

NORTHWESTERN UNIVERSITY

Quantitative Insights into Cell Fate Decisions During Development

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 Insights into Cell Fate Decisions During Development

Sebastian Michal Bernasek

Quantitative analysis methods comprise

I'm honored to be considered for this award as Northwestern ChBE boasts many outstanding researchers whose achievements merit recognition. I feel my work is distinguished by its creative integration of several disciplines and by the broad relevance of its findings. My research falls under the umbrella of quantitative biology. I combine chemical engineering, computer science, and statistics to provide simple explanations for complex biological phenomena by attaching numbers to processes that are notoriously difficult to measure.

The bulk of my efforts are focused on deciphering how cells make reliable decisions during development. Cellular decisions to grow, divide, die, or differentiate are controlled 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.

One of my projects revealed a novel mechanism underlying a specific neuronal differentiation decision in the fruit fly eye. Proteins called transcription factors coordinate the timing and execution of differentiation decisions by binding to target genes and modulating

their expression. The prevailing belief was that virtually all such decisions are triggered by changes in the absolute concentration of relevant transcription factors. In most cases, these beliefs were based on qualitative observations as it is difficult to quantify transcription factor dynamics *in vivo*. Using computer vision and statistical modeling techniques, I extracted quantitative measurements of transcription factor dynamics from microscope images of fruit fly eyes collected by my collaborators. We showed that differentiation is driven by dynamic changes in the ratio between two transcription factors, and is agnostic to changes in their absolute concentrations as long as the ratio remains constant. I developed a general model based on the statistical physics of transcription factor DNA binding to show that this phenomenon is a natural consequence of competition between transcription factors for common binding sites. The study adds a new dimension to our understanding of how transcription factors execute cellular decisions, and showcases the importance of quantification in biology.

Another project addresses the more general question of why many components of regulatory networks appear to serve the same purpose. Networks typically contain multiple negative regulators tasked with attenuating expression of a single transcription factor. Despite serving the same purpose, these redundant regulators are often all essential for normal growth, development, and function of complex organisms. Without them, cells make incorrect decisions and development fails. My collaborators discovered that many essential negative regulators are rendered unnecessary when carbohydrate metabolism is slowed. Their experiments surveyed a broad range of developmental contexts, but offered no insight into the underlying mechanism. I developed a computational framework for probing the molecular behavior responsible for the observed phenomenon. My model suggests the experiments reflect a general principle of dynamic systems; they are more sensitive to perturbation when internal dynamics are fast. In this case, transcription factor activity is more sensitive to

changes in regulation when mRNA and protein biosynthesis rates are high. We successfully validated this theory by quantifying transcription factor activity in one of the experimental systems. The findings suggest that redundant negative regulators enable development to proceed more quickly by mitigating erroneous cellular decisions when cells are rapidly metabolizing. As shorter developmental times confer a selective advantage upon organisms, this likely represents a novel evolutionary driving force for increased redundancy in regulatory networks.

Beyond their biological insights, my projects have spawned computational tools that will likely prove valuable to the broader community. My gene network simulation package has already been adopted by two other researchers at Northwestern. I also plan to distribute my transcription factor binding model and computer vision methods, as these resources are broadly applicable to many different biological contexts. By sharing them I hope to promote the adoption of quantitative methods in biology and continue to embrace the spirit of interdisciplinary collaboration that lured me toward this department.

## Acknowledgements

I would like to thank those that provided support and encouragement over the past few years, particularly the artisanal coffee growers of several South American countries.

## Table of Contents

ABSTRACT	3
Acknowledgements	6
Table of Contents	7
List of Tables	9
List of Figures	10
Chapter 1. Ratiometric sensing of two transcription factors regulates a transit to differentiation	12
1.1. Introduction	12
1.2. PntGFP expression dynamics in the developing eye	17
1.3. The Pnt-to-Yan ratio varies between cells in different states	17
1.4. Cooperative DNA-binding sensitizes transcriptional output to changes in Pnt-to-Yan ratio	23
1.5. Regulation stabilizes the Pnt-to-Yan ratio against varying Pnt and Yan concentrations	27
1.6. Notch signaling lowers the Pnt-to-Yan ratio in progenitor cells	29
1.7. Ras signaling elevates the Pnt-to-Yan ratio in progenitor cells	33
1.8. Discussion	36
Chapter 2. Automated analysis of mosaic eye imaginal discs in <i>Drosophila</i>	41

2.1. Introduction	41
2.2. Image segmentation and quantification of nuclear fluorescence levels	45
2.3. Bleedthrough correction	45
2.4. Automated annotation of clones	48
2.5. Synthetic benchmarking of annotation performance	50
2.6. Discussion	53
References	59
Appendix A. Methods	80
A.1. Automated analysis of mosaic eye imaginal discs in <i>Drosophila</i>	80
A.2. Ratiometric sensing of two transcription factors regulates a transit to differentiation	91
Appendix B. Supporting figures for Chapter 2	103
Appendix C. Supporting figures for Chapter 1	111
Appendix D. Resources	118
D.1. New tools for quantitative biologists	118
D.2. Reproducibility and Supporting Data	120

## List of Tables

D.1	Data Repositories	121
D.2	Code Repositories	121

## List of Figures

1.1	Dynamics of Pnt and Yan expression during eye development.	18
1.2	Pnt-to-Yan ratios differ between cellular states.	24
1.3	Cooperative DNA-binding sensitizes transcriptional output to the Pnt-to-Yan ratio.	27
1.4	The Pnt-to-Yan ratio is stabilized against varying Pnt and Yan concentrations in progenitor cells.	30
1.5	Notch signaling lowers the Pnt-to-Yan ratio in progenitor cells.	34
1.6	Ras signaling elevates the Pnt-to-Yan ratio in progenitor cells.	36
1.7	Conceptual models for regulation of the Pnt-to-Yan ratio.	39
2.1	A framework for conducting quantitative clonal analysis.	55
2.2	Automated correction of fluorescence bleedthrough in the larval eye.	57
2.3	Automated unsupervised annotation of clones in the larval eye.	57
2.4	Synthetic benchmarking of automated annotation performance.	58
B.1	Example clones in the larval fly eye.	103
B.2	Intermediate stages of bleedthrough correction.	104
B.3	Training a clone annotation model.	105
B.4	Label assignment using a trained clone annotation model.	106

B.5	Comparison of automated annotation with manually assigned labels.	107
B.6	Simulated growth of a synthetic cell culture.	108
B.7	Tunable generation of synthetic microscopy data.	109
B.8	Additional results for synthetic benchmarking of annotation algorithm.	110
C.1	PntGFP expression during eye development.	111
C.2	PntGFP and Yan expression dynamics in additional differentiated cell types.	112
C.3	Dynamics of expression variability in progenitors and R8, R2/R5 and R1/R6 cells.	113
C.4	A simple two-species competitive binding model	113
C.5	Thermodynamic model of transcription factor DNA binding.	114
C.6	<i>yan</i> null clones in the eye.	115
C.7	Spatial analysis of Pnt and Yan expression in $N^{ts}$ eye discs.	116
C.8	Pnt and Yan expression dynamics in $EGFR^{ts}$ eye discs.	117

## CHAPTER 1

# Ratiometric sensing of two transcription factors regulates a transit to differentiation

A manuscript closely 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://www.biorxiv.org/content/10.1101/430744v1>.

Please note that all of the experiments detailed in this chapter were conceived, designed, and executed entirely by my colleagues. My personal contributions primarily include all of the computational analysis and modeling. I also produced all of the figures and wrote the majority of the text under the tutelage of Professor Richard Carthew.

### 1.1. Introduction

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 [1–3]. Coordination among transcription factors can thus broaden the spectrum of possible cell

responses to developmental cues [4, 5]. Positive and negative feedback can create bi-stable patterns of mutually exclusive transcription factor expression or activity that render state transitions irreversible [6–9]. Cells can also regulate state transitions by differentially partitioning transcription factors between daughter cells [10]. In all of these models, mRNA and protein expression are believed to depend upon the absolute concentrations of the transcription factors involved [11–14].

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 [15]. Fold-change, rather than absolute levels of  $\beta$ -catenin, dictate Wnt signal transduction in eukaryotic cells [16]. Likewise, aggregation of the social amoeba *Dictyostellium* depends on fold-change detection of extracellular cAMP concentrations [17]. 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 [18]. 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 [19]. Topological features of molecular interaction networks, such as feed-forward loops, can also sense relative changes in molecule abundance [20, 21]. 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 [22, 23], 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 [24–28]. Consistent with their opposing regulatory effects, genetic studies have shown that Pnt and Yan can act antagonistically in guiding numerous cell fate transitions [24, 25, 29, 30]. 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. 1.1A) [31, 32]. Specifically, PntP1 is constitutively active whereas PntP2 requires phosphorylation via the RTK signaling pathway to become a potent activator [29, 30]. Because both Pnt isoforms and Yan bind to a common ETS-binding DNA sequence motif GGA(A/T) [33], competition for occupancy and regulation of common target genes must be precisely orchestrated [25, 27, 28, 30, 34–37].

Pnt and Yan display mutually exclusive expression patterns in several developing tissues [38]. The embryonic ventral ectoderm provides a classic example in which cells unambiguously reside in one of two stable states [6, 24]. 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 [39]. The model posits that RTK signaling triggers a transition from target gene repression to activation by degrading Yan and activating PntP2 [29, 40]. 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 [41]. Sustained PntP1 expression in cells devoid of Yan thereby recapitulates a complementary expression pattern under the control of RTK signaling [38]. However, Pnt and Yan are

also co-expressed during cell fate commitment in several developmental contexts [38]. For example, the two proteins are co-expressed in posterior follicle cells of the early egg chamber and throughout the embryonic mesoderm [38], where they are required for specification of cell fates [25, 26]. Co-expression also occurs in the larval eye [38], despite the repeated use of RTK signaling to designate cell fates [42]. 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 [43]. 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. 1.1B). 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 [44]. 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 [45, 46]. Each R8 cell then locally secretes an RTK ligand that induces R8's multipotent neighbors to differentiate into other photoreceptor types (Fig. 1.1C) [42, 44]. Transitions sequentially occur approximately every two hours to form R2/R5, R3/R4, R1/R6, and R7 photoreceptors [43]. 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 [47].

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 [38]. 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 [48]. 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 [38]. 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 [30, 40].

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 [42, 49], 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.

### **1.2. PntGFP expression dynamics in the developing eye**

Although Pnt expression has been qualitatively studied in the eye [38], 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 [38], *pnt-gfp* rescues *pnt* null mutants to viable, fertile adults with wild type external morphology (Fig. C.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. 1.1D, regions 1 and 2).

### **1.3. 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 [48, 50]. 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 [48, 50]. We mapped each cell's position along the

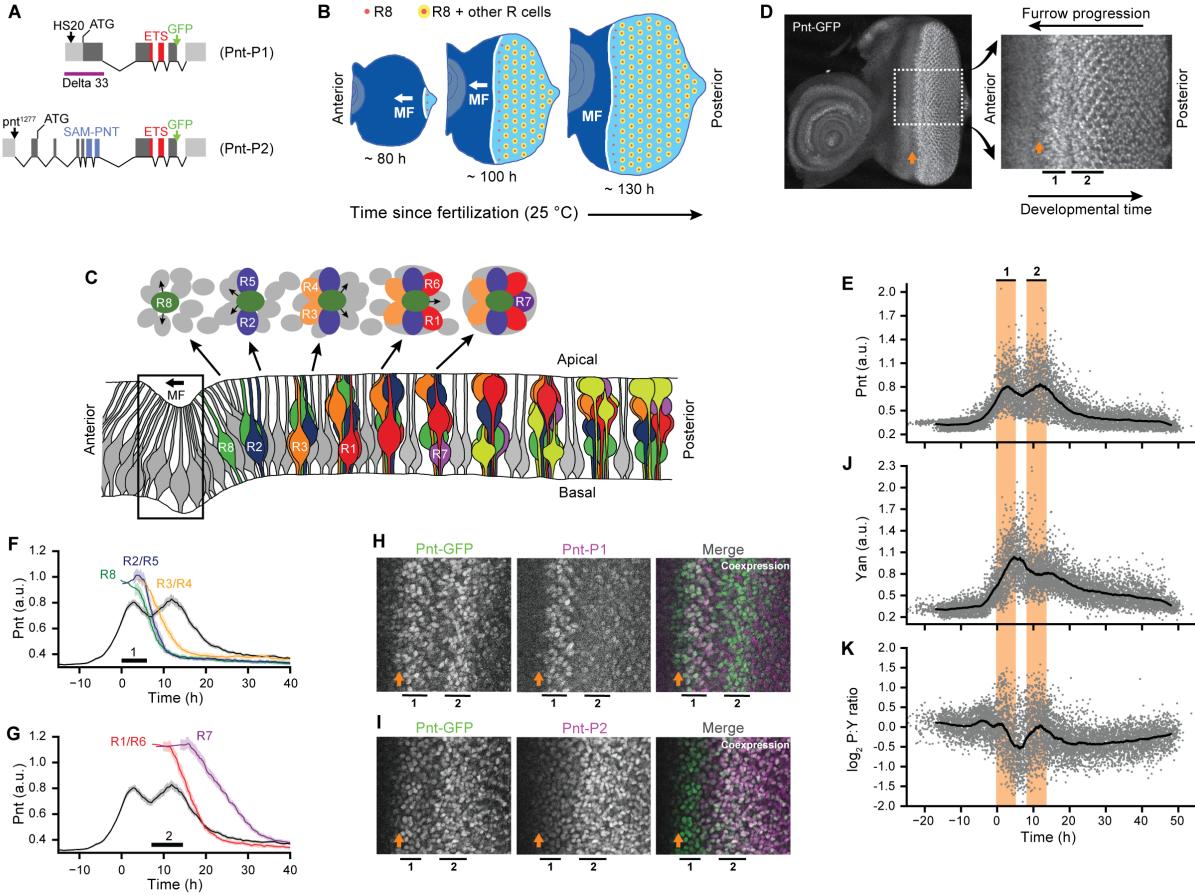


Figure 1.1. (Continued on the next page.)

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 [51, 52]. The distance between a cell and the MF is therefore proportional to the time elapsed since the MF passed. We 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 [43, 48, 50, 53–55]. Combined, these data allowed us to infer a macroscopic view of cell state transition

**Figure 1.1. 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)  $\log_2$ -transformed Pnt-GFP to Yan ratios in progenitor cells. Grey points are individual cells, solid line is a smoothed moving average.

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.

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. 1.1E). 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. 1.1D). 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. 1.1F,G and C.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 [41], 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. 1.1H,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 [50]), 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. 1.1F,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 [48]. 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 [48], validating our approach. Pnt-GFP and Yan were induced at the same time in progenitor cells (Figs. 1.1E,J and C.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. 1.1K). 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 [47]. 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.

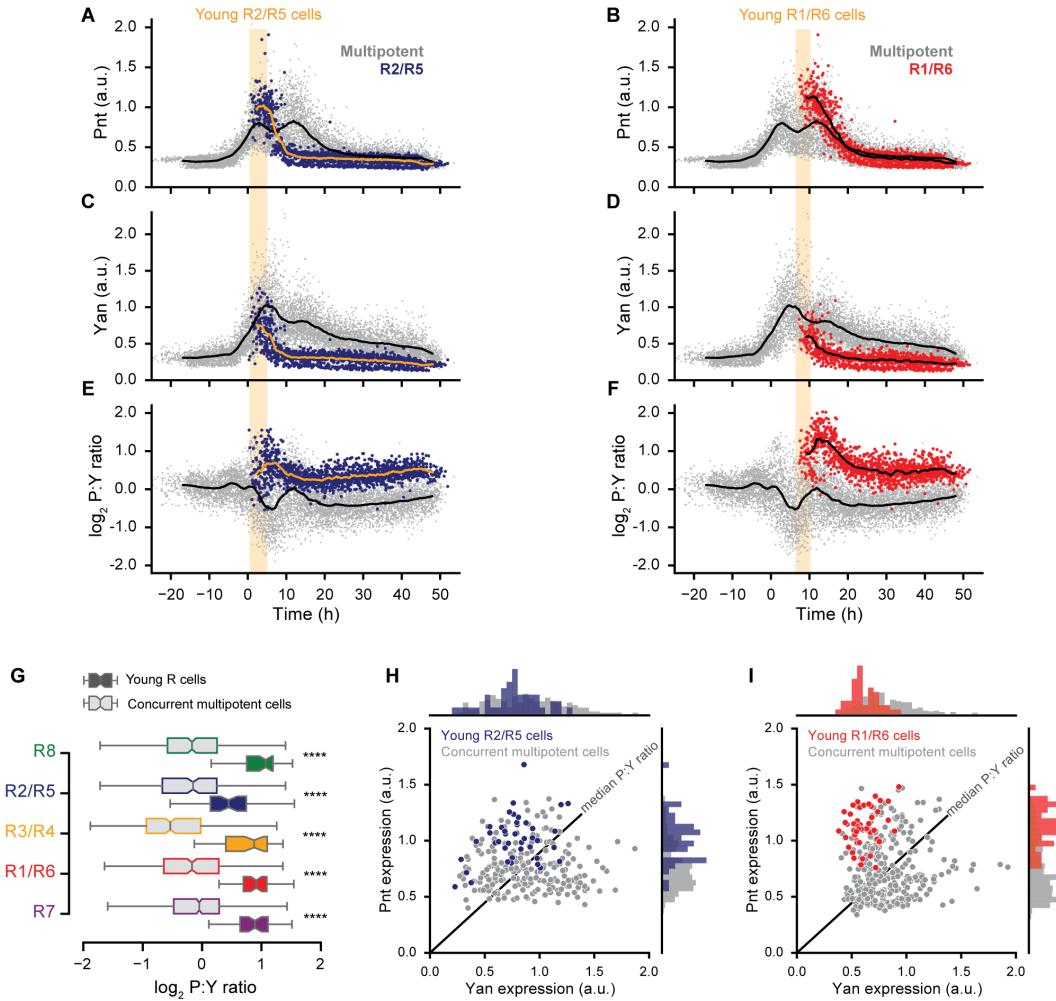
As a first test of this idea, we quantified the levels of Pnt-GFP and Yan in cells that had undergone R cell state transitions. We focused on R2/R5 and R1/R6 cells, since they are representative of transitions of cells derived from groups 1 and 2, respectively. As previously noted, Pnt-GFP levels were elevated in both sets of “young” R cells as soon as we could confidently identify them (Fig. 1.2A,B). In contrast, Yan levels were lower in these young R cells than in progenitor cells of comparable age (Fig. 1.2C,D). This meant that the average ratio of Pnt-to-Yan was elevated 1.5- to 2-fold in young R cells (Fig. 1.2E,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. 1.2G and C.2), suggesting that different state transitions share a common requirement for sustained change in the Pnt-to-Yan ratio relative to progenitor cells (Fig. C.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. 1.2H,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 [48]. 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. C.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 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.

#### **1.4. Cooperative DNA-binding sensitizes transcriptional output 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 [25, 27, 28, 33–35, 56], the underlying mechanism may be a natural consequence of competition for binding sites in target genes. A simple model in which the two transcription factors compete for occupancy of common binding sites shows that if the sites are saturated, then 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. C.4A,B).



**Figure 1.2. 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.

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 [25, 27, 28, 36], Yan binds these enhancers with higher affinity than Pnt [27]. Moreover, recent experiments suggest a scenario in which Pnt and Yan differentially interpret the structural syntax of *cis*-regulatory modules [36]. 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 [36, 57]. These cooperative effects could conceivably bias the competition between Yan and Pnt, which led us to consider a more complex model.

