245] Wenyun Zuo, Melanie E Moses, Geoffrey B West, Chen Hou, and James H Brown. A general model for effects of temperature on ectotherm ontogenetic growth and development. *Proc. R. Soc. B Biol. Sci.*, 279(November):1840–1846, 2011. doi: 10.1098/rspb.2011.2000.
- [246] J D Arendt. Adaptive intrinsic growth rates: An integration across taxa. *Q. Rev. Biol.*, 72(2):149–177, 1997.
- [247] J H Brown, J F Gillooly, A P Allen, V N Savage, and G B West. Toward a metabolic theory of ecology. *Ecology*, 85(7):1771–1789, 2004.
- [248] N B Metcalfe and P Monaghan. Compensation for a bad start: grow now, pay later? *Trends Ecol Evol*, 16(5):254–260, 2001. doi: S0169534701021243[pii].

- [249] Herman Pontzer, Mary H Brown, David A Raichlen, Holly Dunsworth, Brian Hare, Kara Walker, Amy Luke, Lara R Dugas, Ramon Durazo-Arvizu, Dale Schoeller, Jacob Plange-Rhule, Pascal Bovet, Terrence E Forrester, Estelle V Lambert, Melissa Emery Thompson, Robert W Shumaker, and Stephen R Ross. Metabolic acceleration and the evolution of human brain size and life history. *Nature*, 533(7603):390–392, 2016. doi: 10.1038/nature17654.
- [250] S Felkai, J J Ewbank, J Lemieux, J C Labbe, G G Brown, and S Hekimi. CLK-1 controls respiration, behavior and aging in the nematode *Caenorhabditis elegans*. *EMBO J*, 18(7):1783–1792, 1999. doi: 10.1093/emboj/18.7.1783.
- [251] F Levavasseur, H Miyadera, J Sirois, M L Tremblay, K Kita, E Shoubridge, and S Hekimi. Ubiquinone is necessary for mouse embryonic development but is not essential for mitochondrial respiration. *J Biol Chem*, 276(49):46160–46164, 2001. doi: 10.1074/jbc.M108980200M108980200[pii].
- [252] D Nakai, S Yuasa, M Takahashi, T Shimizu, S Asaumi, K Isono, T Takao, Y Suzuki, H Kuroyanagi, K Hirokawa, H Koseki, and T Shirasawa. Mouse homologue of coq7/clk-1, longevity gene in *Caenorhabditis elegans*, is essential for coenzyme Q synthesis, maintenance of mitochondrial integrity, and neurogenesis. *Biochem Biophys Res Commun*, 289(2):463–471, 2001. doi: 10.1006/bbrc.2001.5977S0006291X01959777[pii].
- [253] A Wong, P Boutis, and S Hekimi. Mutations in the clk-1 gene of *Caenorhabditis elegans* affect developmental and behavioral timing. *Genetics*, 139(3):1247–1259, 1995.
- [254] W Brogiolo, H Stocker, T Ikeya, F Rintelen, R Fernandez, and E Hafen. An evolutionarily conserved function of the *Drosophila* insulin receptor and insulin-like peptides in growth control. *Curr Biol*, 11(4):213–221, 2001. doi: S0960-9822(01)00068-9[pii].
- [255] Sophie Layalle, Nathalie Arquier, and Pierre Léopold. The TOR pathway couples nutrition and developmental timing in *Drosophila*. *Dev. Cell*, 15(4):568–577, 2008.
- [256] Eric J Rulifson, Seung K Kim, and Roel Nusse. Ablation of insulin-producing neurons in flies: growth and diabetic phenotypes. *Science (80-. ).*, 296(5570):1118–1120, 2002. doi: 10.1126/science.1070058296/5570/1118[pii].
- [257] A W Shingleton, J Das, L Vinicius, and D L Stern. The temporal requirements for insulin signaling during development in *Drosophila*. *PLoS Biol*, 3(9):e289, 2005. doi: 10.1371/journal.pbio.0030289.
- [258] James. F. Gillooly, Eric L. Charnov, Geoffrey B. West, Van M. Savage, and James H. Brown. Effects of size and temperature on developmental time. *Nature*, 417(6884):70–3, may 2002. doi: 10.1038/417070a.
- [259] D A Cusanovich, J P Reddington, D A Garfield, R M Daza, D Aghamirzaie, R Marco-Ferreres, H A Pliner, L Christiansen, X Qiu, F J Steemers, C Trapnell, J Shendure, and E E M Furlong. The cis-regulatory dynamics of embryonic development at single-cell resolution. *Nature*, 555(7697):538–542, 2018. doi: 10.1038/nature25981nature25981[pii].

- [260] P Lawrence. *The making of a fly: the genetics of animal design.* Blackwell Scientific, Oxford, UK, 1992.
- [261] M Kohwi and C Q Doe. Temporal fate specification and neural progenitor competence during development. *Nat Rev Neurosci*, 14(12):823–838, 2013.
- [262] C Cepko. Intrinsically different retinal progenitor cells produce specific types of progeny. *Nat Rev Neurosci*, 15(9):615–627, 2014. doi: 10.1038/nrn3767nrn3767[pii].
- [263] Richard J Julian. Production and growth related disorders and other metabolic diseases of poultrya review. *Vet. J.*, 169(3):350–369, 2005.
- [264] Colin C Whitehead, Robert H Fleming, Richard J Julian, and P Sørensen. Skeletal problems associated with selection for increased production. *Poult. Genet. Breed. Biotechnol.*, pages 29–52, 2003.
- [265] G Child. The effect of increasing time of development at constant temperature on the wing size of vestigial of *Drosophila melanogaster*. *Biol. Bull.*, 77(2):432–442, 1939.
- [266] T H Morgan. *The mechanism of Mendelian heredity.* Holt, New York, 1915.
- [267] T H Morgan. Variability in eyeless. *Carnegie Inst. Wash. Publ.*, 399:139–168, 1929.
- [268] J H Sang and B Burnet. Environmental Modification of the Eyeless Phenotype in *Drosophila Melanogaster*. *Genetics*, 48:1683–1699, 1963.
- [269] G Child. Phenogenetic Studies on Scute-1 of *Drosophila Melanogaster*. I. The Associations between the Bristles and the effects of Genetic Modifiers and Temperature. *Genetics*, 20(2):109–126, 1935.
- [270] J Krafka. The Effect of Temperature Upon Facet Number in the Bar-Eyed Mutant of *Drosophila* : Part I. *J Gen Physiol*, 2(4):409–432, 1920.
- [271] R A Lewis, B T Wakimoto, R E Denell, and T C Kaufman. Genetic Analysis of the Antennapedia Gene Complex (Ant-C) and Adjacent Chromosomal Regions of DROSOPHILA MELANOGASTER. II. Polytene Chromosome Segments 84A-84B1,2. *Genetics*, 95(2):383–397, 1980.
- [272] C Villee. Phenogenetic studies of the homoeotic mutants of *Drosophila melanogaster*.1. The effects of temperature on the expression of *aristapedia*. *J Exp Zool*, 93:75–98, 1943.
- [273] X J Quan and B A Hassan. From skin to nerve: flies, vertebrates and the first helix. *Cell Mol Life Sci*, 62(18):2036–2049, 2005. doi: 10.1007/s00018-005-5124-1.
- [274] U Banerjee, P J Renfranz, D R Hinton, B A Rabin, and S Benzer. The sevenless+ protein is expressed apically in cell membranes of developing *Drosophila* retina; it is not restricted to cell R7. *Cell*, 51(1):151–158, 1987.

- [275] H Jafar-Nejad, M Acar, R Nolo, H Lacin, H Pan, S M Parkhurst, and H J Bellen. Senseless acts as a binary switch during sensory organ precursor selection. *Genes Dev*, 17(23):2966–2978, 2003.
- [276] Yan Li, Fay Wang, Jin A. Lee, and Fen Biao Gao. MicroRNA-9a ensures the precise specification of sensory organ precursors in Drosophila. *Genes Dev*, 20(20):2793–2805, 2006. doi: 10.1101/gad.1466306.
- [277] Mark Van Doren, Adina M Bailey, Joan Esnayra, Kekoa Ede, and James W. Posakony. Negative regulation of proneural gene activity: Hairy is a direct transcriptional repressor of achaete. *Genes Dev.*, 8(22):2729–2742, 1994. doi: 10.1101/gad.8.22.2729.
- [278] Carl J Neumann and Stephen M Cohen. Sternopleural is a regulatory mutation of wingless with both dominant and recessive effects on larval development of Drosophila melanogaster. *Genetics*, 142(4):1147–1155, 1996.
- [279] A Kozomara and S Griffiths-Jones. miRBase: annotating high confidence microRNAs using deep sequencing data. *Nucleic Acids Res*, 42(Database issue):D68–73, 2014. doi: 10.1093/nar/gkt1181gkt1181[pii].
- [280] Richard W Carthew and Erik J Sontheimer. Origins and Mechanisms of miRNAs and siRNAs, 2009.
- [281] S Pressman, C A Reinke, X Wang, and R W Carthew. A Systematic Genetic Screen to Dissect the MicroRNA Pathway in Drosophila. *G3*, 2(4):437–448, 2012. doi: 10.1534/g3.112.002030GGG\_002030[pii].
- [282] D. Seborg, T. Edgar, D. Mellicamp, and F. Doyle III. *Process Dynamics and Control (3rd Edition)*. John Wiley & Sons, 2011. doi: 10.1007/s13398-014-0173-7.2.
- [283] A M Arias and P Hayward. Filtering transcriptional noise during development: concepts and mechanisms. *Nat Rev Genet*, 7(1):34–44, 2006. doi: nrg1750[pii]10.1038/nrg1750.
- [284] G C Brown. Control of respiration and ATP synthesis in mammalian mitochondria and cells. *Biochem J*, 284 ( Pt 1):1–13, 1992.
- [285] D E Atkinson. *Cellular Energy Metabolism and Its Regulation*. Academic Press, New York, 1977.
- [286] H Lempiainen and D Shore. Growth control and ribosome biogenesis. *Curr Opin Cell Biol*, 21(6):855–863, 2009. doi: 10.1016/j.ceb.2009.09.002S0955-0674(09)00158-6[pii].
- [287] S J Marygold, J Roote, G Reuter, A Lambertsson, M Ashburner, G H Millburn, P M Harrison, Z Yu, N Kenmochi, T C Kaufman, S J Leevers, and K R Cook. The ribosomal protein genes and Minute loci of Drosophila melanogaster. *Genome Biol*, 8(10):R216, 2007. doi: gb-2007-8-10-r216[pii]10.1186/gb-2007-8-10-r216.

