The natural world presents a stunning cornucopia of varied life forms. We tend to distinguish them by the set of physiological traits that confer a stereotyped identity to each species. For instance, penguins have black and white fur, a hard beak, two eyes, and an impressive ability to wobble around on ice. These traits are a culmination of the many complex cellular processes we collectively refer to as \emph{development}.

Development is remarkably robust, as evidenced by its tendency to yield a consistent end product amidst the varied conditions encountered in natural environments. These processes are so reliable that we often take them for granted. It is difficult, for instance, to imagine a scenario in which we might mistake another human for a penguin. However, developmental processes can and do exhibit minor variation with fatal consequences for human health. They have therefore long been a central focus of biological research. Researchers continue to characterize these complex processes in the hope that they might one day be able to control their behavior; either to rectify the diseases that arise when they fail, or exploit them in novel biotechnologies.

\section{The molecular origin of organismal phenotypes}

Biologists have long understood that development begins at the cellular scale. Early experiments showed that all multicellular organisms originate from a single cell, with subsequent growth and division events driving progression toward adulthood. Cells acquire increasingly specific roles and functions as growth proceeds, ultimately culminating in stereotyped adult morphologies. This process, often referred to as lineage restriction, demands that each cell decides to adopt the correct fate at the appropriate time and place.

Early experiments in a variety of model organisms demonstrated that cell fate decisions are executed via the activities of biomolecules, most notably proteins. They also established that the spatiotemporal expression of these proteins is encoded in the genome. Genetics were therefore believed to provide a predefined roadmap for the journey from embryo to adulthood, leading researchers to dedicate much of the past century to elucidating the role of individual genes in coordinating adult phenotypes.

Most of their efforts embrace a common philosophy; break something and see what happens. Nowhere is the legacy of this approach more pronounced than in the fruit fly, *Drosophila melanogaster*, whose genes are predominantly named after the phenotypes that emerge in their absence. Removing \textit{eyeless}, \textit{wingless}, or most disturbingly, \textit{hairy}, may now seem rudimentary, but these genetic perturbations were vital to discovering genetic components and architectures common to all organisms.

Gene-specific reporters provided a localized readout of transcript and protein abundance that enabled researchers to break something and see what happens to gene expression. They discovered that some genes and their products confer pleiotropic functions across many aspects of development, while others only act within a narrow context. They also found that networks of multiple genes may act in tandem to coordinate subsequent developmental events. Furthermore, many genes are only essential for proper development when others are absent, thus rendering them redundant under certain conditions. Perhaps to their initial dismay, they also discovered that cells are remarkably heterogeneous, with expression levels varying considerably within the same developmental context. These and many other discoveries form the foundation of our contemporary systems-level view of development, in which complex networks of regulatory interactions guide the stochastic behavior of cells toward desired outcomes.

\section{Quantifying gene expression during development}

Experimental perturbations continue to supply the most potent tools in the arsenal available to researchers today. However, it has become clear that qualitative descriptions of phenotypes and gene expression do not provide adequate insight into the complex spatiotemporal interactions that govern cell fate decisions. Moreover, any attempts to engineer new decision mechanisms or rectify faulty ones, such as in cancer treatment, demand predictive models driven by quantitative data.

The focus has therefore gradually shifted from \textit{seeing} what happens to \textit{measuring} what happens. The transition has been made possible by simultaneous advances in the resolution with which we can quantify the abundance of gene products during development. The state of the art has progressed from aggregate measurements of transcript and protein content across entire tissues, to \emph{in vivo} measurements of individual cells, to \emph{in situ} quantification of single-cell expression dynamics. These advances enable the quantitative characterization of gene regulatory networks (GRNs) necessary to support the development of useful mathematical models, which may then be used to generate testable predictions for further experimental validation.

Unfortunately, quantitative measurement and analysis techniques have not been universally adopted by the research community, leading to the sustained prevalence of subjective qualitative analysis in the literature. One plausible explanation is that quantitative analysis frequently demands computational proficiency that falls beyond the scope of many experimental labs. Interdisciplinary collaborations should help alleviate this challenge, but they do not provide a permanent solution. However, the advent of open-source automated analysis software has revolutionized several other subdisciplines of biology by lowering the barrier to adoption of high-throughput measurement strategies. For example, RNA-seq would not be a standard approach in transcriptomics without the supporting software for aligning sequence reads. Similar tools are therefore needed to support quantitative measurements of gene expression during development, lowering the barrier to adoption of data-driven modeling of cell fate decisions.