A modeling framework was recently introduced in order to probe the effects of *cis*-regulatory syntax on Yan binding site occupancy [58]. 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. C.5A,B). 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.

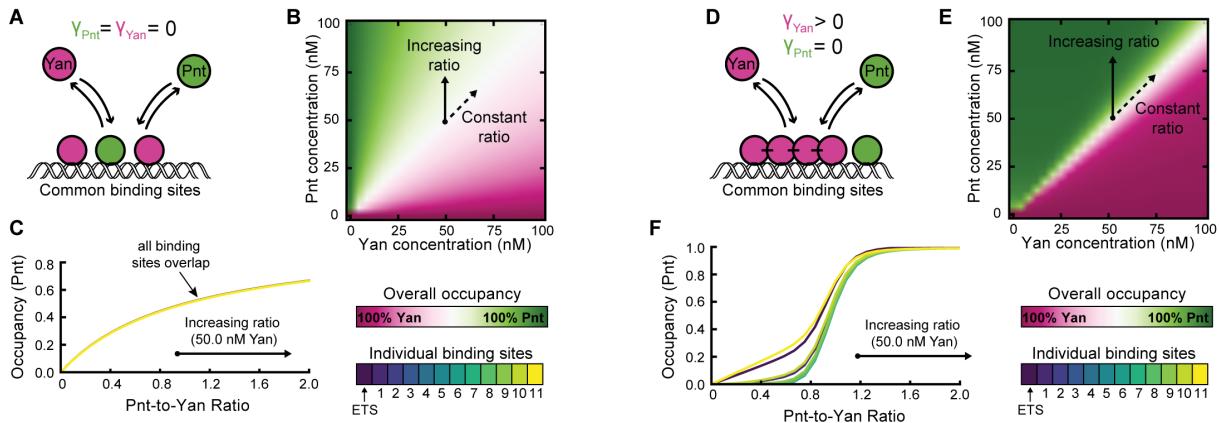
We first considered a scenario in which Yan and Pnt did not exhibit cooperativity (Fig. 1.3A). In the absence of stabilizing SAM-SAM interactions, the landscape of overall binding site occupancy is identical to that obtained with the simple binding model described above (Figs. 1.3B and C.4B). Increasing the Pnt-to-Yan ratio revealed a gradual increase in Pnt

occupancy for all individual binding sites (Fig. 1.3C). This titration contour closely resembles a Langmuir isotherm or Michaelis-Menten saturation curve [59].

We then introduced a stabilizing SAM-interaction for Yan (Fig. 1.3D). The resultant landscape of overall Pnt binding site occupancy is clearly distinguished from the simple binding model by a sharpening of the transition from Yan to Pnt dominance in occupancy (Fig. 1.3E). 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. C.5C).

Increasing the Pnt-to-Yan ratio revealed nonlinear transitions from low to high Pnt occupancy for each individual binding site (Fig. 1.3F). These transitions resemble Hill functional forms [59], 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.



**Figure 1.3. 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).

### 1.5. Regulation stabilizes the Pnt-to-Yan 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 gene [60], validating our strategy. We found that the eyes of adult flies were morphologically indistinguishable across this *pnt* dosage range (Fig. C.1A,B). A similar lack of dosage sensitivity had been previously observed with the *yan* gene [48]. 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. 1.4A,B). Interestingly, average Yan protein concentration also scaled with *pnt-gfp* gene copy number (Fig. 1.4C,D) resulting in an essentially identical Pnt-to-Yan protein ratio (Fig. 1.4E). 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 [48]. 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.

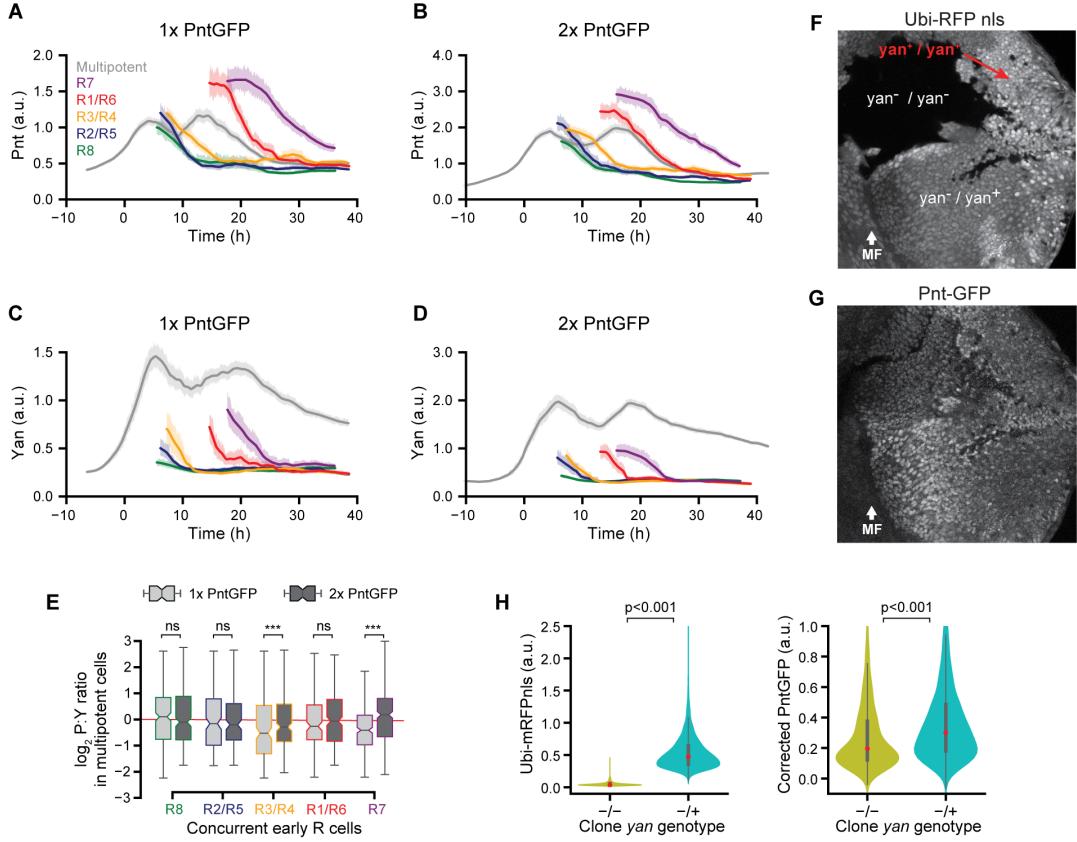
Effective control of the Pnt-to-Yan ratio would require a similar dependence of Pnt protein output on the concentration of Yan. We assessed this prediction by quantifying Pnt-GFP protein in cells with different copy numbers of the *yan* gene. Due to the embryonic lethality of *yan*-null mutations, we 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. 1.4F,G and C.6A,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. 1.4H). 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. 1.1K). 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. 1.1E,J).

### 1.6. 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 [39], and therefore present



**Figure 1.4. 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 2.3). 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.

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 [49], and ensures proper patterning of the first group of R cells by constraining the proximity of adjacent R8 cells [61, 62]. Conversely, RTK signaling is required to initiate each of the subsequent R cell state transitions [42]. Both signaling pathways influence the Pnt-Yan network in multipotent eye cells [29, 40, 63]. Notch stimulates Yan expression [64], while RTK activates PntP2 and stimulates PntP1 expression [29, 41] while attenuating the pulse of Yan expression [40, 48]. 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 [44, 64, 65].

We used a temperature-sensitive *Notch* mutant [66] 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. C.7A, 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. C.7A, thick black lines).

At the non-permissive temperature, progenitor cells had visibly reduced Yan levels (Fig. C.7A,B), consistent with previous reports of Yan expression's dependence on Notch signaling in the developing eye disc [64]. Pnt-GFP levels were also reduced in region 1, but appeared close to normal at later times (Fig. C.7A,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. C.7C). 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. 1.5A,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. 1.5A). 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 [49] 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 [61]. At the permissive temperature, there was a periodic pattern in the Pnt-to-Yan ratio of progenitor cells (Fig. 1.5B). 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. 1.5C). We detected periodic spatial patterns with a constant period of oscillation that was approximately equivalent to the length scale separating adjacent R8 cells (Fig. 1.5D,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. 1.5B). The ratio was more uniform along the dorsoventral axis, and while there were modestly detectable oscillations in the ratio, their period was not stable (Fig. 1.5F,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 [61, 67, 68]. If Notch signaling is inhibited, very few high Pnt-to-Yan ratio clusters and R8 cells are formed because the first step is blocked [67]. 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.

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

RTK signals received by progenitor cells trigger their transition to R cell states [42]. 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 [69]. At high temperatures, the mutant blocks RTK signal transduction, which triggers 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 [48].

At intermediate temperatures, the entire second pulse of Pnt-GFP expression disappeared upon restriction of EGFR activity (Fig. C.8A). Because the second pulse can be predominantly ascribed to PntP2 expression (Fig. 1.1I), this result conflicts with a report that PntP2 expression is not dependent upon RTK signaling [41]. At intermediate signaling, Yan levels were also reduced, resulting in Pnt-to-Yan ratios that were indistinguishable from wildtype during the second wave of state transitions (Fig. C.8B,C).

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 [70, 71]. This

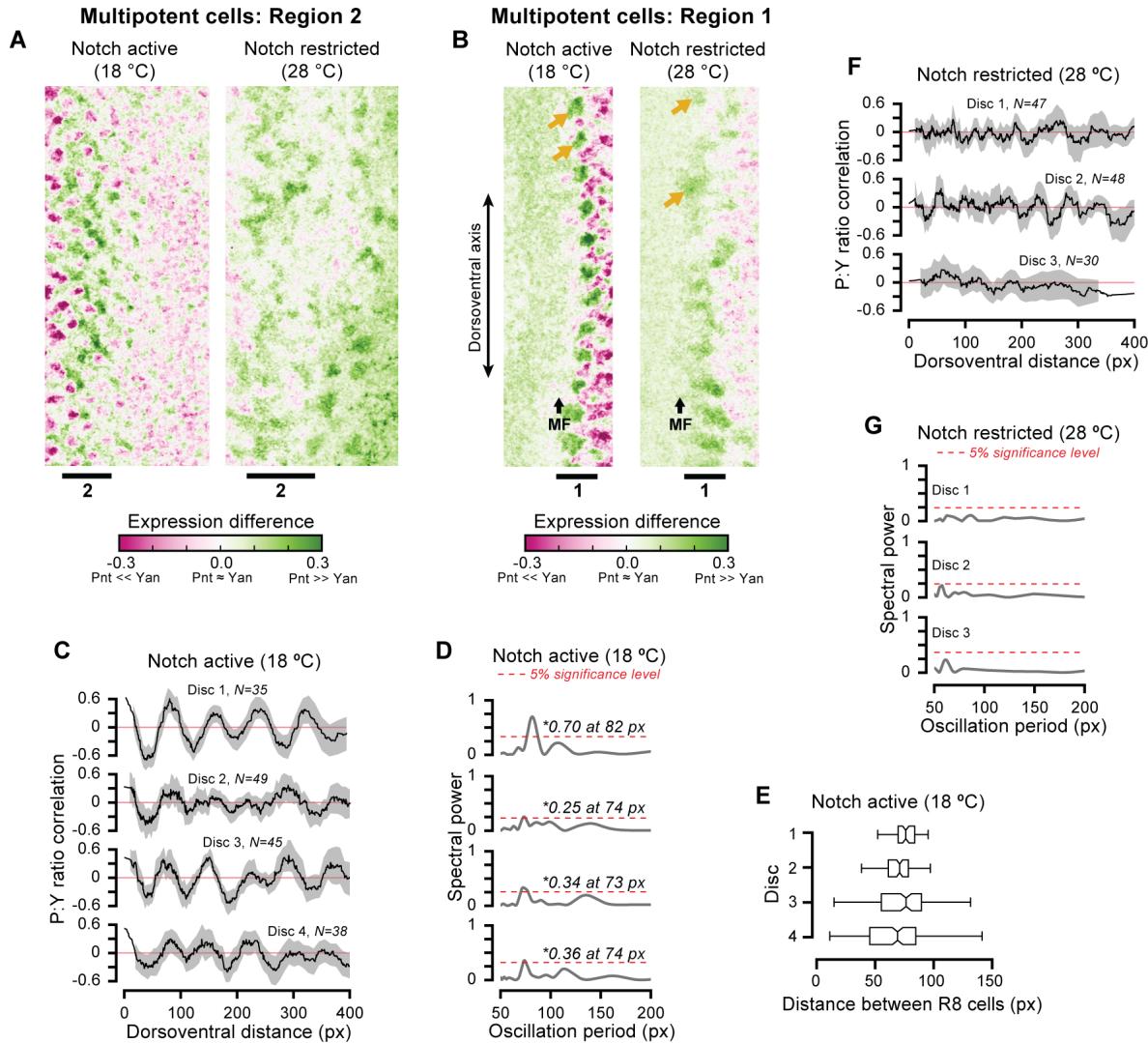


Figure 1.5. (Continued on the next page.)

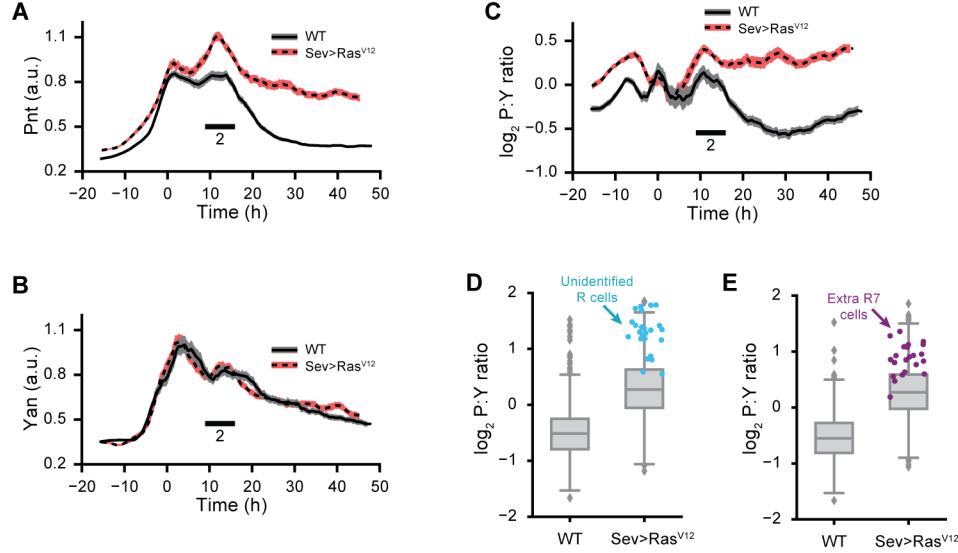
construct uses a *sev* promoter to drive Ras expression, limiting its effects to the second region of Pnt-GFP expression. Constitutive Ras activity dramatically increased the amplitude and duration of the second pulse of Pnt-GFP expression in progenitor cells (Fig. 1.6A) but did not significantly alter Yan expression dynamics (Fig. 1.6B). This yielded a sustained increase in Pnt-to-Yan ratio in progenitors during the second wave of state transitions (Fig. 1.6C), as well as the ectopic differentiation of R cells. As previously reported [71], supernumerary

**Figure 1.5. 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.

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. 1.6D,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. 1.6D,E blue and purple markers). These

observations suggest that abnormally high Pnt-to-Yan ratios in progenitor cells accompany ectopic R cell state transitions.



**Figure 1.6. 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 [48], 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.

## 1.8. Discussion

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 [72]. 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 [27, 73]. 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 [74]. 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 bindings 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.

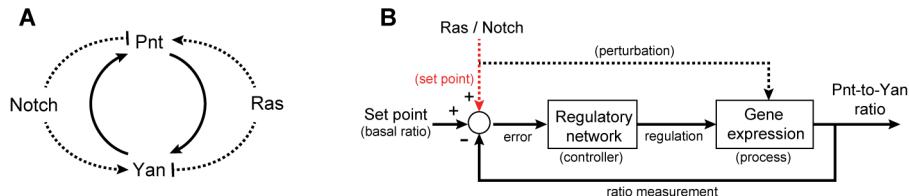
In general, the proposed ratiometric sensing mechanism is well suited when either or both transcription factors bind their target sites in a manner stabilized by cooperative interactions. Like its human ortholog TEL1 [75], DNA-bound Yan monomers enhance recruitment of Yan to adjacent binding sites through stabilizing SAM-SAM polymerization [76]. 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. 1.7A). 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. 1.7B).

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. 1.7B, dashed black arrow). These signals could permanently set the ratio by adjusting the reference value (Fig. 1.7B, 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 1.7. 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 [71, 77]. 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 2

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

A manuscript closely resembling this chapter is currently in preparation for submission to a peer reviewed journal. The work was coauthored with Neda Bagheri and Luís Amaral, using experimental data borrowed from Nicolás Peláez. These data are part of a larger study discussed at length in Chapter 1.

### 2.1. Introduction

Quantification will be essential as biologists study increasingly complex facets of organismal development [78]. 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 [79–81], 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 [82–84], genomics and transcriptomics [85–88], and other subdisciplines of biology [89, 90]. Image analysis has proven particularly amenable to automation, with several computer vision tools having gained traction among biologists [91–94]. These platforms are popular because they increase productivity, improve the consistency and sensitivity of measurements, and obviate the need for specialized computational proficiency [95–97]. 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 [98–100].

Mosaic analysis addressed this challenge in *Drosophila* by limiting perturbations to a subset of cells within the larval eye [101, 102], a model system with enduring relevance [103]. 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 [104, 105]. The timing of these events determines the number and size of the resultant clones [106]. 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. 2.1C). Additional reporters may be used to monitor differences in gene expression or morphology across clones (Fig. 2.1D). Variants of this strategy led to seminal discoveries in neural patterning [107–109] and morphogenesis [110, 111], and remain popular today [112–114].

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. 2.1E). 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 [115–118], 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. 2.1F). Averaging the pixel intensities within each boundary then yields a fluorescence intensity measurement for each reporter in each identified nucleus (Fig. 2.1G). The measurements should then be corrected to account for any fluorescence bleedthrough between reporter channels (Fig. 2.1H). Correction often requires single-reporter calibration experiments to quantify the crosstalk, followed by complex calculations to remedy the data [119, 120]. 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. 2.1I). 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. 2.1I, 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. 2.1J). 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 ClonalTools 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 [121]. 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 [122, 123]. While useful in many other settings, neither of these tools support automated labeling of individual cells or explicit comparison of clones with 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 [116–118]. 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 [48], in addition to a collection of tools for analyzing reporter expression in the segmented nuclei [115]. 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. 2.1F-J). We demonstrate each of these functions by applying them to real confocal images of clones in the eye (Fig. B.1A,B), 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. Image segmentation and quantification of nuclear fluorescence levels

We implemented a segmentation strategy based upon a standard watershed approach [124]. 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 [124, 125]. 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. B.1C).

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 [126], including FlyEye Silhouette [48], 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.

### 2.3. 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. 2.1D, yellow lines) [119, 127]. 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. 2.2A). 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. B.1A). However, the images visibly suffered from bleedthrough between the RFP and GFP channels (Fig. 2.2A,B). Bleedthrough was similarly evident when we compared measured GFP levels across clones. Nuclei labeled mutant, heterozygous, or homozygous for the clonal marker had low, medium, and high expression levels of the control reporter, respectively (Fig. 2.2C, 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 [119, 127]. These methods typically entail characterizing the extent of crosstalk between fluorophores globally [128, 129], on a pixel-by-pixel basis [120], or by experimental calibration [119], 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.