- [288] Stein Sæbøe-Larsen, May Lyamouri, John Merriam, Morten P Oksvold, and Andrew Lambertsson. Ribosomal protein insufficiency and the minute syndrome in drosophila: A dose-response relationship. *Genetics*, 148(3):1215–1224, 1998. doi: 10.1073/PNAS.89.23.11302.
- [289] R Nolo, L A Abbott, and H J Bellen. Senseless, a Zn finger transcription factor, is necessary and sufficient for sensory organ development in Drosophila. *Cell*, 102(3):349–362, 2000. doi: S0092-8674(00)00040-4[pii].
- [290] H Jafar-Nejad, A C Tien, M Acar, and H J Bellen. Senseless and Daughterless confer neuronal identity to epithelial cells in the Drosophila wing margin. *Development*, 133(9):1683–1692, 2006. doi: dev.02338[pii]10.1242/dev.02338.
- [291] K J Venken, Y He, R A Hoskins, and H J Bellen. P[acman]: a BAC transgenic platform for targeted insertion of large DNA fragments in *D. melanogaster*. *Science (80-.)*, 314(5806):1747–1751, 2006.
- [292] N Bushati and S M Cohen. microRNA functions. *Annu Rev Cell Dev Biol*, 23:175–205, 2007.
- [293] R W Carthew, P Agbu, and R Giri. MicroRNA function in *Drosophila melanogaster*. *Semin Cell Dev Biol*, 65:29–37, 2017. doi: S1084-9521(16)30076-3[pii]10.1016/j.semcd.2016.03.015.
- [294] Matthew J Brauer, Curtis Huttenhower, Edoardo M Airoldi, Rachel Rosenstein, John C Mateese, David Gresham, Viktor M Boer, Olga G Troyanskaya, and David Botstein. Coordination of growth rate, cell cycle, stress response, and metabolic activity in yeast. *Mol Biol Cell*, 19(1):352–367, jan 2008. doi: E07-08-0779[pii]10.1091/mbc.E07-08-0779.
- [295] C Lu, M J Brauer, and D Botstein. Slow growth induces heat-shock resistance in normal and respiratory-deficient yeast. *Mol Biol Cell*, 20(3):891–903, 2009. doi: 10.1091/mbc.E08-08-0852E08-08-0852[pii].
- [296] M O Casanueva, A Burga, and B Lehner. Fitness trade-offs and environmentally induced mutation buffering in isogenic *C. elegans*. *Science (80-.)*, 335(6064):82–85, 2012. doi: 10.1126/science.1213491science.1213491[pii].
- [297] D F Jarosz and S Lindquist. Hsp90 and environmental stress transform the adaptive value of natural genetic variation. *Science (80-.)*, 330(6012):1820–1824, 2010. doi: 330/6012/1820[pii]10.1126/science.1195487.
- [298] Marieke Visscher, Sasha De Henau, Mattheus H.E. Wildschut, Robert M. van Es, Ineke Dhondt, Helen Michels, Patrick Kemmeren, Ellen A. Nollen, Bart P. Braeckman, Boudeijn M.T. Burginger, Harmjan R. Vos, and Tobias B. Dansen. Proteome-wide Changes in Protein Turnover Rates in *C. elegans* Models of Longevity and Age-Related Disease. *Cell Rep.*, 16(11):3041–3051, sep 2016. doi: 10.1016/j.celrep.2016.08.025.

- [299] Rodney Tollerson, Anne Witzky, and Michael Ibba. Elongation factor P is required to maintain proteome homeostasis at high growth rate. *Proc. Natl. Acad. Sci.*, 115(43):11072–11077, oct 2018. doi: 10.1073/pnas.1812025115.
- [300] Nikolay Burnaevskiy, Bryan Sands, Soo Yun, Patricia M Tedesco, Thomas E Johnson, Matt Kaeberlein, Roger Brent, and Alexander Mendenhall. Differences in protein dosage underlie nongenetic differences in traits. *bioRxiv*, page 431643, oct 2018. doi: 10.1101/431643.
- [301] Vassilios Sotiropoulos and Yiannis N. Kaznessis. Analytical derivation of moment equations in stochastic chemical kinetics. *Chem. Eng. Sci.*, 66(3):268–277, feb 2011. doi: 10.1016/j.ces.2010.10.024.
- [302] Abhyudai Singh and Joao P. Hespanha. Approximate moment dynamics for chemically reacting systems. *IEEE Trans. Automat. Contr.*, 56(2):414–418, feb 2011. doi: 10.1109/TAC.2010.2088631.
- [303] Brian Munsky and Mustafa Khammash. The finite state projection algorithm for the solution of the chemical master equation. *J. Chem. Phys.*, 124(4):044104, jan 2006. doi: 10.1063/1.2145882.
- [304] Rudolf Gesztelyi, Judit Zsuga, Adam Kemeny-Beke, Balazs Varga, Bela Juhasz, and Arpad Tosaki. The Hill equation and the origin of quantitative pharmacology, jul 2012.
- [305] David L. Shellenbarger and Dawson J. Mohler. Temperature-sensitive mutations of the notch locus in *Drosophila melanogaster*. *Genetics*, 81(1):143–162, 1975.
- [306] R.J. Clifford and T. Schupbach. Coordinately and differentially mutable activities of torpedo, the *Drosophila melanogaster* homolog of the vertebrate EGF receptor gene. *Genetics*, 123(4):771–787, 1989. doi: 10.1016/0168-9525(90)90089-O.
- [307] Steven Fortune. Voronoi diagrams and Delaunay triangulations. *Comput. Euclidean Geom.*, 1992. doi: 10.1016/0098-3004(93)90062-A.
- [308] Abraham Savitzky and Marcel J.E. Golay. Smoothing and Differentiation of Data by Simplified Least Squares Procedures. *Anal. Chem.*, 36(8):1627–1639, 1964. doi: 10.1021/ac60214a047.
- [309] Jacob VanderPlas, Andrew J. Connolly, Zeljko Ivezic, and Alex Gray. Introduction to astroML: Machine learning for astrophysics. In *Proc. - CIDU 2012*, pages 47–54, 2012. doi: 10.1109/CIDU.2012.6382200.
- [310] Jacob VanderPlas. Understanding the Lomb-Scargle periodogram. *Astrophys. J. Suppl. Ser.*, 236(1):1–55, 2018.
- [311] I Lohmann, N McGinnis, M Bodmer, and W McGinnis. The *Drosophila* Hox gene deformed sculpts head morphology via direct regulation of the apoptosis activator reaper. *Cell*, 110(4):457–466, 2002. doi: S0092867402008711[pii].

- [312] D Graifer, A Malygin, D O Zharkov, and G Karpova. Eukaryotic ribosomal protein S3: A constituent of translational machinery and an extraribosomal player in various cellular processes. *Biochimie*, 99:8–18, 2014. doi: 10.1016/j.biochi.2013.11.001S0300-9084(13)00397-0[pii].
- [313] A Ferrus. Parameters of mitotic recombination in minute mutants of *Drosophila melanogaster*. *Genetics*, 79(4):589–599, 1975.
- [314] Mary M Golic and Kent G Golic. A quantitative measure of the mitotic pairing of alleles in *Drosophila melanogaster* and the influence of structural heterozygosity. *Genetics*, 143(1):385–400, 1996.
- [315] R Milo and R Phillips. *Cell Biology by the Numbers*. Garland Science, New York, NY, 2016.
- [316] Daniel T Gillespie. Exact Stochastic Simulation of Coupled Chemical Reactions. *J. Phys. Chem.*, 81(25):2340–2361, 1977. doi: 10.1021/j100540a008.

## APPENDIX A

### Supporting information for Chapter 3

The experiments detailed in this section were conceived, designed, and conducted by my collaborators, most notably Nicol'as Pel'aez and Jean-Francois Boisclair Lachance. They are included here for completeness.

#### A.1. Genetics

The recombineered *pnt-gfp* BAC transgene inserted into the VK00037 landing site was previously described in Boisclair-Lachance et al. (2014). Alleles *pnt*<sup>Δ88</sup> [200] and *pnt*<sup>2</sup> (Bloomington Stock 2222) were used to render the endogenous *pnt* gene null in the presence of *pnt-gfp*. A single copy of *pnt-gfp* rescued *pnt*<sup>Δ88</sup>/*pnt*<sup>2</sup> to full viability and fertility (Fig. A.1A). Cell nuclei of developing eye-antennal discs were marked by recombining *H2Av-mRFP* (Bloomington stock 23651) with *pnt*<sup>2</sup>. Experiments measuring wild type dynamics of *Pnt-GFP* were done by dissecting eye discs from white prepupae carrying *w*<sup>1118</sup>; *pnt-gfp/pnt-gfp*; *pnt*<sup>Δ88</sup>/*pnt*<sup>2</sup>, *H2Av-mRFP*. *Pnt* isoform-specific expression was detected using enhancer traps *HS20* (gift from B. Shilo) and *pnt*<sup>1277</sup> (Bloomington stock 837), which report *PntP1* and *PntP2* transcription respectively by expressing LacZ [202]. *pnt-gfp* and *pnt* isoform-specific expression were compared in white prepupae carrying *w*<sup>1118</sup>; *pnt-gfp/pnt-gfp*; *HS20/+* and *w*<sup>1118</sup>; *pnt-gfp/pnt-gfp*; *pnt*<sup>1277</sup>/*pnt*<sup>1277</sup>. *Pnt* gene dosage experiments were done using *w*<sup>1118</sup>; *pnt-gfp/+*; *pnt*<sup>Δ88</sup>/*pnt*<sup>2</sup> (1x *pnt*) and *w*<sup>1118</sup>; *pnt-gfp/pnt-gfp*; *pnt*<sup>Δ88</sup>/*pnt*<sup>2</sup> (2x *pnt*). Notch activity was conditionally reduced using the *N*<sup>ts1</sup> temperature sensitive allele [305]. *N*<sup>ts1</sup>/*N*<sup>ts1</sup>; *pnt-gfp/+* animals were raised at the permissive temperature (18 °C) and shifted

