Single-Cell Analysis of Organ Development and Regeneration in *Drosophila*

By

Nicholas J. Everetts

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Computational Biology

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Iswar Hariharan, Co-Chair
Professor Nir Yosef, Co-Chair
Professor Sandrine Dudoit
Professor Craig Miller

Fall 2021

Abstract

Single-Cell Analysis of Organ Development and Regeneration in *Drosophila*

by

Nicholas J. Everetts

Doctor of Philosophy in Computational Biology

University of California, Berkeley

Professors Iswar Hariharan & Nir Yosef, Co-Chairs

Organ development is a process that operates at both local and global cellular levels. Locally, individual cells undergo proliferation and are responsible for all biomass production. Furthermore, the function of a cell over the course of development and in the adult animal is largely dictated by the state of its transcriptome and proteome. However, globally, all cells must collectively work together to build an organ of proper size, shape, and function. To this end, distinct types of cells must be present in the correct number and location within the organ. Furthermore, organ development must also be robust, capable of withstanding potential setbacks such as tissue damage and the emergence of aberrant cells. Linking these local and global scales of organ development likely intertwines signaling pathways with emergent properties of the developing organ such as physical forces, cellular heterogeneity, and local heterotypic signaling interactions.

The common fruit fly *Drosophila melanogaster* is a powerful workhorse in the field of developmental biology, and boasts a massive genetic toolkit for *in vivo* experimentation. Specifically for organ development, the *Drosophila* wing-imaginal disc present an ideal system for studying the genetic regulation and heterotypic interactions that govern organ growth and patterning. The wing disc is embryonically derived from approximately 30 cells, and by late larval development, forms a simple sac-like structure consisting of two epithelial layers called the disc proper and peripodial epithelium. Significant literature exists surrounding the disc proper, which develops into the adult dorsal thorax and wing blade. Underneath the proximal-most region of the disc proper are the adult flight muscle precursors (AMPs), the myoblasts that give rise to nearly all adult flight muscles within the dorsal thorax. Despite the variety of morphologically- and physiologically-distinct thoracic muscle fibers created by the AMPs, there exists little knowledge on how the AMPs achieve such diversity. Experiments performed in the 1980s and 1990s in which wing-disc AMPs were transplanted into ectopic locations revealed that these cells readily fused with myoblasts in their new environment, suggesting that the AMPs are a largely naïve population shaped by external cues. Since then, only two epithelial-to-myoblasts interactions have been characterized, involving Wingless and Notch signaling.

The work described in Chapter 2 of my thesis identifies additional epithelial-to-myoblast interactions using single-cell transcriptomics. Single-cell RNA sequencing (scRNAseq) enables sequencing of thousands of individual cells, providing an unprecedented opportunity to understand cellular heterogeneity and infer heterotypic interactions. Using this technique, my colleagues and

I catalogued the cell-type composition of the disc epithelium and AMPs at two developmental time points, mid and late larval 3rd instar (L3) development, and built a multi-layered model of the wing disc. While we observed a consolidation of epithelial identities by mid-L3 larval development, AMP identities were largely established later in development. We examined our data for the expression of receptors and ligands that could be mediating heterotypic interactions between the disc proper and AMPs, and characterized two such pathways. The first, fibroblast growth factor (FGF) signaling, was required to localize the AMPs to their developmental niche along the wing disc. Cell-type specific expression of FGF ligands, Thisbe and Pyramus, restricted the AMPs to the proximal regions of the disc proper, and overexpressing either ligand was sufficient to induce AMP growth and proliferation in ectopic regions. Second, we found that Hedgehog ligand secreted from the posterior disc proper induced a cell-type identity within the posterior-most AMPs, defined by AMP-specific Hedgehog targets *Neurotactin* and *midline*. While perturbing Hedgehog transduction within the AMPs had no dramatic effect on AMP numbers during larval development, we observed severe anomalies in adult muscle formation, suggesting that this Hedgehog-conferred identity is needed for proper muscle development.

Organ development can easily go awry, and *Drosophila* has provided many insights into the process of tissue regeneration. While early imaginal disc regeneration studies dating back into the mid-1900s required transplantation experiments, the *Drosophila* community has developed a number of *in vivo* ablation systems for the wing disc. Wing disc regeneration is accomplished via the formation of a blastema, proliferative region characterized by heightened cellular plasticity that develops in response to damage. The regenerative process is complex, requiring cells surrounding the wound site to respond and undergo some level of cell-fate change to recuperate the lost tissue. The heterotypic interactions responsible for activating downstream genetic programs in both near and distant cells, and those that reestablish cellular repatterning, are not fully understood. Furthermore, many of the genes known to be crucial for regeneration are used in earlier developmental processes. Whether there exists a regulatory network designed solely for tissue regeneration, with a negligible role in normal development, would be fascinating, but such a pathway has not been described.

In Chapter 3 of my thesis, my colleagues and I applied scRNAseq to study wing disc regeneration. We observed unappreciated heterogeneity within the blastema at 24 hours into regeneration and characterized two distinct populations, Blastema1 and Blastema2. Both populations express genetic markers of regeneration, such as *Wingless* and *Insulin-like Peptide 8 (Ilp8)*, but Blastema1 is an inner subset of Blastema2 and uniquely expressed a number of secreted molecules. We examined our scRNAseq for transcription factors that might regulate the regenerative process and identified the gene *Ets21C* as being specifically expressed within both blastema populations. Mutant analysis revealed that while Ets21C is largely dispensable for normal development (having no observable aberrant phenotype), it is absolutely crucial for proper regeneration, suggesting that it regulates a regeneration-specific gene regulatory network. Indeed, loss of *Ets21C* abolishes the expression of Blastema1 markers, decreases Ilp8 expression within the blastema, and induces premature pupariation of regenerating larvae. Additionally, we identified blastema-like cells that upregulated components of the Ets21C-dependent gene network within scRNAseq data collected from wing disc tumors, highlighting a mechanism in which tumors may co-opt regenerative processes.

Table of Contents

List of Figures

Acknowledgements

Throughout my doctoral journey, I have been surrounded and supported by incredibly talented and inspiring people. Without a doubt, I would not be the individual and scientist that I am today without this community. These people are a constant source of motivation, creativity, and comfort. For all of this and more that I cannot adequately express in words, you have all been instrumental in the formation of this thesis.

First and foremost, I want to sincerely thank both of my advisors, Iswar Hariharan and Nir Yosef. Iswar has trained so many biology students in his lab, but I doubt he was expecting a *computational* biology student to walk through his doors back in 2017. I am very appreciative of the opportunity and support that he provided to such an unorthodox student as myself. He constantly pushed me to become a more dedicated and curious scientist, and always encouraged me to take more integrative approaches (with wet and dry lab) to my questions and experiments. Additionally, despite our back-and-forth banter and verbal jabs at each other, Iswar is extremely empathetic towards the needs and challenges of being a graduate student. Naturally, I view Iswar as my scientific father, due to the inspirational aura that he constantly emanates as well as his idiosyncrasies that were (on occasion) reminiscent to that of my actual biological father. Unfortunately, my time in Iswar's lab has made me a much worse League of Legends player, as for what must certainly be a surprise to my advisor, I have not played a single game in years.

While Iswar was the "biology" behind my computational biology training, Nir was the foundation for all aspects computational. I thank Nir for guiding my single-cell and statistical analyses, and for always being on the lookout for methods and papers that could provide insight into analyzing our *Drosophila* data. Despite existing in a distant time zone for the latter half of my doctoral research, Nir always found time to discuss research progress and provide mentorship. Furthermore, Nir sponsored many lab events and celebrations that were among the highlights of my time at Berkeley (in particular, I will always remember the amazing escape room that we did as a lab in San Francisco). Nir, thank you so much for all of your encouragement and feedback that has molded me into an interdisciplinary scientist.

Both Iswar and Nir have cultivated lab environments with incredible scientists and welcoming atmospheres; I am so honored to be an alumnus of both groups, and I want to express my thanks to all of my fantastic lab mates. However, I will start by providing my utmost thanks to lab mate in particular. Words cannot describe the appreciation I have for my long-time collaborator and unofficial third mentor, Melanie Worley. Even before speaking with Iswar in his office, Melanie was the person who first greeted me at the door. Unequivocally, she is *the most passionate developmental biologist and geneticist that I have met*. If not for her husband at home, I am fairly certain that she would have moved into Iswar's lab space. Melanie has taught me a tremendous amount about laboratory techniques, *Drosophila* biology, and developing and exploring hypotheses. She is a beacon of experimental ingenuity. I would not be half the scientist I am today without her help, and I look forward to seeing the development of the Worley lab.

Within the Hariharan lab, I would like to thank Melanie Worley (for a second time), Maya Emmons-Bell, Danielle Spitzer, Sophia Friesen, Jamie Lahvic, Joyner Cruz, Riku Yasutomi, Claire Wingert, Rebecca Chang, Dennis Sun, Linda Setiawan, Robin Harris, Taryn Sumabat, Jo Downes Bairzin, and Abraham Asare. All of you are creative, empathetic, and absolutely extraordinary people. Each of you has helped me become a better scientist, a better communicator, a better teacher, and a better person. I want to extend a special thank you to Riku Yasutomi, for eagerly joining the projects described in this thesis and performing some of the best *Drosophila*

muscle dissections ever seen. Within the Yosef lab, I would like to thank Tal Ashuach, Matt Jones, Adam Gayoso, Alex Khodaverdian, Carlos Buen Abad Najar, Zoë Steier, Alyssa Morrow, Galen Xing, Pierre Boyeau, Allon Wagner, Xiuwei Zhang, Anat Kremier, Elior Rahmani, Mariano Gabitto, Romain Lopez, Michael Cole, David DeTomaso, Shaked Afik, and Jim Kaminski. The Yosef lab is filled with incredibly intelligent people, and I thank all of you for listening to my research, providing fantastic feedback, and exposing me to such a wide diversity of exciting projects. I want to extend special thanks to Tal Ashuach, Adam Gayoso, Galen Xing, and David DeTomaso for listening to my long-winded rants and providing answers to my many research questions.

Outside of my labs, I would like to thank my rotation mentors Nick Ingolia and Hernan Garcia. You both were the earliest mentors I had upon arriving to Berkeley, and provided me with temporary research homes. I express my thanks towards David Bilder and members of the Bilder lab for feedback and discussions during joint lab meetings and in the hallways. I'd also like to thank members of my thesis and qualifying exam committees: Sandrine Dudoit (thesis and QE committees), Craig Miller (thesis committee), David Bilder (QE committee), Aaron Streets (QE committee), and Hernan Garcia (QE committee). You all motivated me to think critically about my research and consider the broader impacts beyond the *Drosophila* field. Additionally, I would like to thank the entire Computational Biology community at Berkeley, as it is filled with wonderful students and faculty. In particular, I would like to acknowledge Kate Chase and Xuan Quach for organizing major events and ensuring the success of students in the program.

Importantly, I want to acknowledge my extended cohort within my program, as one of the greatest treasures of earning a PhD are the close friends you make along the way. Tal Ashuach: I appreciate the countless questions of mine that you have explained *in clear, concise, and intuitive ways* (and we both know how important that last part is). Occasionally, I will ask myself "What would Tal do?", because that is likely to be the best course of action. You have made me a better person. Amanda Mok: You are an incredibly talented individual, and a perfect combination of sass and empathy. I enjoyed my adventures with Callie Tealeaf, and look forward to hearing about the brilliant "Doc Mok". I also apologize for not writing my children's book for you to illustrate, but maybe in the future...? Sandra Hui: Thank you for being the responsible member of the group, as we desperately needed that. Also, thank you also for reacting to my jokes that no one else heard (or simply ignored). You are a wonderful person. Jordi Silvestre-Ryan: You are a beautiful blend of charisma and wit. Thank you for providing the group with an abundance of bad puns... so that I wasn't the only one. Jared Bennett: Thank you for motivating me to be more irresponsible – I didn't need it, but am grateful regardless. You are a beautiful, exquisite human. To all of you: Have fun, be safe, and... always remember why 6 was afraid of 7!

Outside of the university, I want to first thank my closest friends Jeff Huang and Abel Rodriguez. Thank you both for being amazing roommates during my undergraduate years, and supporting me during my transition over to Berkeley. Our travel plans are always highlights of my year, and some of my favorite memories have been our trips to Los Angeles and Korean BBQ. If you have ever been curious about why I haven't played more games with you guys, look no further than this dissertation.

I want to sincerely thank my amazing girlfriend Emily Ramey for being a fantastic partner and a massive source of support. I feel so incredibly lucky to have you in my life, and I cherish all of our *collaborative* adventures. You are unbelievably sweet, and I love you so much! I even appreciate Socks and Chewie, the little demons (cats?) that you have brought into my life. I apologize for working absurd hours and coming home late for dinner; you can blame this thesis.

Additionally, you have a wonderful and kind family. To Tom, Kristi, and Grace: Thank you so much for including me on all of your *Wooooah Ramey!* events.

Lastly, I am indebted to my family for motivating and supporting me in all aspects of life, long before I had any ambitions of graduate school. **To Andrew:** As my brother, thank you for being my first friend and role model. Your tenacity inspires me every day, and I hope I make you proud. **To my father:** Thank you so much for providing advice in all aspects of life. You are a role model for hard work and success that I strive to achieve. I appreciate all the support that you and the National Bank of Dad have provided me, and I will continue to rely on this support (especially from the NBD). And just to be clear, you will now need address me as Dr. Everetts, or you will be hearing from my lawyers. **To my mother:** For as long as I can remember, you have been my biggest fan and advocate. By far, you have been my largest source of support, and I could write a second thesis solely dedicated to how much you've helped me in everything (but one thesis has been quite enough work for me). Thank you for all the late-night phone calls to help me de-stress. Thank you for being there for me, especially when I have needed you the most. You can address me as you'd like; the "Dr." requirement only applies to Dad. To the three of you: I love you so much!

This dissertation is dedicated to my parents for their incredible and lifelong effort into ensuring my success.

Chapter 1

Introduction

There is growing evidence that interactions between different cell types (heterotypic interactions) during development play a crucial role in ensuring that the appropriate number and spatial layout of each cell type are present in the fully formed adult organ. Complex multicellular eukaryotes, including humans, have many organs that serve specialized functions composed of cells of numerous types. These cell types are sometimes derived from different germ layers such as the ectoderm and the mesoderm. Heterotypic interactions are not just observed during organ development *in vivo*, but have also been observed during the development of organoids in culture (Wan, 2016; Dutta & Clevers, 2017). Additionally, aberrant heterotypic interactions have been observed in many pathological conditions such as cancer. For example, interactions between breast cancer cells and myofibroblasts are thought to play an important role in tumor progression (Orimo & Weinberg, 2006; Yamashita et al., 2012; Song et al., 2017).

A platform for studying heterotypic interactions: The Drosophila wing-imaginal disc

The common fruit fly, *Drosophila melanogaster*, provides an ideal system for studying the genetics of organ development and heterotypic interactions. First, *Drosophila* possesses a massive experimental toolkit. Central to this toolkit is the *Gal4/UAS* genetic system (Fischer et al., 1988; Brand & Perrimon, 1993), in which localized expression of the yeast-derived transcription factor Gal4 can drive expression of genes engineered with the Gal4-recognized upstream activation sequence (UAS). Alternatively, this system can be combined with the extensive *UAS-RNAi* and *UAS-Cas9/guideRNA* lines to perturb expression of a vast number of genes (reviewed in Bilder & Irvine, 2017). Second, the life cycle of *Drosophila* is short compared to popular vertebrate model organisms, such as mice and the African clawed frog. At 25 ^{o}C, newly laid *Drosophila* eggs will reach adulthood in approximately 10 days, enabling quick experimental iterations and results. Furthermore, the abundance of eggs laid in conjunction with this fast life cycle aids in performing genetic crosses and collection of numerous tissue samples. Third, *Drosophila* has been a pioneering system for the discovery of many high-conserved features and pathways that are crucial for animal development. For example, the homeobox, a stretch of DNA found in master regulators of body-plan development of nearly all animals, was first described by the groups of Walter Gehring and Matthew Scott in the 1980s when studying structural transformations within *Drosophila* (Garber et al., 1983; McGinnis et al., 1984; Scott & Weiner, 1984). Additionally, components of the Hippo pathway were first discovered in *Drosophila* genetic screens as regulators of organ growth and able to cause massive overgrowth (Justice et al., 1995; Xu et al., 1995; Tapon et al., 2002; Harvey et al., 2003). Homologs to *Drosophila* Hippo genes would later be identified as harboring mutations in mammalian cancers (reviewed in both Chan et al., 2011; Harvey et al., 2013).

Within *Drosophila*, the wing-imaginal disc is a popular platform for organ development research (Cohen, 1993; Neto-Silva et al., 2009) (**Figure 1.1)**. Broadly speaking, imaginal discs are epithelial structures within the larvae of holometabolous insects and give rise to specific adult structures after metamorphosis. As partially suggested by the name, the *Drosophila* wing disc develops into the wing blade as well as the thorax of the adult fly. The epithelium of the wing disc is composed of two layers, the columnar cells of disc proper and squamous cells of the peripodial epithelium (PE), that are connected at the edges to form a continuous monolayer of cells resembling a flattened sac. The proximal-most region of the disc tapers into a tip, referred to as the stalk (**Figure 1.1**), and is attached to the larval body wall. Cells of the wing disc epithelium are specified early in development; during embryogenesis, an estimated 30 cells form the wing disc primordia (Madhavan & Schneiderman, 1977; Worley et al., 2013; Requena et al., 2017).

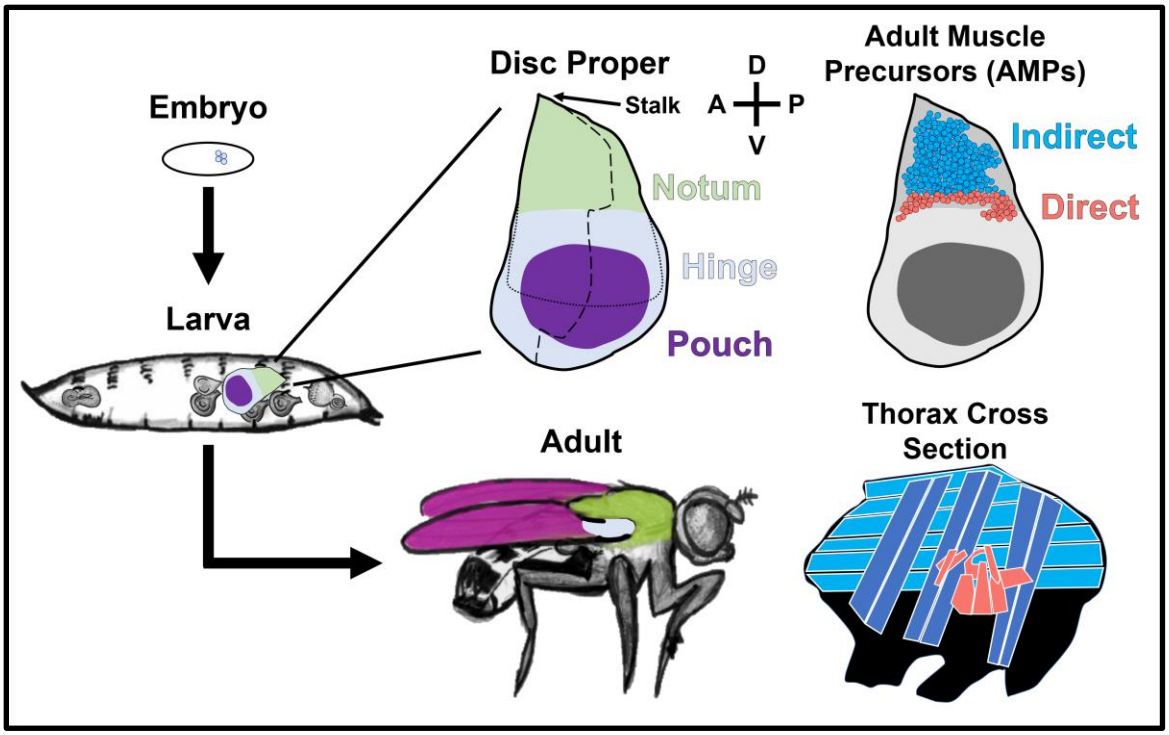

Figure 1.1. Development of the wing disc and adult flight muscle precursors (AMPs) into their adult structures. The disc proper epithelium is composed of three domains: the notum, hinge, and pouch, which give rise to the matching-colored regions within the adult (see main text for additional description). The AMPs reside on the basal surface of the disc epithelium (underlying the notum) and are split into two populations: the indirect AMPs, which produce to the large indirect muscles of the thorax that generate flight power, and the direct AMPs, which develop into the smaller thoracic direct muscles that control wing steering. Long dashed line corresponds to the anterior-posterior compartment boundary; transverse short-dotted line corresponds to the dorsal-ventral compartment boundary. A = anterior; P = posterior; D = dorsal; V = ventral. Figure partially adapted from Everetts et al., 2021.

These cells undergo around 10-11 rounds of cell division to develop into the mature wing disc, approximately 40,000 cells just prior to pupariation (Garcia-Bellido & Merriam, 1971; Martín et al., 2009).

Literature on wing disc biology overwhelmingly focuses on the disc proper layer, cementing it as one of the most well-studied models for organ development (Hariharan, 2015). The disc proper is canonically separated into three cell-type domains (**Figure 1.1**), based on which adult structures they create and distinct patterns of gene expression. These domains are the notum, which develops into the adult thorax; the pouch, which develops into the adult wing blade; and the hinge, which develops into the connective tissue between the thorax and wing blade (reviewed in Held Jr, 2002). Popularity for wing disc research can be partially attributed to the fact that the wing disc produces external structures on the adult fly. This enables genetic screens to be performed with adult phenotypes efficiently assayed, as an irregular wing blade can be easily observed and does not compromise the vitality of the fly (the wing blade is not needed for survival in laboratory settings).

The wing disc is further subdivided into compartments, which are well-defined populations of cells that are related by cell lineage. Despite no obvious physical restrictions, cells from one compartment will not intermingle with cells of the opposing compartment throughout development. The first compartment boundary organizes the disc into anterior (A) and posterior

(P) compartments (Garcia-Bellido et al., 1973) (**Figure 1.1**), and is established during embryogenesis by selective expression of the homeobox transcription factors *engrailed (en)* and *invected (inv)* (paralogs with largely redundant functions) within posterior compartment cells (Tabata et al., 1995; Gustavson et al., 1996). Posterior cells are further marked by the expression of the ligand *hedgehog (hh)*, whereas anterior compartment cells are marked by expression of *cubitus interruptus (ci)*. The juxtaposition of A/P compartments defines a crucial heterotypic interaction for proper wing disc growth, as anterior cells in close proximity to the A/P boundary respond to Hedgehog ligand secreted by posterior compartment cells (Tanimoto et al., 2000; Hatori & Kornberg, 2020). The result of this interaction is high expression of Hedgehog-pathway target genes in anterior cells that straddle the compartment boundary (Tanimoto et al., 2000; Hatori & Kornberg, 2020). These target genes include *patched (ptc)*, which encodes for the Hedgehog-pathway receptor, and *decapentaplegic (dpp)*, which encodes for a BMP morphogen. The subsequent diffusion of Dpp morphogen from its source cells establishes a gradient of protein expression along the A/P axis of the tissue which is critical for regulating organ growth, although the mechanics of which are still actively being researched.

While the A/P compartments define one axis of the wing disc, a second axis is defined by the dorsal (D) and ventral (V) compartments (**Figure 1.1**). The D/V boundary is established around mid-larval development, much later than the A/P boundary (Blair, 1993; Diaz-Benjumea & Cohen, 1993). As with cells divided by the A/P boundary, cells within the D/V compartments become lineage-restricted and do not intermingle with cells of the opposing compartment. The divergence of the D/V compartments is the result of the homeobox transcription factor *apterous (ap)*, which is expressed in all dorsal compartment cells and determines the dorsal cell fate. It is unlikely that a ventral-specifying gene exists; loss of *ap* alone appears sufficient to specify the ventral fate within clones of cells (Blair, 1993; Diaz-Benjumea & Cohen, 1993). Like the A/P compartment boundary, the D/V boundary represents another heterotypic interaction that is vital for wing disc development. Notch-mediated signaling between the D/V compartments induces a variety of genes with distinct expression patterns (Diaz-Benjumea & Cohen, 1995; Kim et al., 1996; Neumann & Cohen, 1997); genes like *wingless* and *cut* are expressed in a thin line pattern that tightly straddles the D/V boundary, whereas genes like *vestigial* and *scalloped* are expressed in a much broader pattern along the D/V boundary. Not only does the D/V boundary form another axis of signaling within the wing disc, roughly orthogonal to that of the A/P boundary, but the genes induced also specify the population of cells that form the wing margin.