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 [130]:

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

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,  $\{\alpha_1, \alpha_2, \dots, \alpha_K\}$ . We estimate these parameters by characterizing their impact on background pixels (see Section A.1.2). When applied to the microscopy data (Fig. B.2), this procedure successfully eliminated any detectable difference in GFP expression across clones (Fig. 2.2C, red boxes).

## 2.4. 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.

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. B.3A). This entails fitting a weighted mixture of three or more bivariate lognormal distributions (components) to a two dimensional set of observations (Fig. B.3B,C, see Section A.1.4 for details). 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. B.3D). 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. B.3E). We then cluster the components into three groups on the basis of their mean values (Fig. B.3F), 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.

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. B.4A). We identify these communities by applying a community detection algorithm to an undirected graph connecting adjacent cells (Fig. B.4B). Edges in this graph are weighted by the similarity of clonal marker expression between neighbors, resulting in communities with similar expression levels (Fig. B.4E, 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 estimate of the probability that an observation belongs to one of the model's components (Fig. B.4E, Step III). We further refine these estimates by allowing the probabilities estimated for each cell to diffuse throughout the graph (Fig. B.4E, 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. B.4E, 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. B.4F,G), resulting in a fully labeled image (Fig. B.4H).

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. B.3D and B.4D), thus averting any need for user input.

We sought to validate the performance of the annotation algorithm by assessing its ability to accurately reproduce human-assigned labels. We manually labeled nuclei in each eye imaginal disc as mutant, heterozygous, or homozygous for the clonal marker, then automatically labeled the same cells (Fig. 2.3A). The two sets of labels showed strong overall agreement (Figs. 2.3B and B.5A). Excluding cells on the border of each clone revealed greater than 97% agreement in seven of the nine annotated images (Fig. B.5B). Upon secondary inspection of the sole instance of substantial disagreement (Fig. B.5C), we are unable to confidently discern which set of labels are more accurate.

## 2.5. Synthetic benchmarking of annotation performance

While it is common practice to use human-labeled data as the gold standard, 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. Synthetic benchmarking provides a powerful alternative. 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 used synthetic microscopy data to benchmark the performance of our annotation strategy. Each synthetic dataset depicts a simulated culture of cells distributed roughly uniformly in space (Fig. B.6A). Cells in this culture contain zero, one, or two copies of a gene encoding an RFP-tagged clonal marker (Fig. B.6B). Our simulation procedure ensures that cells tend to remain proximal to their clonal siblings (Fig. B.6C), thus forming synthetic clones with tunable size and spatial heterogeneity (Fig. B.6D,E). We generated synthetic measurements by randomly sampling fluorescence levels in a dosage-depend manner (Fig. B.7A-C). We varied the similarity of fluorescence levels across clones using an ambiguity parameter,  $\sigma_\alpha$ , that modulates the spread of the distributions used to generate fluorescence levels (Fig. B.7D-F). Using this schema as a template, we generated a large synthetic dataset, annotated each set of measurements, and compared the assigned labels with their true values. We used the mean absolute error as a comparison metric because it provides a stable measure of accuracy for multiclass classification problems in which the labels are intrinsically ordered [131]. Annotation performance is very strong for all cases in which  $\sigma_\alpha \leq 0.3$  (Fig. 2.4). Unsurprisingly, performance suffers as the difficulty of the classification problem is increased. The same trends are evident when performance is graded strictly on accuracy (Fig. B.8A). 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.

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. B.4G). As expected, the variant’s performance exhibited no dependence on clone size (Fig. B.8B). Comparing the two strategies confirmed that spatial context confers the most benefit when clones are large (Fig. B.8C). 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 [132]. 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.

## 2.6. Discussion

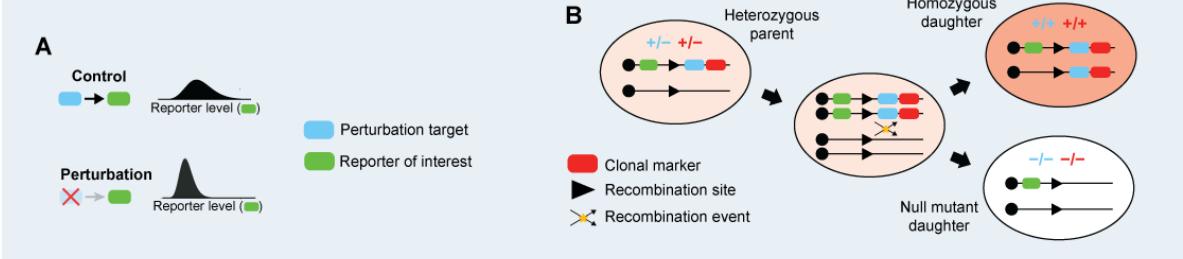
We used synthetic data to survey the performance of our annotation strategy across a much broader range of conditions than would have otherwise been possible with manually labeled data. This included conditions well beyond those of practical use. In particular, experiments designed to compare gene expression 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 [133, 134], but also frequently opt for binary labeling of mutant versus non-mutant clones [135–137] and dichromic labeling of twin-spots [138, 139]. 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 [140–142]. Our framework is thus well suited to many different mosaic analysis platforms deployed in the larval eye.

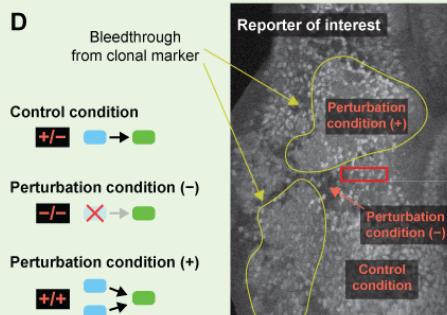
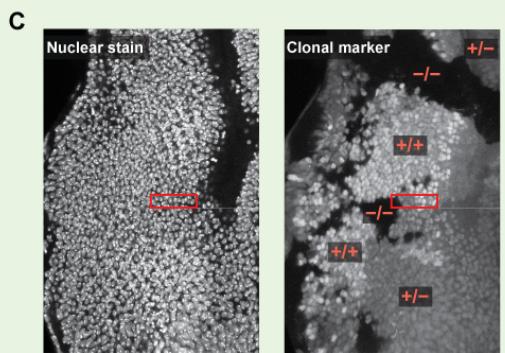
In principle, the framework described here should also be applicable to a wide variety of other tissues [143, 144] and model organisms [145–147] in which mosaics are studied. In practice, application to alternate contexts would require modifying some stages of the analysis. Most notably, image segmentation is strongly context dependent and any attempts

to develop a universally successful strategy are likely to prove futile [148]. 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.

## Experimental framework



## Conventional clonal analysis



## Quantitative clonal analysis

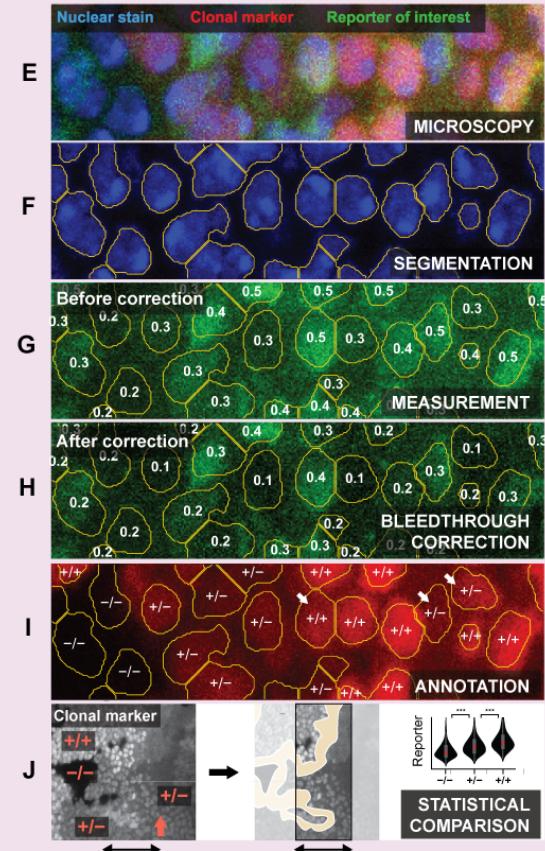
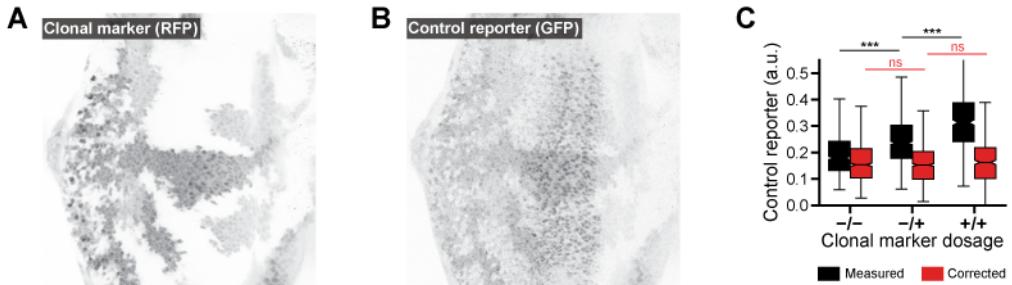
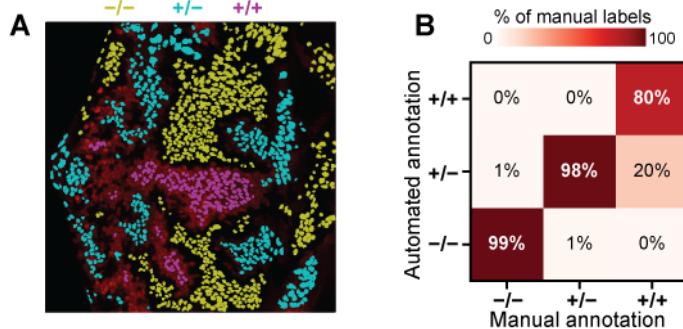


Figure 2.1. (Continued on next page.)

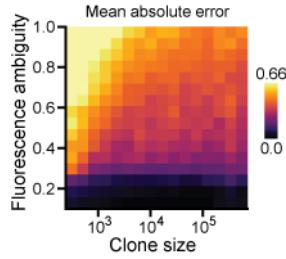
**Figure 2.1. A framework for conducting quantitative clonal analysis.** (A,B) Experimental framework using mitotic clones to test whether or not regulatory interactions occur between a perturbation target and reporter of interest. Blue and green ovals represent the respective genes encoding the perturbation 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. (C,D) Conventional clonal analysis in the larval fruit fly eye. (C) Clones are identified by visual comparison of clonal marker fluorescence among nuclei. (D) 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. (E-J) Quantitative clonal analysis. Panels depict a magnified view of the region enclosed by red rectangles in panels C and D. (E) Raw confocal image of the nuclear stain, clonal marker, and reporter of interest. (F) Segmentation identifies distinct nuclei. (G) Reporter expression is quantified by averaging the pixel intensities within each segment. Numbers reflect measured values. (H) Measurements may be corrected to mitigate fluorescence bleedthrough. (I) Individual nuclei are labeled mutant, heterozygous, or homozygous for the clonal marker. White arrows mark nuclei with ambiguous fluorescence levels. (J) 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).



**Figure 2.2. 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. B.8E). 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,  $p > 0.05$ ).



**Figure 2.3. 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 overlayed on the RFP channel of the image shown in 2.1B. 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.



**Figure 2.4. Synthetic benchmarking of automated annotation performance.** Grid shows the mean absolute error (MAE) between assigned labels and the corresponding ground truth as a function of fluorescence ambiguity and clone size. Values shown reflect the average across 50 replicate simulations. Performance improves with increasing clone size and worsens with increasing fluorescence ambiguity. Clone size reflects the mean number of cells per clone.

## References

- [1] 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. ISSN 00928674. doi: 10.1016/S0092-8674(00)80240-8. URL [https://ac.els-cdn.com/S0092867400802408/1-s2.0-S0092867400802408-main.pdf?{\\_}tid=a1ac0cd4-e6df-46a9-929b-256dc6bb5031{&}acdnat=1521552994{\\_}0bd11e44665e33813f08708743cad4f8http://linkinghub.elsevier.com/retrieve/pii/S0092867400802408](https://ac.els-cdn.com/S0092867400802408/1-s2.0-S0092867400802408-main.pdf?{_}tid=a1ac0cd4-e6df-46a9-929b-256dc6bb5031{&}acdnat=1521552994{_}0bd11e44665e33813f08708743cad4f8http://linkinghub.elsevier.com/retrieve/pii/S0092867400802408).
- [2] 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. ISSN 00928674. doi: 10.1016/S0092-8674(00)80257-3. URL <https://www.sciencedirect.com/science/article/pii/S0092867400802573>.
- [3] James D. McGhee, Tetsunari Fukushige, Michael W. Krause, Stephanie E. Minnema, Barbara Goszczynski, 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. *Developmental Biology*, 327(2):551–565, 2009. ISSN 1095564X. doi: 10.1016/j.ydbio.2008.11.034. URL [https://ac.els-cdn.com/S0012160608013560/1-s2.0-S0012160608013560-main.pdf?{\\_}tid=9124866b-92ab-4aa5-835e-b0b1e01cb022{&}acdnat=1521552088{\\_}ccca6e5d118b216f1b71e3382b6e4263](https://ac.els-cdn.com/S0012160608013560/1-s2.0-S0012160608013560-main.pdf?{_}tid=9124866b-92ab-4aa5-835e-b0b1e01cb022{&}acdnat=1521552088{_}ccca6e5d118b216f1b71e3382b6e4263).
- [4] Mary O’Riordan and Rudolf Grosschedl. Coordinate regulation of B cell differentiation by the transcription factors EBF and E2A. *Immunity*, 11: 21–31, 1999. ISSN 10747613. doi: 10.1016/S1074-7613(00)80078-3. URL [https://ac.els-cdn.com/S1074761300800783/1-s2.0-S1074761300800783-main.pdf?{\\_}tid=21992660-5fde-4241-8e1c-f8e71a18638f{&}acdnat=1521552607{\\_}36c8de119def2a9c57f3edd31a4eeecd](https://ac.els-cdn.com/S1074761300800783/1-s2.0-S1074761300800783-main.pdf?{_}tid=21992660-5fde-4241-8e1c-f8e71a18638f{&}acdnat=1521552607{_}36c8de119def2a9c57f3edd31a4eeecd).
- [5] 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. *Experimental Cell Research*, 282(2):90–100, jan 2003. ISSN 00144827. doi: 10.1016/S0014-4827(02)00015-0. URL <https://www.sciencedirect.com/science/article/>

pii/S0014482702000150.

- [6] Gustavo J. Melen, Sagi Levy, Naama Barkai, and Ben-Zion Shilo. Threshold responses to morphogen gradients by zero-order ultrasensitivity. *Molecular Systems Biology*, pages 1–11, 2005. ISSN 1744-4292. doi: 10.1038/msb4100036. URL <http://msb.embopress.org/content/msb/1/1/2005.0028.full.pdf>.
- [7] 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*, 341:670–673, aug 2013. ISSN 0036-8075. doi: 10.1126/science.1240831. URL <http://science.sciencemag.org/content/sci/341/6146/670.full.pdf><http://www.sciencemag.org/content/341/6146/670.short>{%}5Cn<http://www.sciencemag.org/content/341/6146/670.full.pdf><http://www.jem.org/lookup/doi/10.1084/jem.20041535><http://www.sciencemag.org/cg>.
- [8] Guang Yao, Tae Jun Lee, Seiichi Mori, Joseph R. Nevins, and Lingchong You. A bistable Rb-E2F switch underlies the restriction point. *Nature Cell Biology*, 10(4):476–482, 2008. ISSN 14657392. doi: 10.1038/ncb1711. URL [https://www.researchgate.net/profile/Lingchong\\_{\\_}You/publication/5487129{\\_}A{\\_}bistable{\\_}Rb-E2F{\\_}switch{\\_}underlies{\\_}the{\\_}restriction{\\_}point/\\_links/00463529a47db8a7db000000.pdf](https://www.researchgate.net/profile/Lingchong_{_}You/publication/5487129{_}A{_}bistable{_}Rb-E2F{_}switch{_}underlies{_}the{_}restriction{_}point/_links/00463529a47db8a7db000000.pdf).
- [9] Byung Ouk Park, Robert Ahrends, and Mary N. Teruel. Consecutive Positive Feedback Loops Create a Bistable Switch that Controls Preadipocyte-to-Adipocyte Conversion. *Cell Reports*, 2(4):976–990, oct 2012. ISSN 22111247. doi: 10.1016/j.celrep.2012.08.038. URL <https://www.sciencedirect.com/science/article/pii/S2211124712002793>.
- [10] 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. *Molecular Systems Biology*, 14(9):e8140, jan 2018. doi: 10.1101/137299. URL <https://www.biorxiv.org/content/early/2018/01/26/137299>.
- [11] 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. ISSN 00928674. doi: 10.1016/0092-8674(94)90006-X. URL [http://www.cell.com/cell/pdf/0092-8674\(94\)90006-X.pdf](http://www.cell.com/cell/pdf/0092-8674(94)90006-X.pdf).
- [12] 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. ISSN 00928674. doi: 10.1016/j.cell.2006.06.052. URL <https://ac.els-cdn.com/S0092867406009809/1-s2.0-S0092867406009809-main>.

pdf?{\_}tid=92734ac4-a659-4761-9be4-796f090733d5{&}acdnat=1521556329{\_}7e66697d61f6e4109d31895a6a1a0c24.

- [13] Arjun Raj, Scott A. Rifkin, Erik Andersen, and Alexander van Oudenaarden. Variability in gene expression underlies incomplete penetrance. *Nature*, 463(7283):913–918, 2010. ISSN 1476-4687. doi: 10.1038/nature08781. URL <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2836165/pdf/nihms167246.pdf>.
- [14] Hitoshi Niwa, Jun-Ichi Miyazaki, and Austin G. Smith. Quantitative expression of Oct-3/4 defines differentiation, dedifferentiation or self-renewal of ES cells. *Nature Genetics*, 24(April):372–376, 2000. URL [https://www.nature.com/articles/ng0400{\\_}372.pdf](https://www.nature.com/articles/ng0400{_}372.pdf).
- [15] Christopher L. Frick, Clare Yarka, Harry Nunns, and Lea Goentoro. Sensing relative signal in the Tgf- $\beta$ /Smad pathway. *Proceedings of the National Academy of Sciences*, 114(14):E2975–E2982, mar 2017. ISSN 0027-8424. doi: 10.1073/pnas.1611428114. URL <http://www.ncbi.nlm.nih.gov/pubmed/28320972http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC5389321http://www.pnas.org/lookup/doi/10.1073/pnas.1611428114>.
- [16] Lea Goentoro and Marc W. Kirschner. Evidence that Fold-Change, and Not Absolute Level, of  $\beta$ -Catenin Dictates Wnt Signaling. *Molecular Cell*, 36(5):872–884, dec 2009. ISSN 10972765. doi: 10.1016/j.molcel.2009.11.017. URL <https://www.sciencedirect.com/science/article/pii/S1097276509008582>.
- [17] Keita Kamino, Yohei Kondo, Akihiko Nakajima, Mai Honda-Kitahara, Kunihiro Kaneko, and Satoshi Sawai. Fold-change detection and scale invariance of cellcell signaling in social amoeba. *Proceedings of the National Academy of Sciences*, 114(21):E4149–E4157, may 2017. ISSN 0027-8424. doi: 10.1073/pnas.1702181114. URL <http://www.ncbi.nlm.nih.gov/pubmed/28495969http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC5448168http://www.pnas.org/lookup/doi/10.1073/pnas.1702181114>.
- [18] 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. ISSN 10974172. doi: 10.1016/j.cell.2017.08.015. URL <https://www.sciencedirect.com/science/article/pii/S0092867417309406>.
- [19] 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. *Molecular Systems Biology*, 12:898, dec 2016. ISSN 1744-4292. doi: 10.15252/msb.20166910. URL <http://www.ncbi.nlm.nih.gov/pubmed/28034910http://www.pubmedcentral>.