to the restrictive temperature (28.5 °C) as third instar larvae for 24 h. Animals exposed to the restrictive temperature that were transferred back to the permissive temperature had roughened eye phenotypes and a notched wing phenotype as adults, consistent with effective inhibition of Notch activity. Control larvae of the same genotype were grown at the permissive temperature until dissection. Both control and heat-treated larvae were sexed and only *N* hemizygote males carrying *N<sup>ts1</sup>/Y*; *pnt-gfp/+* were dissected as white prepupae. EGFR activity was conditionally reduced by placing the null allele *egfr<sup>f24</sup>* - also known as *egfr<sup>CO</sup>* [306] *in trans* to the thermo-sensitive allele *egfr<sup>tsla</sup>* [227], as previously described [43]. The genotype was *w<sup>1118</sup>; egfr<sup>tsla</sup>, pnt-gfp/egfr<sup>24</sup>, pnt-gfp*. Ras activation was achieved using a transgene expressing a *Ras1<sup>V12</sup>* mutant and driven by a 3xsev enhancer and promoter [229] as previously described [43]. *Pnt-gfp* in the Ras mutant background was measured using *w<sup>1118</sup>; pnt-gfp, Sev > Ras<sup>v12</sup>/pnt-gfp;pnt<sup>2</sup>,H2Av-mRFP/+*. Controls animals carried *w<sup>1118</sup>; pnt-gfp/pnt-gfp, pnt<sup>2</sup>, H2Av-mRFP/+*. *Yan* mutant eye clones were generated using the *yan<sup>833</sup>* null allele [204], *ey>FLP* and the FRT40 crossover point. *Pnt<sup>+</sup>* tissue was labeled using the clonal marker *Ubi > mRFP<sub>nls</sub>* (Bloomington Stock 34500). Developing eyes were dissected from white prepupae carrying *w, ey > FLP; pnt-gfp, yan<sup>833</sup>, FRT40A/pnt-gfp, Ubi > mRFP<sub>nls</sub>, FRT40A*. Control discs to measure the GFP-mRFP fluorophore bleed-through were obtained from flies carrying *w, ey > FLP; pnt-gfp, Ubi > mRFP<sub>nls</sub>, FRT40A/pnt-gfp, FRT40A or w, ey > FLP; pnt-gfp, Ubi > mRFP<sub>nls</sub>, FRT40A/CyO*.

## A.2. Immunohistochemistry

Unless otherwise noted, Pnt-GFP and Yan were measured in developing animals raised at 21 ° C, selected as white prepupae, and subsequently aged in humid chambers for 5-10 h. Eye-antennal discs were dissected in PBS, and fixed in 4 % (w/v) paraformaldehyde/PBS

for ~45 min. Endogenous Yan protein was detected with the mouse monoclonal anti-Yan antibody 8B12 (Developmental Studies Hybridoma Bank, 1:200 dilution) and the secondary goat anti-mouse Pacific Blue antibody (Life Technologies, 1:200 dilution). Expression from the *HS20* and *pnt*<sup>1277</sup> enhancer traps was detected using mouse anti- $\beta$ -galactosidase 40-1a (Developmental Studies Hybridoma Bank, 1:50 dilution). H2Av-mRFP was used as a nuclear marker as previously described in Peláez et al. (2015). Discs were incubated in 1:1 (v/v) PBS:VectaShield (Vector Laboratories) for 45 min, then in 100% VectaShield for an additional 45 min before mounting.

For experiments using *yan* mutant clones, *N*<sup>ts</sup>, or *EGFR*<sup>ts</sup> alleles, nuclei were stained with a 4',6-diamidino-2-phenylindole (DAPI) nuclear marker. Samples were fixed in 4% paraformaldehyde, rinsed with PBS-Tween 0.5%, and permeabilized with PBS-Triton X-100 0.1% for 20 minutes at room temperature. Permeabilization was important to allow DAPI penetration without perturbing the fluorescence of the Pnt-GFP protein. After permeabilization, eye discs were incubated in a blocking solution containing PBS-Tween 0.1% and 1% normal goat serum for 30 minutes at room temperature. Primary and secondary antibodies were incubated each for 2 hours at room temperature. Antibodies used with DAPI were: mouse anti-Yan 8B12 (DHSB, 1/500) and goat anti-mouse Cy3 (1/2000, Jackson Immunoresearch). Discs were mounted in 0.5% n-propyl-gallate, 0.1M Tris pH 8.0 and 90% glycerol.

Samples were kept in the dark at -20 °C and imaged no later than 18-24 hr after fixation. In all cases, 1024 x 1024 16-bit images were captured using either a Zeiss LSM880 or a Leica SP5 confocal microscope equipped with 40X oil objectives. During imaging, discs were oriented with the equator parallel to the x-axis of the image. Optical slices were set at 0.8m slices (45-60 optical slices) with an additional digital zoom of 1.2-1.4 to completely

image eye discs from basal to apical surfaces. Images recorded a region of at least 6 rows of ommatidia on each side or the dorsal-ventral eye disc equator. All discs for a given condition were fixed, mounted, and imaged in parallel to reduce measurement error.

### A.3. Quantification of expression levels

Expression dynamics were inferred from confocal image stacks using an updated version of an existing segmentation and annotation pipeline [43]. The new pipeline includes *FlyEye Silhouette*; an open-source package for macOS that integrates our image segmentation algorithm with a GUI for cell type annotation. Subsequent analysis and visualization procedures were implemented in Python.

In all cases, cell segmentation was performed using either H2Av-mRFP or DAPI signals as a reference channel for identification of cell nuclei boundaries. Each layer of the reference channel was segmented independently. A single contour containing each unique cell was manually selected and assigned a cell type using a custom graphic user interface. DAPI-stained discs were segmented using a separate script based on the watershed algorithm in order to mitigate the effect of bright spots caused by DAPI accumulation in nucleoli. Further care was taken to avoid annotating contours containing such nucleoli. For each annotated cell contour, expression measurements were obtained by normalizing the mean fluorescence of the Pnt-GFP and Yan antibody channels by the mean fluorescence of the reference channel. This normalization serves to mitigate variability due to potentially uneven sample illumination, segment area, and in the case of His-RFP, differences in protein expression capacity between cells.

#### A.4. Conversion of distance to time

Cell positions along the anterior-posterior axis were mapped to developmental time as described previously [43, 213]. This is predicated on two assumptions: the furrow proceeds at a constant rate of one column of R8 neurons per two hours, and minimal cell migration occurs. We find no reason to discard these assumptions.

For each disc, Delaunay triangulations were used to estimate the median distance between adjacent columns of R8 neurons [307]. We used the median rather than the mean distance, as was used in our previous study, because it minimized the influence of non-adjacent R8s that were falsely identified by the triangulation. Dividing the furrow velocity of 2 h per column by this median distance yields a single conversion factor from position along the anterior-posterior axis to developmental time. This factor was applied to all cell measurements within the corresponding disc, yielding expression time series. Notably, these are not single cell dynamics, but rather aggregate dynamics across the development time course of a cell population.

#### A.5. Computation of moving averages and confidence intervals

Moving averages were computed by first-order Savitzky-Golay filtration [308]. This method augments the simple windowing approach used in our previous study [43] by enabling visualization of expression trends at early time-points that are otherwise obscured by large window sizes. A secondary first-order filtration with one-fifth the original window size was applied to smooth lines for visualization purposes.

None of our conclusions are sensitive to our choice of filtration or smoothing method [43]. Primary window sizes of 250 and 75 cells were used for reporting the expression of

multipotent and differentiated cells, unless noted otherwise. Confidence intervals for the moving average were inferred from the 2.5th and 97.5th percentile of 1000 point estimates of the mean within each window. Point estimates were generated by bootstrap resampling (with replacement) the expression levels within each window.

#### A.6. Alignment of expression data

Cells were aligned with a reference population by shifting them in time. The magnitude of this shift was determined by maximizing the cross-correlation of progenitor Pnt-GFP expression  $Y(t)$  with the corresponding reference time series  $X(t)$ . Rather than raw measurements, moving averages within a window of ten cells were used to improve robustness against noise. This operation amounts to:

$$z = \underset{dt}{\operatorname{argmax}} \ E \left[ \frac{(Y(t + dt) - \mu_Y)(X(t + dt) - \mu_X)}{\sigma_Y \sigma_X} \right] \quad (\text{A.1})$$

where,  $\mu$  and  $\sigma$  are the mean and standard deviation of each time series, and  $dt$  is the time by which the population should be shifted.

For each experimental treatment, a disc was randomly chosen and shifted in time such that time zero corresponds to the first annotated R8 neuron. This disc then served as the reference population for the alignment of all subsequent biological replicates within the treatment. Similarly, different experimental treatments (e.g. control and perturbation) were aligned by first aligning the discs within each treatment, then aggregating all cells within each treatment and repeating the procedure with the first treatment serving as the reference.

This approach differs from the previous implementation of our pipeline in which discs were manually aligned by the inflection point of their Yan-YFP expression profiles [43]. Manual alignment entails arbitrary prioritization of certain dynamic features over others.

Our revised protocol yields consistent, reproducible alignment of expression time series that equally weighs the entire time course. The automated approach is more principled but less robust than the manual approach. Specifically, it fails when dynamic forms qualitatively differ between experimental treatments. In these instances, we revert to manual alignment using the inflection point of Pnt-GFP induction as a reference.

### A.7. Analysis of *yan* clones

We used *ey>FLP* and *FRT40A* to generate *yan*<sup>833</sup> null clones within 23 eye discs carrying the Pnt-GFP transgene (see Section A.1). The chromosome carrying the wildtype *yan* allele was marked with a Ubi-mRFPnls transgene, enabling automated detection of subpopulations with distinct *yan* gene dosages, each characterized by a distinct level of mRFP fluorescence. Discs were dissected, fixed, and co-stained with DAPI prior to confocal imaging. Images of 36 unique vertical cross-sections spanning non-overlapping cells were collected in total.

We deployed the framework developed in Chapter 2 to measure the expression level of each reporter in each nucleus and automatically label each measurement as mutant, heterozygous, or homozygous for the Ubi-mRFPnls clonal marker. For each segment, Ubi-mRFPnls and Pnt-GFP fluorescence was quantified by normalizing the average intensity of all pixels within the respective fluorescence channel by the average DAPI fluorescence. Segments containing less than 250 pixels were removed.