What lies beneath the surface: The disc-associated myoblasts

Underneath the basal surface of the notum region of the disc proper reside the adult muscle precursor cells (AMPs), a population of myoblasts that develop into the thoracic flight muscles of the adult fly (Bate & Martinez Arias, 1991; Fernandes et al., 1991) (**Figure 1.1**). The AMPs become associated with the wing disc epithelium during late embryogenesis (Bate & Martinez Arias, 1991; Fernandes et al., 1991; Bate, 1993), and over the course of larval development, proliferate into a population of approximately 2,500 cells (Gunage et al., 2014). Both AMPs and larval muscle precursors (the latter of which build the musculature of the *Drosophila* larva rather than the adult musculature) originate from a naive population of myoblasts derived from the embryonic mesoderm during gastrulation (Bate, 1993). The divergence of AMPs from their larval muscle precursor siblings is closely tied to the expression of the mesodermal transcription factor *twist (twi)*. Initially, all embryonic myoblasts express high levels of *twi*, but only the larval muscle precursors lose *twi* expression as they further develop (Bate et al., 1991; Laurichesse & Soler,

2020). The small subset of embryonic myoblasts that retain high *twi* expression develop into the AMPs that associate with epithelial imaginal discs, including the wing-disc AMPs.

Canonically, the wing-disc AMPs have been grouped into two subpopulations, the direct and indirect AMPs (Bate, 1993; Roy & VijayRaghavan, 1999; Sudarsan et al., 2001) (**Figure 1.1**). The pioneering markers for the direct and indirect AMPs have been the genes *cut (ct)* and *vestigial (vg)*. Direct AMPs express high levels of Ct and negligible levels of Vg, whereas the indirect AMPs express high levels of Vg and relatively low levels of Ct (Sudarsan et al., 2001). Together, these two subpopulations give rise to the majority of muscle fibers within the adult thorax (Lawrence, 1982; Laurichesse & Soler, 2020). The indirect AMPs will fuse into the largest muscle fibers within the thorax, the indirect adult flight muscles (**Figure 1.1**), which generate the power for wing strokes via thoracic contractions. The indirect adult flight muscles are further divided into two sets: the dorsal ventral muscles (DVMs) and the dorsal longitudinal muscles (DLMs). There are 7 DVMs and 6 DLMs per hemithorax, for a total of 26 indirect flight muscles within the adult fly. The direct AMPs develop into the direct adult flight muscles (**Figure 1.1**), a much smaller set of muscles that control flight steering via fine-tuned positioning of the wing blade. There are 10 direct flight muscles per hemithorax, historically numbered as muscles 49-58, for a total of 20 direct muscle fibers. Notably, clonal lineage tracing experiments suggest that muscle 51 is unrelated to the other direct flight muscles, and likely does not develop from the pool of wing-disc AMPs but rather AMPs associated with the leg discs (Lawrence, 1982). Additionally, the fiber morphology of the direct and indirect flight muscles differs (Deak, 1977; Laurichesse & Soler, 2020). The direct flight muscles are described as tubular, having fibers with aligned myofibrils that form a tube around a central lumen. Conversely, the indirect flight muscles possess unaligned myofibrils with high mitochondria count supporting their high energy needs; they are called fibrillar fibers.

Despite the remarkable diversity of muscle fibers formed, the wing-disc AMPs appear to be relatively naive during larval development. Experiments by Lawrence and Brower in 1982 indicated that when wing-disc AMPs were transplanted to ectopic locations, they readily adapted and indiscriminately incorporated into muscle fibers of the new position (even if this required a potential change in fate) (Lawrence & Brower, 1982). These results were reaffirmed by additional AMP transplantation experiments performed by Roy and VijayRaghavan in 1997 (Roy & VijayRaghavan, 1997). Given that heterogeneity exists within the AMP population (e.g., direct and indirect AMPs) under normal conditions, these findings suggest that AMPs receive information cues from their environment. More recently, single-cell RNA sequencing work by Zappia et al. has uncovered unappreciated heterogeneity within the AMPs, including additional genetic markers of the direct and indirect subpopulations (Zappia et al., 2020). However, relatively few studies have examined the mechanisms by which the wing disc epithelium may influence the growth, development, and heterogeneity of the underlying AMPs.

Two signaling cues operating between the AMPs and epithelial disc proper have been identified for proper patterning and growth of the AMPs, the first of which being the Wingless (Wg) signaling pathway. The Wg pathway and the homologous mammalian Wnt pathway are involved in a vast number of developmental processes across metazoans, including being required for embryogenesis as well as the growth and regeneration of organs (Seto & Bellen, 2004; Bejsovec, 2018). In the context of the wing disc, Wingless (as its name suggests) is necessary for proper patterning of the pouch and formation of the wing blade, and the thin expression pattern of Wg at the D/V boundary specifies the cells of the wing margin (Ng et al., 1996; Swarup & Verheyen, 2012). However, during late larval development of the wing disc (starting at late L2 /

early L3 stages), Wg is expressed at significant levels within the notum of the disc proper, adjacent to the underlying AMPs. Sudarsan et al. in 2001 showed that this notal Wg expression was required to establish the direct and indirect subpopulations within the AMPs (Sudarsan et al., 2001). Wg transduction stabilizes the expression of *vg* at the cost of *ct* expression (as *vg-ct* expression is mutually antagonistic) within most AMPs, thus enforcing the indirect identity within these cells. The distal-most AMPs, unable to receive appreciable amounts of Wg ligand, promote high levels of Ct and become direct AMPs.

The second epithelium-to-myoblast pathway implicated in the development of AMPs was the Notch signaling pathway. Notch signaling represents a highly conserved pathway that regulates a variety of cell-fate decisions; in *Drosophila*, there is a single Notch (N) receptor that interacts with two ligands, Serrate (Ser) and Delta (Dl) (Mumm & Kopan, 2000). Notch-regulated systems include neuroblast development during the embryogenesis, cellular differentiation in the wing and eye imaginal discs, and stem cell maintenance. In 2014, Gunage et al. demonstrated that AMPs undergo symmetric cell division during early wing disc development (L1 and L2 stages) to create a monolayer of AMP stem cells adjacent to the disc epithelium (Gunage et al., 2014); this proliferation required both AMP expression of N and notum expression of Ser. Later in development (L3 stage), notal Wg signaling to the AMPs induced expression of N-inhibitor Numb, associated with a transition to asymmetric cell division to produce a multilayer pool of AMPs. Only AMPs in contact with the epithelium would retain their stem-cell properties, whereas more distal AMPs would remain mitotically silent.

Both Wingless and Notch signaling represent epithelial-to-myoblast interactions that provide identity and proliferative cues to the AMPs, shedding new light on the mechanisms that mold the naive pool of AMPs. But even so, knowledge on the growth and diversification of adult *Drosophila* musculature remains limited. Significant variation exists between adult muscle fibers, seemingly beyond the direct/indirect dichotomy that has been canonically described for the AMPs. One hypothesis is that while the disc epithelium provides some initial and crude patterning of the AMPs, significant muscle resolution is achieved after larval development, during metamorphosis. However, additional heterogeneity identified within the direct and indirect AMPs suggests that more larval mechanisms for patterning AMPs remain to be uncovered (Zappia et al., 2020). Furthermore, the notum of the disc epithelium is the predominant signaling source for the AMPs, forming a developmental niche for these cells. While Notch/Serrate interactions regulate some aspects of AMP growth, the mechanisms that localize the AMPs to the notum niche remain unresolved. The work described in Chapter 2 of this thesis examines these questions of epithelial-to-myoblast communication in greater detail and characterizes new heterotypic interactions that govern AMP development.

Replace and repattern: Studies of regeneration in *Drosophila*

Organismal development is not a deterministic process. The mechanisms that produce an adult animal from a fertilized egg are remarkably complex and subject to considerable noise and potential setbacks. The source of these setbacks can vary, possibly being internal (a result of spontaneous mutations, misexpression of genes, or other genetic shortcomings) or external (due to factors outside of the developing body that create a harsh and nonideal environment). In short, development does not always proceed in an exact and expected manner, and during this process, cells may experience aberrant situations. Thus, the processes of animal development and, more specifically, organ development must be sufficiently robust to handle internal biological noise and external environmental chaos.

Organ damage represents one anomalous outcome of interrupted or misregulated development, necessitating the evolution of regenerative programs to recover lost tissue. The goals of these programs are two-fold. First, the damaged tissue needs to be replaced, which requires induced proliferation from cells adjacent to the site of damage. Second, expression and differentiation patterns that have been lost or interrupted by the damage must be reintroduced, enacting a process that involves some level of rejuvenation and fate change in the cells that will eventually replace the damaged tissue. The extent at which organisms can accomplish these goals varies significantly (reviewed in both Carlson, 2007; Brockes & Kumar, 2008). Many animals, from invertebrates like hydra and planaria (flatworms) to vertebrates such as salamanders and newts, display complete regenerative responses and can replace nearly any lost tissue throughout their life cycle. However, many others display regenerative capabilities that become severely attenuated over the course of their life cycle, often peaking during early developmental stages but limited in many tissues during adulthood. This not only includes mammals like humans and mice, but also invertebrates like *Drosophila*.

Drosophila melanogaster has a rich history of regenerative studies, providing valuable insights into the initial responses of tissue damage and regeneration that are largely conserved among species. Many of the earliest experiments were pioneered by the research group under Ernst Hadorn during the mid-1900s, specifically examining the genital, leg, and wing imaginal discs (Hadorn et al., 1949; Hadorn & Buck, 1962; Hadorn, 1965). First and foremost, by taking fragments of imaginal discs and implanting these into young larvae, mature larvae, or adult abdomens, the Hadorn group demonstrated that these tissues could indeed regenerate damaged or lost structure. However, this process needed sufficient time in these *in vivo* culture experiments; implanting imaginal discs into mature larvae (just prior to pupariation) inhibited regeneration, presumably due to the process being interrupted during metamorphosis. Regeneration achieved greater success when fragmented tissues were implanted into younger larvae or adult abdomens, the latter providing an environment in which discs could regenerate without undergoing cellular differentiation. Under these conditions, serially damaged discs seem to be able to regenerate indefinitely (Hadorn et al., 1949; Hadorn & Buck, 1962; Hadorn, 1965). Second, regeneration was characterized by localized cell death and proliferation in regions adjacent to the site of injury or fragmentation (Kiehle & Schubiger, 1985; O'Brochta & Bryant, 1987; Smith-Bolton et al., 2009). This region of high proliferation would eventually give rise to new tissue and was called the blastema. Importantly, the formation of the blastema occurred rapidly (within the first 24 hours of regeneration) before any continuity or repatterning of tissue can be reestablished. This observation suggests that the initial catalyst for regeneration originates from the wound site itself rather than an assessment of missing cell types or disparate positional identities by the tissue, which had been hypothesized during early experiments. Indeed, more recent experiments have identified waves of intracellular calcium signaling emanating from the wound as one of the earliest observable events during regeneration (Narciso et al., 2015; Restrepo & Basler, 2016).

While incredibly insightful, the early experiments employed by Hadorn and others were tedious and low throughput, necessitating the creation of more modern tools. *Drosophila* regenerative studies received a boon with the implementation of *Gal4/UAS*-based regeneration assays (Smith-Bolton et al., 2009; Bergantiños et al., 2010) (**Figure 1.2**), which alleviated much of these previous limitations. By placing pro-apoptotic genes (e.g., the JNK agonist *eiger* or caspase activator *reaper*) downstream of UAS, cell death can be induced via the expression of the Gal4 transcription factor. Incorporation of a temperature-sensitive Gal80 protein ($Gal80^{ts}$), a repressor of Gal4 function, provides experimental control over the time and duration of induced

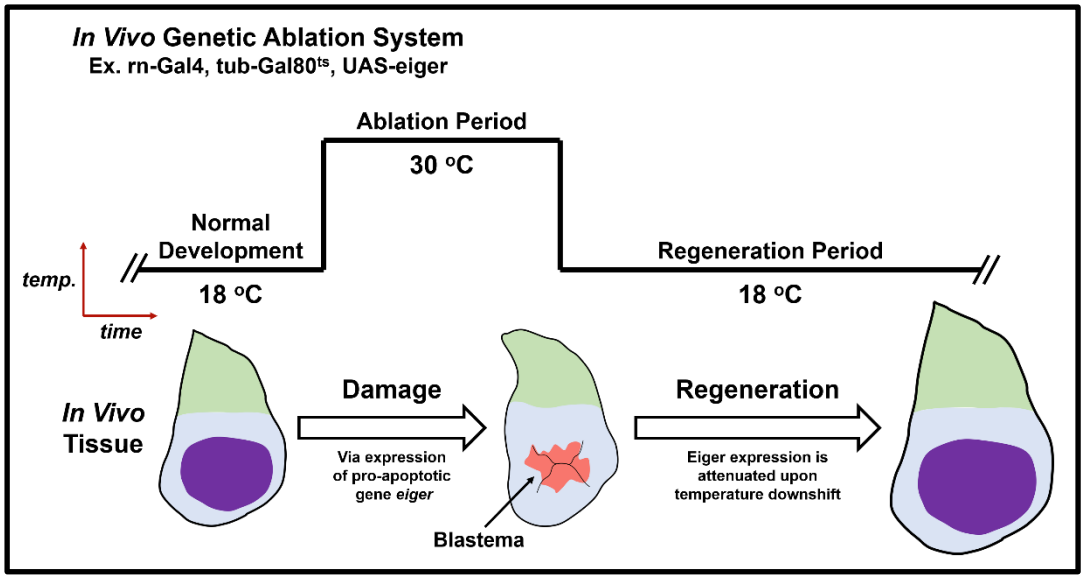

Figure 1.2. *Drosophila in vivo* genetic ablation system. Upon a heat shift to 30 °C, the pouch of the disc proper is ablated via expression of the pro-apoptotic gene *eiger*, driven by the *rotund-Gal4* pouch driver that is normally repressed at lower temperatures by a temperature-sensitive Gal80 protein. The ablated pouch forms a zone of high cellular proliferation and plasticity called the blastema. Downshifting the larvae inactivates the ablation system, enabling the study of the blastema and regenerative programs. Figure partially adapted from Smith-Bolton et al., 2009.

cell death by raising larvae to 30 °C (relieving Gal80^{ts} repressor and enabling Gal4-mediated expression of apoptotic genes). Downshifting animals to 18 °C shuts off damage and allows subsequent tissue regeneration. Furthermore, the vast variety of tissue-specific Gal4 drivers provide localization of cell death within the organism. Importantly, this genetic ablation system provides an *in vivo* method of studying the process of tissue damage and subsequent regeneration without need for implantations (thus completely preserving the native environment), all under the control of a simple mechanism (a temperature shift) that enables efficient and higher-throughput experimentation. The *rotund-Gal4* driver is commonly used to induce apoptosis within most of the pouch of the wing disc epithelium (Smith-Bolton et al., 2009) (**Figure 1.2**). By examining wing blade defects in adult flies, this genetic ablation system allows for regenerative capacity to be easily scored and facilitates genetic screens for the identification of key regeneration components. Variations and improvements on genetic ablation systems provide additional flexibility to experimental design. One such example is the DUAL control system (Harris et al., 2020), which combines both *Gal4/UAS* and *Escherichia coli*-derived *LexA/LexAop* gene drive systems to independently drive tissue damage via *LexA/LexAop* and genetic perturbations via *Gal4/UAS* (taking advantage of the massive number of UAS transgenes generated by the *Drosophila* community). Altogether, modern tools have developed *Drosophila* into an impressive *in vivo* model of regeneration, boasting efficient, high throughput, and replicable experimental tools.

Studies using wing disc ablation systems have demonstrated that regeneration is not an isolated event, but rather involves complex heterotypic interactions between the blastema and cells both local and distant. Indeed, by inducing damage within the wing pouch, the resulting blastema is characterized by the secretion of many signaling molecules, including the JAK/STAT Unpaired (Upd) ligands (Pastor-Pareja et al., 2008; La Fortezza et al., 2016), the EGFR ligand Spitz (Fan et al., 2014), and the Wnt ligand Wingless (Gibson & Schubiger, 1999; Smith-Bolton et al., 2009;

Harris et al., 2016). Locally, the cells within and surrounding the blastema are characterized by high levels of proliferation and promoters of growth such as the proto-oncogene Myc (Kiehle & Schubiger, 1985; Smith-Bolton et al., 2009). This growth response is controlled, as proliferation and Myc expression is downregulated in wing disc regions distant from the blastema (e.g., the notum). The mechanism by which the blastema is able to upregulate nearby growth while suppressing growth in more distant regions of the wing disc is not fully understood. Across the entire disc epithelium, cellular plasticity is increased (Hadorn, 1965; Repiso et al., 2013; Herrera & Morata, 2014), likely to enable proper replacement of the damaged pouch by non-pouch cells. However, in certain genetic backgrounds, this increase in plasticity can destabilize the notum and cause transdetermination, in which notum cells undergo a cell-fate change and develop into an ectopic pouch (resulting in a wing disc with two pouches) (Worley et al., 2018). On an organism-wide scale, tissue damage of the pouch is able to delay the onset of pupariation (Smith-Bolton et al., 2009; Hackney et al., 2012), highlighting that regeneration induced in a small, localized population has much larger, global effect. The mechanism behind this interaction involves the secretion of Insulin-like Peptide 8 (Ilp8) from the blastema, which binds to the relaxin receptor Lgr3 within the brain and suppresses production of the metamorphosis-controlling hormone ecdysone (Vallejo et al., 2015; Colombani et al., 2015; Garelli et al., 2015; Jaszczak et al., 2016). Inhibiting this production of Ilp8 within the blastema compromises the regenerative process.

Regeneration is a complex process, and despite the progress made by *Drosophila* research, there are still major aspects of the process that remain unresolved. Many of the genes involved during regeneration are also used in other non-regenerative developmental contexts. One hypothesis is that regeneration is a replay of genetic programs used during earlier development, and that reactivation of these programs at the site of damage enacts a rejuvenation and rebuilding process. Alternatively, there may exist sets of genes that are activated only under regenerative conditions, and thus largely dispensable for normal organism development. A regeneration-specific program, if one exists, has not been recognized. Furthermore, not all cells respond similarly to regeneration, and these differing responses are likely a reflection of the distance to the blastema as well as cell-type specific attributes. Characterizing the heterogeneity of regenerative responses, such as changes in transcriptional and chromatin signatures, at a cell-type resolution has not been extensively performed but would be valuable in understanding heterotypic interactions stemming from the blastema. Additionally, the blastema is a transient state, and the evolution and resolution of this state has not been fully mapped out. A time-course of blastema development could reveal key insights into regulators of blastema formation and genes involved in tissue repatterning towards the end of the regenerative process. These ideas serve as motivating questions for the work presented in Chapter 3 of this thesis, which explores unappreciated cellular heterogeneity within the blastema and identifies a regeneration-specific gene regulatory network.

Modeling organ development at cellular resolution: Single-cell transcriptomics

A common thread among developmental and regenerative studies is the increasing evidence that cellular heterogeneity and heterotypic interactions play a vital role. Unfortunately, the dominant experimental approaches in *Drosophila* have significant shortcomings for examining tissue heterogeneity. For example, immunohistochemistry and fluorescent genetic reporters, nearly ubiquitous within *Drosophila* literature, have two major problems: these techniques are (1) low-throughput, only able to assess the expression patterns of a few genes per experiment, and (2) require knowledge of the genes of interest, a painstakingly difficult task that inhibits the discovery of novel biological mechanisms. Sequencing technologies can alleviate these issues by providing

genome- or transcriptome-wide measurements without the need for any *a priori* information. There are an abundance of sequencing technologies: DNA sequencing is necessary for *de novo* genome construction and can assay allelic variation and single-nucleotide polymorphisms (SNPs); RNA sequencing measures gene expression, providing insight into cellular states and transcriptional modifications; ATAC (Assay for Transposase-Accessible Chromatin) sequencing assays chromatin accessibility and is vital for uncovering genomic regulatory elements (Buenrostro et al., 2013); and ChIP (Chromatin Immunoprecipitation) sequencing can further specify the roles of regulatory elements by assaying for post-translational modification of histones and identifying the locations of specific protein-DNA interactions (e.g., transcription factor binding sites). Sequencing has historically required the homogenization of whole tissues as its starting material ("bulk" genomics), thus completely abolishing any assessment of heterogeneity within a sample, but massive improvements for sequencing at single-cell resolution ("single-cell" genomics) have revolutionized the field. The most frequently applied single-cell technique is single-cell RNA sequencing (scRNAseq), which can profile the transcriptomes of individual cells derived from a biological sample (e.g., organ tissue). Additional sequencing methods have also been applied at single-cell resolution. CITE-seq (Cellular Indexing of Transcriptomes and Epitopes by Sequencing) combines scRNAseq with antibody-bound oligonucleotides (Stoeckius et al., 2017), which can be sequenced after antibody binding to protein targets supplying both single-cell gene expression and proteome measurements. Single-cell ATAC sequencing (scATACseq) assays chromatin accessibility within diverse cell types (Cusanovich et al., 2015; Buenrostro et al., 2015), and can be performed simultaneously with scRNAseq to infer links between enhancers and the genes they regulate. The research presented in this thesis focuses on scRNAseq, as it is an ideal starting point for understanding cellular heterogeneity, inferring heterotypic interactions, and building models for the genetics that shape organ growth and regeneration.

Early scRNAseq experiments conducted in the 2000s were limited in scope, only collecting measurements for tens to hundreds of cells (Cauli et al., 2000; Tietjen et al., 2003; Svensson et al., 2020). But since then, scRNAseq has absolutely exploded; in the last 20 years, more than 1,500 single-cell datasets have been produced (Svensson et al., 2020). Similarly, the number of cells sequenced in each dataset has increased exponentially, with recent datasets boasting transcriptomes for tens of thousands to even *millions* of sequenced cells (Cao et al., 2019, 2020; Sinha et al., 2021). This massive growth of data has necessitated the development of analysis pipelines to efficiently extract meaningful insights and develop hypotheses, making scRNAseq studies a marriage between experimental and computational sciences. These computational analyses aim to identify transcriptional signatures within the many cells sequenced, providing a catalogue of cell populations, the expression profiles that distinguish them, and trends within the data. Ultimately, this builds a computational model of the sequenced sample, enabling insights into various biological processes such as cellular differentiation, cell-cell interactions, master transcriptional regulators within a tissue, and disease effects on cell populations. Already, scRNAseq has been applied (in abundance) to nearly every organism of interest. In mice, every major organ has been translated into a scRNAseq "atlas" (Schaum et al., 2018; Almanzar et al., 2020), and scRNAseq experiments chronicle mouse development from embryogenesis to adulthood (Cao et al., 2019). In humans, numerous diseases have been dissected via scRNAseq, including cancer (Li et al., 2019; Couturier et al., 2020), multiple sclerosis (Schirmer et al., 2019; Pappalardo et al., 2020), Crohn's disease (Martin et al., 2019), and COVID-19 (Ren et al., 2021; Sinha et al., 2021). Comparatively, there have been far fewer but still notable scRNAseq experiments on *Drosophila*. Many of these focus on developmental gene regulation within samples

such as the embryo (Karaiskos et al., 2017), larval brain (Davie et al., 2018), and imaginal discs (Ariss et al., 2018; Bageritz et al., 2019).

Although scRNAseq is a powerful technique for obtaining a near-comprehensive view of a tissue or biological process, the technique has notable drawbacks. Perhaps the most commonly cited problem with scRNAseq is the low capture efficiency of the technique, with many widely-used methods only capturing an estimated 10-15% of all mRNA transcripts within a cell (Islam et al., 2013; Macosko et al., 2015; AlJanahi et al., 2018). Often referred to as "drop-outs" (although the usage of this term often varies and is poorly defined), many transcripts will remain undetected by the scRNAseq process, leading to an inaccurate and noisy view of the transcriptome with many lowly-expressed genes being completely undetected in cells. Many early computational methods attempt to account for an overabundance of undetected genes in samples, termed "zero-inflation", although the need for explicit zero-inflation modeling has been disputed more recently (Svensson, 2020; Sarkar & Stephens, 2021). Another widely cited issue are batch effects, systematic differences between data obtained from separate experimental runs that are attributed as technical artifacts (Leek et al., 2010; Hicks et al., 2018; Luecken et al., 2020). Many scRNAseq datasets are the aggregation of data collected from multiple experiments as replicates, intended to be integrated together during computational analysis. Even when protocols are carefully maintained between batches, biological replicates will exhibit significant differences that inhibit integration between batches. As such, there exists many computational tools for removing batch effects during data analysis (reviewed and benchmarked in Luecken et al., 2020), harmonizing cells such that biological variation is preserved and perceived technical effects are eliminated. In addition to batch effects, scRNAseq can be sensitive to strong biological signals such as cell cycle and cell sex (the sex of the donor animal) (Buettner et al., 2015; Wagner et al., 2016; Ransick et al., 2019; Everetts et al., 2021). These signals can dominate computational analyses and consequently mask other biological variation, hinder cell-type classification, and decrease statistical power. Correcting for unwanted biological variation can be performed but must be carefully evaluated to avoid overcorrecting the analysis. Lastly, the number of cells sequenced per scRNAseq experiment is growing dramatically (Svensson et al., 2019), increasing the computational resources needed to process these samples. Many larger datasets cannot be run on standard computers, and must be transitioned onto computational clusters for some if not all of the analysis. More efficient computational algorithms are also being developed to offset this computational expense.