- [nih.gov/articlerender.fcgi?artid=PMC5199120http://msb.embopress.org/lookup/doi/10.15252/msb.20166910](http://nih.gov/articlerender.fcgi?artid=PMC5199120http://msb.embopress.org/lookup/doi/10.15252/msb.20166910).
- [20] Lea Goentoro, Oren Shoval, Marc W. Kirschner, and Uri Alon. The incoherent feed-forward loop can provide fold-change detection in gene regulation. *Molecular Cell*, 36:894–899, dec 2009. ISSN 10972765. doi: 10.1016/j.molcel.2009.11.018. URL <https://www.sciencedirect.com/science/article/pii/S1097276509008594>.
  - [21] Miri Adler and Uri Alon. Fold-change detection in biological systems. *Current Opinion in Systems Biology*, 8:81–89, apr 2018. ISSN 2452-3100. doi: 10.1016/j.coisb.2017.12.005. URL <https://www.sciencedirect.com/science/article/pii/S2452310017301889https://doi.org/10.1016/j.coisb.2017.12.005>.
  - [22] Chris Q. Doe. Temporal patterning in the Drosophila CNS. *Annual Review of Cell and Developmental Biology*, 33:219–240, oct 2017. ISSN 1081-0706. doi: 10.1146/annurev-cellbio-111315-125210. URL <http://www.annualreviews.org/doi/10.1146/annurev-cellbio-111315-125210http://www.ncbi.nlm.nih.gov/pubmed/28992439%0Ahttp://www.annualreviews.org/doi/10.1146/annurev-cellbio-111315-125210>.
  - [23] 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*, 541:365–370, jan 2017. ISSN 14764687. doi: 10.1038/nature20794. URL <http://www.nature.com/articles/nature20794>.
  - [24] 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.
  - [25] 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. ISSN 00928674. doi: 10.1016/S0092-8674(00)00105-7. URL <https://www.sciencedirect.com/science/article/pii/S0092867400001057?via%3Dihub>.
  - [26] 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. ISSN 0950-1991.

- [27] Chunyan Xu, Rachelle 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. ISSN 00928674. doi: 10.1016/S0092-8674(00)00107-0. URL [http://www.cell.com/cell/pdf/S0092-8674\(00\)00107-0.pdf](http://www.cell.com/cell/pdf/S0092-8674(00)00107-0.pdf).
- [28] 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. ISSN 00928674. doi: 10.1016/S0092-8674(00)00106-9. URL <https://www.sciencedirect.com/science/article/pii/S0092867400001069>.
- [29] D Brunner, K Dürker, 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. ISSN 0028-0836. doi: 10.1038/370386a0. URL <http://www.ncbi.nlm.nih.gov/pubmed/8047146>.
- [30] 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.
- [31] 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. ISSN 0950-1991.
- [32] 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.
- [33] Gong Hong Wei, Gwenael Badis, Michael F. Berger, Teemu Kivioja, Kimmo Palin, Martin Enge, Martin Bonke, Artti 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 Journal*, 29(13):2147–2160, jul 2010. ISSN 02614189. doi: 10.1038/emboj.2010.106. URL <http://www.ncbi.nlm.nih.gov/pubmed/20517297> <http://www.ncbi.nlm.nih.gov/entrez/f.cgi?artid=PMC2905244>.
- [34] 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. ISSN 00166731. doi: 10.1534/genetics.112.146647.

- [35] 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 & Development*, 27(21):2293–2298, nov 2013. ISSN 08909369. doi: 10.1101/gad.225789.113. URL <http://www.ncbi.nlm.nih.gov/pubmed/24186975><http://www.ncbi.nlm.nih.gov/articlerender.fcgi?artid=PMC3828515>.
- [36] 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 and Development*, 32:389–401, mar 2018. ISSN 15495477. doi: 10.1101/gad.307132.117. URL <http://www.ncbi.nlm.nih.gov/pubmed/29535190><http://www.ncbi.nlm.nih.gov/articlerender.fcgi?artid=PMC5900712>.
- [37] 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. ISSN 0950-1991. doi: 10.1242/dev.165985. URL <http://www.ncbi.nlm.nih.gov/pubmed/29848501><http://www.ncbi.nlm.nih.gov/articlerender.fcgi?artid=PMC6053666><http://dev.biologists.org/lookup/doi/10.1242/dev.165985>.
- [38] 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. *Developmental Biology*, 385:263–278, jan 2014. ISSN 00121606. doi: 10.1016/j.ydbio.2013.11.002. URL <http://linkinghub.elsevier.com/retrieve/pii/S0012160613006027>.
- [39] 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. ISSN 1477-9129. doi: 10.1242/dev.044826. URL <http://dev.biologists.org/content/develop/137/14/2265.full.pdf><http://dev.biologists.org/cgi/doi/10.1242/dev.044826><http://www.ncbi.nlm.nih.gov/pubmed/20570936><http://www.ncbi.nlm.nih.gov/articlerender>.
- [40] 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.
- [41] 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. ISSN 1477-9129. doi: 10.1242/dev.093138. URL <http://dev.biologists.org/content/develop/140/13/2746.full>.

- pdf<http://dev.biologists.org/lookup/doi/10.1242/dev.093138>.
- [42] Matthew Freeman. Reiterative use of the EGF receptor triggers differentiation of all cell types in the Drosophila eye. *Cell*, 87:651–660, nov 1996. ISSN 00928674. doi: 10.1016/S0092-8674(00)81385-9. URL <https://www.sciencedirect.com/science/article/pii/S0092867400813859>.
  - [43] Tanya Wolff and Donald F. Ready. Pattern formation in the Drosophila retina. In M. Bate and A. Martinez-Arias, editors, *In the Development of Drosophila Melanogaster Vol. 2*, pages 1277–1325. Cold Spring Harbor Press, Plainview, NY, 1993.
  - [44] Matthew G. Voas and Ilaria Rebay. Signal integration during development: Insights from the Drosophila eye, jan 2004. ISSN 10588388. URL <http://doi.wiley.com/10.1002/dvdy.10449>.
  - [45] 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. URL <http://www.nature.com/nature/journal/v369/n6479/pdf/369398a0.pdf>.
  - [46] 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 atonal by Ey and So. *Development*, 133(24):4881–4889, dec 2006. ISSN 0950-1991. doi: 10.1242/dev.02669. URL <http://www.ncbi.nlm.nih.gov/pubmed/17108002><http://dev.biologists.org/cgi/doi/10.1242/dev.02669>.
  - [47] T. Wolff and Donalf F. Ready. Cell death in normal and rough eye mutants of Drosophila. *Development*, 113(3):825–839, 1991. ISSN 0950-1991. doi: 8223268.
  - [48] 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. ISSN 2050-084X. doi: 10.7554/eLife.08924. URL <http://www.ncbi.nlm.nih.gov/pubmed/26583752><http://www.ncbi.nlm.nih.gov/articlerender.fcgi?artid=PMC4720516>.
  - [49] 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. ISSN 0028-0836. doi: 10.1038/36555a0. URL <http://www.nature.com/doifinder/10.1038/36555a0>[papers2://publication/doi/10.1038/36555a0](http://www.nature.com/doifinder/10.1038/36555a0).

- [50] Nicolás Peláez. *Dynamics and heterogeneity of fate determinants during transition towards cell differentiation*. PhD thesis, Northwestern University, 2016.
- [51] Konrad Basler and Ernst Hafen. Dynamics of Drosophila eye development and temporal requirements of sevenless expression. *Development*, 107:723–731, 1989.
- [52] 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 Archives of Developmental Biology*, 181:227–245, 1977.
- [53] Donald F. Ready, Thomas E. Hanson, and Seymour Benzer. Development of the Drosophila retina, a neurocrystalline lattice. *Developmental Biology*, 53(2):217–240, oct 1976. ISSN 00121606. doi: 10.1016/0012-1606(76)90225-6. URL <http://linkinghub.elsevier.com/retrieve/pii/0012160676902256>.
- [54] Andrew Tomlinson. The cellular dynamics of pattern formation in the eye of Drosophila. *Development*, 89:313–331, 1985. ISSN 0022-0752. URL <http://www.ncbi.nlm.nih.gov/pubmed/3937883>.
- [55] Andrew Tomlinson and Donald F. Ready. Neuronal differentiation in the Drosophila ommatidium. *Developmental Biology*, 120(2):366–376, apr 1987. ISSN 00121606. doi: 10.1016/0012-1606(87)90239-9. URL <http://linkinghub.elsevier.com/retrieve/pii/0012160687902399>.
- [56] Kazuhiro R. Nitta, Arttu 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. ISSN 2050084X. doi: 10.7554/eLife.04837. URL <https://search.proquest.com/openview/7d6023879cf2bf4acad5238131f7acf3/1?pq-origsite=gscholar&cbl=2045579>.
- [57] 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. ISSN 00928674. doi: 10.1016/j.cell.2004.07.010. URL <https://www.sciencedirect.com/science/article/pii/S0092867404006634> [http://www.ncbi.nlm.nih.gov/pubmed/15238131](https://www.sciencedirect.com/science/article/pii/S0092867404006634).
- [58] Matthew C. Hope, Ilaria Rebay, and John Reinitz. DNA Occupancy of polymerizing transcription factors: A chemical model of the ETS family factor Yan. *Biophysical Journal*, 112(1):180–192, jan 2017. ISSN 15420086. doi: 10.1016/j.bpj.2016.11.901. URL <https://www.sciencedirect.com/science/article/pii/S0006349516319312>.

- [59] H. S. Fogler. *Elements of chemical reaction engineering*. Prentice Hall, 1987. ISBN 0139737855. doi: 10.1016/0009-2509(87)80130-6.
- [60] 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*. *Biochemical Genetics*, 9(1):41–51, may 1973. ISSN 0006-2928. doi: 10.1007/BF00485589. URL <http://link.springer.com/10.1007/BF00485589>.
- [61] David K. Lubensky, Matthew W. Pennington, Boris I. Shraiman, and Nicholas E. Baker. A dynamical model of ommatidial crystal formation. *Proceedings of the National Academy of Sciences*, 108(27):11145–11150, 2011. ISSN 0027-8424. doi: 10.1073/pnas.1015302108. URL <http://www.pnas.org/content/108/27/11145.full.pdf> <http://www.pnas.org/cgi/doi/10.1073/pnas.1015302108>.
- [62] 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. *Nature Communications*, 7:10461, 2016. ISSN 2041-1723. doi: 10.1038/ncomms10461. URL <https://www.nature.com/articles/ncomms10461.pdf> <http://www.nature.com/doifinder/10.1038/ncomms10461>.
- [63] 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. ISSN 0950-1991.
- [64] 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. *Current Biology*, 12:576–581, 2002.
- [65] 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 (Cambridge, England)*, 130(16):3703–14, apr 2003. ISSN 0950-1991. doi: 10.1242/dev.00556. URL <http://www.ncbi.nlm.nih.gov/pubmed/11245584> <http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?artid=2778067&tool=pmcentrez&rendertype=abstract>.
- [66] Ross L. Cagan and Donald F. Ready. Notch is required for successive cell decisions in the developing *Drosophila* retina. *Genes & Development*, 3:1099–1112, 1989. ISSN 08909369. doi: 10.1101/gad.3.8.1099. URL <http://genesdev.cshlp.org/content/3/8/1099.full.pdf>.
- [67] Nicholas E. Baker and Sung-Yun Yu. Proneural function of neurogenic genes in the developing *Drosophila* eye. *Current Biology*, 7:122–132, feb 1997. ISSN 09609822. doi: 10.1016/S0960-9822(06)00056-X. URL <https://www.sciencedirect.com/science/article/pii/S096098220600056X>.

- [68] Yanxia Li and Nicholas E. Baker. Proneural enhancement by Notch overcomes suppressor-of-hairless repressor function in the developing Drosophila eye. *Current Biology*, 11(5):330–338, mar 2001. ISSN 09609822. doi: 10.1016/S0960-9822(01)00093-8. URL <https://www.sciencedirect.com/science/article/pii/S0960982201000938>.
- [69] 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. ISSN 0950-1991.
- [70] 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. ISSN 00928674. doi: 10.1016/0092-8674(91)90065-7. URL <https://www.sciencedirect.com/science/article/pii/0092867491900657>.
- [71] 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. ISSN 0028-0836. doi: 10.1038/355559a0. URL <https://www.nature.com/articles/355559a0.pdf> <http://www.ncbi.nlm.nih.gov/pubmed/1311054> <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC1311054/>
- [72] François Spitz and Eileen E. M. Furlong. Transcription factors: From enhancer binding to developmental control. *Nature Reviews Genetics*, 13(9):613–626, sep 2012. ISSN 14710056. doi: 10.1038/nrg3207. URL <http://www.nature.com/articles/nrg3207>.
- [73] Raghavendra Nagaraj, Jude Canon, and Utpal Banerjee. Cell fate specification in the Drosophila eye. In *Results and Problems in Cell Differentiation*, Vol. 37, pages 73–88. Springer, Berlin, Heidelberg, 2002. doi: 10.1007/978-3-540-45398-7\_6. URL [http://link.springer.com/10.1007/978-3-540-45398-7\\_6](http://link.springer.com/10.1007/978-3-540-45398-7_6).
- [74] 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. ISSN 14764687. doi: 10.1038/nature15710. URL <http://www.nature.com/articles/nature15710>.
- [75] 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 Journal*, 20(15):4173–4182, 2001. ISSN 02614189. doi: 10.1093/emboj/20.15.4173.

- [76] 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. *Molecular and Cellular Biology*, 30(5):1158–1170, mar 2010. ISSN 0270-7306. doi: 10.1128/MCB.01225-09. URL <http://www.ncbi.nlm.nih.gov/pubmed/20048052> <http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?artid=PMC2820895> <http://mcb.asm.org/cgi/doi/10.1128/MCB.01225-09>.
- [77] Lihui Yang and Nicholas E. Baker. Notch activity opposes ras-induced differentiation during the second mitotic wave of the developing Drosophila eye. *BMC Developmental Biology*, 6(8):1–10, feb 2006. ISSN 1471213X. doi: 10.1186/1471-213X-6-8. URL <http://bmcdevbiol.biomedcentral.com/articles/10.1186/1471-213X-6-8>.
- [78] Andrew C. Oates, Nicole Gorfinkiel, Marcos González-Gaitán, and Carl Philipp Heisenberg. Quantitative approaches in developmental biology, aug 2009. ISSN 14710056. URL <http://www.nature.com/articles/nrg2548>.
- [79] Dale Muzzey and Alexander van Oudenaarden. Quantitative Time-Lapse Fluorescence Microscopy in Single Cells. *Annual Review of Cell and Developmental Biology*, 25(1):301–327, nov 2009. ISSN 1081-0706. doi: 10.1146/annurev.cellbio.042308.113408. URL <http://www.annualreviews.org/doi/10.1146/annurev.cellbio.042308.113408>.
- [80] Ernst H K Stelzer. Light-sheet fluorescence microscopy for quantitative biology. *Nature Methods*, 12(1):23–26, dec 2014. ISSN 1548-7091. doi: 10.1038/nmeth.3219.
- [81] Thai V. Truong and Willy Supatto. Toward high-content/high-throughput imaging and analysis of embryonic morphogenesis. *Genesis*, 49(7):555–569, jul 2011. ISSN 1526954X. doi: 10.1002/dvg.20760. URL <http://doi.wiley.com/10.1002/dvg.20760>.
- [82] 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, Tatio 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. *Nature Methods*, 10(3):228–238, mar 2013. ISSN 15487091. doi: 10.1038/NMETH.2365. URL <http://www.nature.com/articles/nmeth.2365>.
- [83] 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. *Clinical Immunology*, 157(2):249–260, apr 2015. ISSN 15217035. doi: 10.1016/j.clim.2014.12.009. URL <https://www.sciencedirect.com/science/article/pii/S1521661614002873>.
- [84] 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. *Proceedings of the National Academy of Sciences*, 106(21):8519–8524, may 2009. ISSN 0027-8424. doi: 10.1073/pnas.0903028106. URL <http://www.ncbi.nlm.nih.gov/pubmed/19443687><http://www.ncbi.nlm.nih.gov/articlerender.fcgi?artid=PMC2682540>.
- [85] 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 Biology*, 9(9):R137, sep 2008. ISSN 1465-6906. doi: 10.1186/gb-2008-9-9-r137. URL <http://genomebiology.biomedcentral.com/articles/10.1186/gb-2008-9-9-r137>.
- [86] 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 Biology*, 8(2), feb 2007. ISSN 1465-6914. doi: 10.1186/gb-2007-8-2-r19. URL <http://genomebiology.biomedcentral.com/articles/10.1186/gb-2007-8-2-r19>.
- [87] Ben Langmead and Steven L Salzberg. Fast gapped-read alignment with Bowtie 2. *Nature methods*, 9(4):357–9, apr 2012. ISSN 1548-7105. doi: 10.1038/nmeth.1923. URL <http://www.nature.com/articles/nmeth.1923><http://www.ncbi.nlm.nih.gov/pubmed/22388286>{%}0A<http://www.ncbi.nlm.nih.gov/articlerender.fcgi?artid=PMC3322381>.
- [88] Cole Trapnell, Lior Pachter, and Steven L. Salzberg. TopHat: Discovering splice junctions with RNA-Seq. *Bioinformatics*, 25(9):1105–1111, may 2009. ISSN 13674803. doi:

- 10.1093/bioinformatics/btp120. URL <https://academic.oup.com/bioinformatics/article-lookup/doi/10.1093/bioinformatics/btp120>.
- [89] 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. *Biophysical Journal*, 86(6):3993–4003, jun 2004. ISSN 00063495. doi: 10.1529/biophysj.103.038422. URL <https://www.sciencedirect.com/science/article/pii/S0006349504744392>.
  - [90] 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. *Nature Protocols*, 10(6):845–858, jun 2015. ISSN 17502799. doi: 10.1038/nprot.2015.053. URL <http://www.nature.com/articles/nprot.2015.053>.
  - [91] 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 Biology*, 7(10):R100, oct 2006. ISSN 14747596. doi: 10.1186/gb-2006-7-10-r100. URL <http://genomebiology.biomedcentral.com/articles/10.1186/gb-2006-7-10-r100>.
  - [92] 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. *Molecular Microbiology*, 99(4):767–777, feb 2016. ISSN 13652958. doi: 10.1111/mmi.13264. URL <http://doi.wiley.com/10.1111/mmi.13264>.
  - [93] 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. ISSN 15487091. URL <http://www.nature.com/articles/nmeth.2019>.
  - [94] Christoph Sommer, Christoph Straehle, Ullrich Kothe, and Fred A. Hamprecht. Ilastik: Interactive learning and segmentation toolkit. In *Proceedings - International Symposium on Biomedical Imaging*, pages 230–233. IEEE, mar 2011. ISBN 9781424441280. doi: 10.1109/ISBI.2011.5872394. URL <http://ieeexplore.ieee.org/document/5872394/>.
  - [95] Florian Jug, Tobias Pietzsch, Stephan Preibisch, and Pavel Tomancak. Bioimage Informatics in the context of Drosophila research. *Methods*, 68(1):60–73, jun 2014. ISSN 10959130. doi: 10.1016/j.ymeth.2014.04.004. URL <https://www.sciencedirect.com/science/article/pii/S1046202314001480>.