Fluorescence bleedthrough between the RFP and GFP channels was visually apparent in these discs. To confirm our suspicion, control clones were generated in six wildtype *yan* eye discs co-expressing Ubi-mRFPnls and Pnt-GFP. These are the same discs as those analyzed in Chapter 2. We used the correction strategy presented in Section 2.4 to systematically correct

for bleedthrough from the Ubi-mRFPnls reporter into the GFP channel. The correction successfully eliminated any detectable difference in Pnt-GFP expression between Ubi-mRFPnls genotypes in the wildtype *yan* control discs (see Fig. 2.4C). The same procedure was applied to all measurements of *yan* null clones included in Fig. 3.7H.

As described in Section 2.4.2, we limited the analysis to cells of similar developmental age that were located immediately posterior to the MF in each eye disc. These restrictions served to buffer against differences in developmental context and focus attention on the region of elevated Pnt-GFP expression. Cells residing on the border of each clone were excluded from all comparisons to mitigate edge effects. The remaining measurements were aggregated across all eye discs for statistical comparison between clonal genotypes.

### A.8. Parameterization of equilibrium binding model

We restricted our focus to a regulatory element comprised of 12 binding sites in which only the first site carried the ETS designation. We retain the same parameterization of Yan binding proposed by Hope, Rebay, and Reinitz (2017):  $\alpha_Y = -9.955 \text{ kcal mol}^{-1}$ ,  $\beta_Y = -5.837 \text{ kcal mol}^{-1}$ , and  $\gamma_Y = -7.043 \text{ kcal mol}^{-1}$ . We parameterized Pnt binding thermodynamics to provide balanced competition between Pnt and Yan in the absence of any SAM-mediated polymerization of Pnt. That is, we set Pnt binding affinities such that the transition from Pnt to Yan occupancy occurs when Pnt and Yan concentrations are approximately equal. While a parameterization using experimentally measured data would improve predictive accuracy, our aim here is primarily to obtain insight. The model used to generate Fig. 3.6D-F assumes that Pnt binds individual sites with elevated affinities  $\alpha_P = 0.96(\alpha_Y + \gamma_Y)$  and  $\beta_P = 0.96(\beta_Y + \gamma_Y)$ . The model used to generate Fig. 3.6A-C uses these same elevated

binding affinities for Yan, while setting  $\gamma_Y = 0$  kcal mol<sup>-1</sup>. Qualitatively, our results are not sensitive to this parameterization.

#### A.9. Visualization of relative Pnt and Yan expression in *Notch* mutant discs

Visualizations were constructed by applying a smoothing operation to maximum intensity projections across confocal layers spanning progenitors, and then mapping the absolute difference in Pnt-GFP and Yan antibody fluorescence to a diverging color scale. The smoothing operation consists of three sequential applications of a grey-closing filter followed by a single pass of a three-pixel wide median filter. This procedure dampens noise. Raw image fluorescence intensities were normalized to a 0-1 scale before application of any filters, so the maximum possible difference between Pnt-GFP and Yan fluorescence channels is unity. The color scale was truncated to a range of -0.3 to 0.3 for visualization purposes. No post-processing was applied to the maximum intensity projections used to generate the visualization.

#### A.10. Analysis of periodic spatial patterns in *Notch* mutant discs

Progenitor cells were selected from a 1.75 h window immediately posterior to the morphogenetic furrow. This window corresponds to approximately one column of eventual ommatidia. The window is identifiable in *Notch* mutant discs because the MF serves as a reference. Digital spatial signals were assembled by sampling progenitor  $\log_2$ -transformed Pnt-to-Yan ratios,  $X$ , as a function of cell position along the dorso-ventral axis,  $y$ . Both autocorrelation analysis and spectral decomposition were applied to these signals.

Autocorrelation functions were assembled by computing the moving average of expression similarity,  $C$ , as a function of dorso-ventral separation distance,  $d$ :

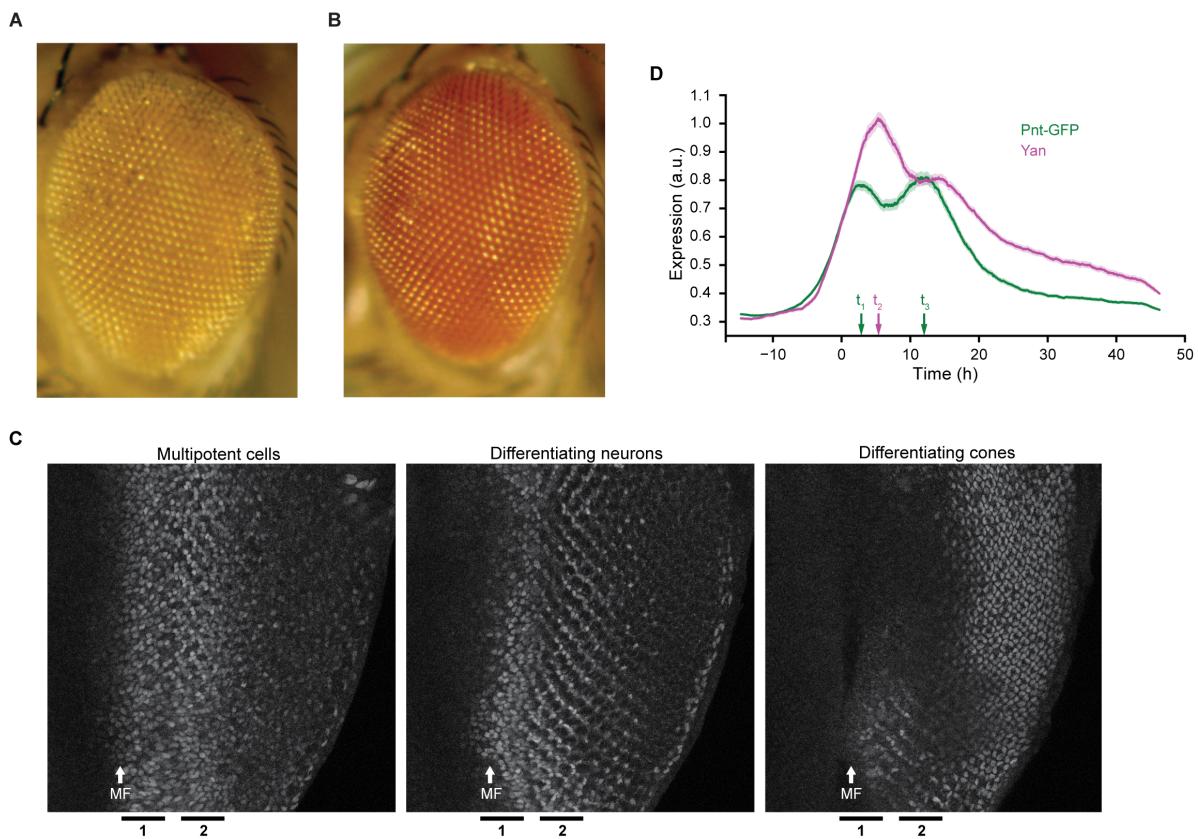
$$C_{ij} = \frac{(X_i - E[X])(X_j - E[X])}{E[X^2] - E[X]^2} \quad (\text{A.2})$$

$$d_{ij} = |y_i - y_j| \quad (\text{A.3})$$

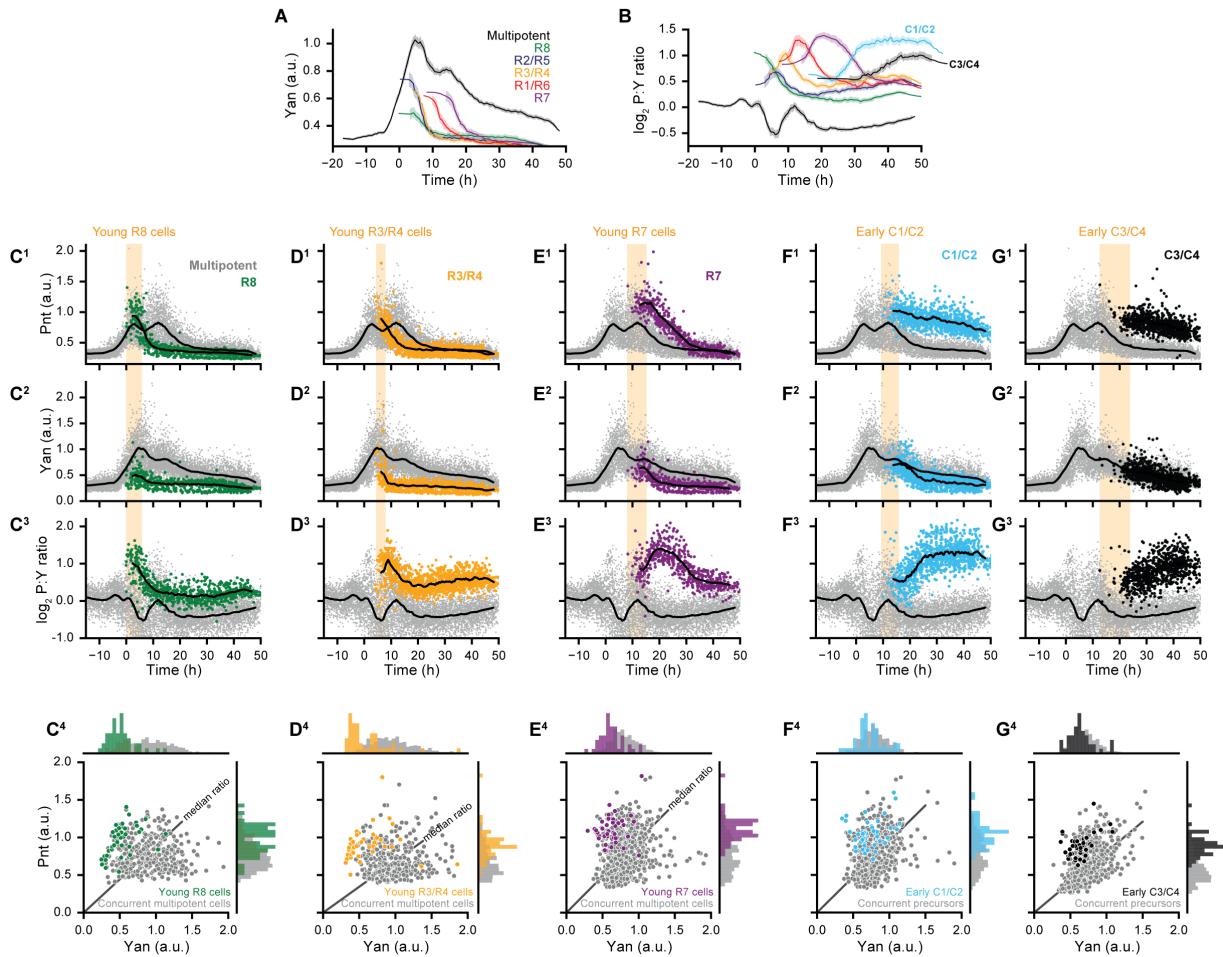
where  $E$  denotes the expected value, and  $i$  and  $j$  are indexed over all cells in order of increasing separation distance. Moving averages and confidence intervals were computed as described previously, with a window size of 50 sequential values.