Experimentally, collecting scRNAseq data from a tissue sample can be separated into three steps. First, the sample must be broken down into a single-cell suspension, typically requiring incubating the tissue with a protease cocktail. No standard cell dissociation procedure exists, but rather the exact method (such as concentrations of proteases used and the intubation time and temperature) must be optimized for the tissue of interest. This is a non-trivial and often tedious process, and a poor dissociation protocol can result in significant cell death or misrepresentation of cell populations. Commonly used proteases include trypsin, collagenase, elastase, and dispase, often with an incubation temperature at 37 ^{o}C (the temperature at which these proteases have optimal activity) or room temperature (~20-22 ^{o}C, to avoid a potential heat-shock response). Occasionally, fluorescent-activated cell sorting (FACS) is employed after dissociation to remove dead cells, cellular debris, and aggregations of multiple cells.

The second experimental step for scRNAseq is the single-cell barcoding, which differentiates scRNAseq from traditional bulk RNA sequencing. During barcoding, individual cells are segregated and subjected to a series of micro-reactions, in which cells are lysed and the RNA is captured. Importantly, cell barcodes are incorporated during the reverse transcription of

each captured RNA transcript such that transcripts from the same cell (in the same micro-reaction partition) are labelled with the same cell barcode. Many recent methods also incorporate unique molecular identifiers (UMIs), which distinguish each individual transcript even from the same cell to correct amplification bias during PCR. Single-cell barcoding techniques can be split into two major categories: droplet-based and plate-based methods. Droplet-based methods, such as Drop-seq and 10x Genomics Chromium (Macosko et al., 2015; Zheng et al., 2017), are the popular choice in the field currently. They utilize microfluidic devices to combine individual cells with RNA-capturing microbeads in water droplets partitioned by oil, processing thousands of cells per run with the incorporation of transcript UMIs (Zhang et al., 2019). 10x Chromium represents the state-of-the-art technology (Zheng et al., 2017), boasting high sensitivity for transcript capture and experimental replicability. However, Drop-seq is an extremely attractive alternative (Macosko et al., 2015), as it is not reliant on proprietary devices or reagents and is thus much more cost-efficient. Plate-based methods, such as Smart-seq2 and CEL-seq2 (Picelli et al., 2013; Hashimshony et al., 2016), require individual cells to be deposited into multi-well plates, often performed manually with micro capillary pipettes or automated via FACS. As a result, these methods are significantly lower-throughput compared to their droplet-based cousins. Despite this drawback, plate-based barcoding is more sensitive to capturing transcripts, and Smart-seq2 notably provides full-length coverage of the transcripts compared to only the 3' end provided by most methods (Ziegenhain et al., 2017).

Finally, after the barcoded library is PCR amplified, the third experimental step involves sequencing the barcoded reads. The overall structure of the barcoded reads varies depending on the protocol used but will contain Illumina sequencing primers on each end of each read. This enables paired-end sequencing of the transcript read, the cell barcode, and UMI (when applicable). Typically, 15,000 reads per cell is sufficient for reasonable cell-type classification (Svensson et al., 2019), although many studies aim for 30,000 to 50,000 reads per cell. The output of sequencing is a series of FASTQ files, a text-based file format that contains data on sequencing reads, demarcating the experimental end of scRNAseq experiments and the beginning of computational data processing.

The standard computational pipeline for analyzing scRNAseq data first begins with alignment and quality control, often referred to as preprocessing steps (Luecken & Theis, 2019) (**Figure 1.3**). After the scRNAseq library has been sequenced, the FASTQ files are parsed for cell barcodes and UMIs, and transcript reads are aligned to a reference genome. The end result of alignment is a cell-by-gene count matrix, in which each matrix element represents the transcript counts for a particular gene in an individual cell (**Figure 1.3**). Methods that utilize UMI incorporation will combine transcripts with identical UMI sequences to prevent overcounting transcripts due to PCR duplication. Next, the distribution of reads and genes detected per cell is examined and the quality of each cell is evaluated. Typically, a threshold is applied to remove cells with unusually high and low number of reads and / or genes detected, as these cells likely correspond to doublets (aggregates of multiple cells) or low-quality cells (e.g., dying or compromised cells). Quality control can also be applied after additional downstream analysis, often in a more supervised manner (e.g., identifying clusters with high expression of pro-apoptotic genes).

Once high-quality cells are isolated, exploratory data analysis is performed by a number of methods, namely dimensionality reduction, clustering, visualization, and differential expression testing (**Figure 1.3**) (Luecken & Theis, 2019). Dimensionality reduction projects cells into a latent space, a low-dimensional embedding of the count matrix. Specifically, the gene space is

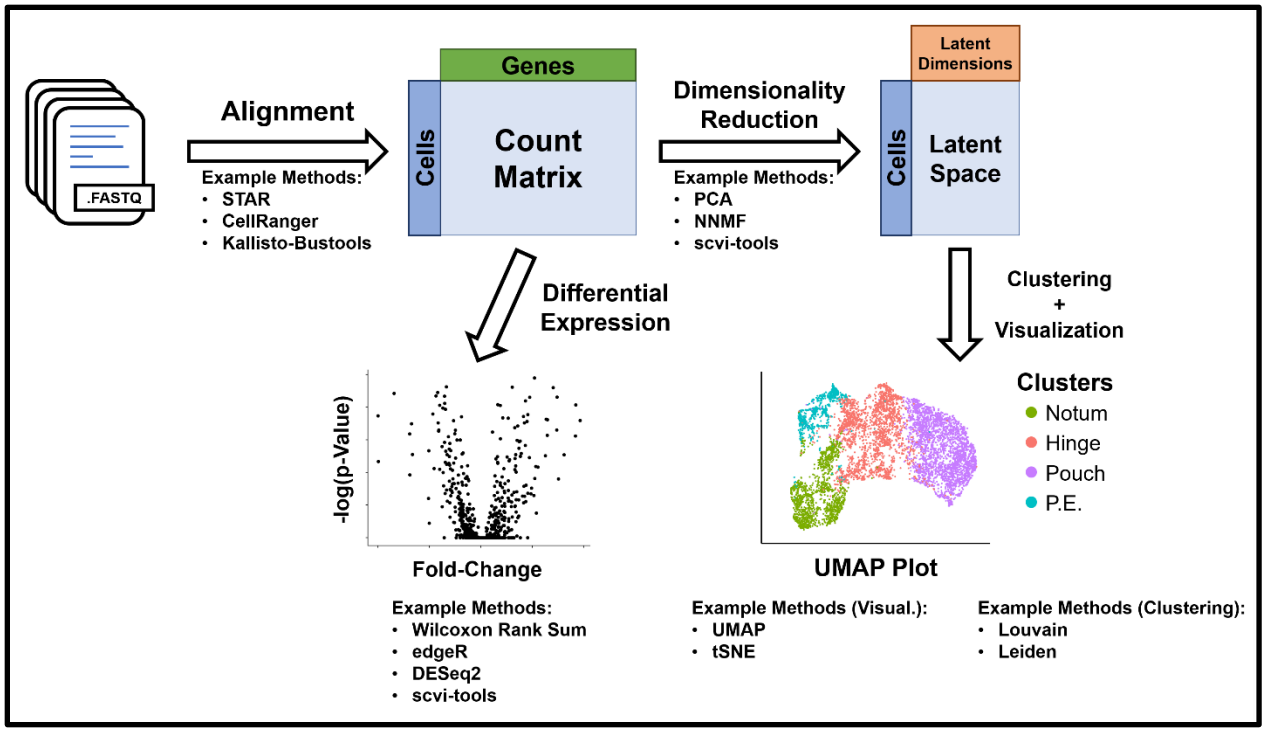

Figure 1.3. Core computational pipeline for exploring scRNAseq data. Diagram connecting standard scRNAseq tasks with example tools for each. Alignment: STAR (Dobin et al., 2013); CellRanger (Zheng et al., 2017); Kallisto-Bustools (Melsted et al., 2019). Dimensionality reduction: principal component analysis (PCA) (Pearson, 1901; Hotelling, 1933); non-negative matrix factorization (NNMF) (Paatero & Tapper, 1994; Lee & Seung, 1999); scvi-tools (Lopez et al., 2018; Gayoso et al., 2021). Differential expression: Wilcoxon Rank Sum test (Wilcoxon, 1945; Mann & Whitney, 1947); edgeR (Robinson et al., 2010), DESeq2 (Love et al., 2014), scvi-tools. Visualization: Uniform Manifold Approximation and Projection (UMAP) (McInnes et al., 2018); t-distributed stochastic neighbor embedding (tSNE) (Van Der Maaten & Hinton, 2008). Clustering: Louvain algorithm (Blondel et al., 2008); Leiden algorithm (Traag et al., 2019).

compressed into a series of signals, called latent dimensions, that best represent the structure and preserve the variation of the original count data (**Figure 1.3**). While a scRNAseq dataset might measure over 10,000 genes, this is reduced to approximately 10-30 latent dimensions during dimensionality reduction. Principal component analysis (PCA) and similar algorithms are dominant approaches for this process. Dimensionality reduction provides several benefits for scRNAseq pipelines, but primarily serves to reduce complexity and noise within the data, making it more amenable to downstream analysis tasks (many of which are computationally expensive). Additionally, batch correction is often paired with dimensionality reduction, such that the resulting latent space represents cells in an integrated embedding (reviewed in Luecken et al., 2020). After dimensionality reduction, clustering is applied to the latent space to group cells with similar transcriptional signatures, providing cells with discrete categorical labels (**Figure 1.3**). The Louvain and Leiden algorithms are widely used (Blondel et al., 2008; Traag et al., 2019), which identify communities within graph-based representations of the latent space, but a plethora of clustering methods have been developed that may be more appropriate for different datasets. Visualization methods utilize the same graph-based representation of the latent space as clustering algorithms in order to build 2-dimensional embeddings of the data (**Figure 1.3**). Nearly all scRNAseq publications use either the tSNE or UMAP algorithms to visualize cluster labels and

gene expression patterns within the data (Van Der Maaten & Hinton, 2008; McInnes et al., 2018). Lastly, differential expression testing, which identifies genes with statistically-significant expression differences between cell populations (such as clusters), is a crucial component of the scRNAseq analysis toolkit (**Figure 1.3**). Popular strategies include the Wilcoxon Rank Sum test applied to the normalized count matrix or methods originally designed for bulk RNA sequencing data, such as edgeR and DESeq2 (Robinson et al., 2010; Love et al., 2014), that have been adapted for scRNAseq data (Dal Molin et al., 2017; Soneson & Robinson, 2018). Notably, Seurat and Scanpy are two of the most popular scRNAseq packages (Wolf et al., 2018; Hao et al., 2021), offering well-documented end-to-end analysis for performing the common scRNAseq tasks along with convenient interfaces for alternative and non-standard analysis methods.

The field of single-cell sequencing is rich with computational packages for efficient data analysis, and notably, machine learning methods have also gained significant prominence. The fusion of scRNAseq and machine learning is perhaps best exemplified by deep-learning software scvi-tools (Lopez et al., 2018; Gayoso et al., 2021). The base model of the software, scVI, uses a variational autoencoder (VAE) to model scRNAseq data (Kingma & Welling, 2013) (**Figure 1.4**). The VAE is composed of two neural networks, the first (the encoder) which compresses the input single-cell data and the second (the decoder) which attempts to reconstruct the original data from the compression. The scVI model is trained over hundreds of iterations, termed epochs, in order to find optimal parameters for the encoder and decoder via a process called gradient descent. At first glance, this may seem like a Sisyphean effort, compressing and then reconstructing the data. But the overall architecture of the VAE is adaptable to many common scRNAseq analysis tasks. First, to find an informative compression of the data, the VAE must identify important signals within the data which form a latent space. Unlike a latent space constructed with PCA, which

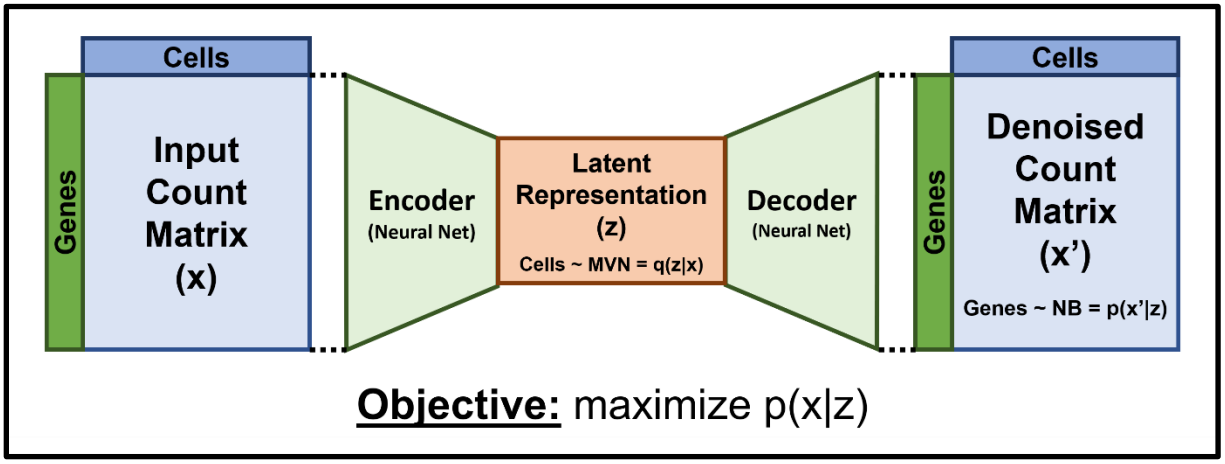

Figure 1.4. Simplified view of scVI's variational autoencoder (VAE) architecture. scVI (Lopez et al., 2018; Gayoso et al., 2021) compresses and then reconstructs scRNAseq data using a VAE (Kingma & Welling, 2013). Two neural networks, the encoder and decoder, link the input data (x), latent representation (z), and reconstructed data (x'). The model creates probabilistic representations of the data, useful for inferring values (such as denoised counts) while quantifying the uncertainty in those values due to noise. Within the latent representation (z), cells are modeled by multivariate normal (MVN) distributions conditioned on the input data (x), formalized as q(z|x). The reconstructed data also contains probabilistic representations, namely that genes are modeled by negative binomial (NB) distributions with parameters estimated by the decoder given the latent representation, formalized as p(x'|z). Optimization is performed via gradient descent to maximize p(x|z), the probability of reconstructing the input data from the model with its learned parameters.

attempts a linear decomposition of the input data, the VAE architecture can model more complex (i.e., non-linear) relationships within single-cell data. Second, because the compression of data into a latent space minimizes noise, the output of the decoder is a denoised reconstruction of the original data (**Figure 1.4**). Given the noise and low-capture efficiency problems of scRNAseq, data denoising or imputation is a desired task that is inherently performed by scVI. Third, the compression performed by the VAE model creates a probabilistic representation of each cell (in some sense, the probability that a cell exists in different transcriptional states), namely that each cell is mapped to a multivariate normal distribution with parameters learned by the model (**Figure 1.4**). Importantly, this probabilistic representation of cells is amenable to tasks such as computational resampling of cells, differential expression, and calculating gene correlations. Fourth, batch information and unwanted biological signals can be provided to the scvi-tools model, enabling the model to correct for these effects. Thus, scVI provides a single model for performing many crucial analysis tasks while minimizing the effects of batch and other unwanted covariates. While scvi-tools was initially designed for analyzing only scRNAseq data, the powerful VAE architecture has been adapted to many types of data, including CITE-seq, scATACseq, and conjoined scRNAseq + scATACseq.

Summary of dissertation research

By providing a comprehensive view of a tissue, scRNAseq is an ideal approach for understanding organ development. In this dissertation, my colleagues and I applied scRNAseq to better understand the cellular heterogeneity and heterotypic interactions that underlie wing disc development under both normal and regenerative conditions.

In Chapter 2 of this thesis, we assayed the single-cell composition of the developing wing disc at 96 and 120 hours after egg lay, corresponding to mid and late third instar development. While the canonical domains within the disc epithelium (notum, hinge, pouch, and peripodial epithelium) were well-preserved and easily identifiable at both developmental time points, we found that many differentially-expressed genes between the direct and indirect AMPs were only distinguishable at late third instar development, supporting the hypothesis that the AMPs are a largely naïve population throughout much of development. From our data, we identified and characterized two epithelial-to-myoblast interactions operating via the fibroblast growth factor (FGF) and Hedgehog signaling pathways. While FGF signaling controls AMP numbers along the wing disc and restricts these cells to the notum region, Hedgehog signaling is necessary to establish a unique cell identity within posterior-localized AMPs that is necessary for proper development of adult muscle structures.

Chapter 3 of this dissertation builds upon our developing wing disc data by adding scRNAseq data collected from regenerating wing discs. By harmonizing our developing and regenerating datasets, we identified two regeneration-specific cell clusters derived from the blastema which we termed Blastema1 and Blastema2. Both clusters expressed known blastema markers, but Blastema1 was particularly distinguishable for its additional expression of genes encoding secreted molecules. We characterized both of these clusters within the tissue, and identified Blastema1 and Blastema2 as representing inner and outer regions of the blastema, respectively, at 24 hours into regeneration. By searching our dataset for transcription factors that may regulate regenerative programs, we identified the gene *Ets21C* as a novel marker of the wing disc blastema. We provide evidence that *Ets21C* regulates a regeneration-specific program; while *Ets21C* mutant animals are healthy and fertile, they display severely compromised regenerative capabilities. *Ets21C* mutant larvae with damaged wing discs undergo pupariation prematurely, and

at 24 hours into the regenerative process, we find decreased levels of *Insulin-like peptide 8 (Ilp8)* and elimination of Blastema1 markers within the wing disc. Furthermore, we identify blastema-like cells within tumorous wing disc scRNAseq data, suggesting that tumors may co-opt regenerative programs to stimulate overgrowth.

Overall, the work described in this thesis builds a foundation for future investigations into how cells respond to damage and cell-cell signaling events that coordinate growth control. Furthermore, this research highlights the power of single-cell analysis in advancing our knowledge of biological processes. Indeed, the data collected for these studies is still rich with untapped insight, and can further be used to ignite exciting research opportunities.

REFERENCES

AlJanahi, A. A., Danielsen, M., & Dunbar, C. E. (2018). An Introduction to the Analysis of Single-Cell RNA-Sequencing Data. Molecular Therapy - Methods & Clinical Development, 10, 189–196. https://doi.org/10.1016/J.OMTM.2018.07.003

Almanzar, N., Antony, J., Baghel, A. S., Bakerman, I., Bansal, I., Barres, B. A., Beachy, P. A., Berdnik, D., Bilen, B., Brownfield, D., Cain, C., Chan, C. K. F., Chen, M. B., Clarke, M. F., Conley, S. D., Darmanis, S., Demers, A., Demir, K., de Morree, A., … Zou, J. (2020). A single-cell transcriptomic atlas characterizes ageing tissues in the mouse. Nature 2020 583:7817, 583(7817), 590–595. https://doi.org/10.1038/s41586-020-2496-1

Ariss, M. M., Islam, A. B. M. M. K., Critcher, M., Zappia, M. P., & Frolov, M. V. (2018). Single cell RNA-sequencing identifies a metabolic aspect of apoptosis in Rbf mutant. Nature Communications, 9(1), 5024. https://doi.org/10.1038/s41467-018-07540-z

Bageritz, J., Willnow, P., Valentini, E., Leible, S., Boutros, M., & Teleman, A. A. (2019). Gene expression atlas of a developing tissue by single cell expression correlation analysis. Nature Methods, 16(8), 750–756. https://doi.org/10.1038/s41592-019-0492-x

Bate, M. (1993). The Mesoderm and its Derivatives. In M. Bate & A. Martinez Arias (Eds.), The Development of Drosophila melanogaster (pp. 1013–1090). Cold Spring Harbor Laboratory Press.

Bate, M., & Martinez Arias, A. (1991). The embryonic origin of imaginal discs in Drosophila. In Development (Vol. 112).

Bate, M., Rushton, E., & Currie, D. A. (1991). Cells with persistent twist expression are the embryonic precursors of adult muscles in Drosophila. Development (Cambridge, England), 113(1), 79–89. http://www.ncbi.nlm.nih.gov/pubmed/1765010

Bejsovec, A. (2018). Wingless Signaling: A Genetic Journey from Morphogenesis to Metastasis. Genetics, 208(4), 1311–1336. https://doi.org/10.1534/genetics.117.300157

Bergantiños, C., Corominas, M., & Serras, F. (2010). Cell death-induced regeneration in wing imaginal discs requires JNK signalling. Development, 137(7), 1169–1179. https://doi.org/10.1242/DEV.045559

Bilder, D., & Irvine, K. D. (2017). Taking stock of the Drosophila research ecosystem. In Genetics (Vol. 206, Issue 3, pp. 1227–1236). Genetics Society of America. https://doi.org/10.1534/genetics.117.202390

Blair, S. S. (1993). Mechanisms of compartment formation: evidence that non-proliferating cells do not play a critical role in defining the D/V lineage restriction in the developing wing of Drosophila. Development, 119(2), 339–351. https://doi.org/10.1242/DEV.119.2.339

Blondel, V. D., Guillaume, J. L., Lambiotte, R., & Lefebvre, E. (2008). Fast unfolding of communities in large networks. Journal of Statistical Mechanics: Theory and Experiment, 2008(10). https://doi.org/10.1088/1742-5468/2008/10/p10008

Brand, A. H., & Perrimon, N. (1993). Targeted gene expression as a means of altering cell fates and generating dominant phenotypes. Development, 118(2), 401–415. https://doi.org/10.1242/DEV.118.2.401

Brockes, J. P., & Kumar, A. (2008). Comparative Aspects of Animal Regeneration. Annual Review of Cell and Developmental Biology, 24, 525–549. https://doi.org/10.1146/ANNUREV.CELLBIO.24.110707.175336

Buenrostro, J. D., Giresi, P. G., Zaba, L. C., Chang, H. Y., & Greenleaf, W. J. (2013). Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. Nature Methods, 10(12), 1213–1218. https://doi.org/10.1038/NMETH.2688

Buenrostro, J. D., Wu, B., Litzenburger, U. M., Ruff, D., Gonzales, M. L., Snyder, M. P., Chang, H. Y., & Greenleaf, W. J. (2015). Single-cell chromatin accessibility reveals principles of regulatory variation. Nature 2015 523:7561, 523(7561), 486–490. https://doi.org/10.1038/nature14590

Buettner, F., Natarajan, K. N., Casale, F. P., Proserpio, V., Scialdone, A., Theis, F. J., Teichmann, S. A., Marioni, J. C., & Stegle, O. (2015). Computational analysis of cell-to-cell heterogeneity in single-cell RNA-sequencing data reveals hidden subpopulations of cells. Nature Biotechnology 2014 33:2, 33(2), 155–160. https://doi.org/10.1038/nbt.3102

Cao, J., O'Day, D. R., Pliner, H. A., Kingsley, P. D., Deng, M., Daza, R. M., Zager, M. A., Aldinger, K. A., Blecher-Gonen, R., Zhang, F., Spielmann, M., Palis, J., Doherty, D., Steemers, F. J., Glass, I. A., Trapnell, C., & Shendure, J. (2020). A human cell atlas of fetal gene expression. Science, 370(6518). https://doi.org/10.1126/SCIENCE.ABA7721

Cao, J., Spielmann, M., Qiu, X., Huang, X., Ibrahim, D. M., Hill, A. J., Zhang, F., Mundlos, S., Christiansen, L., Steemers, F. J., Trapnell, C., & Shendure, J. (2019). The single-cell transcriptional landscape of mammalian organogenesis. Nature, 566(7745), 496–502. https://doi.org/10.1038/s41586-019-0969-x

Carlson, B. M. (2007). Principles of Regenerative Biology. Elsevier. https://books.google.com/books?hl=en&lr=&id=f_Epd5lHTNkC&oi=fnd&pg=PP1&dq=Principles+of+regenerative+biology&ots=1knA7t0XUB&sig=jUX8U9-UsRAPzAjns7t29vlcvXY#v=onepage&q=Principles of regenerative biology&f=false

Cauli, B., Porter, J. T., Tsuzuki, K., Lambolez, B., Rossier, J., Quenet, B., & Audinat, E. (2000). Classification of fusiform neocortical interneurons based on unsupervised clustering. Proceedings of the National Academy of Sciences, 97(11), 6144–6149. https://doi.org/10.1073/PNAS.97.11.6144

Chan, S. W., Lim, C. J., Chen, L., Chong, Y. F., Huang, C., Song, H., & Hong, W. (2011). The hippo pathway in biological control and cancer development. Journal of Cellular Physiology, 226(4), 928–939. https://doi.org/10.1002/JCP.22435

Cohen, S. M. (1993). Imaginal disc development. In M. Bate & A. Martinez Arias (Eds.), The Development of Drosophila melanogaster (pp. 747–841). Cold Spring Harbor Laboratory Press.