- [96] Ivo F. Sbalzarini. Seeing is believing: Quantifying is convincing: Computational image analysis in biology. *Advances in Anatomy, Embryology and Cell Biology*, 219:1–39, 2016. ISSN 03015556. doi: 10.1007/978-3-319-28549-8\_1. URL [http://link.springer.com/10.1007/978-3-319-28549-8{\\_}1](http://link.springer.com/10.1007/978-3-319-28549-8{_}1).
- [97] Johannes Schindelin, Curtis T. Rueden, Mark C. Hiner, and Kevin W. Eliceiri. The ImageJ ecosystem: An open platform for biomedical image analysis, jul 2015. ISSN 10982795. URL <http://doi.wiley.com/10.1002/mrd.22489>.
- [98] Ian T. Simpson and David J. Price. Pax6; a pleiotropic player in development, nov 2002. ISSN 02659247. URL <http://doi.wiley.com/10.1002/bies.10174>.
- [99] 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. ISSN 00166731.
- [100] Ben Zion Shilo and Erez Raz. Developmental control by the *Drosophila* EGF receptor homolog DER, nov 1991. ISSN 01689525. URL <https://www.sciencedirect.com/science/article/pii/016895259190261N>.
- [101] T Xu and G M Rubin. Analysis of genetic mosaics in developing and adult *Drosophila* tissues. *Development (Cambridge, England)*, 117(4):1223–37, apr 1993. ISSN 0950-1991. doi: 8404527. URL <http://www.ncbi.nlm.nih.gov/pubmed/8404527>.
- [102] Tian Xu and Gerald M Rubin. The effort to make mosaic analysis a household tool. *Development*, 139(24):4501–4503, dec 2012. ISSN 0950-1991. doi: 10.1242/dev.085183. URL <http://www.ncbi.nlm.nih.gov/pubmed/23172911><http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3509720/>.
- [103] Jorge V. Beira and Renato Paro. The legacy of *Drosophila* imaginal discs, sep 2016. ISSN 14320886. URL <http://link.springer.com/10.1007/s00412-016-0595-4>.
- [104] T P Newsome, B Asling, and B J Dickson. Analysis of *Drosophila* photoreceptor axon guidance in eye-specific mosaics. *Development (Cambridge, England)*, 127(4):851–60, feb 2000. ISSN 0950-1991. URL <http://www.ncbi.nlm.nih.gov/pubmed/10648243>.
- [105] Nicole A. Theodosiou and Tian Xu. Use of FLP/FRT system to study *Drosophila* development. *Methods: A Companion to Methods in Enzymology*, 14(4):355–365, apr 1998. ISSN 10462023. doi: 10.1006/meth.1998.0591. URL <https://www.sciencedirect.com/science/article/pii/S1046202398905916?via%23Dihub>.
- [106] Gary Struhl and Konrad Basler. Organizing activity of wingless protein in *Drosophila*. *Cell*, 1993. ISSN 00928674. doi: 10.1016/0092-8674(93)90072-X.

- [107] 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 (Cambridge, England)*, 128(9):1687–96, may 2001. ISSN 0950-1991. URL <http://www.ncbi.nlm.nih.gov/pubmed/11290305>.
- [108] A Tomlinson and G Struhl. Delta/Notch and Boss/Sevenless signals act combinatorially to specify the Drosophila R7 photoreceptor. *Molecular cell*, 7(3):487–95, mar 2001. ISSN 1097-2765. URL <http://www.ncbi.nlm.nih.gov/pubmed/11463374>.
- [109] L Yang and N E Baker. Role of the EGFR/Ras/Raf pathway in specification of photoreceptor cells in the Drosophila retina. *Development (Cambridge, England)*, 128(7):1183–91, apr 2001. ISSN 0950-1991. URL <http://www.ncbi.nlm.nih.gov/pubmed/11245584>.
- [110] Jianbin Huang, Shian Wu, Jose Barrera, Krista Matthews, and Duoja 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. ISSN 00928674. doi: 10.1016/j.cell.2005.06.007. URL <https://www.sciencedirect.com/science/article/pii/S0092867405005520>.
- [111] 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. ISSN 00928674. doi: 10.1016/j.cell.2006.07.013. URL <https://www.sciencedirect.com/science/article/pii/S0092867406009536>{#}bib4.
- [112] Mardelle Atkins. Drosophila Genetics: The Power of Genetic Mosaic Approaches. In *Methods in Molecular Biology*, volume 1893, pages 27–42. Humana Press, New York, NY, 2019. doi: 10.1007/978-1-4939-8910-2\_2. URL [http://link.springer.com/10.1007/978-1-4939-8910-2\\_{\\_}2](http://link.springer.com/10.1007/978-1-4939-8910-2_{_}2).
- [113] Masato Enomoto, Carmen Siow, and Tatsushi Igaki. Drosophila as a cancer model. In *Advances in Experimental Medicine and Biology*, volume 1076, pages 173–194. Springer, Singapore, 2018. ISBN 9789811305290. doi: 10.1007/978-981-13-0529-0\_10. URL [http://link.springer.com/10.1007/978-981-13-0529-0\\_{\\_}10](http://link.springer.com/10.1007/978-981-13-0529-0_{_}10).
- [114] Federico Germani, Cora Bergantinos, and Laura A Johnston. Mosaic analysis in Drosophila. *Genetics*, 208(2):473–490, feb 2018. ISSN 19432631. doi: 10.1534/genetics.117.300256. URL <http://www.ncbi.nlm.nih.gov/pubmed/29378809><http://www.ncbi.nlm.nih.gov/entrez/fetch?artid=PMC5788516>.
- [115] 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. URL <https://www.biorxiv.org/content/10.1101/430744v1>.
- [116] Haley Burrous, Thomas Kenney, Wei Dai, Denise J. Montell, and Amy Peterson. Quantitative microscopy of the Drosophila ovary shows multiple niche signals specify progenitor cell fate. *Nature Communications*, 8(1):1244, dec 2017. ISSN 2041-1723. doi: 10.1038/s41467-017-01322-9. URL <http://www.nature.com/articles/s41467-017-01322-9>.
  - [117] 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. ISSN 0950-1991. doi: 10.1242/dev.161117. URL <http://dev.biologists.org/content/145/14/dev161117.abstract>.
  - [118] Ke Li and Nicholas E Baker. Regulation of the Drosophila ID protein Extra macrochaetae by proneural dimerization partners. *eLife*, 7, 2018. ISSN 2050-084X. doi: 10.7554/elife.33967. URL <http://www.ncbi.nlm.nih.gov/pubmed/29687780><http://www.ncbi.nlm.nih.gov/articlerender.fcgi?artid=PMC5915177>.
  - [119] 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. ISSN 14394235. doi: 10.1002/cphc.201100801. URL <http://doi.wiley.com/10.1002/cphc.201100801>.
  - [120] 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. ISSN 10462023. doi: 10.1016/S1046-2023(02)00283-9. URL <https://www.sciencedirect.com/science/article/pii/S1046202302002839>.
  - [121] Richard L. Mort. Quantitative analysis of patch patterns in mosaic tissues with ClonalTools software. *Journal of Anatomy*, 215(6):698–704, dec 2009. ISSN 00218782. doi: 10.1111/j.1469-7580.2009.01150.x. URL <http://doi.wiley.com/10.1111/j.1469-7580.2009.01150.x>.
  - [122] 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. ISSN 14712105. doi: 10.1186/1471-2105-11-372. URL <http://bmcbioinformatics.biomedcentral.com/articles/10.1186/1471-2105-11-372>.
  - [123] 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. ISSN 14712105. doi: 10.1186/1471-2105-14-349. URL <http://>

[bmcbioinformatics.biomedcentral.com/articles/10.1186/1471-2105-14-349](http://bmcbioinformatics.biomedcentral.com/articles/10.1186/1471-2105-14-349).

- [124] 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. ISSN 2167-8359. doi: 10.7717/peerj.453.
- [125] Nobuyuki Otsu. A Threshold Selection Method from Gray-Level Histograms. *IEEE Transactions on Systems, Man, and Cybernetics*, 1979.
- [126] 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. *Nature Protocols*, 9(3):586–596, mar 2014. ISSN 1754-2189. doi: 10.1038/nprot.2014.037. URL <http://www.nature.com/articles/nprot.2014.037>.
- [127] 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. ISSN 0044-5991. doi: 10.1267/ahc.07002. URL <http://joi.jlc.jst.go.jp/JST.JSTAGE/ahc/07002?from=CrossRef>.
- [128] 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. *Scientific Reports*, 7(1), dec 2017. ISSN 20452322. doi: 10.1038/s41598-017-15411-8. URL <http://www.nature.com/articles/s41598-017-15411-8>.
- [129] 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. *Journal of Optics*, 15(9), sep 2013. ISSN 20408978. doi: 10.1088/2040-8978/15/9/094011. URL <http://www.ncbi.nlm.nih.gov/pubmed/26185614><http://www.ncbi.nlm.nih.gov/articlerender.fcgi?artid=PMC4501387>.
- [130] Patrick D. McMullen, Richard I. Morimoto, and Luís A. Nunes Amaral. Physically grounded approach for estimating gene expression from microarray data. *Proceedings of the National Academy of Sciences*, 107(31):13690–13695, aug 2010. ISSN 1091-6490. doi: 10.1073/pnas.1000938107. URL <http://www.ncbi.nlm.nih.gov/pubmed/20643961><http://www.ncbi.nlm.nih.gov/articlerender.fcgi?artid=PMC2922271><http://dx.doi.org/10.1073/pnas.1000938107>{%}5Cn<http://www.pnas.org/content/107/31/13690>.abstract{%}5Cn<http://www.pnas.org/content/107/31/13690.ful>.
- [131] Lisa Gaudette and Nathalie Japkowicz. Evaluation methods for ordinal classification. In *Lecture Notes in Computer Science*, volume 5549 LNAI, pages 207–210. Springer,

- Berlin, Heidelberg, 2009. ISBN 3642018173. doi: 10.1007/978-3-642-01818-3\_25. URL [http://link.springer.com/10.1007/978-3-642-01818-3{\\_}25](http://link.springer.com/10.1007/978-3-642-01818-3{_}25).
- [132] Thanh Minh Nguyen and Q. M.Jonathan Wu. Gaussian mixture-model-based spatial neighborhood relationships for pixel labeling problems. *IEEE Transactions on Systems, Man, and Cybernetics*, 42(1):193–202, feb 2012. ISSN 10834419. doi: 10.1109/TSMCB.2011.2161284. URL <http://ieeexplore.ieee.org/document/5983453/>.
- [133] Alexis Gambis, Pierre Dourlen, Hermann Steller, and Bertrand Mollereau. Two-color in vivo imaging of photoreceptor apoptosis and development in Drosophila. *Developmental Biology*, 351(1):128–134, mar 2011. ISSN 1095564X. doi: 10.1016/j.ydbio.2010.12.040. URL <https://www.sciencedirect.com/science/article/pii/S0012160610012960>.
- [134] 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. *Journal of Visualized Experiments*, 79, sep 2013. ISSN 1940-087X. doi: 10.3791/50610. URL <http://www.ncbi.nlm.nih.gov/pubmed/24084155http://www.ncbi.nlm.nih.gov/articlerender.fcgi?artid=PMC3923918>.
- [135] 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. ISSN 2050084X. doi: 10.7554/eLife.22279. URL <http://www.ncbi.nlm.nih.gov/pubmed/28211790http://www.ncbi.nlm.nih.gov/articlerender.fcgi?artid=PMC5342825>.
- [136] Joy S Wu and Liqun Luo. A protocol for mosaic analysis with a repressible cell marker (MARCM) in Drosophila. *Nature Protocols*, 1(6):2583–2589, jan 2007. ISSN 1754-2189. doi: 10.1038/nprot.2006.320. URL <http://www.ncbi.nlm.nih.gov/pubmed/17406512http://www.nature.com/doifinder/10.1038/nprot.2006.320>.
- [137] 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. ISSN 1526968X. doi: 10.1002/dvg.22984. URL <http://doi.wiley.com/10.1002/dvg.22984>.
- [138] 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. *Nature Methods*, 6(8):600–602, aug 2009. ISSN 1548-7091. doi: 10.1038/nmeth.1349. URL <http://www.ncbi.nlm.nih.gov/pubmed/19633664http://www.ncbi.nlm.nih.gov/articlerender.fcgi?artid=PMC2720837>.

- [139] 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 Biology*, 8(8):39–40, aug 2010. ISSN 15449173. doi: 10.1371/journal.pbio.1000461. URL <https://dx.plos.org/10.1371/journal.pbio.1000461>.
- [140] 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. ISSN 0950-1991. doi: 10.1242/dev.102913. URL <http://dev.biologists.org/content/141/2/472.short>.
- [141] 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*. *Nature Methods*, 8(3):260–266, mar 2011. ISSN 15487091. doi: 10.1038/nmeth.1567. URL <http://www.nature.com/articles/nmeth.1567>.
- [142] 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. *Nature Methods*, 8(3):253–259, mar 2011. ISSN 1548-7091. doi: 10.1038/nmeth.1566. URL <http://www.nature.com/articles/nmeth.1566>.
- [143] 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. ISSN 00928674. doi: 10.1016/S0092-8674(00)81462-2. URL <https://www.sciencedirect.com/science/article/pii/S0092867400814622>.
- [144] 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. ISSN 00166731.
- [145] Russell T Collins, Claudia Linker, and Julian Lewis. MAZe: A tool for mosaic analysis of gene function in zebrafish. *Nature Methods*, 7(3):219–223, mar 2010. ISSN 15487091. doi: 10.1038/nmeth.1423. URL <http://www.nature.com/doifinder/10.1038/nmeth.1423>.
- [146] 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. ISSN 19432631. doi: 10.1534/genetics.117.201012. URL <http://www.ncbi.nlm.nih.gov/pubmed/28646043> <http://www.ncbi.nlm.nih.gov/articlerender.fcgi?artid=PMC5560786>.

- [147] Wei Wang, Madhuri Warren, and Allan Bradley. Induced mitotic recombination of p53 in vivo. *Proceedings of the National Academy of Sciences*, 104(11):4501–4505, mar 2007. ISSN 0027-8424. doi: 10.1073/pnas.0607953104. URL <http://www.ncbi.nlm.nih.gov/pubmed/17360553><http://www.ncbi.nlm.nih.gov/articlerender.fcgi?artid=PMC1838630>.
- [148] Erik Meijering. Cell Segmentation: 50 Years Down the Road. *IEEE Signal Processing Magazine*, 29(5):140–145, sep 2012. ISSN 1053-5888. doi: 10.1109/msp.2012.2204190. URL <http://ieeexplore.ieee.org/document/6279591/>.
- [149] David L. Shellenbarger and Dawson J. Mohler. Temperature-sensitive mutations of the notch locus in *Drosophila melanogaster*. *Genetics*, 81(1):143–162, 1975. ISSN 1943-2631.
- [150] 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. ISSN 00166731. doi: 10.1016/0168-9525(90)90089-O.
- [151] Steven Fortune. Voronoi diagrams and Delaunay triangulations. *Computing in Euclidean geometry*, 1992. ISSN 00983004. doi: 10.1016/0098-3004(93)90062-A.
- [152] Abraham Savitzky and Marcel J.E. Golay. Smoothing and Differentiation of Data by Simplified Least Squares Procedures. *Analytical Chemistry*, 36(8):1627–1639, 1964. ISSN 15206882. doi: 10.1021/ac60214a047.
- [153] Zhi-Xin Wang. An exact mathematical expression for describing competitive binding of two different ligands to a protein molecule. *FEBS Letters*, 360: 111–114, 1995. ISSN 00145793. doi: 10.1016/0014-5793(95)00062-E. URL [http://ac.els-cdn.com/001457939500062E/1-s2.0-001457939500062E-main.pdf?{\\_}tid=2fc90d24-7fd1-11e7-94e3-0000aacb35e{&}acdnat=1502592388{\\_}8237163457ade068f84572379522ea03](http://ac.els-cdn.com/001457939500062E/1-s2.0-001457939500062E-main.pdf?{_}tid=2fc90d24-7fd1-11e7-94e3-0000aacb35e{&}acdnat=1502592388{_}8237163457ade068f84572379522ea03).
- [154] Jacob VanderPlas, Andrew J. Connolly, Zeljko Ivezić, and Alex Gray. Introduction to astroML: Machine learning for astrophysics. In *Proceedings - CIDU 2012*, pages 47–54, 2012. ISBN 9781467346252. doi: 10.1109/CIDU.2012.6382200. URL <http://ieeexplore.ieee.org/ielx5/6373897/6382172/06382200.pdf?tp={&}arnumber=6382200{&}isnumber=6382172>.
- [155] Jacob VanderPlas. Understanding the Lomb-Scargle periodogram. *The Astrophysical Journal Supplement Series*, 236(1):1–55, 2018. URL <http://arxiv.org/abs/1703.09824>.
- [156] Chikara Furusawa, Takao Suzuki, Akiko Kashiwagi, Tetsuya Yomo, and Kunihiko Kaneko. Ubiquity of Log-normal Distributions in Intra-cellular Reaction Dynamics.

*Biophysics*, 1:25–31, 2005. ISSN 1349-2942. doi: 10.2142/biophysics.1.25. URL <http://www.ncbi.nlm.nih.gov/pubmed/27857550><http://www.ncbi.nlm.nih.gov/articlerender.fcgi?artid=PMC5036630><http://arxiv.org/abs/q-bio/0503040>.

- [157] Jacob Beal. Biochemical complexity drives log-normal variation in genetic expression. *Engineering Biology*, 1(1):55–60, jun 2017. ISSN 2398-6182. doi: 10.1049/enb.2017.0004. URL <https://digital-library.theiet.org/content/journals/10.1049/enb.2017.0004>.
- [158] Henry Teicher. Identifiability of Finite Mixtures. *The Annals of Mathematical Statistics*, 34(4):1265–1269, dec 1963. ISSN 0003-4851. doi: 10.1214/aoms/1177703862. URL <http://projecteuclid.org/euclid.aoms/1177703862>.
- [159] M. Rosvall, D. Axelsson, and C. T. Bergstrom. The map equation. *European Physical Journal: Special Topics*, 2009. ISSN 19516355. doi: 10.1140/epjst/e2010-01179-1.
- [160] Leo Katz. A new status index derived from sociometric analysis. *Psychometrika*, 1953. ISSN 00333123. doi: 10.1007/BF02289026.
- [161] Tomihisa Kamada and Satoru Kawai. An algorithm for drawing general undirected graphs. *Information Processing Letters*, 31(1):7–15, apr 1989. ISSN 00200190. doi: 10.1016/0020-0190(89)90102-6. URL <https://www.sciencedirect.com/science/article/pii/0020019089901026?via%23Dihub>.

## APPENDIX A

### Methods

#### A.1. Automated analysis of mosaic eye imaginal discs in *Drosophila*

##### A.1.1. Genetics and microscopy of *Drosophila* eye imaginal discs

We borrowed an experimental control dataset from an unrelated study of neuronal fate commitment during retinal patterning in *Drosophila* [115]. The data consist of six eye imaginal discs dissected and fixed during the third larval instar of *Drosophila* development. Within each disc, *ey<sub>d</sub>FLP* and *FRT40A* were used to generate mitotic clones. The chromosome targeted for recombination was marked with a *Ubi-mRFPnls* transgene, enabling automated detection of subpopulations characterized by distinct levels of mRFP fluorescence. 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.2.1 for additional details regarding genetics and experimental conditions.