Spatial signals were decomposed into spectral components via the Lomb-Scargle periodogram using the AstroML software package [309]. These periodograms were used rather than Fourier decomposition because they enable spectral decomposition of irregularly sampled signals [310]. Significance thresholds were inferred from the 95th percentile of peak spectral powers detected during repeated decomposition of 1000 null signals. Null signals were constructed by resampling signal intensities while maintaining constant sampling times.

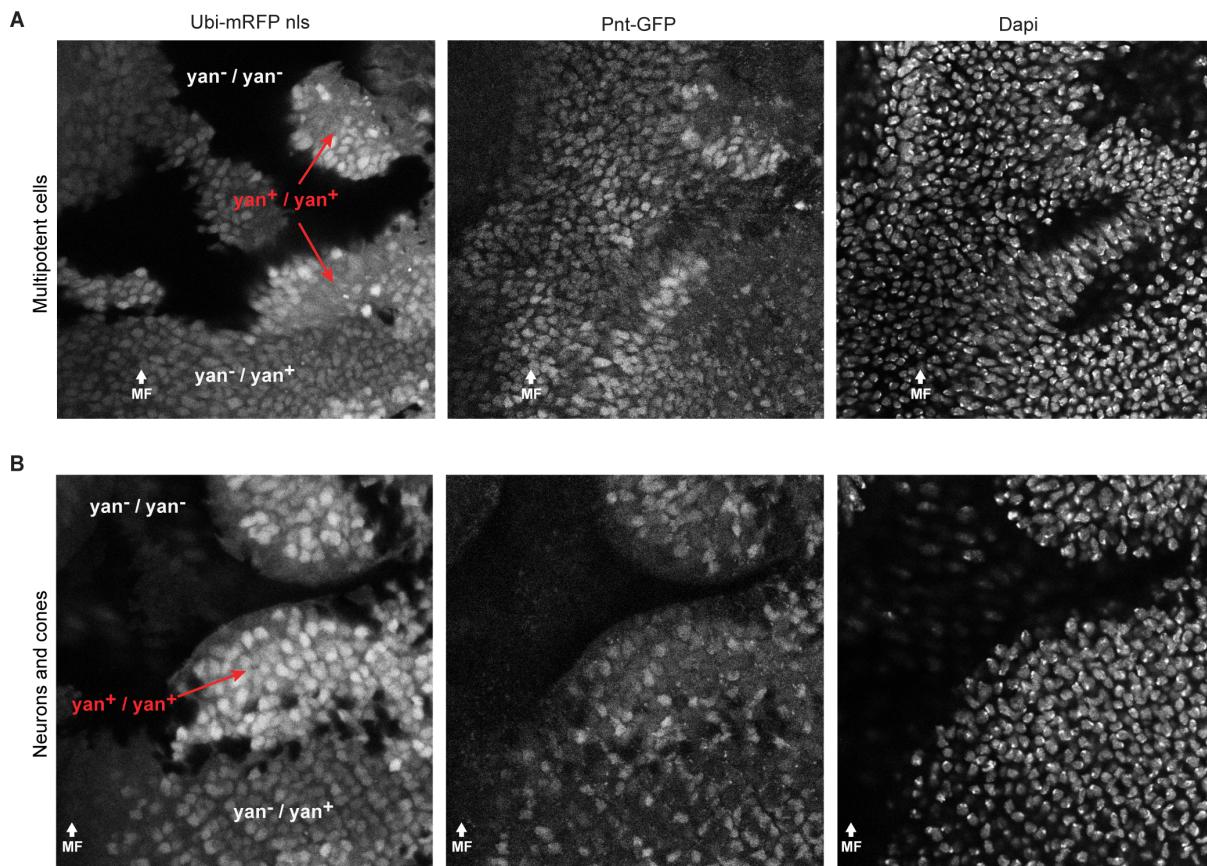
### A.11. Supplementary figures



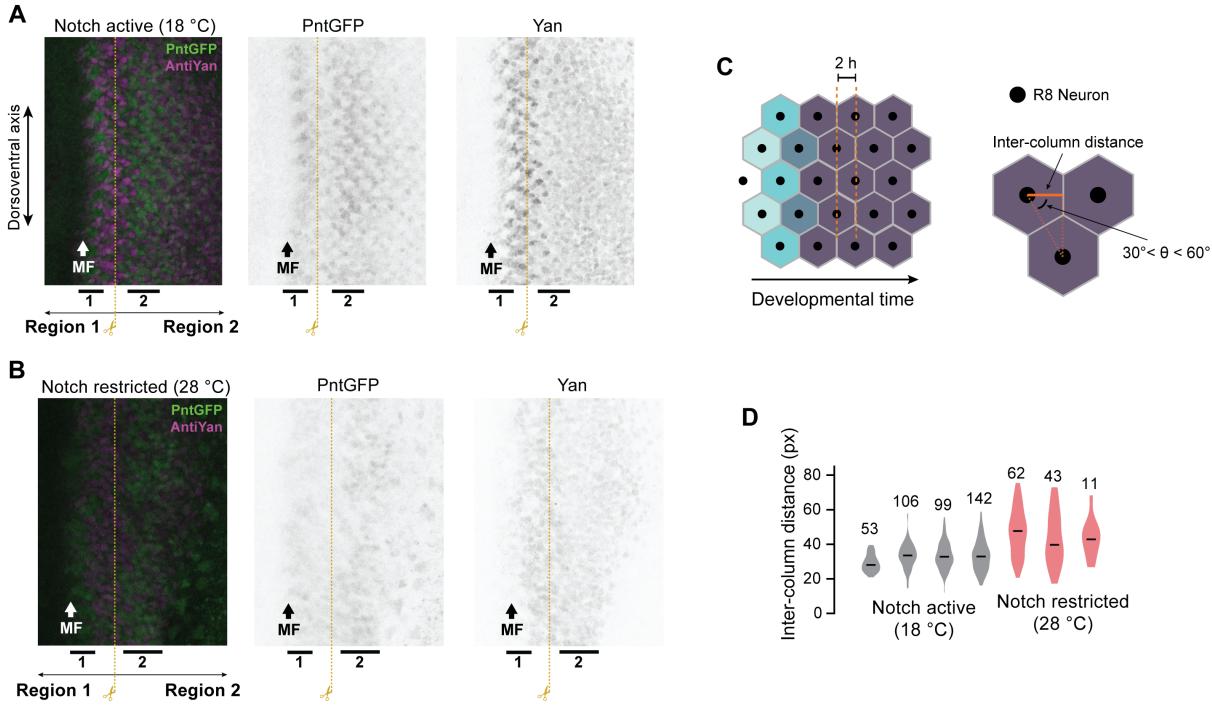
**Figure A.1. PntGFP expression during eye development.** (A,B) Adult eyes of flies carrying (A) one or (B) two copies of the recombined *pnt-gfp* transgene under a *pnt* null mutant background. Note the wildtype retina patterning under both rescue conditions. (C) Maximum intensity projections of Pnt-GFP fluorescence across layers spanning multipotent cells, differentiating R cells, and differentiating cones. White arrow denotes morphogenetic furrow. Black bars denote first and second periods of elevated Pnt-GFP expression. (D) Simultaneous Pnt-GFP (green) and Yan (magenta) expression dynamics in progenitor cells. Lines are smoothed moving averages across 500 sequential progenitors, shaded regions are bootstrapped 95% confidence intervals for the mean. Arrows indicate the times at which local maxima occur.



**Figure A.2. PntGFP and Yan expression dynamics in additional differentiated cell types.** (A,B) Measured expression dynamics for all annotated cell types. Solid lines are moving averages across 250 and 75 sequential cells for progenitors and differentiating cells, respectively. Shading denotes bootstrapped 95% confidence interval for the moving average. Colors denote cell type. (C-E) Expression dynamics and joint Pnt-Yan distributions for differentiating R8, R3/R4, and R7 cells. Joint distributions are limited to progenitor and R cells drawn from the shaded yellow region spanning the first ten R cells of the specified type in each disc. Solid lines are smoothed moving averages across 250 and 50 samples for progenitor and R cells, respectively. (F-G) Expression dynamics and joint Pnt-Yan distributions for differentiating C1/C2 and C3/C4 cone cells.



**Figure A.3. *yan* null clones in the eye.** (A, B) Confocal images of (A) progenitor cells and (B) photoreceptors and cone cells in *yan* null and heterozygote clones. Regions of Ubi-mRFPnls expression are manually labeled by *yan* genotype. White text indicates regions of reduced Yan abundance, red denotes wildtype. DAPI visualizes all nuclei.



**Figure A.4. Spatial analysis of Pnt and Yan expression in *N<sup>ts</sup>* eye discs.** (A,B) Maximum intensity projections across confocal layers spanning progenitor cells when Notch signaling is (A) active and (B) restricted. Middle and right panels show Pnt-GFP and Yan expression, left panel shows merge in which Pnt-GFP is green and Yan is magenta. Black bars denote first and second periods of elevated Pnt-GFP expression. Dashed yellow line indicates crop boundary used to construct Figures 3.8A and 3.8B. (C,D) Distances between adjacent ommatidial columns in Notch mutant discs. (C) Procedure used to estimate inter-column distance. Neighboring R8 cells are identified by Delaunay triangulation, with an added constraint that edges must fall within 30 to 60 degrees of the anterior-posterior axis. The inter-column distance is estimated by averaging the anterior-posterior distance between neighbors (solid orange line). (D) Inter-column distances are more variable when Notch signaling is restricted (red) than under wildtype (grey) conditions. Black bars denote median, numbers above violins indicate the number of neighboring R8 cells. High variability prevents accurate estimation of MF velocity and precludes conversion of spatial positions to developmental times.