Colombani, J., Andersen, D. S., Boulan, L., Boone, E., Romero, N., Virolle, V., Texada, M., & Léopold, P. (2015). Drosophila Lgr3 Couples Organ Growth with Maturation and Ensures Developmental Stability. Current Biology, 25(20), 2723–2729. https://doi.org/10.1016/J.CUB.2015.09.020

Couturier, C. P., Ayyadhury, S., Le, P. U., Nadaf, J., Monlong, J., Riva, G., Allache, R., Baig, S., Yan, X., Bourgey, M., Lee, C., Wang, Y. C. D., Wee Yong, V., Guiot, M. C., Najafabadi, H., Misic, B., Antel, J., Bourque, G., Ragoussis, J., & Petrecca, K. (2020). Single-cell RNA-seq reveals that glioblastoma recapitulates a normal neurodevelopmental hierarchy. Nature Communications 2020 11:1, 11(1), 1–19. https://doi.org/10.1038/s41467-020-17186-5

Cusanovich, D. A., Daza, R., Adey, A., Pliner, H. A., Christiansen, L., Gunderson, K. L., Steemers, F. J., Trapnell, C., & Shendure, J. (2015). Multiplex single-cell profiling of chromatin accessibility by combinatorial cellular indexing. Science, 348(6237), 910–914. https://doi.org/10.1126/SCIENCE.AAB1601

Dal Molin, A., Baruzzo, G., & Di Camillo, B. (2017). Single-cell RNA-sequencing: Assessment of differential expression analysis methods. Frontiers in Genetics, 8(MAY), 62. https://doi.org/10.3389/FGENE.2017.00062/BIBTEX

Davie, K., Janssens, J., Koldere, D., De Waegeneer, M., Pech, U., Kreft, Ł., Aibar, S., Makhzami, S., Christiaens, V., Bravo González-Blas, C., Poovathingal, S., Hulselmans, G., Spanier, K. I., Moerman, T., Vanspauwen, B., Geurs, S., Voet, T., Lammertyn, J., Thienpont, B., … Aerts, S. (2018). A Single-Cell Transcriptome Atlas of the Aging Drosophila Brain. Cell, 174(4), 982-998.e20. https://doi.org/10.1016/J.CELL.2018.05.057

Deak, I. I. (1977). A histochemical study of the muscles of Drosophila melanogaster. Journal of Morphology, 153(2), 307–316. https://doi.org/10.1002/JMOR.1051530209

Diaz-Benjumea, F. J., & Cohen, S. M. (1993). Interaction between dorsal and ventral cells in the imaginal disc directs wing development in Drosophila. Cell, 75(4), 741–752. https://doi.org/10.1016/0092-8674(93)90494-B

Diaz-Benjumea, F. J., & Cohen, S. M. (1995). Serrate signals through Notch to establish a Wingless-dependent organizer at the dorsal/ventral compartment boundary of the Drosophila wing. Development, 121(12), 4215–4225. https://doi.org/10.1242/DEV.121.12.4215

Dobin, A., Davis, C. A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., & Gingeras, T. R. (2013). STAR: ultrafast universal RNA-seq aligner. Bioinformatics, 29(1), 15–21. https://doi.org/10.1093/BIOINFORMATICS/BTS635

Dutta, D., & Clevers, H. (2017). Organoid culture systems to study host–pathogen interactions. Current Opinion in Immunology, 48, 15–22. https://doi.org/10.1016/J.COI.2017.07.012

Everetts, N. J., Worley, M. I., Yasutomi, R., Yosef, N., & Hariharan, I. K. (2021). Single-cell transcriptomics of the Drosophila wing disc reveals instructive epithelium-to-myoblast interactions. ELife, 10. https://doi.org/10.7554/eLife.61276

Fan, Y., Wang, S., Hernandez, J., Yenigun, V. B., Hertlein, G., Fogarty, C. E., Lindblad, J. L., & Bergmann, A. (2014). Genetic Models of Apoptosis-Induced Proliferation Decipher Activation of JNK and Identify a Requirement of EGFR Signaling for Tissue Regenerative Responses in Drosophila. PLOS Genetics, 10(1), e1004131. https://doi.org/10.1371/JOURNAL.PGEN.1004131

Fernandes, J., Bate, M., & Vijayraghavan, K. (1991). Development of the indirect flight muscles of Drosophila. Development (Cambridge, England), 113(1), 67–77. http://www.ncbi.nlm.nih.gov/pubmed/1765009

Fischer, J. A., Giniger, E., Maniatis, T., & Ptashne, M. (1988). GAL4 activates transcription in Drosophila. Nature 1988 332:6167, 332(6167), 853–856. https://doi.org/10.1038/332853a0

Garber, R. L., Kuroiwa, A., & Gehring, W. J. (1983). Genomic and cDNA clones of the homeotic locus Antennapedia in Drosophila. The EMBO Journal, 2(11), 2027. https://doi.org/10.1002/j.1460-2075.1983.tb01696.x

Garcia-Bellido, A., & Merriam, J. R. (1971). Parameters of the wing imaginal disc development of Drosophila melanogaster. Developmental Biology, 24(1), 61–87. https://doi.org/10.1016/0012-1606(71)90047-9

Garcia-Bellido, A., Ripoll, P., & Morata, G. (1973). Developmental Compartmentalisation of the Wing Disk of Drosophila. Nature New Biology, 245(147), 251–253. https://doi.org/10.1038/newbio245251a0

Garelli, A., Heredia, F., Casimiro, A. P., Macedo, A., Nunes, C., Garcez, M., Dias, A. R. M., Volonte, Y. A., Uhlmann, T., Caparros, E., Koyama, T., & Gontijo, A. M. (2015). Dilp8 requires the neuronal relaxin receptor Lgr3 to couple growth to developmental timing. Nature Communications 2015 6:1, 6(1), 1–14. https://doi.org/10.1038/ncomms9732

Gayoso, A., Lopez, R., Xing, G., Boyeau, P., Wu, K., Jayasuriya, M., Melhman, E., Langevin, M., Liu, Y., Samaran, J., Misrachi, G., Nazaret, A., Clivio, O., Xu, C., Ashuach, T., Lotfollahi, M., Svensson, V., Da Veiga Beltrame, E., Talavera-López, C., … Yosef, N. (2021). Scvi-Tools: a Library for Deep Probabilistic Analysis of Single-Cell Omics Data. BioRxiv, 2021.04.28.441833. https://doi.org/10.1101/2021.04.28.441833

Gibson, M. C., & Schubiger, G. (1999). Hedgehog is required for activation of engrailed during regeneration of fragmented Drosophila imaginal discs. Development, 126(8), 1591–1599. https://doi.org/10.1242/DEV.126.8.1591

Gunage, R. D., Reichert, H., & VijayRaghavan, K. (2014). Identification of a new stem cell population that generates Drosophila flight muscles. ELife, 3(August2014), 1–25. https://doi.org/10.7554/eLife.03126

Gustavson, E., Goldsborough, A. S., Ali, Z., & Kornberg, T. B. (1996). The Drosophila engrailed and invected Genes: Partners in Regulation, Expression and Function. Genetics, 142(3), 893–906. https://doi.org/10.1093/GENETICS/142.3.893

Hackney, J. F., Zolali-Meybodi, O., & Cherbas, P. (2012). Tissue Damage Disrupts Developmental Progression and Ecdysteroid Biosynthesis in Drosophila. PLOS ONE, 7(11), e49105. https://doi.org/10.1371/JOURNAL.PONE.0049105

Hadorn, E. (1965). Problems of determination and transdetermination. Brookhaven Symposia in Biology.

Hadorn, E., Bertani, G., & Gallera, J. (1949). Regulative capacity and field organization of male genital discs in Drosophila melanogaster. Roux' Archiv Fur Entwicklungsmechanik.

Hadorn, E., & Buck, D. (1962). On the differentiation of transplanted wing imaginal disc fragments of Drosophila melanogaster. Rev. Suisse Zool.

Hao, Y., Hao, S., Andersen-Nissen, E., Mauck, W. M., Zheng, S., Butler, A., Lee, M. J., Wilk, A. J., Darby, C., Zager, M., Hoffman, P., Stoeckius, M., Papalexi, E., Mimitou, E. P., Jain, J., Srivastava, A., Stuart, T., Fleming, L. M., Yeung, B., … Satija, R. (2021). Integrated analysis of multimodal single-cell data. Cell, 184(13), 3573-3587.e29. https://doi.org/10.1016/J.CELL.2021.04.048

Hariharan, I. K. (2015). Organ Size Control: Lessons from Drosophila. In Developmental Cell (Vol. 34, Issue 3, pp. 255–265). Cell Press. https://doi.org/10.1016/j.devcel.2015.07.012

Harris, R. E., Setiawan, L., Saul, J., & Hariharan, I. K. (2016). Localized epigenetic silencing of a damage-activated WNT enhancer limits regeneration in mature Drosophila imaginal discs. ELife, 5(FEBRUARY2016). https://doi.org/10.7554/ELIFE.11588

Harris, R. E., Stinchfield, M. J., Nystrom, S. L., McKay, D. J., & Hariharan, I. K. (2020). Damage-responsive, maturity-silenced enhancers regulate multiple genes that direct regeneration in drosophila. ELife, 9, 1–26. https://doi.org/10.7554/eLife.58305

Harvey, K. F., Pfleger, C. M., & Hariharan, I. K. (2003). The Drosophila Mst Ortholog, hippo, Restricts Growth and Cell Proliferation and Promotes Apoptosis. Cell, 114(4), 457–467. https://doi.org/10.1016/S0092-8674(03)00557-9

Harvey, K. F., Zhang, X., & Thomas, D. M. (2013). The Hippo pathway and human cancer. Nature Reviews Cancer 2013 13:4, 13(4), 246–257. https://doi.org/10.1038/NRC3458

Hashimshony, T., Senderovich, N., Avital, G., Klochendler, A., de Leeuw, Y., Anavy, L., Gennert, D., Li, S., Livak, K. J., Rozenblatt-Rosen, O., Dor, Y., Regev, A., & Yanai, I. (2016). CEL-Seq2: Sensitive highly-multiplexed single-cell RNA-Seq. Genome Biology, 17(1). https://doi.org/10.1186/s13059-016-0938-8

Hatori, R., & Kornberg, T. B. (2020). Hedgehog produced by the Drosophila wing imaginal disc induces distinct responses in three target tissues. Development (Cambridge, England), 147(22). https://doi.org/10.1242/dev.195974

Held Jr, L. I. (2002). Imaginal Discs: The Genetic and Cellular Logic of Pattern Formation. In Imaginal Discs. Cambridge University Press. https://doi.org/10.1017/CBO9780511529733

Herrera, S. C., & Morata, G. (2014). Transgressions of compartment boundaries and cell reprogramming during regeneration in Drosophila. ELife, 2014(3). https://doi.org/10.7554/ELIFE.01831.001

Hicks, S. C., Townes, F. W., Teng, M., & Irizarry, R. A. (2018). Missing data and technical variability in single-cell RNA-sequencing experiments. Biostatistics, 19(4), 562–578. https://doi.org/10.1093/BIOSTATISTICS/KXX053

Hotelling, H. (1933). Analysis of a complex of statistical variables into Principal Components. The Journal of Educational Psychology, 24, 417–441.

Islam, S., Zeisel, A., Joost, S., La Manno, G., Zajac, P., Kasper, M., Lönnerberg, P., & Linnarsson, S. (2013). Quantitative single-cell RNA-seq with unique molecular identifiers. Nature Methods 2013 11:2, 11(2), 163–166. https://doi.org/10.1038/nmeth.2772

Jaszczak, J. S., Wolpe, J. B., Bhandari, R., Jaszczak, R. G., & Halme, A. (2016). Growth coordination during Drosophila melanogaster imaginal disc regeneration is mediated by signaling through the relaxin receptor Lgr3 in the prothoracic gland. Genetics, 204(2), 703–709. https://doi.org/10.1534/GENETICS.116.193706/-/DC1

Justice, R. W., Zilian, O., Woods, D. F., Noll, M., & Bryant, P. J. (1995). The Drosophila tumor suppressor gene warts encodes a homolog of human myotonic dystrophy kinase and is required for the control of cell shape and proliferation. Genes & Development, 9(5), 534–546. https://doi.org/10.1101/GAD.9.5.534

Karaiskos, N., Wahle, P., Alles, J., Boltengagen, A., Ayoub, S., Kipar, C., Kocks, C., Rajewsky, N., & Zinzen, R. P. (2017). The Drosophila embryo at single-cell transcriptome resolution. Science (New York, N.Y.), 358(6360), 194–199. https://doi.org/10.1126/science.aan3235

Kiehle, C. P., & Schubiger, G. (1985). Cell proliferation changes during pattern regulation in imaginal leg discs of Drosophila melanogaster. Developmental Biology, 109(2), 336–346. https://doi.org/10.1016/0012-1606(85)90460-9

Kim, J., Sebring, A., Esch, J. J., Kraus, M. E., Vorwerk, K., Magee, J., & Carroll, S. B. (1996). Integration of positional signals and regulation of wing formation and identity by Drosophila vestigial gene. Nature 1996 382:6587, 382(6587), 133–138. https://doi.org/10.1038/382133a0

Kingma, D. P., & Welling, M. (2013). Auto-Encoding Variational Bayes. ArXiv. https://arxiv.org/abs/1312.6114v10

La Fortezza, M., Schenk, M., Cosolo, A., Kolybaba, A., Grass, I., & Classen, A. K. (2016). JAK/STAT signalling mediates cell survival in response to tissue stress. Development (Cambridge), 143(16), 2907–2919.

https://doi.org/10.1242/DEV.132340/264081/AM/JAK-STAT-SIGNALLING-MEDIATES-CELL-SURVIVAL-IN

Laurichesse, Q., & Soler, C. (2020). Muscle development : a view from adult myogenesis in Drosophila. https://doi.org/10.1016/j.semcdb.2020.02.009

Lawrence, P. A. (1982). Cell Lineage of the Thoracic Muscles of Drosophila. In Cell (Vol. 29).

Lawrence, P. A., & Brower, D. L. (1982). Myoblasts from Drosophila wing disks can contribute to developing muscles throughout the fly. Nature 1982 295:5844, 295(5844), 55–57. https://doi.org/10.1038/295055a0

Lee, D. D., & Seung, H. S. (1999). Learning the parts of objects by non-negative matrix factorization. Nature 1999 401:6755, 401(6755), 788–791. https://doi.org/10.1038/44565

Leek, J. T., Scharpf, R. B., Bravo, H. C., Simcha, D., Langmead, B., Johnson, W. E., Geman, D., Baggerly, K., & Irizarry, R. A. (2010). Tackling the widespread and critical impact of batch effects in high-throughput data. Nature Reviews Genetics 2010 11:10, 11(10), 733–739. https://doi.org/10.1038/nrg2825

Li, H., van der Leun, A. M., Yofe, I., Lubling, Y., Gelbard-Solodkin, D., van Akkooi, A. C. J., van den Braber, M., Rozeman, E. A., Haanen, J. B. A. G., Blank, C. U., Horlings, H. M., David, E., Baran, Y., Bercovich, A., Lifshitz, A., Schumacher, T. N., Tanay, A., & Amit, I. (2019). Dysfunctional CD8 T Cells Form a Proliferative, Dynamically Regulated Compartment within Human Melanoma. Cell, 176(4), 775-789.e18. https://doi.org/10.1016/J.CELL.2018.11.043

Lopez, R., Regier, J., Cole, M. B., Jordan, M. I., & Yosef, N. (2018). Deep generative modeling for single-cell transcriptomics. Nature Methods, 15(12), 1053–1058. https://doi.org/10.1038/s41592-018-0229-2

Love, M. I., Huber, W., & Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biology, 15(12), 1–21. https://doi.org/10.1186/S13059-014-0550-8/FIGURES/9

Luecken, M. D., Büttner, M., Chaichoompu, K., Danese, A., Interlandi, M., Mueller, M. F., Strobl, D. C., Zappia, L., Dugas, M., Colomé-Tatché, M., & Theis, F. J. (2020). Benchmarking atlas-level data integration in single-cell genomics. BioRxiv, 1, 2020.05.22.111161. https://doi.org/10.1101/2020.05.22.111161

Luecken, M. D., & Theis, F. J. (2019). Current best practices in single-cell RNA-seq analysis: a tutorial. Molecular Systems Biology, 15(6), e8746. https://doi.org/10.15252/MSB.20188746

Macosko, E. Z., Basu, A., Satija, R., Nemesh, J., Shekhar, K., Goldman, M., Tirosh, I., Bialas, A. R., Kamitaki, N., Martersteck, E. M., Trombetta, J. J., Weitz, D. A., Sanes, J. R., Shalek, A. K., Regev, A., & McCarroll, S. A. (2015). Highly Parallel Genome-wide Expression Profiling of Individual Cells Using Nanoliter Droplets. Cell, 161(5), 1202–1214. https://doi.org/10.1016/J.CELL.2015.05.002

Madhavan, M., & Schneiderman, H. A. (1977). Histological analysis of the dynamics of growth of imaginal discs and histoblast nests during the larval development of Drosophila melanogaster. Wilhelm Roux's Archives of Developmental Biology, 183(4), 269–305. https://doi.org/10.1007/BF00848459

Mann, H. B., & Whitney, D. R. (1947). On a Test of Whether one of Two Random Variables is Stochastically Larger than the Other. Https://Doi-Org.Libproxy.Berkeley.Edu/10.1214/Aoms/1177730491, 18(1), 50–60. https://doi.org/10.1214/AOMS/1177730491

Martín, F. A., Herrera, S. C., & Morata, G. (2009). Cell competition, growth and size control in the Drosophila wing imaginal disc. Development, 136(22), 3747–3756. https://doi.org/10.1242/dev.038406

Martin, J. C., Chang, C., Boschetti, G., Ungaro, R., Giri, M., Grout, J. A., Gettler, K., Chuang, L. shiang, Nayar, S., Greenstein, A. J., Dubinsky, M., Walker, L., Leader, A., Fine, J. S., Whitehurst, C. E., Mbow, M. L., Kugathasan, S., Denson, L. A., Hyams, J. S., … Kenigsberg, E. (2019). Single-Cell Analysis of Crohn's Disease Lesions Identifies a Pathogenic Cellular Module Associated with Resistance to Anti-TNF Therapy. Cell, 178(6), 1493-1508.e20. https://doi.org/10.1016/J.CELL.2019.08.008

McGinnis, W., Garber, R. L., Wirz, J., Kuroiwa, A., & Gehring, W. J. (1984). A homologous protein-coding sequence in drosophila homeotic genes and its conservation in other metazoans. Cell, 37(2), 403–408. https://doi.org/10.1016/0092-8674(84)90370-2

McInnes, L., Healy, J., & Melville, J. (2018). UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction. ArXiv E-Prints. http://arxiv.org/abs/1802.03426

Melsted, P., Booeshaghi, A. S., Gao, F., Beltrame, E., Lu, L., Eldjárn, K., 4, H., Gehring, J., & Pachter, L. (2019). Modular and efficient pre-processing of single-cell RNA-seq. BioRxiv, 673285. https://doi.org/10.1101/673285

Mumm, J. S., & Kopan, R. (2000). Notch Signaling: From the Outside In. Developmental Biology, 228(2), 151–165. https://doi.org/10.1006/DBIO.2000.9960

Narciso, C., Wu, Q., Brodskiy, P., Garston, G., Baker, R., Fletcher, A., & Zartman, J. (2015). Patterning of wound-induced intercellular Ca2+ flashes in a developing epithelium. Physical Biology, 12(5), 056005. https://doi.org/10.1088/1478-3975/12/5/056005

Neto-Silva, R. M., Wells, B. S., & Johnston, L. A. (2009). Mechanisms of Growth and Homeostasis in the Drosophila Wing. Annual Review of Cell and Developmental Biology, 25(1), 197–220. https://doi.org/10.1146/annurev.cellbio.24.110707.175242

Neumann, C. J., & Cohen, S. M. (1997). Long-range action of Wingless organizes the dorsal-ventral axis of the Drosophila wing. Development, 124(4), 871–880. https://doi.org/10.1242/DEV.124.4.871

Ng, M., Diaz-Benjumea, F. J., Vincent, J. P., Wu, J., & Cohen, S. M. (1996). Specification of the wing by localized expression of wingless protein. Nature 1996 381:6580, 381(6580), 316–318. https://doi.org/10.1038/381316a0

O'Brochta, D. A., & Bryant, P. J. (1987). Distribution of S-phase cells during the regeneration of Drosophila imaginal wing discs. Developmental Biology, 119(1), 137–142. https://doi.org/10.1016/0012-1606(87)90215-6

Orimo, A., & Weinberg, R. A. (2006). Stromal Fibroblasts in Cancer: A Novel Tumor-Promoting Cell Type. Cell Cycle, 5(15), 1597–1601. https://doi.org/10.4161/cc.5.15.3112

Paatero, P., & Tapper, U. (1994). Positive matrix factorization: A non-negative factor model with optimal utilization of error estimates of data values. Environmetrics, 5(2), 111–126. https://doi.org/10.1002/ENV.3170050203

Pappalardo, J. L., Zhang, L., Pecsok, M. K., Perlman, K., Zografou, C., Raddassi, K., Abulaban, A., Krishnaswamy, S., Antel, J., van Dijk, D., & Hafler, D. A. (2020). Transcriptomic and clonal characterization of T cells in the human central nervous system. Science Immunology, 5(51), 8786. https://doi.org/10.1126/SCIIMMUNOL.ABB8786

Pastor-Pareja, J. C., Ming, W., & Tian, X. (2008). An innate immune response of blood cells to tumors and tissue damage in Drosophila. Disease Models & Mechanisms, 1(2–3), 144–154. https://doi.org/10.1242/DMM.000950

Pearson, K. (1901). On lines and planes of closest fit to systems of points in space. Https://Doi.Org/10.1080/14786440109462720, 2(11), 559–572. https://doi.org/10.1080/14786440109462720

Picelli, S., Björklund, Å. K., Faridani, O. R., Sagasser, S., Winberg, G., & Sandberg, R. (2013). Smart-seq2 for sensitive full-length transcriptome profiling in single cells. Nature Methods, 10(11), 1096–1100. https://doi.org/10.1038/nmeth.2639

Ransick, A., Lindström, N. O., Liu, J., Zhu, Q., Guo, J. J., Alvarado, G. F., Kim, A. D., Black, H. G., Kim, J., & McMahon, A. P. (2019). Single-Cell Profiling Reveals Sex, Lineage, and Regional Diversity in the Mouse Kidney. Developmental Cell, 51(3), 399-413.e7. https://doi.org/10.1016/J.DEVCEL.2019.10.005

Ren, X., Wen, W., Fan, X., Hou, W., Su, B., Cai, P., Li, J., Liu, Y., Tang, F., Zhang, F., Yang, Y., He, J., Ma, W., He, J., Wang, P., Cao, Q., Chen, F., Chen, Y., Cheng, X., … Zhang, Z. (2021). COVID-19 immune features revealed by a large-scale single-cell transcriptome atlas. Cell, 184(7), 1895-1913.e19. https://doi.org/10.1016/J.CELL.2021.01.053

Repiso, A., Bergantiños, C., & Serras, F. (2013). Cell fate respecification and cell division orientation drive intercalary regeneration in Drosophila wing discs. Development, 140(17), 3541–3551. https://doi.org/10.1242/DEV.095760

Requena, D., Álvarez, J. A., Gabilondo, H., Loker, R., Mann, R. S., & Estella, C. (2017). Origins and Specification of the Drosophila Wing. Current Biology, 27(24), 3826-3836.e5. https://doi.org/10.1016/j.cub.2017.11.023