During eye development, the PntGFP reporter is predominantly expressed in two narrow stripes of progenitor cells [115]. 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.

### A.1.2. Characterization of fluorescence bleedthrough coefficients

For each image, we morphologically dilate the foreground until no features remain visible (Fig. B.2A). We then extract the background pixels and resample them such that the distribution of pixel intensities is approximately uniform (Fig. B.2B). 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. B.2C). 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. B.2D).

### A.1.3. Statistical comparison of fluorescence levels

To mitigate edge effects, cells residing on the periphery of each clone were excluded from all comparisons (Fig. B.2E). Border cells were identified by using the undirected graph constructed during annotation to find all cells connected to a neighbor within a different clone (Fig. B.4B). Comparisons were limited to the region of elevated GFP expression near the morphogenetic furrow (Fig. B.2F), then further limited to cells undergoing similar stages of development (Fig. B.2G). 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 by two-sided Mann-Whitney  $U$  test.

#### A.1.4. Clone annotation algorithm

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:

$$x_i \sim p_m(x) \quad (\text{A.1})$$

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

$$p_m(\ln x) = \sum_{n=1}^N \lambda_n \mathcal{N}(\ln x | \theta_n) \quad (\text{A.2})$$

$$\sum_{n=1}^N \lambda_n = 1 \quad (\text{A.3})$$

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 [156, 157]. 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:

$$p(\ln x) = \sum_{m=0}^2 \alpha_m p_m(\ln x) = \sum_{m=0}^2 \alpha_m \sum_{n=1}^N \lambda_n \mathcal{N}(\ln x | \theta_n) = \sum_{k=1}^K \lambda_k \mathcal{N}(\ln x | \theta_k) \quad (\text{A.4})$$

$$\sum_{k=1}^K \lambda_k = 1 \quad (\text{A.5})$$

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:

$$p(X) = \sum_{k=1}^K \lambda_k \mathcal{N}(X | \theta_k) \quad (\text{A.6})$$

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):

$$BIC(K) = \ln(N)q_K - 2\ln(\hat{L}_K) \quad (\text{A.7})$$

$$q_K = K - 1 + 2^K \quad (\text{A.8})$$

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:

$$p(k|X_i) = \frac{p(X_i|k)p(k)}{p(X_i)} = \frac{p(X_i|k)\lambda_k}{p(X_i)} \quad (\text{A.9})$$

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  $m$ :

$$f : \{0, 1, \dots K\} \rightarrow \{0, 1, 2\} \quad (\text{A.10})$$

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 [158], 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  $Y_i$ :

$$Y_i = \frac{1}{M_i} \sum_{j=0}^{M_i} X_j \quad (\text{A.11})$$

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,  $\psi(\delta)$ :

$$\psi(\delta) = \frac{<(X_i - \mu_X)(X_j - \mu_X)>_{i,j \in \delta}}{\sigma_X^2} \quad (\text{A.12})$$

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:

$$p(X, Y) = \sum_{k=1}^K \lambda_k \mathcal{N}(X, Y | \theta_k) \quad (\text{A.13})$$

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:

$$p(k|X_i, Y_i) = \frac{p(X_i, Y_i|k)\lambda_k}{p(X_i, Y_i)} \quad (\text{A.14})$$

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

$$p(k|X_i) = \sum_j p(k|X_i, Y_j) \quad (\text{A.15})$$

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

$$f(\operatorname{argmax}_k p(k|X_i)) \quad (\text{A.16})$$

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  $j$ :

$$w_{ij} = \exp\left(\frac{-E_{ij}}{\langle E \rangle}\right) \quad (\text{A.17})$$

$$E_{ij} = |X_i - X_j| \quad (\text{A.18})$$

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 [159]. 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 [160], initialized by  $p(k | X_i, Y_i^c)$ :

$$\hat{p}(k | X_i, Y_i^c) = \alpha \sum_j w_{ij} \hat{p}(k | X_i, Y_i^c) + \beta \quad (\text{A.19})$$

$$\beta = (1 - \alpha)p(k | X_i, Y_i^c) \quad (\text{A.20})$$

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:

$$\hat{p}(k | X, Y^c) = (I - \alpha W)^{-1} (1 - \alpha)p(k | X, Y^c) \quad (\text{A.21})$$

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  $\hat{p}(k \mid X_i, Y_i^c)$ :

$$f(\operatorname{argmax}_k \hat{p}(k \mid X_i, Y_i^c)) \quad (\text{A.22})$$

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

$$\hat{P}(m_i) = \sum_{\{k \mid f(k)=m_i\}} \hat{p}(k \mid X_i, Y_i^c) \quad (\text{A.23})$$

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.

### A.1.5. Measurement curation

Our framework includes a simple graphical user interface that permits manual curation of which regions of the image field are included in subsequent analyses. This tool allows for the exclusion of certain aspects of the segmented data on a context-dependent basis. For example, when studying *Drosophila* eye imaginal discs it is common to restrict analysis to a narrow column of cells in which the cellular processes of interest are taking place (Fig. B.2F). Alternatively, users may exclude regions of the image field marked by extensive tissue deformation or cell death. This functionality is convenient but is not a necessary component of a quantitative clonal analysis workflow.

### A.1.6. Cell growth simulations

We simulated the two dimensional growth of a cell culture seeded with a single cell. In these simulations, growth proceeds through sequential division of cells (Fig. B.6A). 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 are simultaneously subject to mitotic recombination (Fig. B.6C). Each cell contains either zero, one, or two copies of a gene encoding an RFP-tagged clonal marker. 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.

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. B.6D). Early recombination events generally entail larger clones, while shorter recombination periods limit the extent of clone formation (Fig. B.6E).

After each round of cell division, all cells are repositioned in order to preserve approximately uniform spatial density (Fig. B.6C). 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 clonal lineages remain spatially collocated after repositioning. Cell positions are then updated using a force-directed graph drawing algorithm [161]. Alternating cell division and repositioning steps are then repeated until a predefined population size is reached.

We used this platform to generate all synthetic data presented in the manuscript. 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. Except where stated otherwise, all simulations limited recombination to the first sixteen cell division events.

#### A.1.7. Generation of synthetic microscopy data

Cell growth simulations yield a list of spatial coordinates and gene dosages for each nucleus (Fig. B.6B). The corresponding fluorescence levels  $\{x_1, x_2, \dots, x_{i=N}\}$  were sampled from a lognormal distribution conditioned upon the corresponding gene dosage (Fig. B.7A-C):

$$\ln x \sim \mathcal{N}_n(\theta_n) \quad (\text{A.24})$$

where the subscript  $n$  denotes the gene copy number and  $\theta_n = (\mu_n, \sigma_\alpha^2)$  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:

$$\mu_n = \ln(2^{n-1}) \quad (\text{A.25})$$

We refer to  $\sigma_\alpha$  as the fluorescence ambiguity coefficient 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. B.7D,E), and consequently increases the difficulty of the annotation task (Fig. B.7F).

#### A.1.8. Synthetic benchmarking of automated annotation performance

We created a synthetic microscopy dataset by varying the timing of recombination events to generate cell cultures spanning a range of sixteen average clone sizes (Fig. B.6D, only half are shown). We performed 50 replicate simulations for each condition. The mixture model was independently trained and applied to each replicate. Training a single model on all replicates yields stronger performance on average (not shown), but also yields more variable results across the parameter space because all labels are dependent upon the outcome of a single expectation maximization routine.

We then labeled each cell and quantified annotation performance by evaluating the mean absolute difference between the predicted labels and their respective true values. This metric preserves ordinality, meaning egregious misclassifications are penalized more severely than mild ones [131]. The entire procedure was repeated for sixteen different fluorescence ambiguity coefficient values.

## A.2. Ratiometric sensing of two transcription factors regulates a transit to differentiation

### A.2.1. 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> [30] 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. C.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 [32]. *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 [149]. *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  $N^{ts1}/Y$ ; *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>* [150] *in trans* to the thermo-sensitive allele *egfr<sup>tsla</sup>* [69], as previously described [48]. The genotype was *w<sup>1118</sup>; egfr<sup>tsla</sup>, pnt-gfp/egfr<sup>f24</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 [71] as previously described [48]. *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 [34], *ey<sub>d</sub>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.2.2. 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 0.8μm 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 of the dorsal-ventral eye disc equator. All discs for a given condition were fixed, mounted, and imaged in parallel to reduce measurement error.

### A.2.3. Quantification of expression levels

Expression dynamics were inferred from confocal image stacks using an updated version of an existing segmentation and annotation pipeline [48]. 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 (Figs. 1.1, 1.2, C.1, C.2, C.3, and C.6) or DAPI (Figs. 1.4, 1.5, and C.8) 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.2.4. Conversion of distance to time

Cell positions along the anterior-posterior axis were mapped to developmental time as described previously [48, 50]. 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 [151]. 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.2.5. Computation of moving averages and confidence intervals

Moving averages were computed by first-order Savitzky-Golay filtration [152]. This method augments the simple windowing approach used in our previous study [48] 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 [48]. 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.2.6. 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:

$$z = \operatorname{argmax}_{dt} \hat{\gamma}_{X(t), Y(t)} \quad (\text{A.26})$$

$$\hat{\gamma}_{X(t), Y(t)}(dt) = E\left[\frac{(Y(t + dt) - \mu_Y)(X(t + dt) - \mu_X)}{\sigma_Y \sigma_X}\right] \quad (\text{A.27})$$

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.

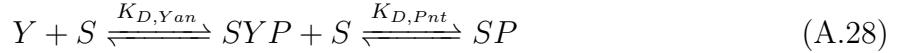
For each experimental treatment, a disc was randomly chosen and shifted in time such that time zero corresponds to the first annotated R8 neuron. This disc then served as the reference population for the alignment of all subsequent biological replicates within the treatment. Similarly, different experimental 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 [48]. 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.2.7. Simple competitive binding model

Figure C.4 presents results for an equilibrium model of 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:

$$Y_0 = Y + SYP_0 = P + SPS_0 = S + SY + SP \quad (\text{A.29})$$

yields an analytically tractable system of nonlinear equations [153]. For each pair of absolute protein abundances ( $Y_0, P_0$ ) in Figure C.4B, the Pnt binding site occupancy is simply  $SP/S_0$ .

#### A.2.8. Competitive binding model with cooperativity

The model presented in Figure 1.3 expands upon the work of Hope, Rebay, and Reinitz (2017). 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 were 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. C.5A). A unique configuration of binding states

for all  $n$  binding sites constitutes a single microstate,  $k$ . The thermodynamic potential of each microstate was taken to be the sum of thermodynamic potentials for each of its constituent binding sites. For each microstate, the stabilizing effect of polymerization was 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 C.5B. The statistical frequencies of each microstate were evaluated by constructing a canonical ensemble:

$$p_k = \frac{\exp\left(\frac{-\Delta G_k}{RT}\right)[P]^{a_P(k)}[Y]^{a_Y(k)}}{\sum_k \exp\left(\frac{-\Delta G_k}{RT}\right)[P]^{a_P(k)}[Y]^{a_Y(k)}} \quad (\text{A.30})$$

in which  $p_k$  is the statistical 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 statistical frequency of all microstates in which the site is occupied by a given transcription factor. Overall fractional occupancies are similarly evaluated across all sites within the element.

We consider regulatory elements comprised of 12 binding sites in which only the first site carries 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. 1.3D-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. 1.3A-C uses these same elevated binding affinities for Yan, while setting  $\gamma_Y = 0 kcal mol^{-1}$ . Qualitatively, our results are not sensitive to this parameterization.

#### A.2.9. Analysis of *yan* clones

We used *ey<sub>d</sub>FLP* and *FRT40A* to generate *yan*<sup>833</sup> null clones within 23 eye discs carrying the Pnt-GFP transgene (see Section ??). 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 methods 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 (See Sections 2.2 and 2.4). 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 presented in Chapter 2. We used the correction strategy presented in Section 2.3 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 (not shown). The same procedure was therefore applied to all measurements of *yan* null clones (Fig. 1.4H).

We used the selection tool described in Section A.1.5 to limit the analysis to cells taken from regions immediately posterior to the MF in each eye disc. We further limited the comparison to clonal genotypes that overlap in developmental time. These restrictions served to buffer against differences in developmental context and focus attention on the region of elevated Pnt-GFP expression. Using the approach described in Section A.1.3, 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.2.10. 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 (Fig. C.5A,B), 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 in Figures C.5A and C.5B.

#### A.2.11. 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,  $d$ :

$$C_{ij} = \frac{(X_i - E[X])(X_j - E[X])}{E[X^2] - E[X]^2} \quad (\text{A.31})$$

$$d_{ij} = |y_i - y_j| \quad (\text{A.32})$$

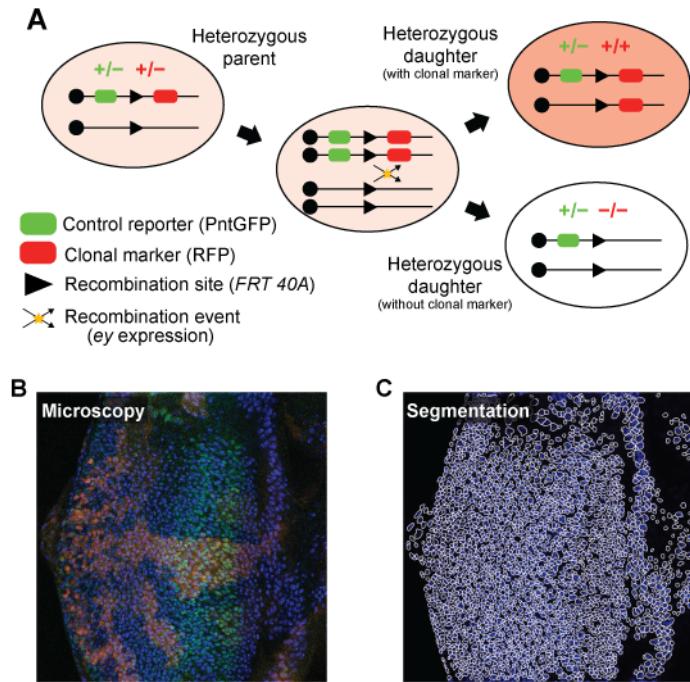
where  $E$  denotes the expected value, and  $i$  and  $j$  are indexed over all cells in order of increasing separation distance. Moving averages and confidence intervals were computed as described previously, with a window size of 50 sequential values.

Spatial signals were decomposed into spectral components via the Lomb-Scargle periodogram using the AstroML software package [154]. These periodograms were used rather

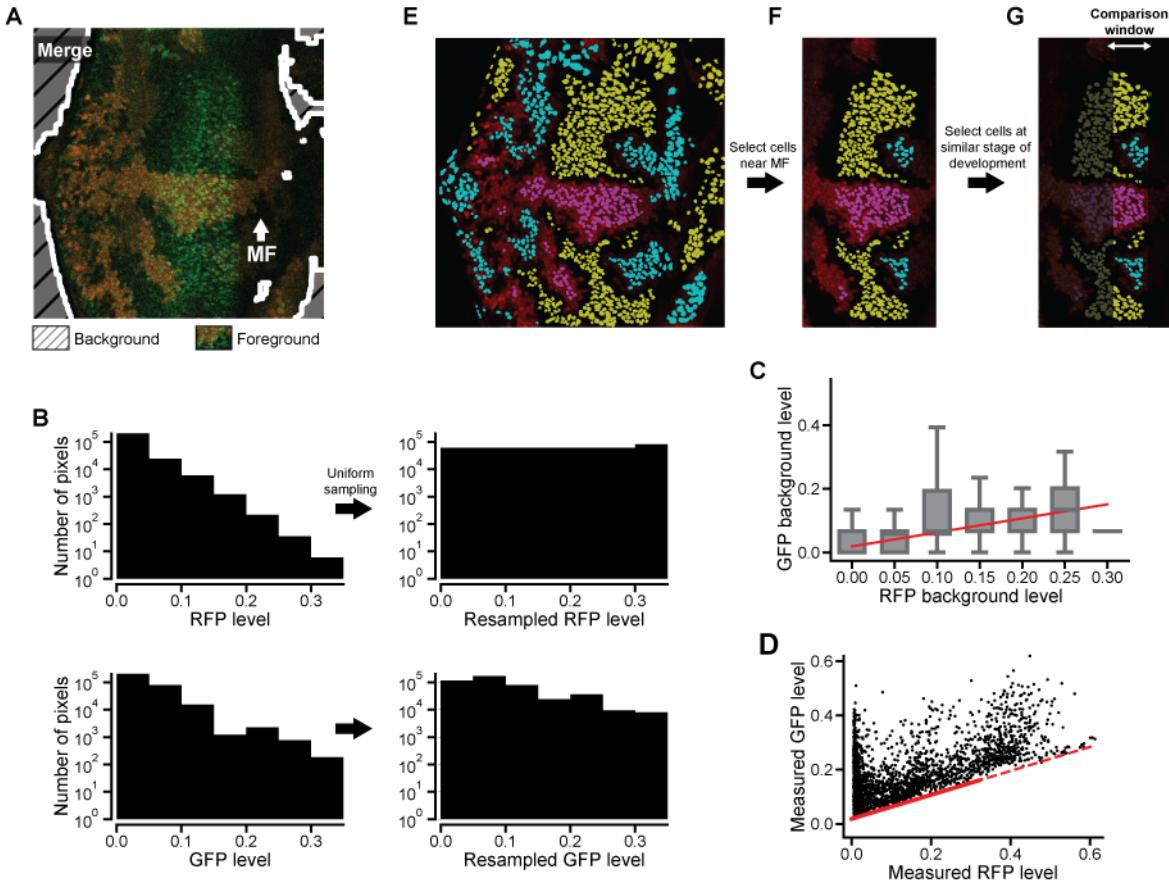
than Fourier decomposition because they enable spectral decomposition of irregularly sampled signals [155]. Significance thresholds were inferred from the 95th percentile of peak spectral powers detected during repeated decomposition of 1000 null signals. Null signals were constructed by resampling signal intensities while maintaining constant sampling times.

## APPENDIX B

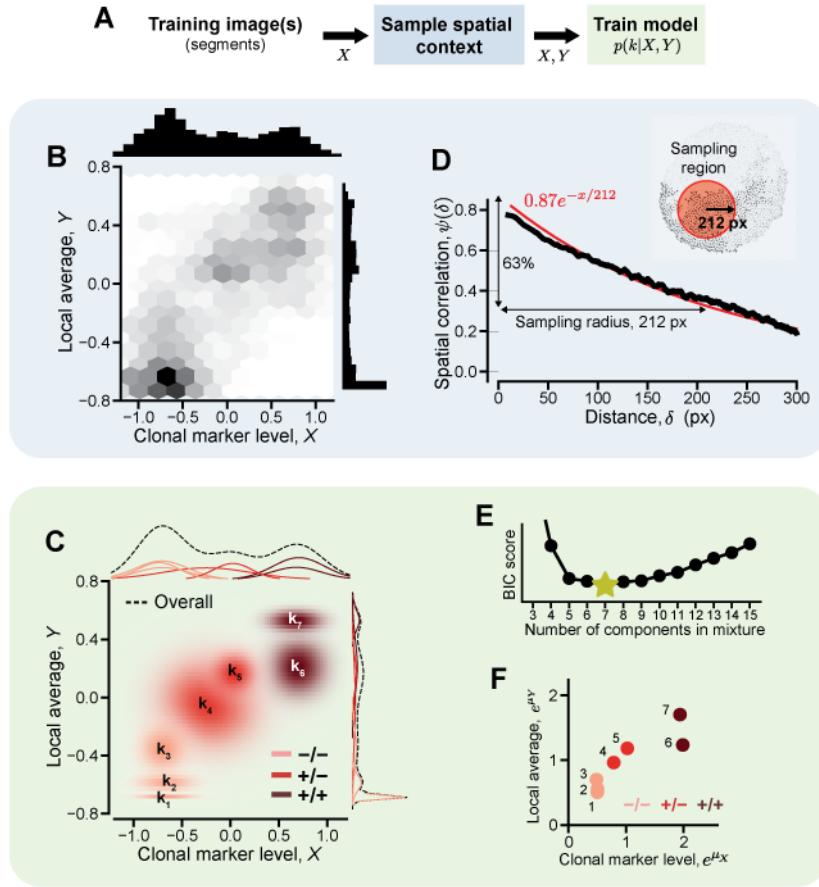
## Supporting figures for Chapter 2



**Figure B.1. 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.



**Figure B.2. Intermediate stages of bleedthrough correction.** (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.



**Figure B.3. 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 B.3C. (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.