## APPENDIX B

**Supporting information for Chapter 4****B.1. Model system**

All experiments were conducted in and by the lab of Professor Richard Carthew at Northwestern University. Yan-YFP expression dynamics in wildtype and *yan*<sup>Δ*miR-7*</sup> animals were measured by Rachael Bakker, while sfGFP-Sens expression levels in the wing disc were measured and analyzed by Ritika Giri. All other experiments were conceived, designed, executed, and analyzed by Justin Cassidy. This section explicitly details these experiments for purposes of reproducibility, and should not be mistaken for contributions of my own.

For all experiments, *Drosophila melanogaster* was raised using standard lab conditions and food. Stocks were either obtained from the Bloomington Stock Center, from listed labs, or were derived in the Carthew laboratory. All experiments used female animals unless stated otherwise.

**B.2. Genetics**

Experiments were performed using either homozygous mutant animals or trans-heterozygous mutants. Table B.1 lists each of the trans-heterozygous allele combinations that were used.

**Table B.1.** Mutant and transgenic alleles

<i>miR-9a</i> <sup>E39</sup> / <i>miR-9a</i> <sup>J22</sup>	<i>hairy</i> <sup>1</sup> / <i>hairy</i> <sup>41</sup>
<i>glass</i> <sup>2</sup> / <i>glass</i> <sup>60j</sup>	<i>wg</i> <sup>S<sup>p</sup>-1</sup> / <i>wg</i> <sup>+</sup>
<i>miR-7</i> <sup>Δ1</sup> / <i>Df(2R)exu1</i>	<i>dcr-1</i> <sup>+</sup> / <i>dcr-1</i> <sup>Q1147X</sup>
<i>dcr-1</i> <sup>K43X</sup> / <i>dcr-1</i> <sup>Q1147X</sup>	<i>dcr-1</i> <sup>W94X</sup> / <i>dcr-1</i> <sup>Q1147X</sup>
<i>dcr-1</i> <sup>Q396X</sup> / <i>dcr-1</i> <sup>Q1147X</sup>	<i>ago1</i> <sup>+</sup> / <i>ago1</i> <sup>Q127X</sup>
<i>ago1</i> <sup>W894X</sup> / <i>ago1</i> <sup>Q127X</sup>	<i>ago1</i> <sup>T908M</sup> / <i>ago1</i> <sup>Q127X</sup>
<i>ago1</i> <sup>E808K</sup> / <i>ago1</i> <sup>Q127X</sup>	<i>ago1</i> <sup>R937C</sup> / <i>ago1</i> <sup>Q127X</sup>

### B.2.1. IPC ablation

To genetically ablate the insulin producing cells (IPCs) of the brain, *yw* animals were constructed bearing an *ILP2-GAL4* gene on chromosome III and a *UAS-Reaper (Rpr)* gene on chromosome I or II. *Rpr* is a pro-apoptotic gene that is sufficient to kill cells in which it is expressed [311]. *ILP2-GAL4* fuses the *insulin-like peptide 2* gene promoter to GAL4, and specifically drives its expression in brain IPCs [256]. Examination of *ILP2-GAL4 UAS-Rpr* larval brains showed that they almost completely lacked IPCs (data not shown). Previous studies found that IPC-deficient adults are normally proportioned but of smaller size [256]. It takes almost twice the length of time to complete juvenile development, and juveniles have a 40% elevation in blood glucose, consistent with these insulin-like peptides being essential regulators of energy metabolism in *Drosophila* [256]. We confirmed that this method of IPC ablation results in small but normally proportioned adults, and it takes almost twice the normal time to develop into adults (Fig. 4.1B,C). For all wildtype controls, we tested animals bearing either the *ILP2-GAL4* or *UAS-Rpr* gene in their genomes.

### B.2.2. Ribosomopathy

To reduce levels of cytoribosomes in cells, we made use of loss-of-function mutations in genes encoding various ribosomal proteins (RPs), which cause the “Minute” syndrome of dominant, haploinsufficient phenotypes, including prolonged development [288]. A total of 64 *RP* genes

exhibit a Minute syndrome when mutated [287]. We selected a subset of these genes to reduce ribosomes. Since one of these, *RpS3*, encodes an RP that also functions in DNA repair [312], we tested it along with other *RP* genes in certain genetic experiments. The mutations used were: *RpS3<sup>Plac92</sup>* [288], *RpS3<sup>2</sup>* [313], *RpS13<sup>1</sup>* [288], and *RpS15<sup>M(2)53</sup>* [314]. Wildtype control animals were *w<sup>1118</sup>*.

### B.2.3. *yan*<sup>ΔmiR-7</sup>-YFP

The recombineered *Yan-YFP* BAC transgene was previously described [204]. We modified the gene by site-directed recombineering to mutate the four identified miR-7 binding sites within the *yan* (*aop*) gene [16]. The binding sites and mutations are shown in Figure B.1. The mutated transgene (*Yan<sup>ΔmiR-7</sup>-YFP*) was shuttled into the P[acman] vector [291], and inserted into the same genomic landing site on chromosome 3 (attP2) as *Yan-YFP*. One copy of the *His2Av-mRFP* transgene was recombined with the *Yan<sup>ΔmiR-7</sup>-YFP* or *Yan-YFP* transgene in order to normalize YFP expression to a housekeeping protein, in this case histone H2A [43]. The *His2Av-mRFP Yan-YFP* (*Yan<sup>ΔmiR-7</sup>-YFP*) chromosome was homozygosed, and placed in a *yan<sup>ER443</sup>* / *yan<sup>E884</sup>* mutant background so that the endogenous *yan* gene did not make any protein.

<b>Site 1 chr2L:</b> 2,158,120 - 2,158,124
5'-TCACCGCACTACATCC <b>ATCTTCCA</b> -3'
<b>AGGAA</b>
<b>Site 2 chr2L:</b> 2,157,048 - 2,157,052
5'-ATAGAAAAACATT <b>GGCTTCCA</b> -3'
<b>AGGAA</b>
<b>Site 3 chr2L:</b> 2,156,854 - 2,156,858
5'-ACGATCTTACCA <b>CACTTCCA</b> -3'
<b>AGGAA</b>
<b>Site 4 chr2L:</b> 2,156,564 - 2,156,568
5'-TCGCCACAATACCTG <b>TTCTTCCA</b> -3'
<b>AGGAA</b>

**Figure B.1. Mutation of the four identified miR-7 binding sites in the *yan* transcript.**  
The seed sequence is highlighted in red. The sequence of the mutations, which are localized to the seeds, are shown in green.

#### B.2.4. sfGFP-sens and sfGFP-sens<sup>m1m2</sup>

The recombineered *sfGFP-sens* BAC transgene was generated as described [94], and the transgene was landed in the genome at VK37 (22A3). The transgene was mutated by site-directed recombineering as described [94] to delete the two miR-9a binding sites within the *sens* gene (*sfGFP-sens<sup>m1m2</sup>*). This transgene was also landed at VK37. The *sfGFP-sens* (*sfGFP-sens<sup>m1m2</sup>*) chromosome was homozygosed, and placed in a *sens<sup>E1</sup>* null mutant background to ensure that endogenous *sens* did not make any protein.

### B.3. Analysis of mutant phenotypes

#### B.3.1. Eye mispatterning

Genetic mosaic animals bearing *miR-7<sup>Δ1</sup>* homozygous mutant eyes were generated using the FLP-FRT system. The animals' genotype was: *w ey-FLP; FRT42D miR-7<sup>Δ 1</sup> / FRT42D GMR-Hid cl.* Matching wildtype control animals' genotype was: *w ey-FLP; FRT42D P[w<sup>+</sup>] / FRT42D GMR-Hid cl.* Individuals also contained either *ILP2-GAL4* alone (control) or *ILP2-GAL4 UAS-Rpr* (IPC ablated) transgenes. All individuals were raised at 29 °C. Eye roughening was scored as previously described [95]. For *RpS3* interactions with *miR-7*, transheterozygous *miR-7* mutants and matched wildtype controls (*Df(2R)exu1/+*) were raised at 29 °C to adulthood. The *RpS3<sup>2</sup>* allele was combined with *miR-7* alleles. Eye roughening was scored as previously described [95]. Genetic mosaic animals bearing *ago1<sup>W894</sup>* homozygous mutant eyes were generated using the FLP-FRT system. The animals' genotype was: *w ey-FLP; FRT42D ago1<sup>W894</sup> / FRT42D GMR-Hid cl.* Matching wildtype control animals' genotype was: *w ey-FLP; FRT42D P[w<sup>+</sup>] / FRT42D GMR-Hid cl.* Individuals also contained either *ILP2-GAL4* alone (control) or *ILP2-GAL4 UAS-Rpr* (IPC ablated) transgenes. For

experiments with Yan transgenics, animals bearing one copy of either the *Yan<sup>ACT</sup>* or *Yan<sup>WT</sup>* [82] transgene also contained either *ILP2-GAL4* alone (control) or *ILP2-GAL4 UAS-Rpr* (IPC ablated) transgenes.

### B.3.2. R7 cell analysis in the eye

Individuals were synchronized at the larval-pupal transition, and incubated for a further 48 hours at 23 °C. Eyes were dissected from pupae, and were fixed for 40 min in 4% paraformaldehyde/PBS. They were permeabilized by incubation in PBS + 0.1% Triton-X100 (PBST) and co-incubated with mouse anti-Prospero (1:10 in PBST, MR1A MAb, Developmental Studies Hybridoma Bank) to stain R7 and bristle cells plus rat anti-Elav (1:10 in PBST, 7E8A10 MAb, Developmental Studies Hybridoma Bank) to stain all R cells. After 60 min, eyes were washed 3 times in PBST and incubated for 60 min in goat anti-mouse Alexa546 and goat anti-rat Alexa633 (1:100 in PBST, Invitrogen). Eyes were washed 3 times in PBST, cleared in Vectashield (Vector Labs), and mounted for microscopy. Samples were scanned and imaged in a Leica SP5 confocal microscopy system. *Drosophila* compound eyes have approximately 800 ommatidia. We scored all ommatidia for each imaged eye sample. The number of scored ommatidia per sample ranged between 481 and 837 (with a median of 594). Fewer than 800 ommatidia were scored per sample because in most cases, some eye tissue was lost during dissection and handling.

**B.3.2.1. Bristle scoring.** Animals of the correct genotype were allowed to age for 3 days after eclosion. The number of scutellar bristles was counted for each individual. Since these large bristles are positioned with high regularity and number on the scutellum, there was no ambiguity in counting the scutellar bristle number. For *wg* experiments, the number of

sternopleural bristles was counted for each individual. Again, the position and number of these bristles is highly regular.