Restrepo, S., & Basler, K. (2016). Drosophila wing imaginal discs respond to mechanical injury via slow InsP3R-mediated intercellular calcium waves. Nature Communications 2016 7:1, 7(1), 1–9. https://doi.org/10.1038/ncomms12450

Robinson, M. D., McCarthy, D. J., & Smyth, G. K. (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics, 26(1), 139–140. https://doi.org/10.1093/BIOINFORMATICS/BTP616

Roy, S., & VijayRaghavan, K. (1999). Muscle pattern diversification in Drosophila: the story of imaginal myogenesis. BioEssays, 21(6), 486–498. https://doi.org/10.1002/(SICI)1521-1878(199906)21:6<486::AID-BIES5>3.0.CO;2-M

Sarkar, A., & Stephens, M. (2021). Separating measurement and expression models clarifies confusion in single-cell RNA sequencing analysis. Nature Genetics. https://doi.org/10.1038/s41588-021-00873-4

Schaum, N., Karkanias, J., Neff, N. F., May, A. P., Quake, S. R., Wyss-Coray, T., Darmanis, S., Batson, J., Botvinnik, O., Chen, M. B., Chen, S., Green, F., Jones, R. C., Maynard, A., Penland, L., Pisco, A. O., Sit, R. V., Stanley, G. M., Webber, J. T., … Weissman, I. L. (2018). Single-cell transcriptomics of 20 mouse organs creates a Tabula Muris. Nature 2018 562:7727, 562(7727), 367–372. https://doi.org/10.1038/s41586-018-0590-4

Schirmer, L., Velmeshev, D., Holmqvist, S., Kaufmann, M., Werneburg, S., Jung, D., Vistnes, S., Stockley, J. H., Young, A., Steindel, M., Tung, B., Goyal, N., Bhaduri, A., Mayer, S., Engler, J. B., Bayraktar, O. A., Franklin, R. J. M., Haeussler, M., Reynolds, R., … Rowitch, D. H. (2019). Neuronal vulnerability and multilineage diversity in multiple sclerosis. Nature 2019 573:7772, 573(7772), 75–82. https://doi.org/10.1038/s41586-019-1404-z

Scott, M. P., & Weiner, A. J. (1984). Structural relationships among genes that control development: sequence homology between the Antennapedia, Ultrabithorax, and fushi tarazu loci of Drosophila. Proceedings of the National Academy of Sciences, 81(13), 4115–4119. https://doi.org/10.1073/PNAS.81.13.4115

Seto, E. S., & Bellen, H. J. (2004). The ins and outs of Wingless signaling. Trends in Cell Biology, 14(1), 45–53. https://doi.org/10.1016/J.TCB.2003.11.004

Sinha, S., Rosin, N. L., Arora, R., Labit, E., Jaffer, A., Cao, L., Farias, R., Nguyen, A. P., N de Almeida, L. G., Dufour, A., Bromley, A., McDonald, B., Gillrie, M. R., Fritzler, M. J., Yipp, B. G., & Biernaskie, J. (2021). Dexamethasone modulates immature neutrophils and interferon programming in severe COVID-19. Nature Medicine 2021, 1–11. https://doi.org/10.1038/s41591-021-01576-3

Smith-Bolton, R. K., Worley, M. I., Kanda, H., & Hariharan, I. K. (2009). Regenerative Growth in Drosophila Imaginal Discs Is Regulated by Wingless and Myc. Developmental Cell, 16(6), 797–809. https://doi.org/10.1016/J.DEVCEL.2009.04.015

Soneson, C., & Robinson, M. D. (2018). Bias, robustness and scalability in single-cell differential expression analysis. Nature Methods 2018 15:4, 15(4), 255–261. https://doi.org/10.1038/nmeth.4612

Song, Y. H., Warncke, C., Choi, S. J., Choi, S., Chiou, A. E., Ling, L., Liu, H. Y., Daniel, S., Antonyak, M. A., Cerione, R. A., & Fischbach, C. (2017). Breast cancer-derived extracellular vesicles stimulate myofibroblast differentiation and pro-angiogenic behavior of adipose stem cells. Matrix Biology, 60–61, 190–205. https://doi.org/10.1016/J.MATBIO.2016.11.008

Stoeckius, M., Hafemeister, C., Stephenson, W., Houck-Loomis, B., Chattopadhyay, P. K., Swerdlow, H., Satija, R., & Smibert, P. (2017). Simultaneous epitope and transcriptome measurement in single cells. Nature Methods 2017 14:9, 14(9), 865–868. https://doi.org/10.1038/nmeth.4380

Sudarsan, V., Anant, S., Guptan, P., VijayRaghavan, K., & Skaer, H. (2001). Myoblast diversification and ectodermal signaling in Drosophila. Developmental Cell, 1(6), 829–839. https://doi.org/10.1016/s1534-5807(01)00089-2

Svensson, V. (2020). Droplet scRNA-seq is not zero-inflated. Nature Biotechnology 2020 38:2, 38(2), 147–150. https://doi.org/10.1038/s41587-019-0379-5

Svensson, V., Beltrame, E. da V., & Pachter, L. (2019). Quantifying the tradeoff between sequencing depth and cell number in single-cell RNA-seq. BioRxiv, 762773. https://doi.org/10.1101/762773

Svensson, V., da Veiga Beltrame, E., & Pachter, L. (2020). A curated database reveals trends in single-cell transcriptomics. Database, 2020. https://doi.org/10.1093/DATABASE/BAAA073

Swarup, S., & Verheyen, E. M. (2012). Wnt/wingless signaling in drosophila. Cold Spring Harbor Perspectives in Biology, 4(6), 1–15. https://doi.org/10.1101/cshperspect.a007930

Tabata, T., Schwartz, C., Gustavson, E., Ali, Z., & Kornberg, T. B. (1995). Creating a Drosophila wing de novo, the role of engrailed, and the compartment border hypothesis. Development, 121(10), 3359–3369. https://doi.org/10.1242/DEV.121.10.3359

Tanimoto, H., Itoh, S., Ten Dijke, P., & Tabata, T. (2000). Hedgehog creates a gradient of DPP activity in Drosophila wing imaginal discs. Molecular Cell, 5(1), 59–71. https://doi.org/10.1016/S1097-2765(00)80403-7

Tapon, N., Harvey, K. F., Bell, D. W., Wahrer, D. C. R., Schiripo, T. A., Haber, D. A., & Hariharan, I. K. (2002). salvador Promotes Both Cell Cycle Exit and Apoptosis in Drosophila and Is Mutated in Human Cancer Cell Lines. Cell, 110(4), 467–478. https://doi.org/10.1016/S0092-8674(02)00824-3

Tietjen, I., Rihel, J. M., Cao, Y., Koentges, G., Zakhary, L., & Dulac, C. (2003). Single-cell transcriptional analysis of neuronal progenitors. Neuron, 38(2), 161–175. https://doi.org/10.1016/S0896-6273(03)00229-0

Traag, V. A., Waltman, L., & van Eck, N. J. (2019). From Louvain to Leiden: guaranteeing well-connected communities. Scientific Reports 2019 9:1, 9(1), 1–12. https://doi.org/10.1038/s41598-019-41695-z

Vallejo, D. M., Juarez-Carreño, S., Bolivar, J., Morante, J., & Dominguez, M. (2015). A brain circuit that synchronizes growth and maturation revealed through Dilp8 binding to Lgr3. Science, 350(6262). https://doi.org/10.1126/science.aac6767

Van Der Maaten, L., & Hinton, G. (2008). Visualizing Data using t-SNE. Journal of Machine Learning Research, 9, 2579–2605.

Wagner, A., Regev, A., & Yosef, N. (2016). Revealing the vectors of cellular identity with single-cell genomics. Nature Biotechnology 2016 34:11, 34(11), 1145–1160. https://doi.org/10.1038/nbt.3711

Wan, A. C. A. (2016). Recapitulating Cell–Cell Interactions for Organoid Construction – Are Biomaterials Dispensable? Trends in Biotechnology, 34(9), 711–721. https://doi.org/10.1016/J.TIBTECH.2016.02.015

Wilcoxon, F. (1945). Individual Comparisons by Ranking Methods. Biometrics Bulletin, 1(6), 80. https://doi.org/10.2307/3001968

Wolf, F. A., Angerer, P., & Theis, F. J. (2018). SCANPY: Large-scale single-cell gene expression data analysis. Genome Biology, 19(1), 1–5. https://doi.org/10.1186/S13059-017-1382-0/FIGURES/1

Worley, M. I., Alexander, L. A., & Hariharan, I. K. (2018). CtBP impedes JNK- and Upd/STAT-driven cell fate misspecifications in regenerating Drosophila imaginal discs. ELife, 7. https://doi.org/10.7554/ELIFE.30391

Worley, M. I., Setiawan, L., & Hariharan, I. K. (2013). TIE-DYE: a combinatorial marking system to visualize and genetically manipulate clones during development in Drosophila melanogaster. Development, 140(15), 3275–3284. https://doi.org/10.1242/dev.096057

Xu, T., Wang, W., Zhang, S., Stewart, R. A., & Yu, W. (1995). Identifying tumor suppressors in genetic mosaics: the Drosophila lats gene encodes a putative protein kinase. Development, 121(4), 1053–1063. https://doi.org/10.1242/DEV.121.4.1053

Yamashita, M., Ogawa, T., Zhang, X., Hanamura, N., Kashikura, Y., Takamura, M., Yoneda, M., & Shiraishi, T. (2012). Role of stromal myofibroblasts in invasive breast cancer: Stromal expression of alpha-smooth muscle actin correlates with worse clinical outcome. Breast Cancer, 19(2), 170–176. https://doi.org/10.1007/s12282-010-0234-5

Zappia, M. P., Castro, L., Ariss, M. M., Jefferson, H., Islam, A. B., & Frolov, M. V. (2020). A cell atlas of adult muscle precursors uncovers early events in fibre-type divergence in Drosophila. EMBO Reports, 21(10), e49555. https://doi.org/10.15252/embr.201949555

Zhang, X., Li, T., Liu, F., Chen, Y., Yao, J., Li, Z., Huang, Y., & Wang, J. (2019). Comparative Analysis of Droplet-Based Ultra-High-Throughput Single-Cell RNA-Seq Systems. Molecular Cell, 73(1), 130-142.e5. https://doi.org/10.1016/J.MOLCEL.2018.10.020

Zheng, G. X. Y., Terry, J. M., Belgrader, P., Ryvkin, P., Bent, Z. W., Wilson, R., Ziraldo, S. B., Wheeler, T. D., McDermott, G. P., Zhu, J., Gregory, M. T., Shuga, J., Montesclaros, L., Underwood, J. G., Masquelier, D. A., Nishimura, S. Y., Schnall-Levin, M., Wyatt, P. W., Hindson, C. M., … Bielas, J. H. (2017). Massively parallel digital transcriptional profiling of single cells. Nature Communications 2017 8:1, 8(1), 1–12. https://doi.org/10.1038/ncomms14049

Ziegenhain, C., Vieth, B., Parekh, S., Reinius, B., Guillaumet-Adkins, A., Smets, M., Leonhardt, H., Heyn, H., Hellmann, I., & Enard, W. (2017). Comparative Analysis of Single-Cell RNA Sequencing Methods. Molecular Cell, 65(4), 631-643.e4. https://doi.org/10.1016/j.molcel.2017.01.023

The following chapter was originally published as an article in
eLife 10 (2021): e61276

Chapter 2

Single-cell transcriptomics of the Drosophila wing disc reveals instructive epithelium-to-myoblast interactions

ABSTRACT

In both vertebrates and invertebrates, generating a functional appendage requires interactions between ectoderm-derived epithelia and mesoderm-derived cells. To investigate such interactions, we used single-cell transcriptomics to generate a temporal cell atlas of the *Drosophila* wing disc from two developmental time points. Using these data, we visualized gene expression using a multi-layered model of the wing disc and catalogued ligand-receptor pairs that could mediate signaling between epithelial cells and adult muscle precursors (AMPs). We found that localized expression of the FGF ligands, Thisbe and Pyramus, in the disc epithelium regulates the number and location of the AMPs. In addition, Hedgehog ligand from the epithelium activates a specific transcriptional program within adjacent AMP cells, defined by AMP-specific targets *Neurotactin* and *midline*, that is critical for proper formation of direct flight muscles. More generally, our annotated temporal cell atlas provides an organ-wide view of potential cell-cell interactions between epithelial and myogenic cells.

INTRODUCTION

The development of multicellular eukaryotes gives rise to organs that are composed of cells of many types, typically derived from different germ layers such as the ectoderm and the mesoderm. There is increasing evidence that signaling during development between these distinct cell types plays an important role in ensuring the appropriate identity and number of cells in the fully formed adult organ (Ribatti & Santoiemma, 2014). A particularly well-studied example of such heterotypic interactions occurs during the development of the vertebrate limb, where signals are exchanged between the apical ectodermal ridge and the underlying mesoderm (Delgado & Torres, 2017).

While vertebrate limbs are relatively complex structures, the *Drosophila* wing-imaginal disc, the larval primordium of the adult wing and thorax, is ideally suited to the study of cell-cell interactions in the context of organ development because of its relative simplicity and amenability to genetic analysis (Waddington, 1940; Cohen, 1993; Neto-Silva et al., 2009). The wing-imaginal disc is composed of epithelial cells that form a sac-like structure (comprised of the columnar cells of the disc proper and the squamous cells of the peripodial epithelium) and a population of adult muscle precursors (AMPs) that resides between the epithelial cells of the disc proper and the underlying basement membrane (**Figure 2.1A**). The epithelial portion of the disc derives from a primordium of approximately 30 cells from the embryonic ectoderm that are specified during embryogenesis (Madhavan & Schneiderman, 1977; Worley et al., 2013; Requena et al., 2017). The AMPs, originally referred to as adepithelial cells (Poodry & Schneiderman, 1970), represent a subset of cells from the embryonic mesoderm that generate the adult flight muscles (Bate et al., 1991; Fernandes et al., 1991). The AMPs underlie the dorsal portion of the wing disc epithelium, the notum, which is the primordium of the dorsal thorax. During metamorphosis, these AMPs generate multiple muscle fibers which comprise the direct and indirect flight muscles (DFMs and IFMs, respectively) (reviewed by Bothe & Baylies, 2016; Gunage et al., 2017; Laurichesse & Soler, 2020).

The mechanisms that influence a seemingly uniform population of AMPs to generate different types of flight muscles, each composed of multiple distinct fibers, are not known. The AMPs are generated by an asymmetric division of a muscle founder cell during embryogenesis; one daughter cell becomes an AMP while the other generates precursors of larval muscles (Bate et al., 1991). In the second thoracic segment, dorsal clusters of AMPs, which express a segment-specific combination of Hox genes (Roy & VijayRaghavan, 1997), become associated with the wing disc, remain in the notum region, and proliferate via symmetric cell divisions. At the onset of the third larval instar (L3), the AMP cells switch to a pattern of asymmetric cell division as a result of Wingless (Wg; Wnt ligand) and Notch signaling interactions. (Gunage et al., 2014). The precursors of the indirect and direct flight muscles can be distinguished by higher levels of expression of the transcriptional regulators Vestigial or Cut, respectively (Sudarsan et al., 2001). The elevated Vestigial expression in the IFM precursors is maintained by expression of Wg ligand from the notum epithelium (Sudarsan et al., 2001).

Two large questions central to the mechanisms that regulate proliferation and cell-fate specification in the AMPs, however, remain largely unanswered. First, what regulates the location and number of AMPs? It has been suggested that the notum acts as a "dynamic niche" that both regulates the survival of AMPs and guides their specification (Gunage et al., 2017). However, the signals emanating from the epithelial cells to either regulate AMP numbers or maintain them in the notum region of the wing disc have not been identified. Second, it is known that the AMPs

contribute to the formation of several complex muscle types: How are different AMP types specified? It has been proposed that extrinsic signals from the disc epithelium function during the larval stage to direct subsets of the AMPs to become precursors of specific types of muscles (Gunage et al., 2017). However, with the exception of Wg, such signals remain unidentified.

Single-cell transcriptomics provides a powerful paradigm for mapping the cellular composition of developing organs (Schier, 2020), including vertebrate appendages (Fabre et al., 2018; Cao et al., 2019; Feregrino et al., 2019), and has been successfully utilized for characterizing the *Drosophila* wing disc at single snapshots during development (Deng et al., 2019; Bageritz et al., 2019; Zappia et al., 2020). Beyond cataloging cell types, transcriptome-scale analysis of single cells opens the way for a comprehensive evaluation of how interactions between cells may facilitate development (Satija et al., 2015; Karaiskos et al., 2017; Deng et al., 2019; Bageritz et al., 2019). Since spatial information is lost in most prevalent scRNAseq protocols, computational methods are used to infer it, usually based on the expression of landmark genes (Satija et al., 2015; Karaiskos et al., 2017; Deng et al., 2019; Bageritz et al., 2019). Complementary to spatial localization are approaches that have examined the expression of receptors, ligands, and downstream molecules to predict which cell subsets interact and by what mechanisms (Vento-Tormo et al., 2018; Browaeys et al., 2020).

The combination of these approaches, namely estimating the physical context of each cell and investigating the expression of extracellular cues such as receptor-ligand pairs, provides a powerful tool for studying organ development. Here, we couple these two approaches to identify heterotypic interactions that are crucial for disc development, focusing on signaling between the disc epithelium and the AMPs. To this end, we collected single cell RNA-sequencing (scRNAseq) data from two developmental time points, derived a comprehensive view of cell subsets and their spatial organization, and examined the expression of receptors and ligands. We show that FGF ligands emanating from the disc epithelium create an AMP niche that regulates AMP numbers and restricts them to the region of the notum. Furthermore, we find that Hedgehog ligand from the disc epithelium specifies a unique subpopulation of AMPs and identified novel Hh-target genes, *Neurotactin (Nrt)* and *midline* (*mid)*, which are induced during the last phase of larval development. Beyond these examples, our annotated dataset provides a resource for spatiotemporal cellular composition in the developing wing disc and points to additional potential heterotypic interactions between epithelial cells and AMPs.

RESULTS

Generation of a temporal cell atlas of the developing wing-imaginal disc

The wing disc is composed of multiple cell types, including the columnar cells of the disc proper, the squamous cells of the peripodial epithelium, and the mesoderm-derived AMPs (**Figure 2.1A**). In addition, the wing disc is in intimate contact with branches of the tracheal system and circulating blood cells called hemocytes. With the goal of generating a spatiotemporal atlas of the developing wing disc, we used single-cell RNA sequencing to collect transcriptional profiles of cells at mid and late 3rd instar, which correspond to 96h and 120h after egg lay (AEL) (**Figure 2.1A**). Two biological replicates were obtained at each time point that, after filtering for low-quality cells, generated data from 6,922 and 7,091 cells in the 96h samples and 7,453 and 5,550 cells in the 120h samples. Harmonization of the different samples and dimensionality reduction was performed using scVI (Lopez et al., 2018). Clustering and differential expression analysis was done with the Seurat v3 R package (Stuart et al., 2019), and two-dimensional visualization of the data was performed with UMAP (McInnes et al., 2018) (see **Materials and Methods**).

Our single-cell analysis identified four major cell types via known gene markers (**Figure 2.1B-E; Figure 2.2A-G**): the AMPs, the wing disc epithelial cells (disc proper and peripodial epithelium), and small numbers of tracheal cells and hemocytes. We observed expression of a similar set of marker genes in each of these cell types to those described by others (Deng et al., 2019; Bageritz et al., 2019; Zappia et al., 2020). Altogether, we recovered profiles for 19,885 adult muscle precursors (AMPs), 7,104 wing disc epithelial cells, 15 tracheal cells, and 12 hemocytes. Notably, our dataset shows an overrepresentation of AMPs, probably the result of our collagenase-based dissociation protocol (see **Materials and Methods**) that dissociated AMPs far more effectively than epithelial cells. This enrichment of AMPs has enabled an especially detailed analysis of this cell type, which has previously received less attention. Unlike bulk RNA-sequencing approaches which would average changes in gene expression across multiple cell types, we were able to observe expression changes between 96h and 120h that occurred in both the epithelium and the AMPs together, as well as those changes that were confined to either the epithelial cells or the AMPs (**Figure 2.2H** and **Supplementary File 2.1**).

Major transcriptional differences between epithelial cells reflect their proximodistal position

To search for the signals that might be exchanged between the epithelium and the AMPs, we first characterized the cell types within each of these populations separately. The wing disc epithelium is often divided into four broad domains – the notum, hinge, pouch, and peripodial epithelium (PE) – based both on morphology and gene expression patterns. The genes encoding the transcription factors *nubbin* (*nub*) and *Zn finger homeodomain 2* (*zfh2*) were used to define the pouch and hinge, respectively (Zirin & Mann, 2007; Terriente et al., 2008; Ayala-Camargo et al., 2013). Expression of these proteins is shown for wing discs at 96h and 120h AEL (**Figure 2.1F, G**). To characterize the epithelial compartment, we clustered the cells and then classified these clusters as originating from the larger domains of notum, hinge, pouch, and peripodial epithelium (PE) based on the expression patterns of marker genes (**Figure 2.1H-K; Figure 2.3A**). Many of these genes were already expressed in a domain-specific manner by 96h, suggesting that cells in the epithelium had already been partitioned into these domains (**Figure 2.1I-K, Figure 2.3B**).

The proximodistal axis was a primary feature in stratifying cells within our analysis, while the anteroposterior axis separated the data to a lesser degree. This is somewhat surprising because anterior and posterior cells arise from distinct embryonic subpopulations (Garcia-Bellido et al.,

1973; Madhavan & Schneiderman, 1977; Worley et al., 2013; Requena et al., 2017). We classified cells as anterior or posterior cells based on expression of compartment marker genes (e.g., *cubitus interruptus*, *hedgehog*, and *engrailed*) (**Figure 2.4A-E**), and found that the concise representation of the data in two dimensions (with UMAP) resulted in stratification of the cells based on the proximodistal axis first, with secondary stratification of cells based on anteroposterior identity (**Figure 2.4E**). Additional analysis indicated that there are more differentially-expressed genes across the proximodistal axis than the anteroposterior axis (**Figure 2.4F-H**).

To investigate how cells changed over developmental time, we analyzed transcriptional changes that occurred within epithelium cell clusters between 96h and 120h AEL and found that 337 and 408 genes were significantly downregulated and upregulated, respectively, within at least one cluster (see **Materials and Methods; Figure 2.3C**; **Supplementary File 2.2**). One example is *string* (*stg*) (Edgar & O'Farrell, 1990), which encodes a regulator of the cell cycle and is upregulated within the wing margin while being downregulated in other regions of the disc. Thus, even though the major cell types are established by 96h of development (**Figure 2.1I-K; Figure 2.3B**), we still find evidence of further pattern refinement via highly-localized gene expression changes (**Figure 2.3C**).

Cell-type identities among the AMPs are consolidated later than in the epithelium

Initial analysis of the AMPs showed a clear partition of the cells with respect to two primary features: cell sex (discs were collected from both male and female larvae) and cell cycle phase (**Figures 2.5A-G and 2.6A-G**). We utilized scVI to suppress the effects of these covariates, which enabled us to obtain a clearer view of other aspects of AMP cell biology (see **Materials and Methods**).

The AMPs are known to differentiate into either direct or indirect flight muscles of the adult fly (Bate, 1993; Roy & VijayRaghavan, 1999; Sudarsan et al., 2001). The precursors of these two populations can be identified by their location within the tissue, and are canonically classified by their relative expression of two transcription factors, Vestigial (Vg) and Cut (Ct), at the late third-instar larval (L3) stage (corresponding to our 120h time point) (Sudarsan et al., 2001). The precursor cells of the indirect flight muscle are localized more dorsally (closer to the wing disc stalk) and display relatively high Vg and low Ct protein expression (Sudarsan et al., 2001). The direct flight muscle cell precursors are localized more ventrally (closer to the wing hinge) and are identifiable by little or no Vg and high Ct protein expression (Sudarsan et al., 2001). Cut expression at both time points is shown in **Figure 2.1L, M**. After unsupervised clustering (**Figure 2.1N**), we found that one cluster was characterized by high levels of *ct* and low levels of *vg*, and the remaining cells displayed relatively elevated levels of *vg* and low levels of *ct* (**Figure 2.1O-Q; Figure 2.6E, F**). Based on this distinction, we classified cells as representing the direct or indirect AMPs (**Figure 2.1N; Figure 2.6F**), obtaining 17,604 indirect AMPs and 2,281 direct AMPs.