\section{Quantitative models of cell fate decisions}

Quantitative measurements and mathematical modeling have been combined to elucidate several novel cell fate decisions mechanisms in recent years.

These models

Proteins called transcription factors coordinate the timing and execution of cell fate 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.

Development is an inherently dynamic process, as each stage of gene expression is responsible for triggering subsequent events in the developmental program. Unsurprisingly, many genes involved in the coordination of cell fate decisions have been shown to be transiently expressed within a given context.

round of cell fate decisions must trigger subsequent events in the developmental program.

and it becoming increasingly obvious that gene expression dynamics play

Mathematical models have also revealed that

\section{Stochasticity in cell fate decisions }

\section{Cell fate decisions as a dynamic process}

Raj et al

Oodenarden

Anderson

Yi et al

\section{Photoreceptor specification the \emph{Drosophila} eye imaginal disc}

History of eye dev

Yan/Pnt papers

Graham + bistability

Shilo et al

Pelaez

\section{Evolutionary drivers of robust cell fate decisions}

Developmental processes show a remarkable robustness to variation that has also been a focal point of prior research. These studies are often traced back to C.H. Waddington’s claim that “there is scarcely a mutant that is comparable in constancy with the wild type” \cite{Waddington1942}. His comment is supported by the consistency of developmental outcomes amidst tolerable levels of genetic and environmental variation, as well as an observable increase in variability when severe perturbations are introduced \cite{Bateman1959,Rendel1959,Rendel1966,Scharloo1991}. Robustness has since been attributed to the interactions of genes and their products \cite{Dun1958,Hogness1996,Rutherford1998}, as well as the systems-level architectures that they comprise \cite{Rutherford1998,Paulsen2011,Li2009,Eldar2002,Denby2012,Cassidy2013,Cassidy2016}. A handful of conserved regulatory motifs, and the strategies they implement, are now understood to pervade most developmental processes \cite{Freeman2000,Hartman2001,Alon2007,Marciano2014}. Robustness has thus come to be accepted as a fundamental organizational principle underlying the evolution of all biological systems \cite{Kitano2004}. That is, selection for genotypes that confer robustness is widely assumed to shape the evolution of gene regulatory network topologies.

The evolutionary origins of robustness remain an ongoing topic of debate \cite{Siegal2014}. Citing phenomenological examples, Waddington attributed robustness to evolutionary selection for optimal traits. More recently, Siegal et al. argued that developmental robustness can arise without stabilizing selection for any particular phenotype \cite{Siegal2002}. They used a gene interaction network model to demonstrate that robustness is an emergent property of complex networks subject only to selection on the basis of their own functional stability. The same authors used a similar in silico evolution approach to show that most genes in developmental networks buffer phenotypic variation, suggesting that robustness is an emergent byproduct of network architecture \cite{Bergman2003}. More broadly, additional computational studies have questioned the notion that global properties of biological networks are a consequence of adaptive evolution \cite{Lynch2007,Wagner2003}. These results suggest that local properties of genetic circuits can drive the emergence of macroscopic features, such as those that ensure robust development.

Redundancy is one of many strategies that confer robustness upon developmental processes \cite{Hartman2001,McAdams1999}. This may seem intuitive; when a gene or its products are absent or fail to perform their functions, the simplest contingency is to have others ready to compensate. Nowak et al. showed that genetic redundancy is particularly evolutionarily stable in developmental systems \cite{Nowak1997}. Using a simple binary outcome simulation procedure, the authors demonstrated that increasing the probability with which each gene in a functionally overlapping pair fails to carry out its function places increasing selection pressure upon both redundant genes. They reasoned that the resultant increase in evolutionary stability explains the prevalence of redundancy in developmental systems, where failure rates are high and erroneous cellular decisions propagate as deleterious phenotypes. Their suggestion is consistent with the extensive functional redundancy documented in developmental systems throughout the literature \cite{Kitano2004}. Redundancy may arise through gene duplication or convergent evolution. A. Wagner argued against the former by associating gene sequence and expression data in yeast, showing that the suppression of phenotypic severity in response to loss-of-function mutations does not correlate with the presence of functionally similar genes elsewhere in the genome \cite{Wagner2000}. He concluded that functional redundancy is not the primary cause of robustness to genetic variation, instead favoring epistatic interactions between otherwise unrelated genes. Combined, these results hint that the prevalence of functional redundancy in biological systems may not solely be the product of direct stabilizing selection for robust cell fate decisions.