### B.3.3. Relative viability

Females bearing either a  $dcr-1^{Q1147X}$  or  $ago1^{Q127X}$  mutant chromosome over a balancer chromosome were crossed to males bearing mutant  $dcr-1$  or  $ago1$  chromosomes over a balancer chromosome. F1 progeny were raised and the numbers of animals that reached either pupal or adult stage were tallied. If the non-balancer chromosome is 100% viable when homozygous, then 33.33% of the F1 progeny would not carry a balancer chromosome. We calculated viability in this manner, relative to balancer viability. Replicate crosses were performed and analyzed. Between 457 and 776 F1 animals (median = 647) were counted in the replicate  $ago1$  crosses. Between 234 and 380 F1 animals (median 285) were counted in the replicate  $dcr-1$  crosses.

### B.3.4. Population statistics

Population proportions were compared using a Chi-square test with Yates' correction and Fisher's exact test. Both gave similar results. All tests involving multiple experimental groups were Bonferroni corrected. In *sev* experiments, R7 cell counts were compared via one-way ANOVA with Bonferroni correction. Relative viabilities were compared using a Mann-Whitney-Wilcoxon test with Bonferroni correction. These tests were performed using Prism 7 (GraphPad) software. P-values shown in figures are presented from tests with the most conservative value shown if more than one test was performed on data. \*  $p < 0.05$ ; \*\*  $p < 0.01$ ; \*\*\*  $p < 0.001$ ; \*\*\*\*  $p < 0.0001$

#### B.4. Quantification of sfGFP-Sens in the wing disc

Wing discs from white-prepupal females were dissected in ice-cold Phosphate Buffered Saline (PBS). Discs were fixed in 4% paraformaldehyde in PBS for 20 minutes at 25 °C and washed with PBS containing 0.3% Tween-20. Then they were stained with 0.5 µg/ml 4,6-diamidino-2-phenylindole (DAPI) and mounted in Vectashield. Discs were mounted apical side up and imaged with identical settings using a Leica TCS SP5 confocal microscope. All images were acquired at 100x magnification at 2048 x 2048 resolution with a 75 nm x-y pixel size and 0.42 µm z separation. Scans were collected bidirectionally at 400 MHz and 6x line averaged. Wing discs of different genotypes were mounted on the same microscope slide and imaged in the same session for consistency in data quality.

For each wing disc, five optical slices containing Sens-positive cells along the anterior wing margin were chosen for imaging and analysis. A previously documented custom MATLAB script was used to segment nuclei in each slice of the DAPI channel [43]. High intensity nucleolar spots were smoothed out to merge with the nuclear area to prevent spurious segmentation. Next, cell nuclei were identified by thresholding based on DAPI channel intensity. Segmentation parameters were optimized to obtain nuclei with at least 100 pixels and no more than 4000 pixels.

The majority of cells imaged did not reside within the proneural region and therefore displayed background levels of fluorescence scattered around some mean level. We calculated the “mean background” in the green channel of each disc individually. We did this by fitting a Gaussian distribution to the population and finding the mean of that fit. In order to separate sfGFP-Sens-positive cells, we chose a cut-off percentile based on the normal distribution, below which cells were deemed sfGFP-Sens-negative. We set this cut-off at the 84<sup>th</sup> percentile for all analysis since empirically it provided the most accurate identification of proneural

cells. To normalize measurements across tissues and experiments, this value was subtracted from the total measured fluorescence for all cells in that disc. Only cells with values above the threshold for sfGFP fluorescence were assumed Sens positive (usually 30% of total cells) and carried forward for further analysis.

Analysis of sfGFP-Sens fluorescence was performed using two independent approaches.

1) For each genotype, 1000 point-estimates were made of the median fluorescence level in cells. Point estimates were generated by bootstrap resampling with replacement of the cell samples within each genotype. Point estimates from wildtype sfGFP-Sens were then randomly paired with point estimates from miR-9a-resistant sfGFP-Sens to derive a set of 1000 point-estimates of the fold-change in median sfGFP-Sens expression. Confidence intervals for the average fold-change in sfGFP-Sens expression were inferred from the 0.5<sup>th</sup> and 99.5<sup>th</sup> percentile of these point estimates. 2) The distributions of fluorescence from wildtype sfGFP-Sens and mutant *sfGFP-Sens<sup>m1m2</sup>* cell populations were compared using a Mann-Whitney-Wilcoxon test implemented in R. By calculating the difference between all randomly paired cell samples from wildtype versus mutant, the location shift is estimated as the median of the difference between a sample from sfGFP-Sens and a sample from *sfGFP-Sens<sup>m1m2</sup>*. Confidence intervals for the shift were inferred from the 2.5<sup>th</sup> and 97.5<sup>th</sup> percentile of the set of differences.

We analyzed >10 replicate wing discs for each treatment. In total, we measured wildtype *sfGFP-Sens* expression in 4,518 cells from wildtype *RpS13* discs and 4,379 cells from discs heterozygous mutant for *RpS13*<sup>1</sup>. We measured mutant *sfGFP-Sens<sup>m1m2</sup>* expression in 4,518 cells from wildtype *RpS13* discs and 4,379 cells from discs heterozygous mutant for *RpS13*<sup>1</sup>.

### B.5. Quantification of Yan-YFP dynamics in the eye

White-prepupal eye discs were dissected, fixed, and imaged by confocal microscopy for YFP and RFP fluorescence, as previously described [43]. Briefly, samples fixed in 4% paraformaldehyde were kept in the dark at -20 °C and imaged no later than 18-24 h after fixation. In all cases, 1024 x 1024 16-bit images were captured using a Leica SP5 confocal microscope equipped with 40X oil objective. During imaging, discs were oriented with the equator parallel to the x-axis of the image. Optical slices were set at 0.8  $\mu\text{m}$  slices (45-60 optical slices) with an additional digital zoom of 1.2-1.4 to completely image eye discs from basal to apical surfaces. Images recorded a region of at least 6 rows of ommatidia on each side or the dorsal-ventral eye disc equator. All discs for a given condition were fixed, mounted, and imaged in parallel to reduce measurement error. Sample preparation, imaging, and analysis were not performed under blind conditions. Image data was processed for automatic segmentation and quantification of RFP and YFP nuclear fluorescence as previously described [43]. Briefly, cell segmentation was performed using a H2Av-mRFP marker as a reference channel for identification of cell nuclei boundaries. Each layer of the reference channel was segmented independently. A single contour containing each unique cell was manually selected and assigned a cell type using a custom graphic user interface. For each annotated cell contour, expression measurements were obtained by normalizing the mean pixel fluorescence of the YFP channel by the mean fluorescence of the His-RFP channel. This normalization serves to mitigate variability due to potentially uneven sample illumination, segment area, and differences in protein expression capacity between cells. We assigned cell-type identities to segmented nuclei by using nuclear position and morphology, two key features that enable one to unambiguously identify eye cell types without the need for cell-specific markers [209]. This task was accomplished using *FlyEye Silhouette*; an open-source package for macOS that

integrates our image segmentation algorithm with a GUI for cell type annotation. Subsequent analysis and visualization procedures were implemented in Python.

Cell positions along the anterior-posterior axis were mapped to developmental time as described previously [43]. This depends on two assumptions that have been extensively validated in the literature. One, the furrow proceeds at a constant velocity of one column of R8 neurons per two hours, and two, minimal cell migration occurs. For each disc, Delaunay triangulations were used to estimate the median distance between adjacent columns of R8 neurons. Dividing the furrow velocity by the median distance yields a single conversion factor from position along the anterior-posterior axis to developmental time. This factor was applied to all cell measurements within the corresponding disc. This method does not measure single cell dynamics, but rather aggregate dynamics across the developmental time course of cells in the eye.

Moving averages were computed by evaluating the median value among a collection of point estimates for the mean generated within a sliding time window. Confidence intervals were inferred from the 2.5<sup>th</sup> and 97.5<sup>th</sup> percentile of the same point estimates. Each point estimate was generated via a hierarchical bootstrapping technique in which we resampled the set of eye discs, then resampled the aggregate pool of cell measurements between them. This novel method enhances our existing approach [43] by capturing variation due to the discretized nature of eye disc sample collection. Using the existing method, the error bars are considerably narrower (not shown). A window size of 500 sequential progenitor cells was used in all cases, but our conclusions are not sensitive to our choice of window size.

Yan level measurements were pooled across multiple replicate eye discs. An automated approach was used to align these replicate samples in time. First, a disc was randomly chosen to serve as the reference population for the alignment of all subsequent replicates. Cells from

each replicate disc were then aligned with the reference population by shifting them in time (see Section A.6).

Different experimental treatments (e.g. wildtype and miR-7 null) were aligned by first aligning the discs within each treatment, then aggregating all cells within each treatment and repeating the procedure with the first treatment serving as the reference. We analyzed four to seven replicate eye discs for each treatment in two separate experiments. In total, we measured wildtype *Yan-YFP* levels in 4,518 cells in normally metabolizing samples and 4,379 cells in slowly metabolizing samples. We measured mutant *Yan<sup>ΔmiR-7</sup>-YFP* levels in 5,382 cells in normally metabolizing samples and 6,716 cells in slowly metabolizing samples.

## B.6. Simulation procedure

### B.6.1. Simulation of protein expression dynamics

Default parameter values were based on approximate transcript and protein synthesis and turnover rates for animal cells reported in the literature [315], while gene activation and decay rates were arbitrarily set to a significantly faster timescale. Default feedback strengths for repressors acting at the gene, transcript, or protein levels were chosen such that ~25-50% of simulations failed to reach the threshold under normal conditions when one of two identical repressors was lost. Population-wide expression dynamics were estimated by simulating 5000 output trajectories in response to a three-hour transient step input to the gene activation rate. Simulations were performed using a custom implementation of the stochastic simulation algorithm [316] (see Appendix C). The algorithm constrains solutions to the set of discrete positive values, consistent with linearization about a basal level of zero gene activity. This simplifying assumption is based on the near-zero basal activities expected in the experimental systems, but is not required to support the conclusions of the model (Figs. 4.16D and 4.18D).