Although the expression of Ct and Vg was less distinct at 96h, our analysis still classified cells from 96h into both direct and indirect populations. This prompted us to investigate if there were more subtle differences besides Ct and Vg at 96h that distinguished the direct and indirect AMPs. We identified a small number of genes that at both time points, showed differential expression between direct and indirect AMPs (**Figure 2.7A; Supplementary File 2.3**), possibly indicative of pathways that are necessary both for the initial establishment and subsequent maintenance of AMP cell types. This included several predicted targets of Wg signaling, specifically *naked cuticle* (*nkd*) and *vg*, and the modulator of Notch signaling *fringe* (*fng*) (**Figure 2.7A, B**). However, the majority of differentially-expressed genes that we identified were

developmentally-regulated. For these genes, differential expression between the direct and indirect AMPs was only observed at 120h. Thus, unlike the epithelium, the subdivision of the AMPs into direct and indirect pools (as assessed by differential expression of genes) mostly occurs only by 120h.

From 96h to 120h, we noticed the activation and refinement of expression of many genes previously implicated in axon guidance (**Figure 2.7A**). These include the receptor-ligand pairs *roundabout2 / slit* and *Neurotactin / Amalgam* (Kidd et al., 1999; Frémion et al., 2000), as well as the synaptic partner-matching genes *Ten-a* and *Ten-m* (Hong et al., 2012). We confirmed that Ten-m protein increases in expression from 96h to 120h by staining discs with anti-Ten-m antibody (**Figure 2.7C**). These observations suggest that pathways known to function in axon guidance could also function in myoblasts.

A three-layered virtual wing disc and expression of ligand-receptor pairs predict cell-cell interactions

Since patterning of the epithelium precedes that of the AMPs, we looked for signals from the epithelium to the AMPs. First, to predict and visualize the location of gene expression within the wing disc, we generated a three-layered model, with gene expression levels inferred at different spatial positions within the AMP, disc proper, and peripodial epithelium layers. This enabled us to discover the expression of genes in regions of the disc epithelium that are closest to the AMPs as well as in the AMPs themselves. Second, we examined the expression of ligand-receptor pairs to identify those with complementary expression patterns between the disc epithelium and the AMPs.

To generate a three-layered transcriptomic map of the wing disc we mapped our single-cell data to a reference model. The cells were mapped with the R package DistMap (Karaiskos et al., 2017) and the reference model was assembled from a manually curated set of gene expression patterns (**Figure 2.8A** and **Materials and Methods**). This spatial mapping of cells was largely consistent with our manual cluster annotations based on marker genes (**Figure 2.9**). This included the four broad epithelial domains (notum, hinge, pouch, peripodial epithelium; **Figure 2.8B**) and more refined cell groups, such as the wing margin, posterior notum, outer pouch, and anterior hinge (**Figure 2.8C**). To test how well our virtual wing disc would predict novel gene expression patterns, we predicted the virtual *in situ* expression patterns of *grain* (*grn*) and *pou domain motif 3* (*pdm3*), neither of which were included in generating the disc model. The predicted patterns largely matched the expressions of transcriptional reporters for *grn* and *pdm3* within wing discs (**Figure 2.8D-G**), indicating that our virtual wing disc can successfully predict novel gene expression patterns.

To look for potential cell communication between the different cell layers of the wing disc, we examined the expression of ligand-receptor pairs within the major domains of the disc epithelium and AMPs (**Figure 2.8H**) (for details on receptor-ligand pairs examined, see **Materials and Methods**). Interestingly, we observed high levels of expression of the genes encoding two FGF-family ligands, *thisbe* (*ths*) and *pyramus* (*pyr*) (Stathopoulos et al., 2004), in the notum region of the epithelium, whereas the gene encoding their receptor, *heartless* (*htl*) (Beiman et al., 1996), was specifically expressed in the AMPs. Similarly, the ligand *hedgehog* (*hh*) appears to be expressed only in the disc epithelium, while its receptor *patched* (*ptc*) and signal transducer *smoothened* (*smo*) are both expressed in the epithelium as expected, but also unexpectedly in the AMPs. Our detailed investigations of the FGF and Hedgehog pathways are presented in this study.

FGF signaling from the epithelium creates a niche that regulates AMP number and localization

Ths and Pyr are the ligands for one of the *Drosophila* fibroblast growth factor (FGF) signaling pathways, and both interact with the receptor Htl (Stathopoulos et al., 2004) (**Figure 2.10A).** While *htl* is detected in nearly all of the AMPs, the three-layered disc map predicts that *ths* and *pyr* are expressed primarily in the epithelial cells of the notum, with considerable overlap (**Figure 2.10B-D**). Specifically, *ths* is localized to the most proximal region of the notum, while *pyr* has a broader expression pattern extending into the posterior hinge (compare **Figures 2.10C** and **2.10D**). The expression patterns of *ths* and *htl* reporters were consistent with the expression predicted using our virtual wing disc (**Figure 2.10E, I**).

During embryogenesis, *htl*, *ths*, and *pyr* are known to influence mesoderm spreading along the embryonic ectoderm and formation of cardiac progenitor cells (Beiman et al., 1996; Stathopoulos et al., 2004; Kadam et al., 2009). Notably, in *htl* mutants and in *ths* and *pyr* double mutants, mesoderm cells are still present within the embryo, but they accumulate in multilayered arrangements instead of a monolayer along the ectoderm (as observed in wild-type embryos) (Beiman et al., 1996; Stathopoulos et al., 2004; Kadam et al., 2009). These observations suggest that FGF signaling might be primarily needed for proper mesoderm spreading, rather than for cell proliferation and survival. By analogy, Ths and Pyr may regulate the localization of AMPs relative to the epithelium.

To examine the consequences of interfering with FGF signaling within the larval wing disc, we perturbed the expression of Pyr and Htl. To disrupt Pyr expression within the epithelial tissue, we utilized an *apterous* (*ap*) Gal4 driver (*ap-Gal4*) that expresses in the entire dorsal compartment of the disc proper, including all of the epithelial cells that overlie the AMPs (**Figure 2.10F, G**). Expressing an RNAi that targets *pyr* with this *ap-Gal4* driver resulted in a reduction in AMPs and an obvious increase in apoptosis (visualized using anti-Dcp1), primarily observed in the more ventral- and posterior-localized AMPs (**Figure 2.10H, compare with 3F**). A likely explanation for this result is that *pyr* knockdown within the wing disc restricts AMP survival to the Ths-expressing region of the dorsal notum, as Pyr and Ths have been noted to have partially redundant functions (Stathopoulos et al., 2004; Kadam et al., 2009). Moreover, this result suggests that sufficient Ths cannot reach this region of the notum to compensate for the knockdown of Pyr at physiological levels of Ths expression. We next tested if the Ths and Pyr receptor Htl is required within the AMPs. Using the AMP-specific driver *15B03-Gal4* (**Figure 2.10J**), we expressed an RNAi for *htl* and observed an obvious decrease in the number of AMPs following knockdown of FGF signal transduction (**Figure 2.10K**). Altogether, we conclude that FGF signaling between AMPs and the disc epithelium is necessary for proper AMP survival.

To determine if the location and level of FGF signaling controls the position and number of the AMPs, we assessed the effects of increasing the levels of FGF ligands and also expressing them ectopically. First, we used a *dpp-Gal4* driver that is expressed in a stripe of cells just anterior to the anterior-posterior compartment boundary of the epithelium, including in the notum (**Figure 2.11A**). Expression of either *pyr* or *ths* in this domain caused a massive increase in the number of AMPs, not only beneath the notum epithelium where AMPs are normally present, but throughout the entire *dpp-Gal4* expression domain, including underneath the wing pouch (**Figure 2.11B; Figure 2.12A**). AMPs were observed along the entire stripe of ectopic FGF expression, all the way to the ventral hinge and even extending on the ventral side to the peripodial epithelium. From this, we conclude that FGF signaling does not just increase AMP number, but can also induce AMP spreading beyond the epithelial notum.

We further investigated the role of FGF signaling in AMP migration by generating a separate patch of *pyr* expression, discontinuous from the domain of endogenous expression. To this end, we ectopically expressed *pyr* within the wing pouch using the TRiP-Overexpression VPR toolkit (Lin et al., 2015) with a *nub* driver. At the onset of *nub* expression, the pouch and the notum are separated by multiple cell diameters. We observed a large number of AMPs basal to the epithelium of the wing pouch (**Figure 2.11C, D**), suggesting that ectopic Pyr expression can even recruit AMPs from a distance of several cell diameters.

To ascertain whether the AMPs were capable of proliferation at ectopic locations, we assessed whether the cells were progressing through the cell cycle. AMPs throughout the A-P axis of the disc were found to be in both S- and M-phase of the cell cycle, as indicated by the incorporation of the thymidine analog EdU and staining for phospho-histone-H3 (PHH3) (**Figure 2.11E, F**; **Figure 2.12D,E**). This shows that when an FGF source is provided in the epithelium, AMPs are capable of cell-cycle entry and progression even in portions of the disc that are distant from the notum, consistent with the large increase in the number of AMPs. Thus, expression of Pyr and Ths in epithelial cells appears sufficient to both attract AMPs to that location and sustain their local proliferation.

To determine whether AMPs can be induced to migrate at different stages of larval development, we restricted the ectopic expression of FGF ligands to the later stages of larval development. The expression of *dpp>pyr* or *dpp>ths* was controlled by a temperature-sensitive Gal80, and when expression was initiated in mid-L3 (48h prior to dissection), AMPs were observed throughout the dorsal portion of the wing pouch (**Figure 2.11G; Figure 2.12B**). In contrast, when expression was initiated later in L3 (24h prior to dissection) AMPs were observed in the dorsal hinge but not the pouch (**Figure 2.11H; Figure 2.12C**). Taken together, these experiments indicate that the myoblasts found associated with other regions of the disc are derived from those underlying the notum. Additionally, these experiments show that AMPs can be induced to emigrate from the notum region even in the later stages of L3, suggesting that the localized expression of the FGF ligands in the notum both localizes AMPs to that region of the disc and also sustains their survival and proliferation (**Figure 2.11I**). Thus, FGF signaling effectively defines the AMP niche.

Hedgehog signaling regulates gene expression in a subset of posterior localized AMPs

Our analysis indicated that *patched* (*ptc*), which encodes the transmembrane receptor for the ligand Hh and is also a transcriptional target for Hh signaling, is expressed at low levels in most AMPs and at a much higher level in a subset of the direct AMPs (**Figure 2.13A**). Moreover, Hh signaling pathway components *smoothened (smo)* and *cubitus interruptus (ci)* are expressed in most AMPs at uniform levels (**Figure 2.13B, C**). However, we detected negligible levels of *hh* transcripts within our AMP data, and furthermore did not detect expression of *hh-Gal4* within the AMPs (**Figure 2.13D; Figure 2.14B**). In contrast, *hh* transcripts were detected in approximately 32% of cells in the epithelium, roughly the size of the Hh-producing posterior compartment (**Figure 2.4B**). Together, these observations support the possibility that Hh from posterior cells of the disc epithelium activates Hh signaling in a subset of myoblasts which respond by expressing higher levels of *ptc*.

We detected Ptc protein in a subpopulation of AMPs, localized primarily beneath the posterior compartment of the disc epithelium and extending approximately 20-40 μm into the region underlying the anterior compartment (**Figure 2.13E; Figure 2.14**). Consistent with our scRNAseq data, Ptc was observed mostly in a group of direct AMPs, but also in two additional

smaller groups of cells that are located more dorsally among indirect AMPs. The proximity of these Ptc-expressing AMPs to Hh-secreting epithelial cells suggests that they are responding to the Hh secreted by these cells rather than a circulating pool of Hh that should be available to all AMPs. Indeed, a recent study suggests that Hh ligand from the epithelium is transported via cytonemes to nearby AMPs (Hatori & Kornberg, 2020).

To determine if Ptc expression in the AMPs closest to the Hh-producing epithelial cells is a result of Hh pathway activation, we reduced *smo* expression in all AMPs and observed that Ptc expression in AMPs was abolished (**Figure 2.13F**). This indicated that, as in the disc epithelium, Hh signal transduction within the AMPs is required to establish high Ptc expression in the posterior-localized AMPs. To address whether all AMPs are capable of this response, we expressed an activated form of the transcription factor Ci (Ci^{3m}), which is resistant to proteolytic cleavage (Price & Kalderon, 1999), in all AMPs and observed that this lead to an uniformly high levels of Ptc protein expression (**Figure 2.13G**). Thus, all AMPs appear capable of responding to Hh, but during normal development, only AMPs with close proximity to the posterior compartment of the disc epithelium receive the signal. Neither smo^{RNAi}-knockdown nor ci^{3m} overexpression caused obvious changes in AMP numbers (compare **Figure 2.13E** with **Figures 2.13F, G**), suggesting that Hh signaling is not controlling AMP proliferation but instead is likely important for patterning.

To investigate a possible role of Hh signaling in AMP cell fate specification, we examined adult flies after genetic perturbations for flight muscle defects. During the pupal phase, the AMPs give rise to three distinct muscle fiber types within the adult thorax: dorsal longitudinal muscles (DLM), dorsoventral muscles (DVM), and direct flight muscles (DFM) (**Figure 2.13H**). While DLMs and DVMs are indirect flight muscles that generate the mechanical movement required for flight by compressing the thorax, the DFMs are responsible for flight steering by fine-tuning the position of the wing blades (reviewed by Bate, 1993). Both DLMs and DVMs are formed from indirect AMPs, whereas the direct AMPs develop into the DFMs. Control adults displayed wild-type posture (**Figure 2.13I**), while after Hh signaling was downregulated in AMPs with smo^{RNAi}, we found that a majority of adults displayed an "outstretched" wing posture phenotype (**Figure 2.13J, L**). When ci^{3m} was expressed in all AMPs, we observed that many adults displayed a "downtilted" wing posture (**Figure 2.13K, L**)**.** These wing posture phenotypes were reproducible with multiple smo^{RNAi} and in both sexes (**Figure 2.15A**)**.** Adults with either the outstretched or downtilted phenotypes were incapable of flight. These observations suggest a crucial role of Hedgehog signaling within the AMPs for the formation of functional adult flight muscles.

To examine if Hh signaling perturbations affected the structure of adult muscle fibers, we dissected adult thoraxes (**Figure 2.13M-R**). When we reduced *smo* expression in the AMPs, we observed a misalignment of the DFM fibers (**Figure 2.13U**). In particular, the more posterior DFMs 52-57 (Miller, 1950; Bate, 1993; Ghazi et al., 2000) displayed improper position and overall disorganization (**Figure 2.15B-M**). Muscle 53, for example, inappropriately projects to the dorsal attachment site of muscle 54. In contrast, the DLM and DVM muscle fibers appeared relatively normal (**Figure 2.13S, T**). This indicates that the loss of Hh-signaling within the AMPs causes defects in the adult muscles, specifically a subset of muscles formed by the direct AMPs.

Conversely, when the Hh-signaling pathway was constitutively active via the expression of ci^{3m} within the AMPs, we observed elimination of the DVMs (**Figure 2.13W**) and the DFMs were often severely disorganized and malformed (**Figure 2.13X**). Importantly, muscle 51, which is derived from a separate group of AMPs not associated with the wing disc (Lawrence, 1982), is unaffected by these manipulations. The DLMs had no noticeable defects (**Figure 2.13V**), likely

because the DLMs do not arise in *de novo*, unlike the other adult flight muscles, but rather by the fusion of AMPs with histolyzing larval muscles that act as templates (Fernandes et al., 1991). Overall, our data suggest that Hh-signaling is important for proper specification of a subset of the direct AMPs and that excessive Ci activity causes inappropriate patterning that perturbs the development of the indirect flight muscles.

Neurotactin and Midline are AMP-specific downstream targets of Hedgehog signaling

What are the downstream targets that are activated by Hh-signaling in the AMPs? Since the canonical target of Hh signaling in the disc epithelium, *dpp*, is not expressed in the AMPs, we searched for candidate genes that were specifically expressed within the subpopulation of AMPs that express high levels of *ptc*. We found that *midline* (*mid*) and *Neurotactin* (*Nrt*) displayed relatively high correlation with that of *ptc* (Pearson correlation of 0.44 and 0.34 with *mid* and *Nrt*, respectively) (**Figure 2.16A-C**). Mid, also known as Neuromancer 2, is a T-box transcription factor most related to mouse Tbx-20 (Buescher et al., 2004) that regulates cell fate in the developing nervous system (Leal et al., 2009). Nrt encodes a single-pass transmembrane protein expressed on the cell surface (Hortsch et al., 1990). Dimers of the secreted protein Amalgam (Ama), which are expressed in the direct AMPs, are able to bind to Nrt on two different cells and promote their adhesion (Frémion et al., 2000; Zeev-Ben-Mordehai et al., 2009).

Nrt and *mid* expression increased dramatically in direct AMPs from 96h to 120h based on our single-cell data (**Figure 2.16A, B**) and antibody staining (**Figure 2.16D-G**), whereas *ptc* is expressed at similar levels at both time points (**Figure 2.16C, H, I**). Surprisingly, while Nrt and Mid expression patterns included all posterior-localized direct AMPs, the expression of both genes extended into anterior-localized direct AMPs which do not currently express Ptc. Due to their high expression levels in the posterior-localized AMPs, we hypothesized that expression of Nrt and Mid, at least in the posterior AMPs, was influenced by Hh signaling.

To determine if *Nrt* and *mid* are downstream Hh-signaling targets, we examined if perturbing the Hh pathway within the AMPs would alter their expression. The knockdown of *smo* and consequent reduction in Hh signaling resulted in a dramatic decrease in both Nrt and Mid expression in the direct AMPs at 120h AEL (**Figure 2.16J, K, N**). Remarkably, this was observed in both the posterior- and anterior-localized AMPs alike. We tested if increased Hh signaling would be sufficient to induce expression of Nrt and Mid by driving ci^{3m} in the AMPs. This resulted in the ectopic expression of both Nrt and Mid in all of the AMPs, although expression was higher within direct AMPs (**Figure 2.16L-N**). Thus, all AMPs are capable of inducing Nrt and Mid expression in response to Hh signaling, but only the AMPs closest to Hh-secreting epithelial cells do so under physiological conditions. This relationship between *ptc*, *mid*, and *Nrt* is specific to the AMPs as we did not observe correlation between *ptc* and either *Nrt* or *mid* within the epithelium. These experiments indicate that Hh signaling is important for proper patterning of the AMPs. However, Hh-signaling is not required to regulate all aspects of the direct AMP cell fate, as high Ct expression was unaffected following the manipulation of Hh signaling by *smo* knockdown (**Figure 2.17A, B**). This suggests that Hh signaling is acting independent from Ct to specify the fate of a subpopulation of the AMPs.

To evaluate the functional consequences of reducing *Nrt* and *mid* expression, we used RNAi to reduce their expression. The RNAi line directed against *mid*, however, failed to significantly reduce Mid protein levels. In contrast, *Nrt* knockdown reduced Nrt levels (**Figure 2.16O**) and resulted in adults with defects in the posterior DFMs (**Figure 2.16P;** compare to

Figure 2.13R), albeit not as severely as knockdown of *smo*. This result shows that the Hh-signaling downstream target *Nrt* is critical for proper DFM development.

The AMPs with high Ptc expression are in the vicinity of Hh-expressing epithelial cells, which is consistent with the notion that Hh is a short-range morphogen capable of acting on cells within 40 μm from the source. However, we observed expression of the two target genes, Nrt and Mid, in anterior-localized AMPs far beyond 40 μm from the Hh source. Knockdown of Hh signaling via smo^{RNAi} expression abolished the expression of both genes in anterior- and posterior-localized AMPs alike (**Figure 2.16J, K**), ruling out the possibility of their activation in a Hh-independent manner. This raised the possibility that these cells were initially closer to the Hh source earlier in development, and had subsequently moved anteriorly either as a result of passive displacement or active migration. In this scenario, the expression of Nrt and Mid would be expected to perdure for longer than the expression of Ptc. To test whether anterior-localized AMPs had past activation of the Hh-signaling pathway, we used a lineage-tracing method to identify cells that had previously expressed *ptc* ($ptc^{ts}>FLP, GFP^{FLP\text{-}Out}$; see **Materials and Methods**). We labeled all cells that displayed activity of *ptc-Gal4* during late 2nd and early 3rd instar, and observed lineage-labelling (indicated by GFP expression) of all direct AMPs underlying the posterior compartment of the epithelium as well as a trail of anterior AMPs that recapitulate the domain of anti-Nrt staining (**Figure 2.16Q**). These results indicate that these anterior GFP-positive AMPs likely descended from cells that previously expressed high levels of *ptc*, and that past activation of the Hh-pathway is likely to be responsible for the expression of both anterior- and posterior-localized Nrt and Mid (**Figure 2.16R**).

DISCUSSION

Heterogeneity and Diversification of Cell Types

Data obtained from scRNAseq experiments can provide both spatial and temporal information that give us a better understanding of how cells diversify and then stabilize their transcriptomes during development, and point to ways in which they interact with each other. Moreover, we have been able to visualize the expression of genes from our dataset in a three-layered virtual wing disc – a format that will be useful for developmental biologists.

One interesting observation from our data is that spatial positioning within the wing disc is highly informative of the transcriptional state of cells. In particular, the proximo-distal axis of the disc epithelium is one of the primary stratifying features within our single-cell data. Epithelial cell clusters were easily mapped back to sub-regions within the notum, hinge, pouch, and peripodial epithelium. In contrast, although the cells of the anterior and posterior compartments have been separated by lineage since early in embryogenesis, we observe less differential expression between the two compartments. Thus, position along the proximodistal axis has a far greater influence on the transcriptome of a cell than its ancestry.

In the epithelium, we observe that most of the major cell types observed at late L3 (120h AEL) are already present at mid L3 (96h AEL). However, the transcriptomes of the two major populations of AMPs, those that give rise to the direct and indirect flight muscles, diverge significantly during this time interval. At 96h, AMPs appear to be in a relatively naive state; canonical markers for the direct and indirect flight precursors, *ct* and *vg*, both show relatively uniform expression at the mid L3 stage. At the late L3 stage, we observe more distinguishable differences between the transcriptomes of direct and indirect cell types. Both *ct* and *vg* have greater differential expression in the AMPs at this time point. The earlier stabilization of epithelial cell fates then provides a stable platform for the generation of spatially-localized signals that regulate myoblast numbers and provide instructive signals for fate specification.

FGF signaling regulates the number and location of AMPs

Ths and Pyr have previously shown to regulate the spreading of mesodermal cells during embryogenesis, but a role for these ligands in regulating myoblast numbers was not previously appreciated. We show these ligands are necessary for AMP survival and that increased levels of these ligands can promote AMP proliferation even at sites distant from the notum. Indeed, increased expression can induce a dramatic overproliferation of the AMPs, even as late as 3rd instar. Ectopic FGF signaling is sufficient to support AMP viability and cell proliferation at other locations in the wing disc, including underneath the pouch. Thus, the localized expression and level of Ths and Pyr secreted by epithelial cells in the notum could provide sufficient trophic support to generate the appropriate number of AMPs during normal development. While this work was in preparation, another group independently showed that the *ths-Gal4* line is expressed in the notum epithelium and that reducing *ths* function reduces AMP numbers (Vishal et al., 2020).

We have also demonstrated that ectopic and elevated levels of expression of Ths or Pyr can draw AMPs out of the notum region, all the way to the ventral hinge and around the ventral edge of the disc proper onto the peripodial epithelium. With constitutive FGF ligand expression driven by *dpp-Gal4*, we did not observe AMP migration to the lateral regions of the disc, which are far from the ectopic FGF source. However, when FGF expression is initiated in L3 by controlled $Gal80^{ts}$ repression, we observed a large number of laterally-located AMPs. We attribute the difference in these two scenarios to ectopic AMPs along the *dpp* stripe possibly serving as an FGF

sink. When FGF is turned on later in development, there are no ectopic AMPs. In their absence, FGF could reach more lateral portions of the disc, thus allowing AMP emigration to those regions. Furthermore, the expression of Pyr in the wing pouch, which is separated from the notum by the dorsal hinge, was sufficient to promote colonization of the pouch region by AMPs. The AMPs can be induced to colonize new domains late into larval development by the expression of either FGF ligand. These results suggest that under physiological conditions, AMPs remain beneath the notum epithelium because there is insufficient FGF outside of this region. Altogether, our work illustrates that a source of epithelial FGF is critical for forming the AMP niche and that levels of FGF regulate both the location and number of the AMP cells.