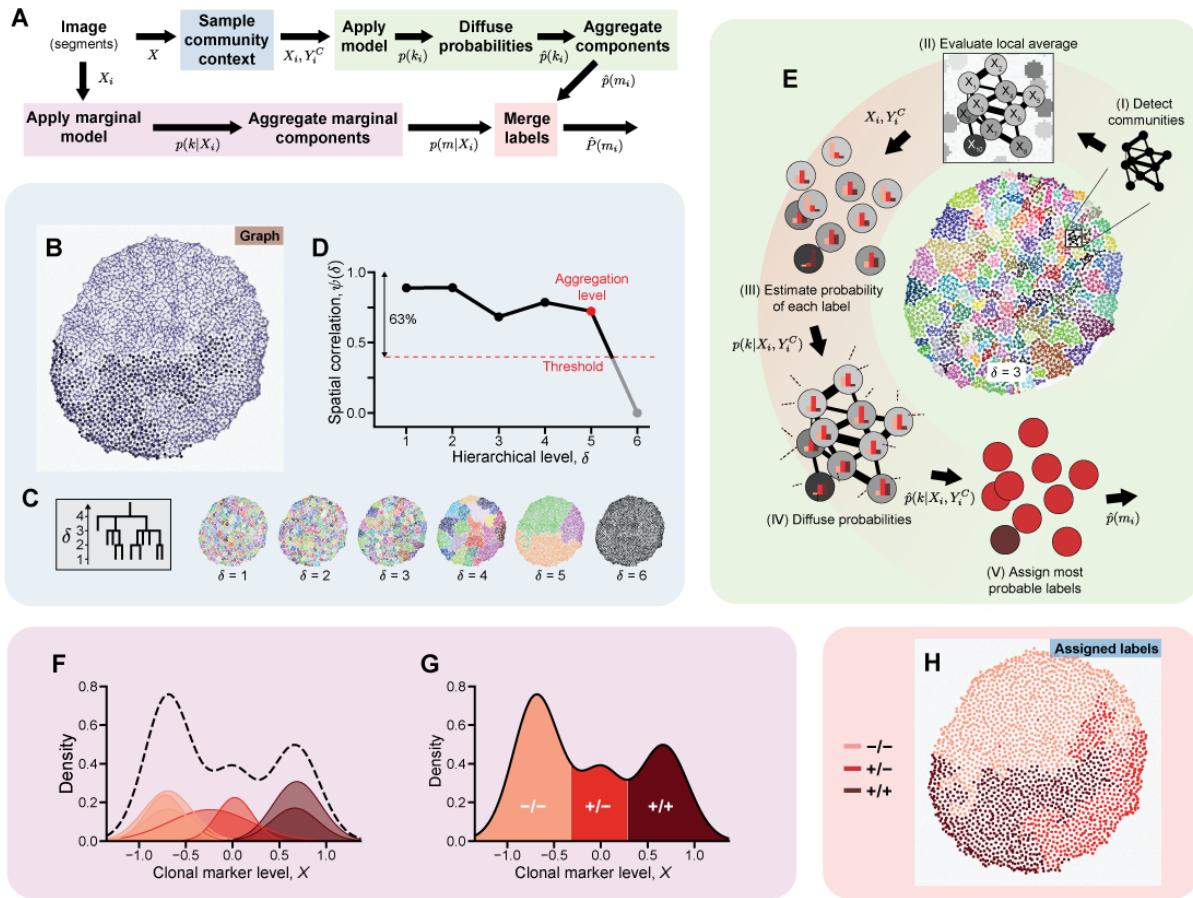


Figure B.4. Label assignment using a trained clone annotation model.

Figure B.4. (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 B.3C. (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.

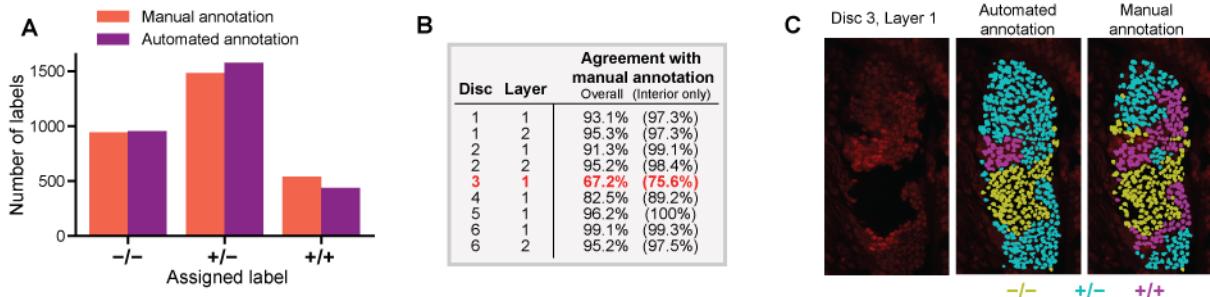
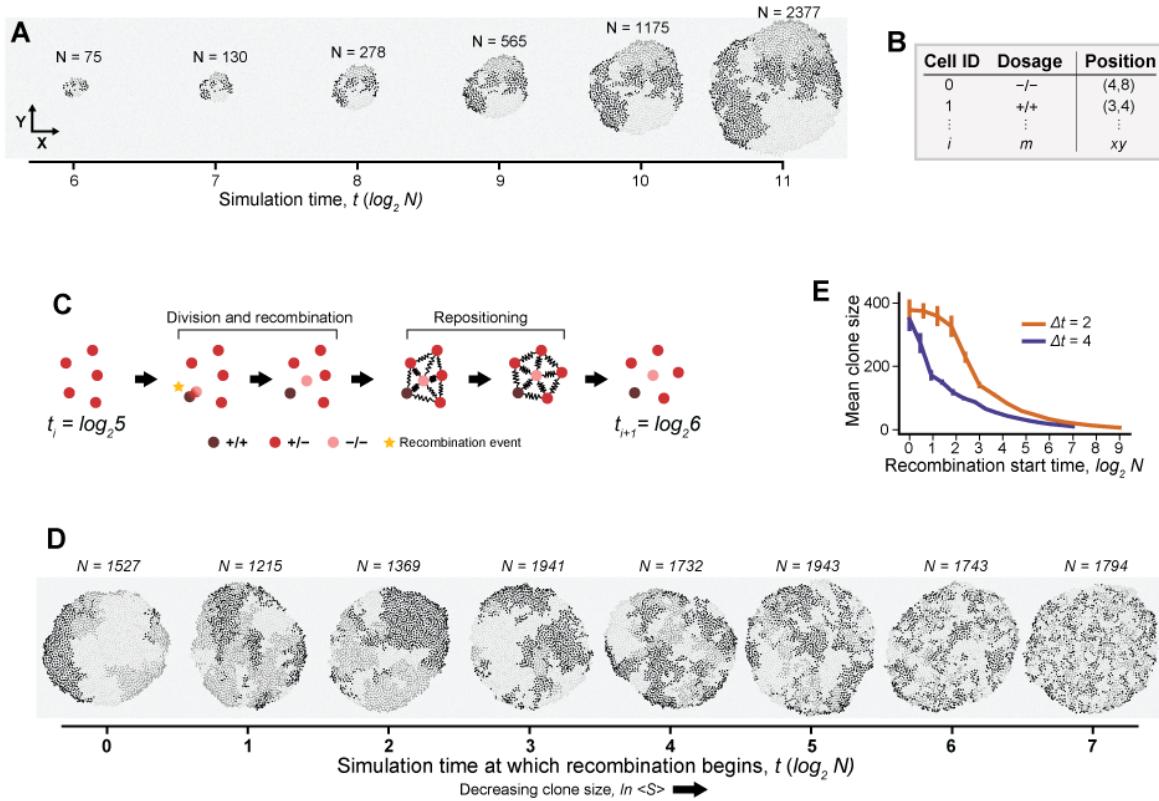
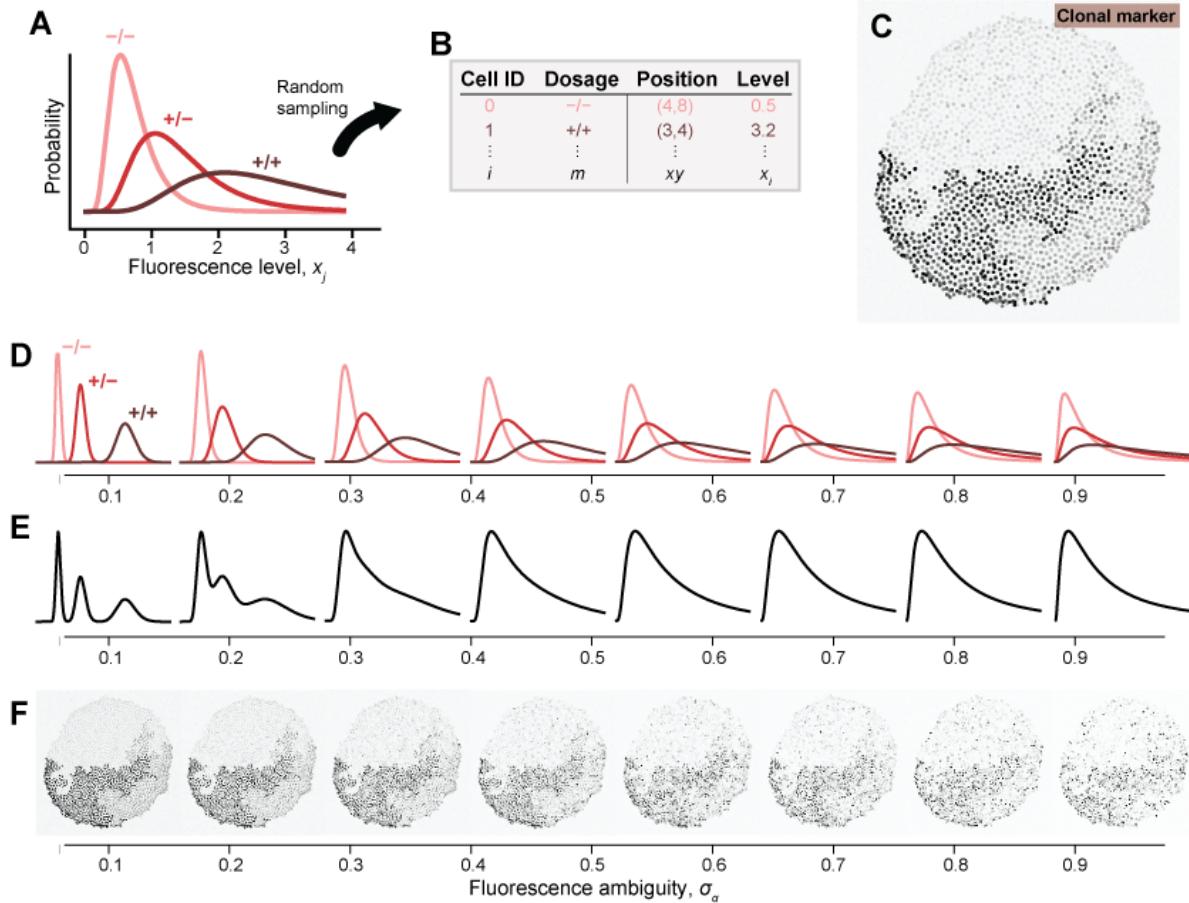


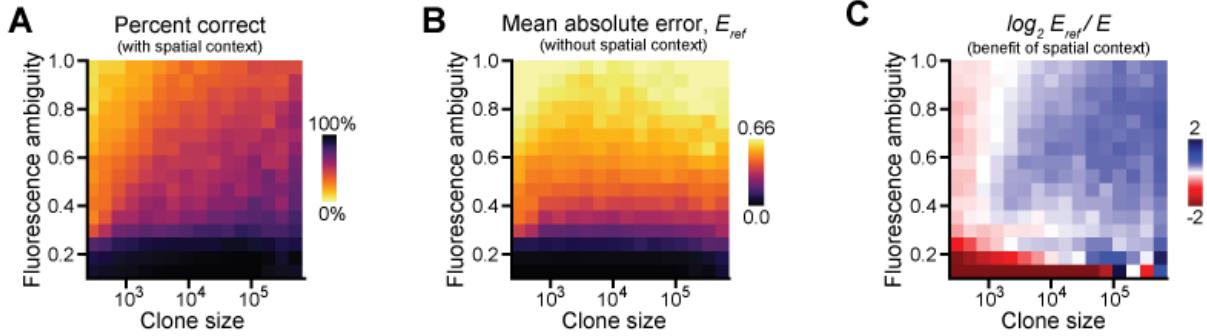
Figure B.5. **Comparison of automated annotation with manually assigned labels.** (A) Distribution of labels among each possible value. (B) Agreement between automated and manual annotation. Values in parentheses exclude cells on the periphery of each clone. The sole instance of low agreement is marked 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.



**Figure B.6. 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.



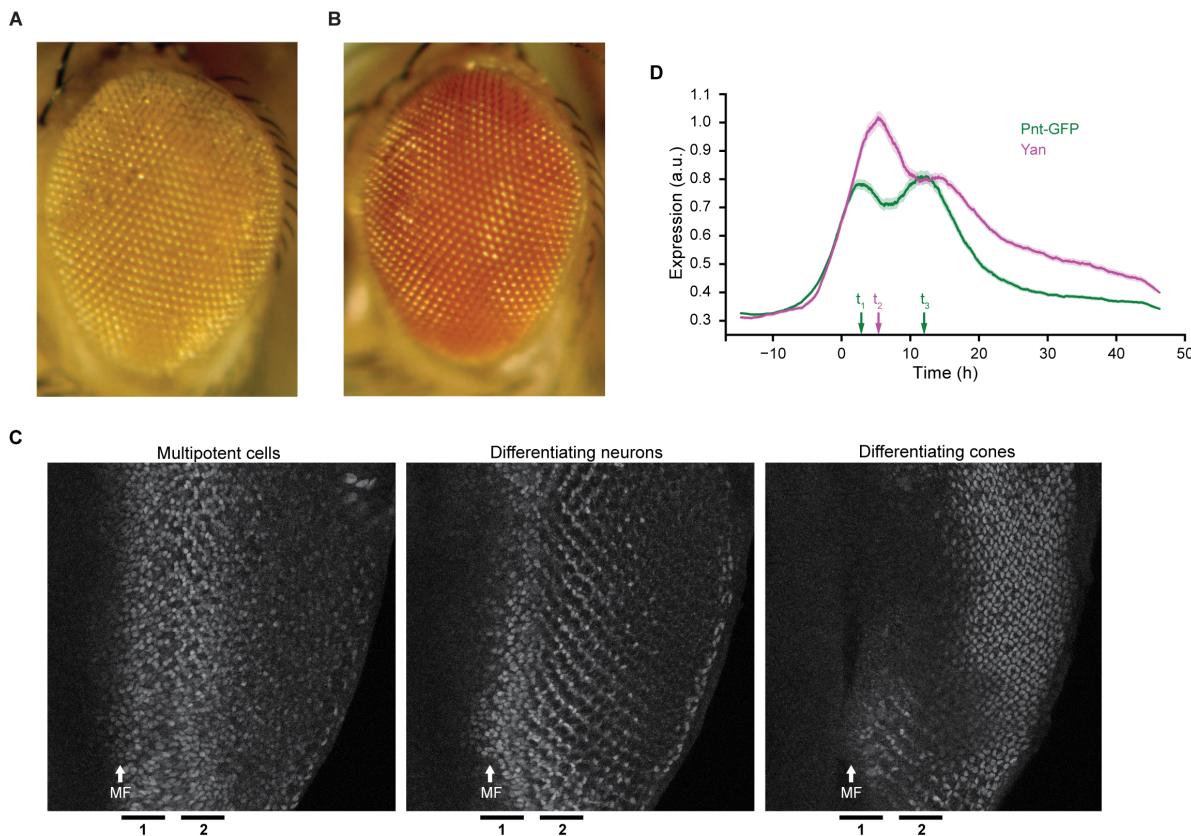
**Figure B.7. Tunable generation of synthetic microscopy data.** (A) Fluorescence levels are sampled from lognormal distributions conditioned upon gene dosage. (B) Synthetic data include a measured fluorescence level for each reporter in each cell. Text color reflects the generative distribution in A. (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.



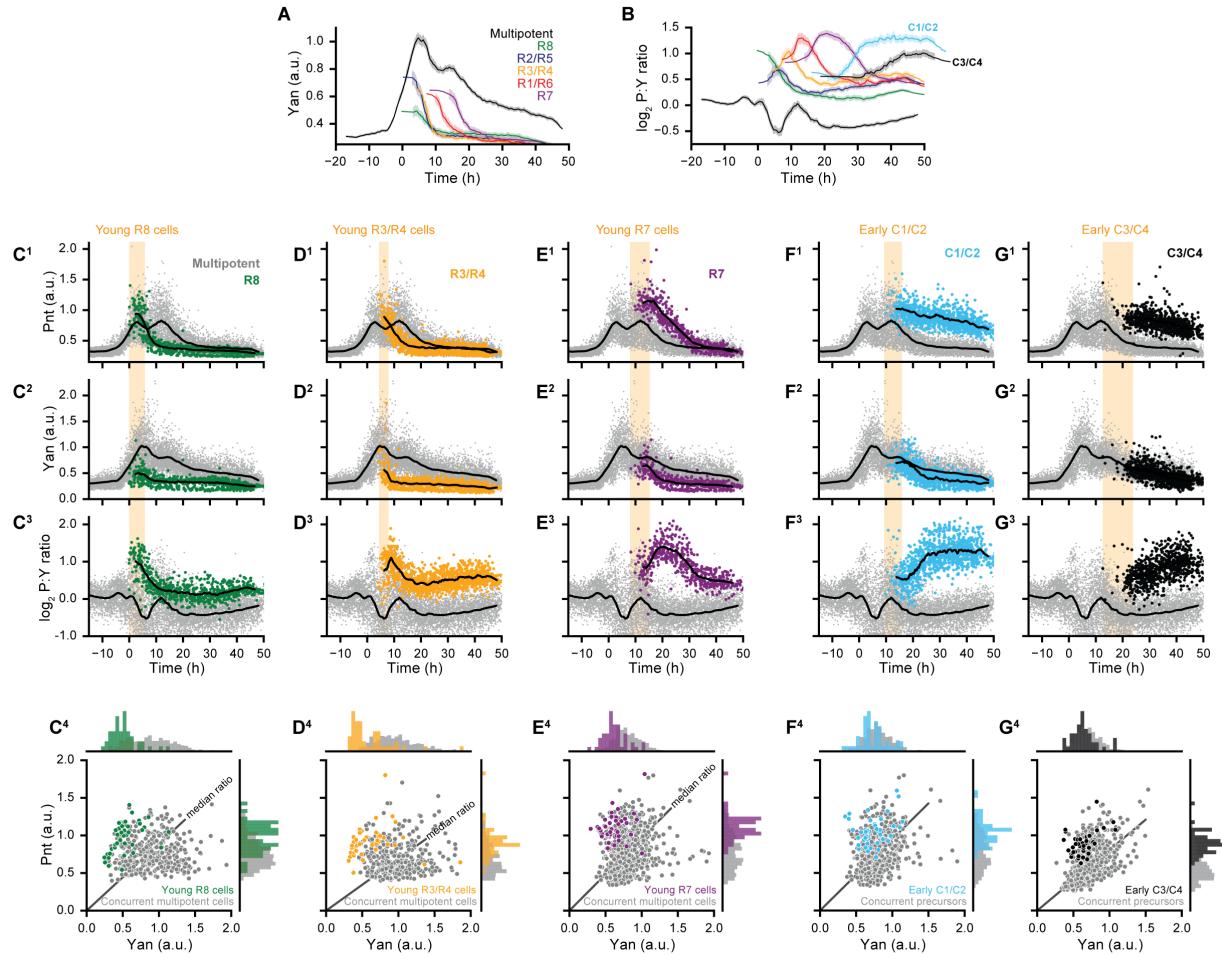
**Figure B.8. Additional results for synthetic benchmarking of annotation algorithm.** (A) Annotation accuracy as a function of fluorescence ambiguity and clone size. Accuracy is defined as the fraction of cells that were correctly labeled. Performance improves with increasing clone size and worsens with increasing fluorescence ambiguity. (B) Annotation performance of a marginal classifier that neglects spatial context. Each point in the grid reflects the mean absolute error (MAE) between assigned labels and the corresponding ground truth. Performance worsens with increasing fluorescence ambiguity but does not depend upon clone size. (C) Annotation performance relative to the marginal classifier. Color scale reflects the  $\log_2$  fold-change in MAE when spatial context is ignored. Blue indicates that spatial context improves performance. Spatial context is informative for larger clones, particularly when fluorescence levels are ambiguous.