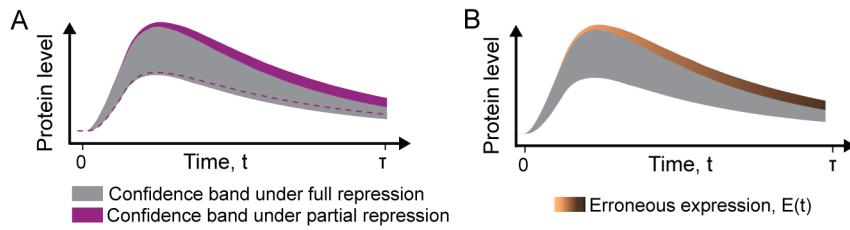
### B.6.2. Evaluation of error frequencies and changes in expression dynamics

Gene expression trajectories were simulated both with (full repression) and without (partial repression) a second repressor. The time point at which the full-repression simulations mean level reached 30% of its maximum value was taken to be the commitment time. At this time, a threshold for developmental success was set at the 99<sup>th</sup> percentile of protein levels subject to full-repression. Error frequencies were obtained by evaluating the fraction of simulated protein levels that exceeded this threshold. Per this definition, the minimum possible error frequency is one percent. For simplicity we subtracted this percentage point from all reported error frequencies.

Protein expression dynamics were compared by evaluating the fraction of partially-repressed simulation trajectories in excess of the 99<sup>th</sup> percentile of fully-repressed trajectories at each point in time. These fractions were then averaged across the time course, beginning with the reception of the input signal and ending at the previously defined commitment time. Each fraction may be thought of as the instantaneous error frequency, and their average reflects the extent to which the expression dynamics differ between the two sets of simulated trajectories (Fig. B.2).

## B.7. Dependence of model parameters on metabolic conditions

IPC ablation reduces cellular glucose consumption. Presumably this would affect either the production and consumption of ATP or the production and consumption of substrates for RNA and protein synthesis (or both). The precise effects are unknown, so we independently modeled each scenario. Since ATP concentration remains fairly constant when respiration is limited [284], ATP flux (and ATP synthesis) is assumed to decrease. Because transcription, translation, and protein degradation all require ATP turnover, we halved their rate parameters



**Figure B.2. Graphical depiction of the method used to quantify the impact of repressor loss on protein expression dynamics.** (A) Confidence bands span the 1<sup>st</sup> to 99<sup>th</sup> quantiles of protein levels simulated under full repression (grey) and under partial repression (purple), where a repressor is lost. The dashed purple line denotes the lower bound of the purple confidence band. The symbol  $\tau$  denotes the commitment time as defined previously. (B) Loss of a repressor causes protein overexpression  $E(t)$ , which is calculated as the fraction of simulations that exceed the confidence band observed under full repression (grey) at a given time point. Orange-brown color scale reflects the value of  $E(t)$  for each time point in the time course. Percent overexpression is calculated as a percent of simulations that exceed the confidence band integrated over the entire time course. A maximum of 100% overexpression would occur when all simulations exceed the confidence band at all timepoints.

under conditions of reduced glucose consumption. Under conditions of reduced substrate availability for RNA/protein synthesis, we assumed that only transcription and translation rates are affected by limiting fluxes of nucleotides and amino acids. We assumed only the translation rate is affected under conditions of reduced ribosome number. These assumptions are encoded in the rate parameters as shown in Table B.2.

**Table B.2. Reaction rate parameter dependence on environmental conditions**

<b>Reaction</b>	<b>Condition</b>			
	Normal	Reduced ATP	Reduced substrates	Reduced ribosomes
Transcription	$k_2$	$k_2/2$	$k_2/2$	$k_2$
Translation	$k_3$	$k_3/2$	$k_3/2$	$k_3/2$
Protein decay	$\gamma_3$	$\gamma_3/2$	$\gamma_3$	$\gamma_3$

In all cases, feedback strengths were reduced in order to account for the intermediate processes abstracted by each feedback element. Feedback strength parameters  $\eta_i$  were reduced four-fold under conditions of reduced energy metabolism and reduced RNA/protein substrate availability. This scaling assumes that both transcription and translation occur within the arbitrarily complex regulatory motifs represented by each repressor. This is a reasonable assumption for repressor proteins such as transcription factors and kinases. For

RNA repressors such as microRNAs, feedback strength parameters could instead be reduced only two-fold to account for their reduced transcription rates. However, microRNAs must be transcribed, processed, and act with effector proteins in order to repress their targets. These fourfold reductions in feedback strength correspond to fourfold reduction of the transcriptional feedback gain  $K_{C1}$  and twofold reduction in the post-transcriptional and post-translational feedback gains  $K_{C2}$  and  $K_{C3}$ . Feedback strength parameters  $\eta_i$  were only reduced two-fold under reduced protein synthesis conditions. This implies that the transcriptional and post-transcriptional feedback gains  $K_{C1}$  and  $K_{C2}$  decrease twofold while the post-translational feedback gain  $K_{C3}$  remains constant. Each of these dependencies are summarized in Table B.3. The corresponding dependencies for our two alternate models based upon nonlinear transcription kinetics were analogous to those used in the linear model, and are listed in Tables B.4 and B.5. The half-maximal occupancy level and Hill coefficients of transcriptional repressors were assumed to be independent of growth rate.

**Table B.3. Feedback strength dependence on environmental conditions**

<b>Feedback strengths</b>	<b>Condition</b>			
	Normal	Reduced ATP	Reduced substrates	Reduced ribosomes
Transcriptional	$\eta_1$	$\eta_1/4$	$\eta_1/4$	$\eta_1/2$
Post-transcriptional	$\eta_2$	$\eta_2/4$	$\eta_2/4$	$\eta_2/2$
Post-translational	$\eta_3$	$\eta_3/4$	$\eta_3/4$	$\eta_3/2$

**Table B.4. Two-state model dependence on environmental conditions**

<b>Reaction</b>	<b>Condition</b>		
	Normal	Reduced ATP	Reduced ribosomes
Transcription	$k_R$	$k_R/2$	$k_R$
Translation	$k_P$	$k_P/2$	$k_P/2$
Protein decay	$\gamma_P$	$\gamma_P/2$	$\gamma_P$
Transcriptional feedback	$\eta_G$	$\eta_G/4$	$\eta_G/2$
Feedback on mRNA	$\eta_R$	$\eta_R/4$	$\eta_R/2$
Feedback on protein	$\eta_P$	$\eta_P/4$	$\eta_P/2$

**Table B.5.** Hill kinetics model dependence on environmental conditions

Reaction	Condition		
	Normal	Reduced ATP	Reduced ribosomes
Transcription	$k_R$	$k_R/2$	$k_R$
Translation	$k_P$	$k_P/2$	$k_P/2$
Protein decay	$\gamma_P$	$\gamma_P/2$	$\gamma_P$
Feedback on mRNA	$\eta_R$	$\eta_R/4$	$\eta_R/2$
Feedback on protein	$\eta_P$	$\eta_P/4$	$\eta_P/2$

## APPENDIX C

## Computational tools for quantitative biologists

Several computational tools were developed in support of this work presented in this thesis. These resources are available online under open license for unrestricted use and future development. The list below describes each tool and its high level functions. With the exception of FlyEye Silhouette, all software is accessible via GitHub repositories mirrored between both my personal account and the Amaral and Bagheri lab accounts. These repositories contain high level API documentation in addition to a series of Jupyter notebooks that walk the user through a series of usage examples.

- **FlyEye Silhouette:** <http://silhouette.amaral.northwestern.edu>

GUI-based MacOS application for segmentation, quantification, and annotation of cell nuclei in the *Drosophila* eye imaginal disc. Developed by Helio Tejedor in the Amaral lab.

- **FlyEye Analysis:** <https://github.com/sebastianbernasek/flyeye>

Python framework for analyzing data generated using FlyEye Silhouette. Core features include inference of cell developmental ages and analysis of the resultant expression dynamics. Also provides tools to quantify expression heterogeneity and spatial patterns.

- **FlyEye Clones:** <https://github.com/sebastianbernasek/clones>

Python framework for automated mosaic analysis of *Drosophila* eye imaginal discs. Core features will be integrated with future versions of FlyEye Silhouette.

- **FlyEye SyClones:** <https://github.com/sebastianbernasek/syclones>

Python framework for generating synthetic microscopy data that mimic key features of mosaic eye imaginal discs.

- **PolyTF Binding:** <https://github.com/sebastianbernasek/binding>

Python framework for simulating the equilibrium occupancy of DNA binding sites by one or more polymerizing transcription factors. Utilizes a C backend that efficiently enumerates all possible microstates in a recursive fashion, enabling nested parallelization of the main computational bottleneck. For systems of two or more transcription factors, the implementation confers a major performance advantage over the sequential enumeration strategy proposed by the manuscript that inspired the model [220].

- **GeneSSA:** <https://github.com/sebastianbernasek/genessa>

Python framework for exact stochastic simulation of gene regulatory network dynamics [316]. Simulations are executed by a C backend optimized for performance on networks with a narrow scope of pre-defined reaction propensity functions. The limited scope is by design; GeneSSA prioritizes computational efficiency at the expense of flexibility by explicitly hard coding a set of functional forms. This design places GeneSSA among the most performant implementations of the exact stochastic simulation algorithm for several common types of GRNs. The framework may be (and has been) extended to include additional kinetic formulations as they are required.

## APPENDIX D

### **Reproduction data and code**

The work presented in this thesis is supported by large volumes of both experimental and simulated data. All of these datasets have been deposited along with their associated documentation in the Northwestern Arch data repository. See Table D.1 for a list of the available data.

**Table D.1. Reproduction data**

Topic	DOI
Chapter 2	<a href="https://doi.org/10.21985/N24Q81">https://doi.org/10.21985/N24Q81</a>
Chapter 3	<a href="https://doi.org/10.21985/N24Q81">https://doi.org/10.21985/N24Q81</a>
Chapter 4	<a href="https://doi.org/10.21985/N2J464">https://doi.org/10.21985/N2J464</a>

In addition, all code used to analyze and visualize the data is publicly available via GitHub, mirrored between both my personal account and the Amaral and Bagheri lab accounts. Table D.2 lists the relevant repository for each chapter of this thesis, along with the software required to execute the corresponding code (see Appendix C). Each repository contains collections of Jupyter notebooks that walk users through the process of generating all of the figures presented in the corresponding chapter. Combined, these resources provide a means to reproduce all of the results presented in both this thesis and the various manuscripts derived from it.

**Table D.2. Reproduction code**

Topic	Repository	Required software
Chapter 2	/clones	FlyEye Clones
Chapter 3	/pnt_yan_ratio	FlyEye Analysis, TF Binding
Chapter 4	/gram	FlyEye Analysis, GeneSSA

<https://github.com/sebastianbernasek/>