Instructive Hedgehog signaling from the epithelium to the myoblasts

We have shown that the anteroposterior identities of the disc epithelium are important for proper specification of gene expression within the underlying AMPs. One powerful advantage of scRNAseq as opposed to its bulk sample predecessor is the ability to measure gene co-expression within subpopulations of a tissue. We leveraged this advantage to identify two novel Hh-signaling targets, *Nrt* and *mid*, within AMPs. While we currently do not know whether *Nrt* and *mid* are direct targets of Ci, both genes do have consensus Ci-binding sites within potential regulatory regions. Nrt is a single-pass transmembrane protein. Its extracellular ligand, Amalgam, has more widespread expression in the direct myoblasts and is expressed at comparable levels at both 96h and 120h AEL. Two molecules of Amalgam can form homodimers and each is capable of binding to Nrt on different cells (Frémion et al., 2000; Zeev-Ben-Mordehai et al., 2009). Thus, an effect of Hh-induced expression of Nrt in a subset of the direct AFMs might be to promote aggregation of Nrt-expressing cells at a later stage of development.

An unexpected observation was that AMPs beneath the anterior compartment, distant from the epithelial source of Hh, express both identified Hh targets Nrt and Mid. However, this expression is dependent upon Hh-signaling since knockdown of *smo* blocks gene expression. Although we cannot completely exclude the possibility that a second signal from posterior AMPs activates *Nrt* and *mid* expression in these cells, our lineage-tracing experiments favor a model where a subset of the direct AMPs are generated posteriorly and move anteriorly during the course of development. Such movement could be due to a process of active migration in response to hitherto unknown external cues or to displacement as a result of oriented cell division. Understanding the mechanistic basis of AMP migration would represent an exciting avenue of future research.

Concluding remarks

Our work has provided a base for the study of heterotypic interactions in the developing wing disc during conditions of normal growth and demonstrate that such interactions can have a major effect on cell number, cell migration, and cell fate in the wing disc. By examining receptor-ligand expression patterns in conjunction with spatial mapping of our single-cell data, our analysis provides many hints of signaling pathways that may function between the disc epithelium and the AMPs and also within subsets of cells with each of these populations that provide multiple avenues for future investigations.

MATERIALS AND METHODS

Generation of single-cell suspension, barcoding, and sequencing

For each sample, approximately 250 staged *Drosophila* wing-imaginal discs were dissected within 1 hr. The collected tissue was then transferred to a microcentrifuge tube and incubated within a dissociation cocktail consisting of 2.5 mg/mL collagenase (Sigma #C9891) and 1X TrypLE (Thermo Fisher #A1217701) in Rinaldini solution (modified from Ariss et al., 2018). The sample tube was placed horizontally on a shaker machine operating at 225 rpm for 25 minutes at room temperature (method modified from Ariss et al., 2018). At the 10, 20, and 25 minute marks, the tube was flicked 20 times for additional mechanical dissociation. Dissociation was halted by centrifuging the sample at 5,000 rpm for 3 minutes, aspirating the dissociation cocktail, and then adding in 1 mL of cold PBS-10% FBS. The cell pellet was mixed by pipetting up-and-down approximately 25 times with a 1 mL pipette for additional mechanical dissociation, and then centrifuged again at 5,000 rpm for 3 minutes. The media was replaced with cold PBS-1% FBS, and the cell pellet was resuspended in preparation for FACS.

FACS of the sample was performed on a BD FACSAria Fusion flow cytometer. Dead cells were identified and removed via the addition of propidium iodide to the sample, and high-quality single cells were sorted into cold PBS-10% FBS. Cell concentration of the post-FACS sample was assessed by a hemocytometer, and adjusted 1,000 cells per uL.

Single-cell suspensions were barcoded for single-cell RNA sequencing with the 10X Chromium Single Cell platform (v2 chemistry). Barcoded samples were sequenced on an Illumina NovaSeq (S2 flow cell) to over 60% saturation.

Single-cell data processing and analysis

The 10X Genomics Cell Ranger pipeline (v2.2.0) was used to align the sequencing reads to the *Drosophila melanogaster* transcriptome (version 6.24). The data was analyzed using the R and Python programming languages, primarily utilizing the packages Seurat v3 (Stuart et al., 2019) and scVI v0.4.1 (Lopez et al., 2018).

Our standard analysis pipeline is as follows: First, each dataset was analyzed separated using the standard Seurat pipeline, with no cells filtered, 30 principal components calculated, and clustering resolution set to 2.0 (all other parameters remained default). We then removed cell clusters with an abundance of low-quality cells (defined as clusters with mean number of genes detected per cell [nGene] was less than one standard deviation below the mean nGene of all cells in the dataset). Additionally, we found that each dataset had a cluster with markers for both AMP and epithelial cell types (e.g., *SPARC* and *Fas3*) and unusually high mean nGene; this cluster was suspected to be AMP-epithelial doublets, and was also removed. Clusters were then split into AMP and epithelial cell subsets based on the expression of known marker genes. Cells within each subset were subsequently filtered if either (1) their nGene count that was outside the mean nGene of the subset +/- 1.5 standard deviations, or (2) their percentage of reads for mitochondrial genes that was greater than 1.5 standard deviations above the mean mitochondrial read percentage of the subset.

Data subsets were harmonized into collective AMP or epithelium datasets using scVI. The scVI VAE model consisted of 2 layers (n_layers=2) and 20 latent dimensions (n_latent=20), with a negative-binomial reconstruction loss (reconstruction_loss='nb'). The model was trained on variable genes selected by Seurat's variance-stabilizing transformation method; 1,000 (for epithelial subsets) or 2,000 (for AMP subsets) variable genes were calculated for each inputted

batch, and then the union of these genes was supplied to scVI. The following parameters were used for model training: train_size=0.75, n_epochs=400, and lr=1e-3 (other parameters were left as default). Cell clustering and UMAP was performed using Seurat on the latent space derived from the scVI model. After harmonization, clusters were re-examined for doublet characteristics; clusters with a mean nGene count greater than one standard deviation above the mean nGene count of all cells were removed, as were clusters that displayed markers for both AMP and epithelial cell types. Identified hemocyte and tracheal cells were also separated out. scVI and Seurat were both re-run on the datasets to generate our final AMP and epithelial cell atlases (**Figure 2.1H, N**).

To generate our full cell atlas consisting of all cell types (**Figure 2.1B, C**), we merged and harmonized the cells in the AMP and epithelium cells atlases along with the separated hemocyte and tracheal cells. scVI and Seurat were run as previously described, with the scVI model trained on the union of the top 2,000 variable genes for each batch as calculated by Seurat. No additional cell filtering was performed after harmonization.

For visualizing data on UMAPs and dot plots, we calculated normalized and scaled expression counts using Seurat's NormalizeData and ScaleData functions, respectively, with default parameters. For the normalized data, raw counts were normalized by total UMIs per cell, multiplied by 10,000. Natural-log normalized data is used for expression levels visualized with UMAP. For the scaled data, the natural-log normalized data is scaled for each gene, such that the mean expression is 0 with a standard deviation of 1. Scaled data is used for expression visualization on the dot plots.

Cell sex and cell cycle correction with AMP data

Cells were classified as male or female by their expression levels of the dosage compensation complex genes *lncRNA:roX1* and *lncRNA:roX2* (Franke & Baker, 1999; Meller & Rattner, 2002), which are both expressed almost exclusively in male cells. For both genes, we examined the natural-log normalized expression counts (calculated by Seurat's NormalizeData function), computed the density over the data, and identified the first local minima as a threshold (see **Figure 2.5C, D**). Cells that were above the threshold for either *lncRNA:roX1* or *lncRNA:roX2* were classified as male; otherwise, they were classified as female. From this, we assigned 8,097 cells as male and 11,788 cells as female, which roughly matches the size ratio between male and female wing discs given that female discs are larger. We removed cell sex stratification by processing male and female AMPs as separate batches (for each actual batch) within scVI (see **Figure 2.5E, F** for comparison of data before and after cell sex correction).

We observed significant data stratification that correlated with a number of cell cycle-related genes, such as *Proliferating cell nuclear antigen* (*PCNA*) and *Cyclin B* (*CycB*) (Yamaguchi et al., 1990; Lehner & O'Farrell, 1990), indicating that our data was split between S phase and non-S phase (**Figure 2.6A-C**). Definitive classification of cells into cell cycle stages is difficult because expression of these genes is not typically demarcated sharply into specific cell cycle stages. To remove cell cycle stratification from our data, we examined the correlation of each scVI latent dimension with the expression levels of highly variable cell-cycle genes, and found that one latent dimension was strongly related (**Figure 2.6D**). By masking this latent dimension from our downstream analysis (e.g., clustering and UMAP), we effectively diminished cell cycle stratification. **Figure 2.1N** shows the UMAP of our AMP data after subtraction of cell sex and cell cycle stratification, which allowed us to focus our analysis on different cell types within the AMPs (to see how each correction affected the AMP data, see **Figure 2.6A-C**).

Determining differentially-expressed genes

When examining clusters, genes were considered to have significant differential expression if they had (1) a false discovery rate < 0.05 (as calculated via Wilcoxon test), (2) a natural-log fold-change of 0.15 or more, and (3) a percent expression of at least 15% in one of the two populations in the comparison. This test was performed with Seurat's FindMarkers function, using a Wilcoxon test (test.use = "wilcox").

When evaluating differential expression between clusters of epithelial cells (e.g., **Figure 2.3A**), we used a one cluster vs. all analysis. When evaluating differential expression between direct and indirect AMPs (e.g., **Figure 2.6E**), we compared cells of the two groups as classified in **Figure 2.6F**. In these cases, differential expression statistics (i.e., FDR and fold change) are obtained by combining the cells (across batches) in each group and conducting a single comparison. When evaluating differential expression between time points (e.g., **Figure 2.2H**, **Figure 2.3C**, and **Figure 2.7A**) (which would be inherently confounded with batch effects, since time points were collected across separate sequencing experiments), we took a conservative approach and only considered genes that were consistently significant (by the criteria defined above) in each temporal pairwise comparison (i.e., DE analysis was conducted between all temporal pairs: 96h1 vs. 120h1, 96h2 vs. 120h1, 96h2 vs.120h1, 96h2 vs.120h2). We report the natural-log of the average value for these pairwise comparisons, and the maximum FDR calculated (see **Supplementary Files 2.1-2.3**).

Generating a virtual model of the wing disc

We assembled reference gene expression patterns from a number of sources (Held Jr, 2002; Butler et al., 2003) and based our starting geometry on the disc proper from images in Bageritz et al., 2019. The images were processed in Adobe Photoshop and assembled in R with EBimage (Pau et al., 2010) to generate binarized gene expression reference for the AMPs, disc proper, and peripodial epithelium. The geometry of the three-layered model is provided in **Supplementary File 2.4** and the binarized reference gene expression patterns are provided in **Supplementary File 2.5**. We used DistMap (Karaiskos et al., 2017) to statistically map single cells back to the reference. With this virtual wing disc model we used DistMap to calculate a 'virtual *in situ'* or a prediction of gene expression patterns. This is based on the detected gene expression with the single-cell data and the mapping location to calculate relative expression values for our model. We mapped the AMP and epithelial cells separately, as this improved how well the model predicted genes with known expression patterns. In addition, we used the scVI imputed gene expression values when mapping the cells to the reference.

Examination of receptor-ligand expression

From FlyBase, we assembled a list of genes encoding receptors and ligands from the following 19 pathways of interest: Wnt/Wingless, FGF, Hedgehog, PDGF/VEGF, JAK-STAT, Activin, BMP, Fat-Ds, Slit-Robo, Ephrin, Toll/Toll-Like, Semaphorin, Notch, Insulin-Like, Fog, Torso, Miple, EGFR, and TNF. For our analysis, we only examined pathways in which at least one receptor or ligand was either (1) differentially-expressed within one of the major domains of the epithelium (notum, hinge, pouch, or PE) when compared to all other epithelial cells, (2) differentially-expressed within one of the major domains of the AMPs (direct or indirect cells) when compared to each other, or (3) differentially-expressed between all epithelial cells vs. all AMP cells. These pathways (and their receptors and ligands) are shown in **Figure 2.11H**.

Drosophila stocks and husbandry

The stocks used in this study include the following lines from the Bloomington Stock Center : *R15B03-GAL4* (BL49261); *G-TRACE* (BL28280, 28281) (Evans et al., 2009); *UAS-FLP, Ubi-FRT-stop-FRT-GFP^{nls}* (BL28282); *pdm3-GFP* (BL60560); *grn-GFP* (BL58483); *ptc-GAL4* (BL2017); *tub-GAL80^{ts}* (BL7108); *dpp-GAL4* (BL1553); *dpp-GAL4, tub-GAL80^{ts}, UAS-dCas9.VPR* (BL67066); *nub-GAL4* (BL25754)*; hh-GAL4* (Tanimoto et al., 2000)*; ap-GAL4* (BL3041); *fng-Gal4* (BL9891); *htl-GAL4* (*GMR93H07-GAL4,* BL40669) is an enhancer within the first intron of the *htl gene*; *UAS-smo^{RNAi}* (primarily BL43134, but also BL27037, 62987 in **Figure 2.15**); *UAS-Nrt^{RNAi}* (BL28742). *Drosophila* stocks from other labs: *UAS-ths* and *UAS-pyr* (A Stathopoulos); *UAS-ci^{3m}* (D Kalderon). TRiP-CRISPR driven overexpression of *pyr* was conducted with a guide RNA that targets the upstream transcriptional start site, *P{TOE.GS00085}attP40* (BL67537), and works together with a nuclease-dead Cas9 fused with a transcriptional activator domain, *UAS-dCas9.VPR* to cause gene activation (BL67055) (Lin et al., 2015).

Dissections, immunohistochemistry, and microscopy

Imaginal discs, unless otherwise noted, were fixed in 4% paraformaldehyde (PFA) for 15 min, permeabilized in PBS plus 0.1% Triton X-100 and blocked in 10% Normal Goat Serum. For anti-Nrt antibody staining, we substituted the Triton X-100 for 0.05% Saponin. The following antibodies were used from the Developmental Studies Hybridoma Bank (DSHB): mouse anti-Cut (1:200, 2B10); mouse anti-Ptc (1:50, Apa-1); mouse anti-Nrt (BP 106 anti-Neurotactin); mouse anti-Wg (1:100, 4D4). The following antibodies were gifted: rat anti-Twist (1:1000, Eric Wieschaus), rabbit anti-Midline (1:500, James Skeath), and rat anti-Zfh2 (1:100, Chris Doe (Tran et al., 2010)). The following antibodies are from commercial sources: rabbit anti-Dcp1 (1:250, Cell Signaling); rabbit anti-GFP (1:500, Torrey Pines Laboratories, Secaucus, NJ); chicken anti-GFP (1:500, ab13970 Abcam, Cambridge, UK); rabbit anti-beta-galactosidase (1:1000, #559762; MP Biomedicals, Santa Ana, CA); rabbit anti-PHH3 (1:500, Millipore-Sigma). Secondary antibodies were from Cell Signaling. Nuclear staining with DAPI (1:1000).

For EdU staining, we incubated live discs in fluorescent EdU incorporation solution for 1 hour, following the protocol for the Click-iT EdU Cell Proliferation Kit, Alexa Fluor 555 (ThermoFisher C10338). After the incubation, discs were fixed in 4% PFA for 15 min, before proceeding with standard antibody stainings as detailed above.

To ectopically express FGF ligands starting at mid or late 3rd instar development (e.g., **Figure 2.11G, H**), we used a temperature-sensitive *dpp-Gal4* stock (*dpp-GAL4, tub-GAL80^{ts}, UAS-dCas9.VPR*) (BL67066) with *UAS-pyr* and *UAS-ths* lines. In-vial egg lays were collected over 8 hours, and larvae were initially raised at 18 ^{o}C. Larvae were shifted to 30 ^{o}C (to relieve Gal80^{ts} repression) at either 48 hours or 24 hours prior to dissection, corresponding to mid and late 3rd instar FGF activation, respectively.

To lineage-trace cells that expressed *ptc-Gal4* during the second and early third instars, we used a temperature-sensitive, Ptc-dependent FLP-Out system (*ptc-Gal4, tub-Gal80^{ts}, UAS-FLP, Ubi-FRT-stop-FRT-GFP^{nls}*). Under this system, inactivation of *Gal80^{ts}* repression permits cells that express *ptc-Gal4* (and their descendant cells) to become permanently labelled with GFP by an FLP-FRT recombination event. In the experiment shown in **Figure 2.16Q**, first instar larvae were collected and reared at 18 ^{o}C. At 5 days post-egg-lay, larvae were shifted to 30 ^{o}C for 24 hours, then shifted back to 18 ^{o}C for 48 hours prior to dissection.

Wing discs were imaged on a Zeiss Axioplan microscope with Apotome attachment, using 10x and 20x objectives. Image files were processed with ImageJ software. For each of the genotypes examined, we examined at least 8 discs and have reported representative results in this paper.

Adult muscle preparations

To image adult flight muscles, male flies aged a minimum of 2 days after eclosion were anesthetized and submerged in 70% ethanol with dry ice. The thorax was isolated by removing the head, wings, legs, and abdomen. Thoraces were bisected sagittally with a 11-blade scalpel blade. For DVMs, the DLMs, leg muscles, and excess cuticle were removed from hemithoraces. For DFMs, the DLMs, DVMs, leg muscles, and excess cuticle were removed from hemithoraces. The DLMs, DVMs, and DFMs were fixed in 4% PFA for 2 hours. Muscles were then rinsed 3 times and permeabilized in 0.3% PBST for 3 cycles, 15 minutes each on a nutator. Hemithoraces were incubated in Rhodamine Phalloidin (1:200) and DAPI (1:500) in 0.3% PBST, then rinsed 3 times and washed in 0.3% PBST for 3 cycles, 15 minutes each on a nutator. Hemithoraces were mounted in a depression slide using antifade mountant. DLMs and DVMs were imaged with a 10x objective using a Zeiss Axioplan microscope. DFMs were imaged with the 20x and 63x objectives using confocal microscopy.

Data Availability

The single-cell RNA sequencing data presented in this study (both raw, unaligned reads and processed count matrices) is available from the NCBI Gene Expression Omnibus (https://www.ncbi.nlm.nih.gov/geo/) under the accession number GSE155543.

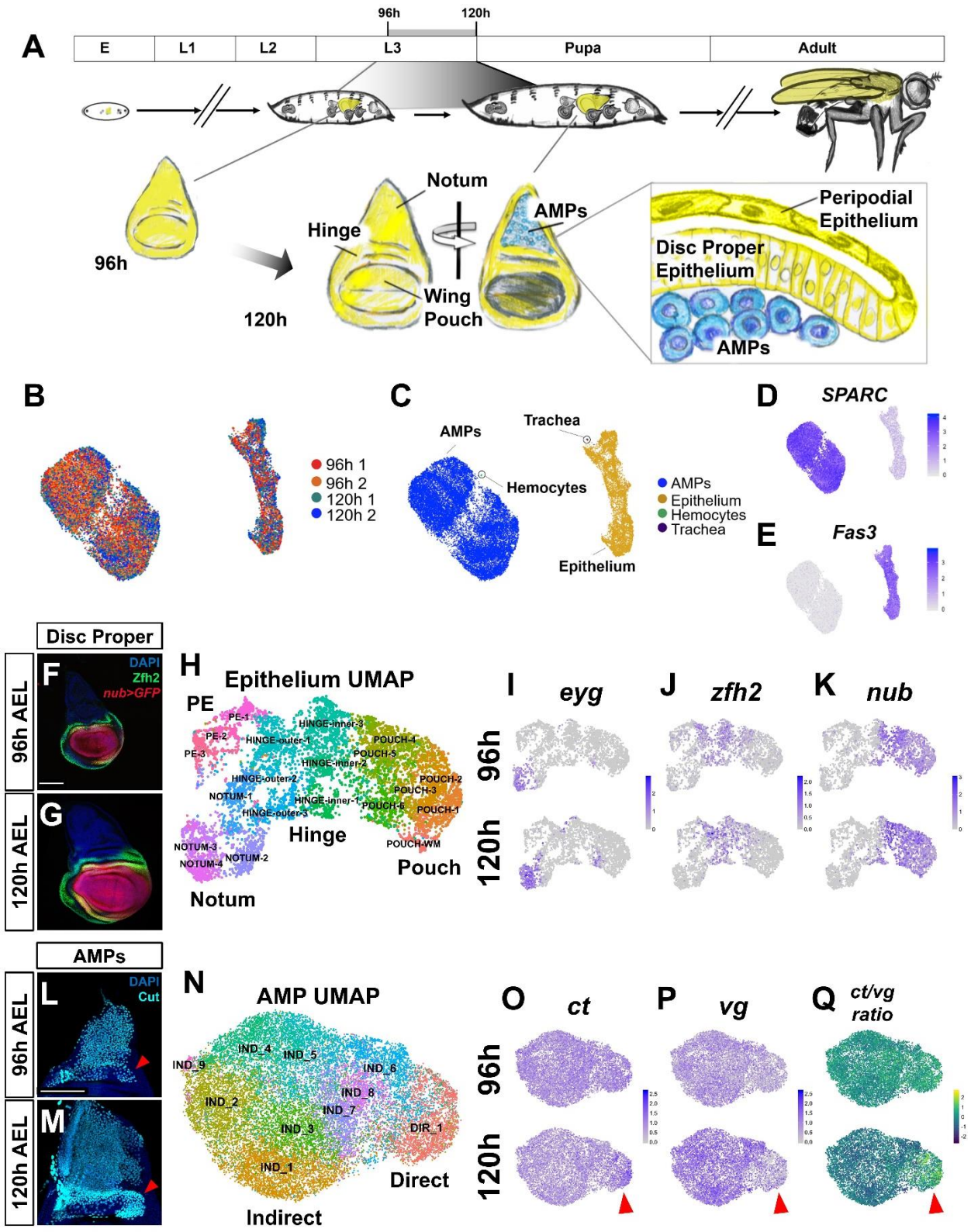

Figure 2.1. Temporal cell atlas of the developing wing-imaginal disc.

(**A**) Timeline of *Drosophila* development: embryo (E), larval phases (L1-L3), pupa, and adult. Diagram of the wing-imaginal disc within the 3rd larval instar (L3) from mid (96h) and late (120h) time points AEL. The epithelial cells of the wing disc become the adult wing blade, hinge, and the majority of the dorsal thorax (shown in yellow). The myoblasts associated with the basal surface of the disc proper epithelium are the adult muscle precursors (AMPs) (shown in blue), which generate the adult flight muscles. (**B, C**) UMAP of harmonized single-cell datasets with cells colored by batch (two time points with biological replicates) (**B**) and by major cell type (**C**). The AMPs and epithelial cells are distinguished by expression of *SPARC* (**D**) and *Fas3* (**E**) respectively. (**F, G**) Mid and late 3rd instar wing-imaginal discs with the pouch marked by *nub-GAL4* driving the expression of GFP (red) and the hinge marked with anti-Zfh2 (green). DAPI (blue) stains all nuclei. (**H**) UMAP of harmonized epithelium cells, colored by unsupervised clustering identities and manually labeled by the expression of marker genes. (**I-K**) UMAPs with cells colored by their expression levels of *eyg* (**I**), *zfh2* (**J**), and *nub* (**K**) at both developmental time points. (**L, M**) Wing discs from 96h (**L**) and 120h (**M**) AEL, stained with anti-Cut to visualize the AMPs. Red arrowheads indicate location of direct AMPs, identifiable at 120h by higher anti-Cut staining and inferred by location at 96h. (**N**) UMAP of harmonized AMPs, colored by unsupervised clustering after cell cycle and cell sex stratification correction (see **Figures 2.5, 2.6**). Cell clusters were classified as Direct (DIR_1) or Indirect AMPs (IND_1-9) based on marker gene expression (see **Figure 2.6E, F**). (**O-Q**) UMAPs separated by developmental time points showing the expression of canonical markers of direct and indirect AMPs, *ct* (**O**) and *vg* (**P**), and the ratio of these two genes within cells (**Q**). Red arrowheads highlight differential expression of these genes in the direct AMPs at 120h. UMAP scale bars correspond to normalized transcript counts on a natural-log scale. Microscopy scale bars = 100 μm.