## APPENDIX C

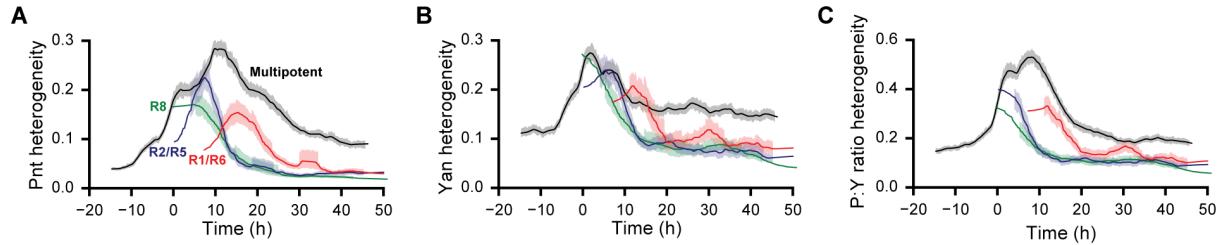
## Supporting figures for Chapter 1



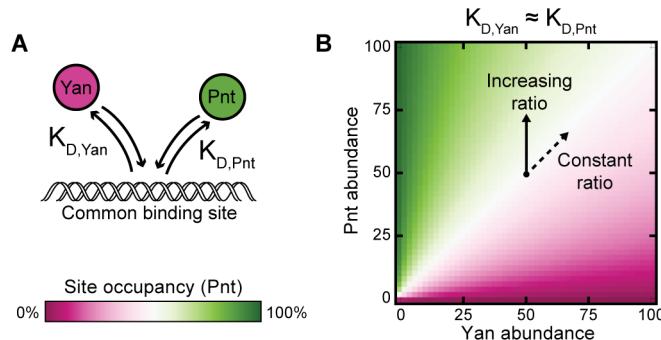
**Figure C.1. PntGFP expression during eye development.** (A,B) Adult eyes of flies carrying (A) one or (B) two copies of the recombined *pnt-gfp* transgene under a *pnt* null mutant background. Note the wildtype retina patterning under both rescue conditions. (C) Maximum intensity projections of Pnt-GFP 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.



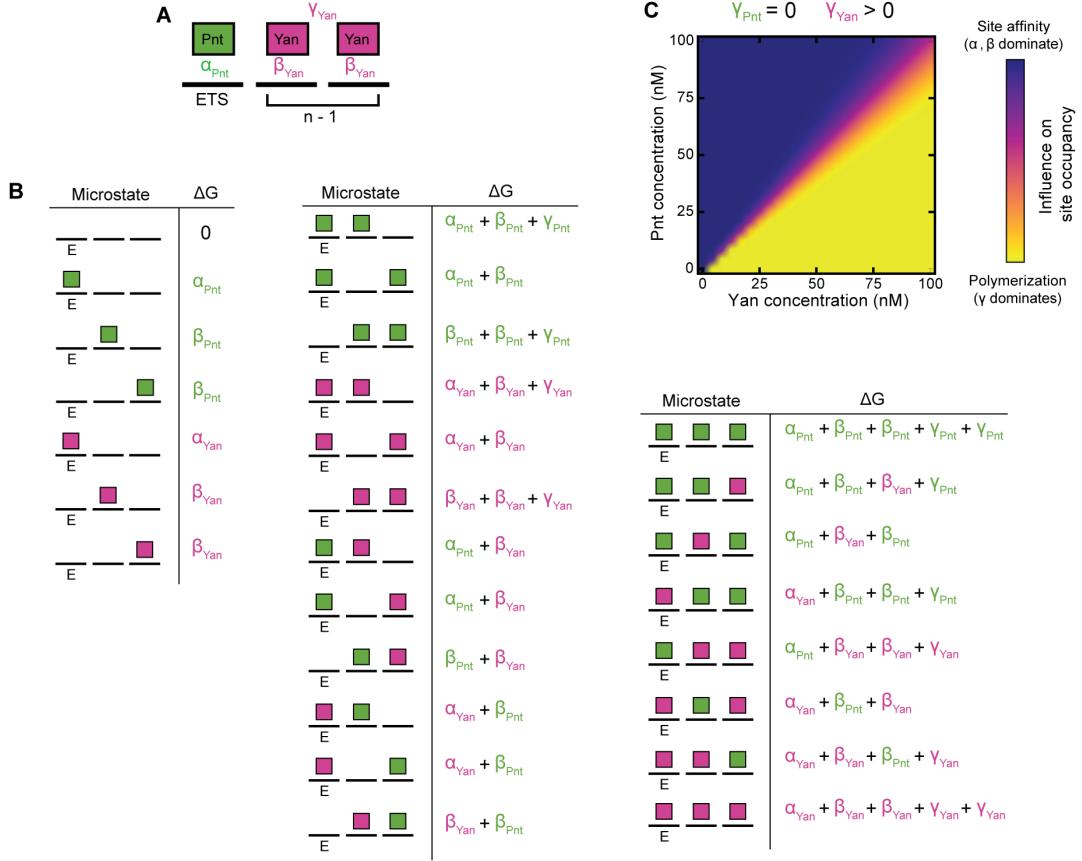
**Figure C.2. PntGFP and Yan expression dynamics in additional differentiated cell types.** (A,B) Measured expression dynamics for all annotated cell types. Solid lines are moving averages across 250 and 75 sequential cells for progenitors and differentiating cells, respectively. Shading denotes bootstrapped 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 C.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  $\log_2$ -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.



**Figure C.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.



**Figure C.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 [58]. (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.

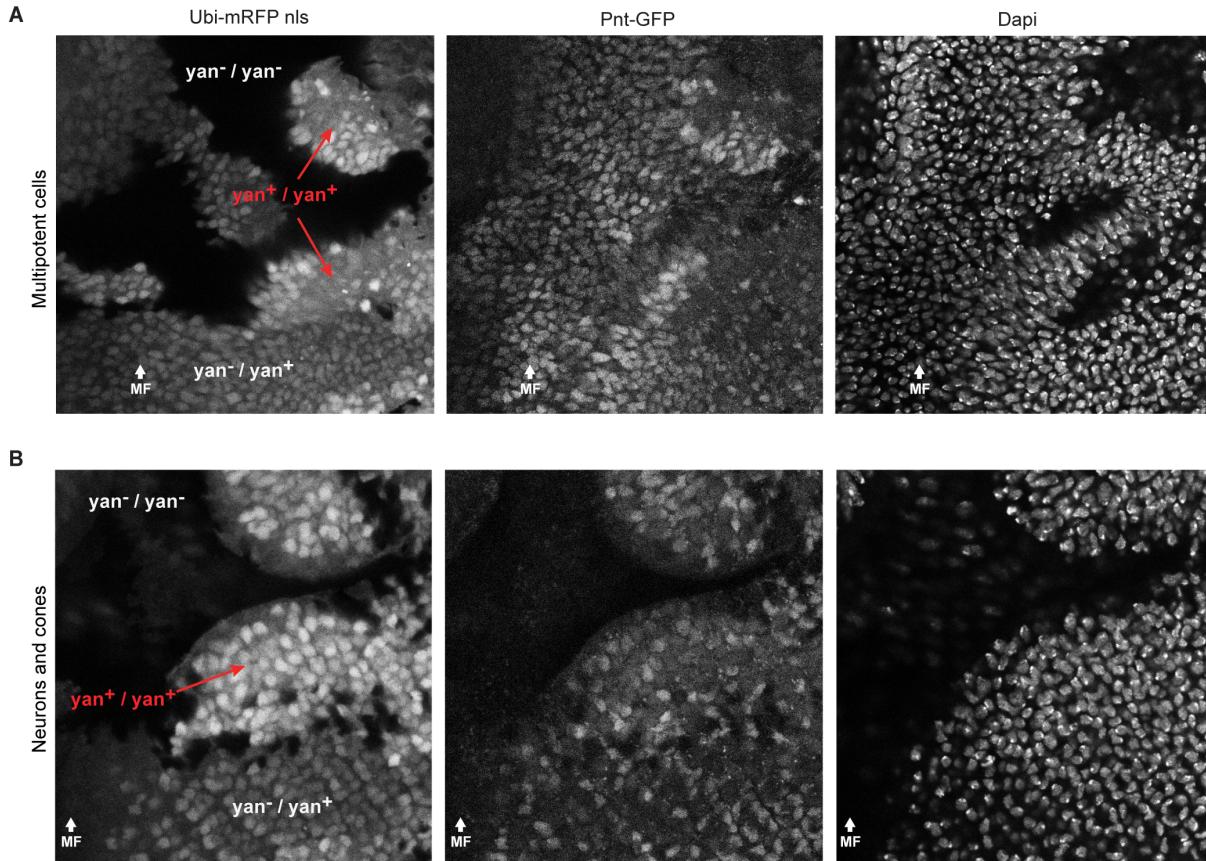
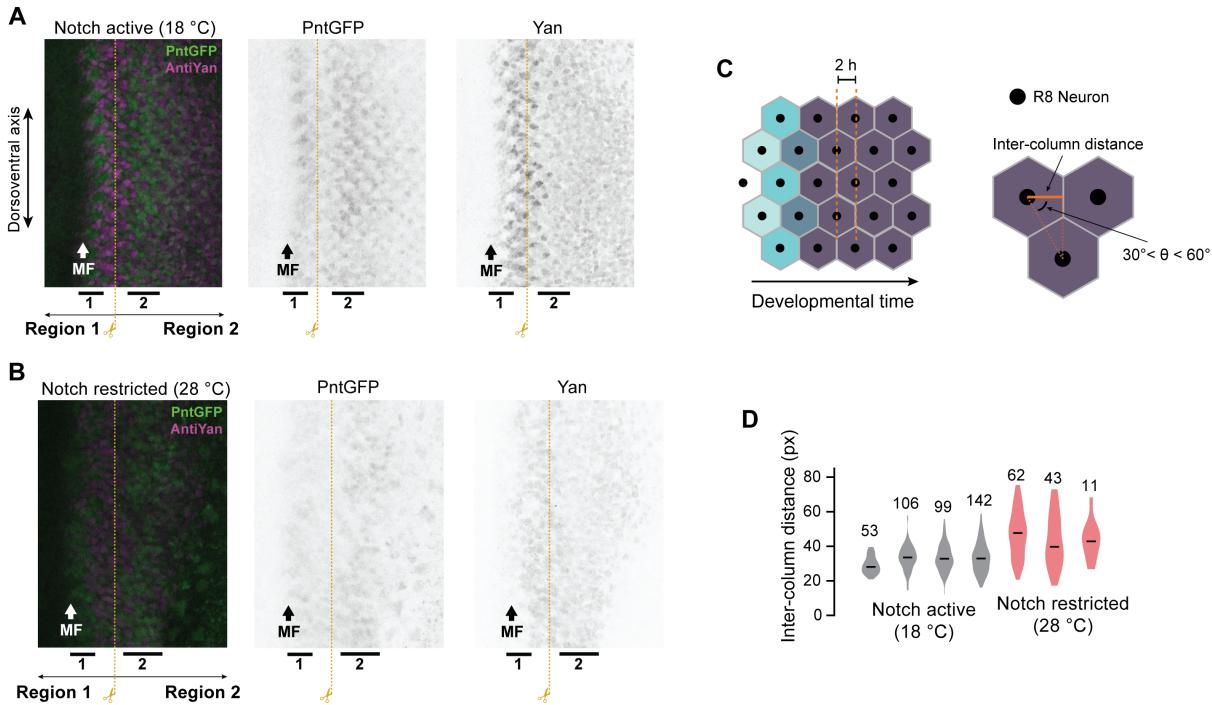
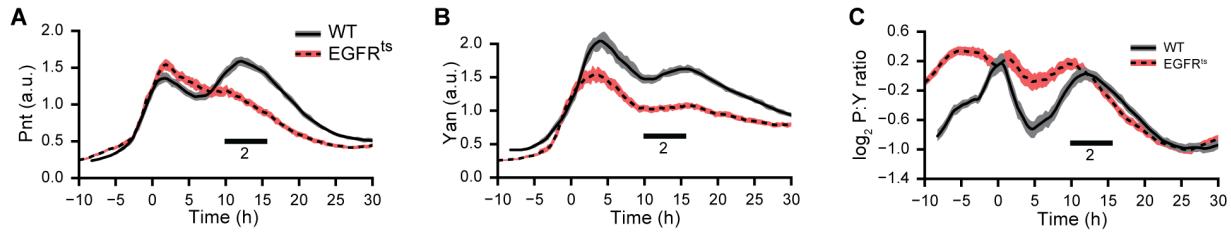


Figure C.6. ***yan* null clones in the eye.** (A, B) Confocal images of (A) progenitor cells and (B) photoreceptors and cone cells in *yan* null and heterozygote clones. Regions of Ubi-mRFPnls expression are manually labeled by *yan* genotype. White text indicates regions of reduced Yan abundance, red denotes wildtype. DAPI visualizes all nuclei.



**Figure C.7. 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 1.5A and 1.5B. (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.



**Figure C.8. 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.

## APPENDIX D

### Resources

#### D.1. New tools for quantitative biologists

Several computational tools were developed in support of the work presented in this thesis. These resources are available online under open license for unrestricted use and future development. By contributing them to the open-source software ecosystem, we aim to help foster the adoption of novel quantitative and computational analysis strategies among the broader community of developmental and *in vivo* cell biologists.

The following sections describe each of these tools and their high level functions. Unless otherwise stated, all tools are freely available in code repositories hosted by GitHub and mirrored between both my personal account and the Amaral and Bagheri lab accounts. These repositories generally contain high level API documentation in addition to a series of Jupyter notebooks that walk the user through a series of usage examples.

##### D.1.1. FlyEye Clones: <https://github.com/sebastianbernasek/clones>

**FlyEye Clones** is a framework for automated quantitative mosaic analysis of *Drosophila* eye imaginal discs. Its many features are detailed throughout Chapter 2. The current implementation is a standalone Python package, but we also intend to incorporate its core features into future versions of *FlyEye Silhouette*, our open-source platform for quantitative analysis of the larval eye that will soon be freely available on the Mac App Store.

#### D.1.2. FlyEye Analysis: <https://github.com/sebastianbernasek/flyeye>

**FlyEye Analysis** is a python-based framework for quantitative analysis of protein expression dynamics in *Drosophila* eye imaginal discs. The framework provides a suite of methods to analyze and visualize measurements obtained using both *FlyEye Clones* and the *FlyEye Silhouette* platform for macOS. The current implementation was used to perform all of the analysis presented in Chapter 1. The core functionality of the framework enables users to infer the approximate developmental age of each measured nucleus, query the measurements by both developmental age and cell type, and visualize the resultant data. It also provides a number of analysis features that allow users to quantify heterogeneity and spatial patterns of protein expression in the developing eye.

#### D.1.3. Binding: <https://github.com/sebastianbernasek/binding>

**Binding** is a cython-based framework for simulating the equilibrium occupancy of DNA binding sites by one or more polymerizing transcription factors. The package provides a high-level Python interface to a C backend that efficiently enumerates all possible microstates. Microstates are enumerated in a recursive fashion, enabling large-scale parallelization of the primary computational bottleneck. When simulating systems comprised of more than one binding species, this implementation confers a substantial performance advantage over the sequential bit-wise implementation proposed by the authors of the original study that inspired the model [58].

#### D.1.4. SyClones: <https://github.com/sebastianbernasek/syclones>

**SyClones** is a python-based framework for generating synthetic microscopy data that mimics key features of mosaic eye imaginal discs. The synthetic data provide a reliable standard

that may be used to objectively compare and benchmark the performance of mosaic analysis platforms.

#### D.1.5. GeneSSA: <https://github.com/sebastianbernasek/genessa>

**GeneSSA** is our cython-based framework for stochastic simulation of gene regulatory network dynamics. It uses the stochastic simulation algorithm to generate exact solutions to the corresponding chemical master equation [? ]. Simulations are executed using a C backend optimized for performance on networks whose reaction propensity functions fall within a narrow scope of pre-defined options (e.g. mass action or Hill kinetic forms). This narrow scope is by design; the framework prioritizes computational efficiency at the expense of flexibility by explicitly hard coding a handful of functional forms for the reaction propensities. This design places GeneSSA among the most performant implementations of the exact stochastic simulation algorithm for the range of systems that it encompasses. The framework may be (and has been! Please see example notebooks) extended to include additional kinetic formulations on an as-needed basis. However, doing so requires a firm command of the cython language.

## D.2. Reproducibility and Supporting Data

The work presented in this thesis is supported by large volumes of both experimental and simulated data. All of these datasets have been made freely available online via the Northwestern Arch Data Repository.

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. My personal account is accessible via: <https://github.com/sebastianbernasek/>.

Table D.1. Data Repositories

Topic	DOI
Chapter 1	<a href="https://doi.org/10.21985/N24Q81">https://doi.org/10.21985/N24Q81</a>
Chapter 2	<a href="https://doi.org/10.21985/N24Q81">https://doi.org/10.21985/N24Q81</a>
Chapter ??	<a href="https://doi.org/10.21985/N2J464">https://doi.org/10.21985/N2J464</a>

Table ?? lists the relevant repository for each chapter of this thesis, along with the software required to execute the corresponding code (see Appendix D.1). 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. Code Repositories

Topic	Repository	Required software
Chapter 1	/pnt_yan_ratio	FlyEye Analysis, Binding
Chapter 2	/clones	FlyEye Clones
Chapter ??	/gram	GeneSSA, FlyEye Analysis