\section{Summary of the current work}

The research presented in this dissertation falls under the umbrella of quantitative biology. It combines chemical engineering, computer vision, and statistics to quantify complex biological processes that are notoriously difficult to measure, all in an effort to decipher how cells make reliable fate decisions during development.

Chapter \ref{ch:clones} introduces a computational framework for automated analysis of genetic mosaics; a class of experiments designed to probe gene expression \textit{in vivo}. The framework combines computer vision and statistics to measure protein levels in individual cells and predict their respective genotypes. It thereby facilitates systematic quantitative comparison of expression levels between cells subject to control and perturbation conditions. The accompanying software eliminates each of the labor-intensive steps of a quantitative workflow, and will hopefully lower the barrier to adoption of quantitative analysis strategies.

Chapter \ref{ch:ratio} explores a novel cell fate decision mechanism underlying photoreceptor specification in the developing fruit fly eye. Computer vision and statistical modeling techniques are used to extract quantitative measurements of transcription factor dynamics from a wealth of confocal microscopy data. These data show 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. A general model based on the statistical physics of transcription factor DNA binding shows that this phenomenon is a natural consequence of competition between the two transcription factors for common binding sites. The findings add a new dimension to our understanding of how transcription factors coordinate cell fate decisions, and showcase the importance of both quantitative and dynamic measurements for characterizing developmental systems.

Chapter \ref{ch:metabolism} proposes a novel theory to explain why the regulatory networks that coordinate cell fate decisions often contain several repressors tasked with attenuating expression of a single target gene. The theory posits that auxiliary negative regulators enable development to proceed more quickly by mitigating erroneous cell fate decisions when cells are rapidly metabolizing. It is supported by a robust collection of qualitative experiments showing that a broad variety of repressor loss-of-function phenotypes are reversed when biosynthesis rates are artificially slowed. A quantitative modeling framework is used survey a hypothesized mechanistic origin of this effect. Namely, that auxiliary repressors help avert erroneous decisions by expanding cells capacity to buffer excess protein expression. Quantitative measurements of transcription factor activity serve to validate the predictions made by the modeling framework. As shorter developmental times confer a selective advantage upon organisms, these findings may represent a novel evolutionary driving force for increased robustness of cell fate decisions.

Beyond their insights into the mechanics of cell fate decisions, these efforts have spawned several computational tools that may prove valuable to the broader community (see \ref{appendix:software}). All such resources have been made freely available, and their continued development should help contribute toward the sustained prevalence of quantitative analysis of cell fate decisions.

My goal is to predict how removal of a negative feedback mechanism causes protein levels within affected cells to differ from their unperturbed counterparts. As nuanced changes in expression yield abnormal morphologies *in vivo*, differences in protein levels manifest as differences in phenotype penetrance(Raj et al., 2010). I aim

Yan is an ETS-domain transcriptional repressor12. A pulse of Yan expression governs recruitment of photoreceptor neurons from a disordered collection of progenitors in the developing *Drosophila* eye13,14. Yan is regulated at the gene, transcript, and protein levels12. The microRNA miR-7 is one such inhibitor of Yan translation15. Yan represses differentiation by competing with the P2 isoform of a concurrently expressed transcriptional activator, *pointed* (Pnt), for occupancy of downstream binding sites13,16. Prolonged Yan expression disrupts timely recruitment of neurons and leads to a roughened eye morphology17. The Carthew Lab has shown that the requirement of miR-7 for successful eye development is relaxed by limiting biosynthesis rates (see Figures 1E, 1F). My hypothesis suggests that this phenomenon occurs because the impact of miR-7 activity on Yan expression and neuronal differentiation diminishes upon restriction of biosynthesis. I will use a kinetic model of the Yan network to investigate these effects.