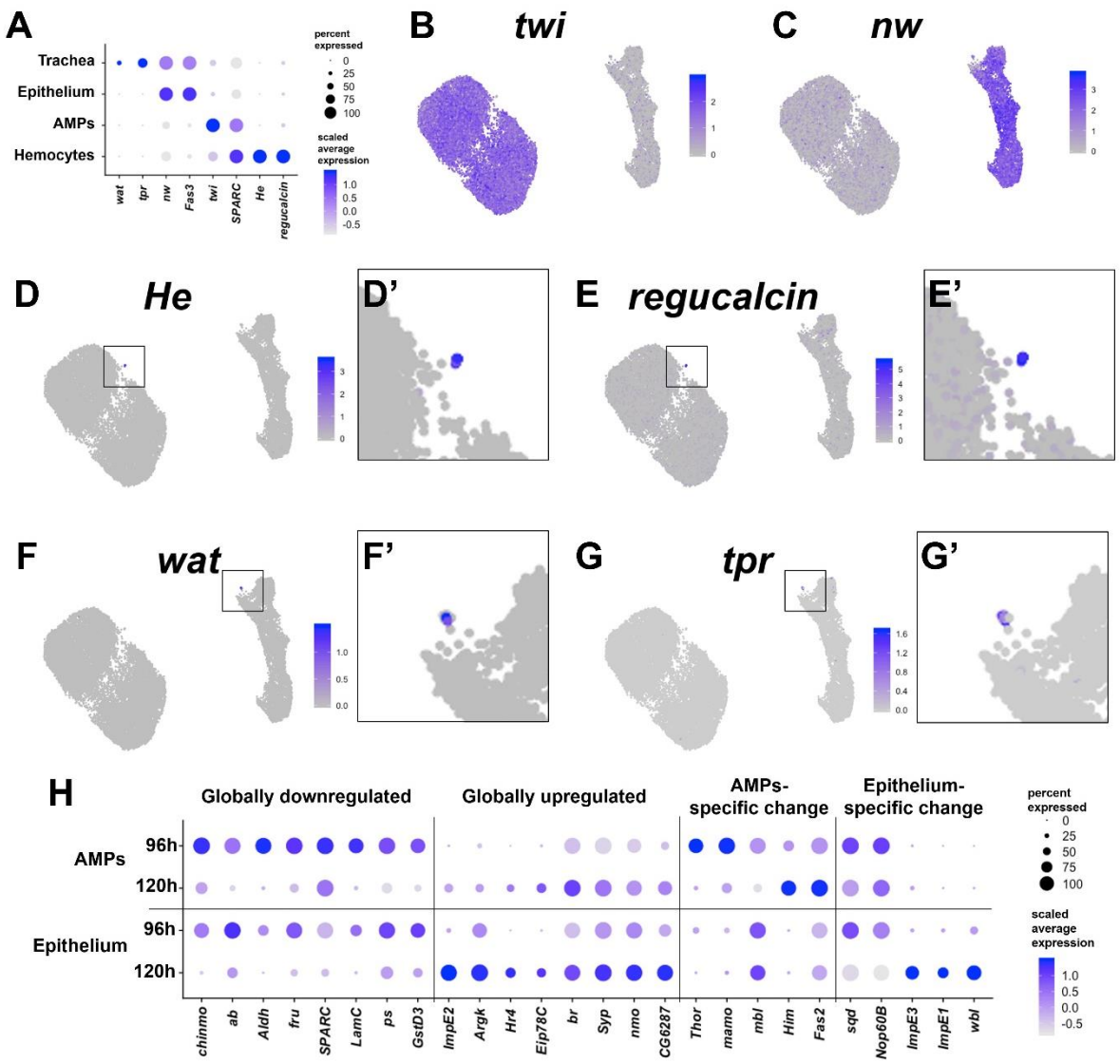

Figure 2.2. Cell type identification and temporal changes within major cell types.
(**A**) Dot plot summarizing gene expression for known markers of each of the identified cell types. For a gene of interest, the size of the dot indicates the percent of cells that express the gene and the color of the dot indicates the relative gene expression level (see **Materials and Methods**). UMAPs of full dataset showing expression levels for the following genes: AMP cell marker *twist (twi)* (**B**), epithelial cell marker *narrow wing (nw)* (**C**), hemocyte markers *Hemese* (*He)* (**D**) and *regucalcin* (**E**), and tracheal cell markers *waterproof* (*wat)* (**F**) and *tracheal-prostasin (tpr*) (**G**). Boxes in **D'**, **E'**, **F'** and **G'** are magnifications of indicated regions of the UMAP to show the cells that express the given marker gene. Color scales for UMAPs correspond to normalized (by total UMI) counts on a natural-log scale. (**H**) Dot plot summarizing gene expression changes from 96h to 120h, that were observed within both the epithelium and AMPs ("Globally downregulated / upregulated") or mostly within a single cell type ("AMP-specific / Epithelium-specific changes"). Genes shown are differentially-expressed between the two time points in the indicated populations (FDR < 0.05 for all pairwise comparisons).

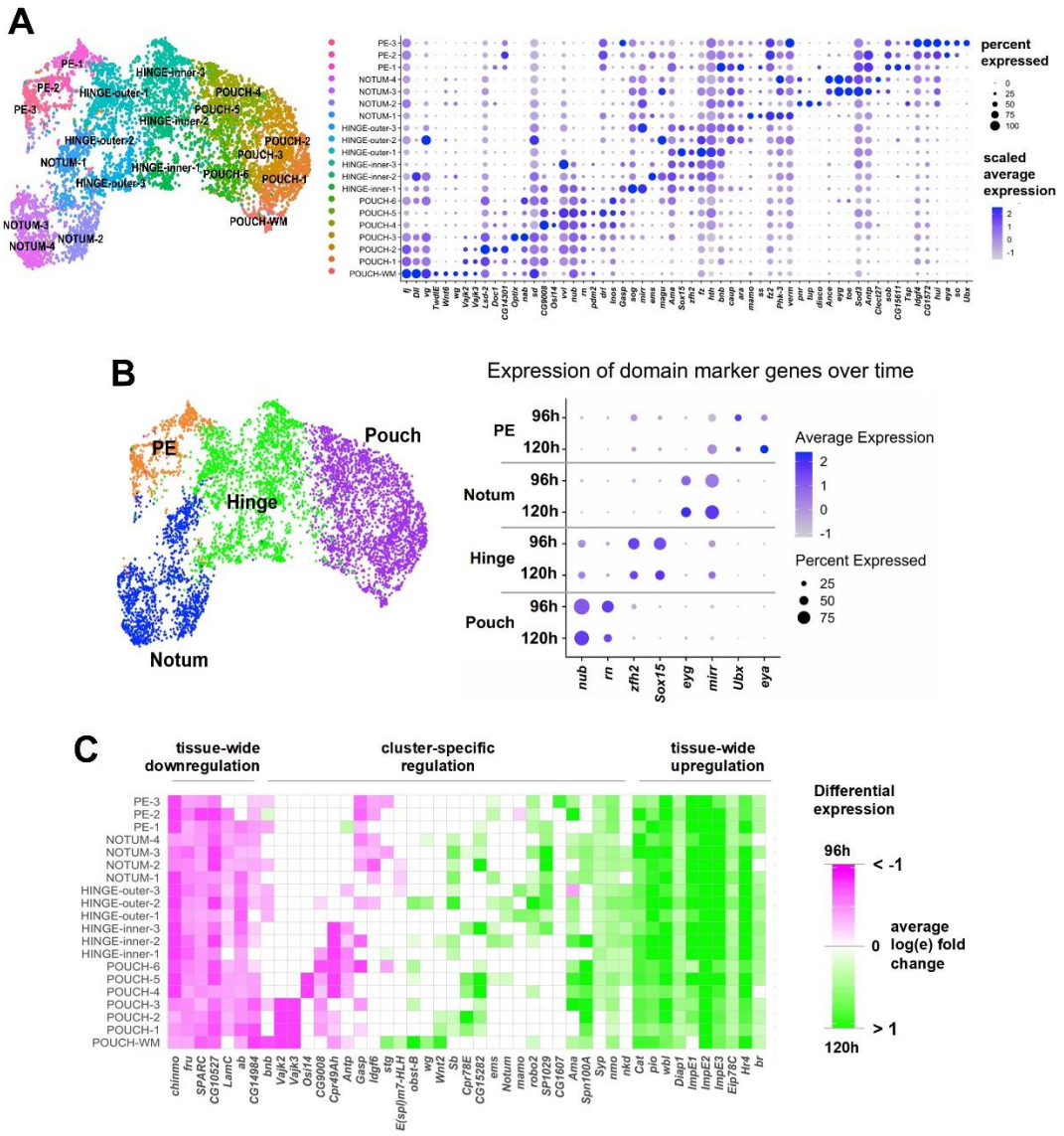

Figure 2.3. Temporal cell atlas of the wing disc epithelium.

(**A**) UMAP of epithelium cells, colored by cluster identities (same as in **Figure 2.1H**). Dot plot showing the level and percent of differentially-expressed (FDR < 0.05) marker gene expression for the cell clusters. Note that there are many genes that are expressed in a gradient. (**B**) UMAP of epithelium cells, where cell clusters have been merged into the major domains (pouch, hinge, notum, and peripodial epithelium (PE)). Dot plot summarizing the expression over time of marker genes for the major domains of the disc epithelium. Note that the expression of these marker genes remains relatively constant over this developmental window. (C) Plot showing differential gene expression for developmentally-regulated genes for the epithelium cell clusters. All genes shown were differentially-expressed between time points in at least once cluster (FDR < 0.05 for all pairwise comparisons, see **Materials and Methods**). Non-significant fold-changes were set to have a value of zero within the plot. Natural-log fold-change of expression was calculated for each of the cell clusters between cells of 96h and 120h and is capped at +/- 1 for better visualization. Genes were selected to highlight different expression dynamics, as either changing throughout most of the clusters ("tissue-wide downregulation / upregulation") or only having cluster-specific changes ("cluster-specific regulation").

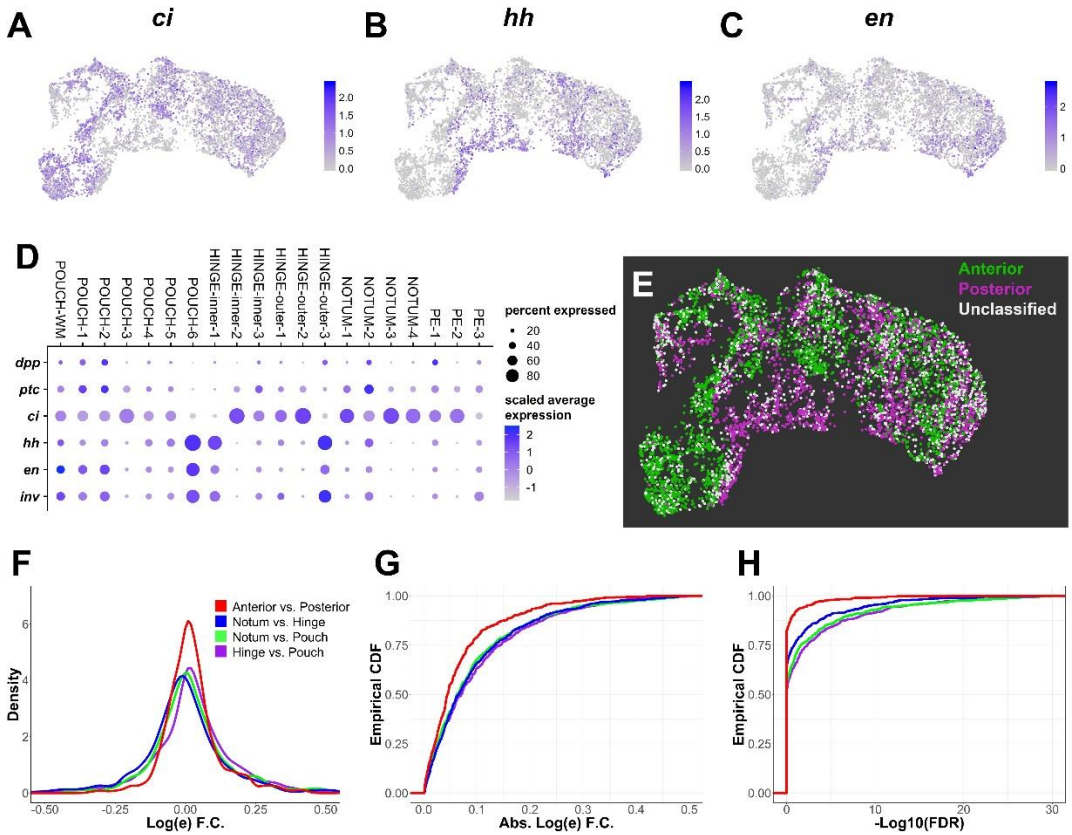

Figure 2.4. Proximodistal axis stratifies the data more than anteroposterior axis.
(**A-C**) UMAPs of disc epithelium single-cell data showing expression levels of anterior compartment marker *ci* (**A**) and posterior compartment markers *hh* (**B**) and *en* (**C**). (**D**) Dot plot summarizing expression of known anterior and posterior markers within the unsupervised clusters of the disc epithelium data. Note that most clusters express both anterior and posterior markers, except for NOTUM-3, NOTUM-4, and HINGE-inner-2, which are primarily anterior, and POUCH-6 and HINGE-inner-1, which are primarily posterior (see also **Figure 2.9**). (**E**) UMAP of anterior-posterior cell classification of disc epithelium data. Anterior cells have at least one transcript of anterior-compartment marker *ci*, whereas posterior cells have at least one transcript of posterior-compartment markers *hh*, *en*, or *inv*. Cells with transcripts for markers of both compartments or cells that lacked markers for either compartment were labeled as "Unclassified". (**F-H**) Distributions of log fold-changes (**F, G**) and false discovery rate (FDR) values (**H**) of differential gene expression between regions of the disc epithelium. Differential expression was limited to variable genes within the disc epithelium. Panel **F** displays the density (or distribution) of log fold-changes of variable genes between regions of the epithelium. Panels **G** and **H** display the empirical cumulative distribution function (ECDF) of log fold-changes (**G**) and FDR (**H**) of variable genes between regions of the epithelium. These ECDF plots are calculated as the percentage of variable genes (y-axis) below a particular log-fold change magnitude (**G**) or FDR threshold (**H**). Note that differential gene expression between anterior and posterior cells is less dramatic than all other comparisons; log fold-changes of variable genes between anterior and posterior cells are less extreme (the density is heavier around 0 in **F** and the ECDF has a steeper rise in **G**), and the associated FDR values are less significant (the ECDF in **H** has a steeper rise). UMAP color scales correspond to normalized (by total UMI) counts on a natural-log scale.

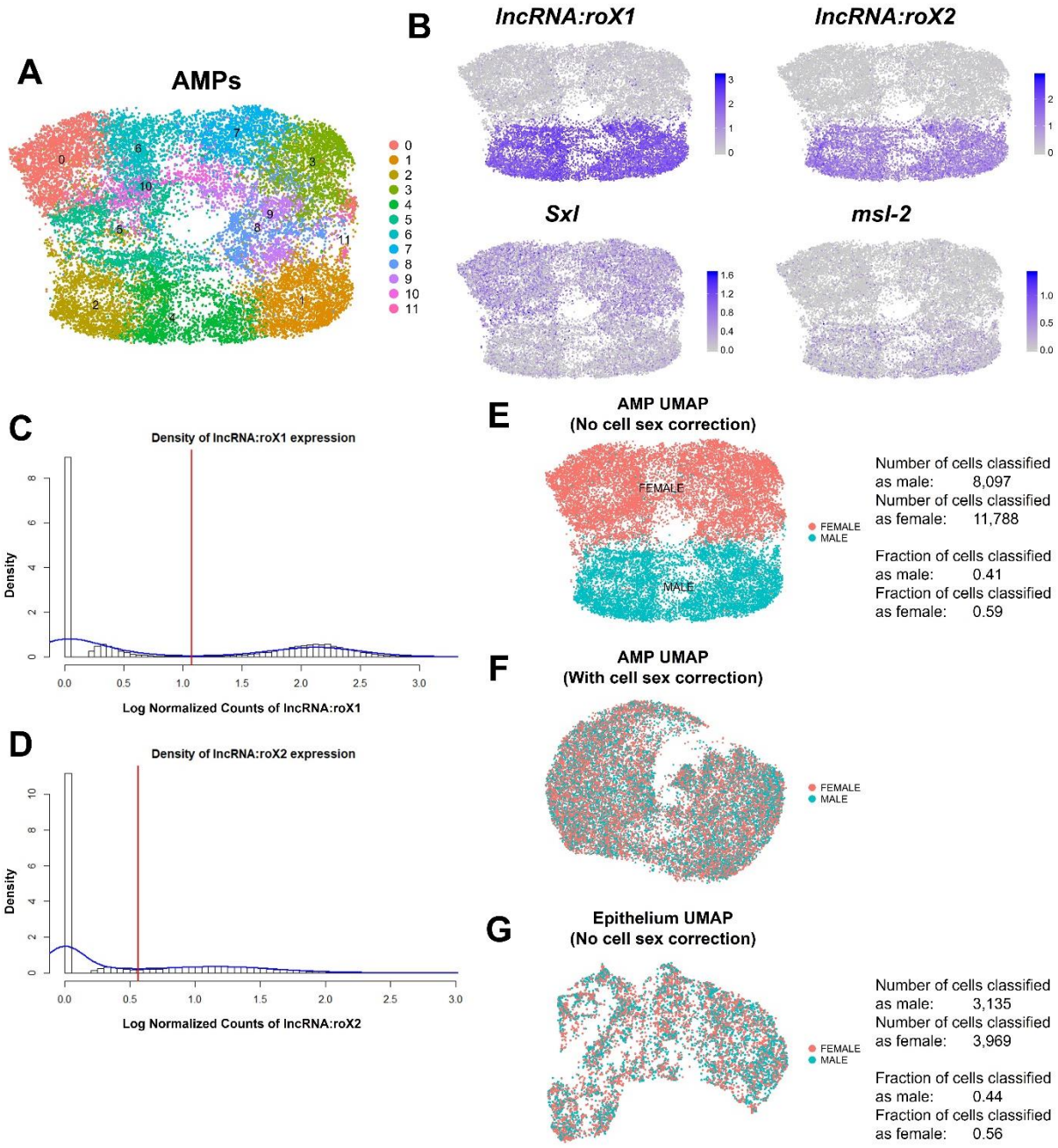

Figure 2.5. Correction of cell sex stratification within AMP scRNAseq data.
(**A**) UMAP of AMP single-cell data, as processed by our standard computational pipeline (without correction for cell sex or cell cycle). Colors correspond to 12 computational cell clusters as determined by unbiased clustering. (**B**) UMAPs showing expression of sex-specific genes *lncRNA:roX1*, *lncRNA:roX2*, *Sex lethal (Sxl)*, and *male-specific lethal 2 (msl-2)*. Note that the UMAPs show significant stratification based on the expression of these sex markers. (**C, D**) Probability histogram plots of the natural-log normalized expression counts for *lncRNA:roX1* (**C**) and *lncRNA:roX2* (**D**) within all cells. Density curves for the data are shown in blue. Red lines are drawn on the first local minima within the density of the data, and serve as a cutoff for classifying cells as having either high or low expression of either gene. Cells with high expression of either *lncRNA:roX1* or *lncRNA:roX2* were classified as male-originating; otherwise, cells were

designated as female-originating. (**E**) Classification of cell sex shown on our standard-analysis UMAP of AMPs. (**F**) Classification of cell sex shown via UMAP after computationally-correcting for cell sex stratification. Correction was performed by treating cell sex as a batch effect during data processing. Note that compared to **E**, male-originating and female-originated cells are now interspersed within the UMAP. (**G**) Classification of cell sex shown on our standard-analysis UMAP of epithelial cells. Note that male-originating and female-originated cells are interspersed within the UMAP without need for correction. Quantifications for the male-female classification within AMPs and epithelial cells are provided next to the UMAPs in (**E**) and (**G**), respectively. UMAP color scales correspond to normalized (by total UMI) counts on a natural-log scale.

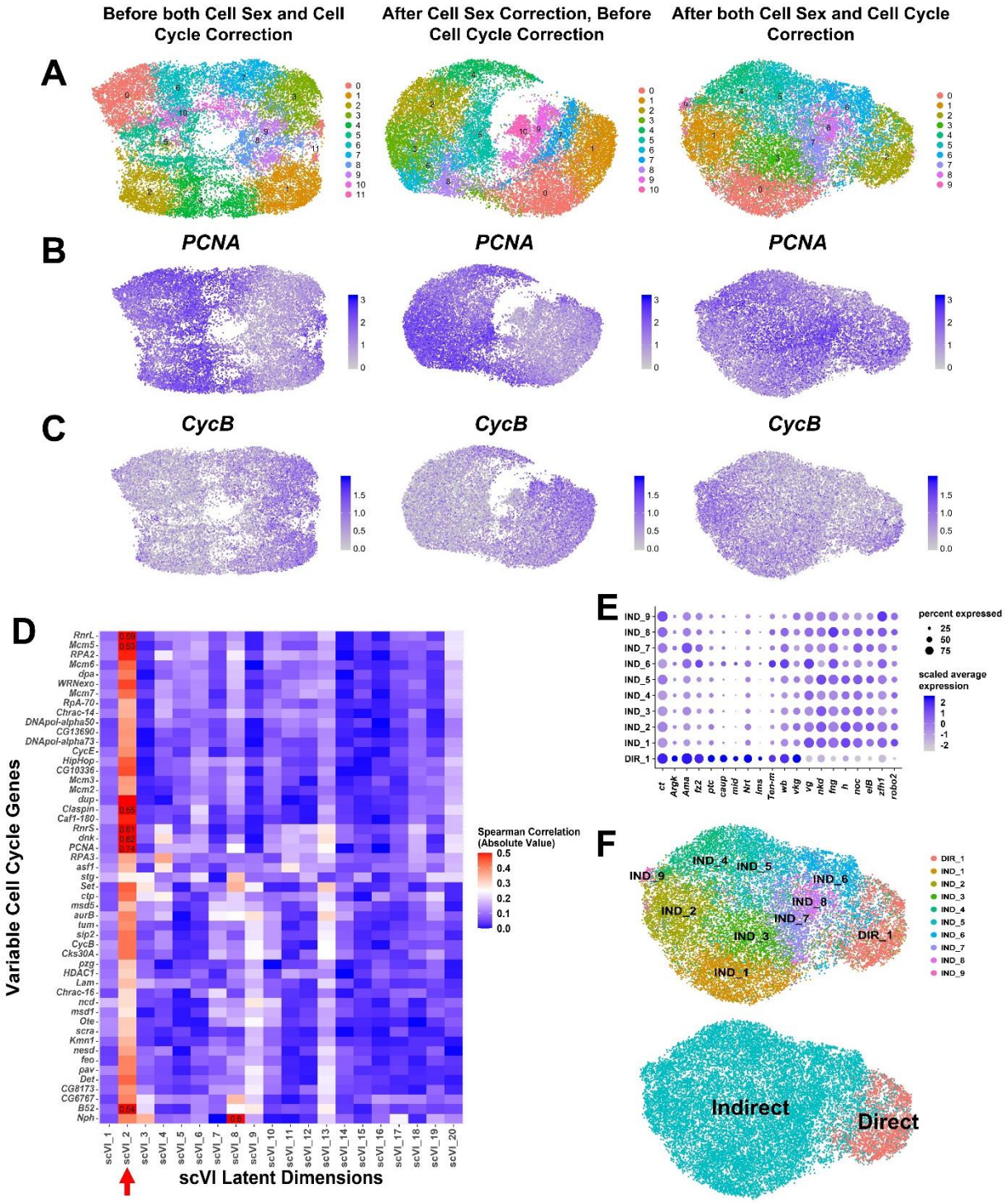

Figure 2.6. Correction of cell cycle stratification within AMP scRNAseq data.
(**A-C**) UMAPs of AMP single-cell data without any data correction (first column), after cell sex correction only (second column) (see **Materials and Methods** and **Figure 2.5** for details), and after both cell sex and cell cycle correction (third column). Cell cycle correction was performed by removing latent dimension scVI_2, which showed high correlation magnitude with variable cell cycle genes, from downstream analysis (e.g., data clustering and visualization) (see panel **D** of this figure and **Materials and Methods** for details). Cell colors correspond to unbiased

clustering identities (**A**), or expression levels of cell-cycle genes *PCNA* (**B**) or *Cyclin B (CycB)* (**C**). Note that after we correct for cell cycle stratification, we observe better mixing of cell cycle markers *PCNA* and *CycB* throughout the data. (**D**) Magnitude of Spearman correlation between the scVI latent dimensions and the expression of variable cell cycle genes within the data. Color scale for correlation magnitude is capped at 0.5 for better visualization; correlation magnitudes that exceed 0.5 are written in the corresponding box. Note that scVI_2 (indicated on the x-axis by the red arrow) displays noticeably high correlation magnitude with many variable cell cycle genes, highlighting that it is capturing most cell cycle variation within the data. (**E, F**) Classification of unsupervised clusters into direct and indirect identities. (**E**) Dot plot that summarizes the expression of canonical marker genes *ct* and *vg* in all clusters, in addition to other genes that displayed differential expression between cluster DIR_1 and all other clusters (FDR < 0.05, see **Materials and Methods**). Note the higher expression of *ct* and lower expression of *vg* in DIR_1 compared to other clusters. After data correction, we identify one cluster as representing direct AMPs (DIR_1), whereas all other clusters are classified as indirect AMPs (IND_1-9). When performing differential expression analysis on DIR_1 vs. all other clusters, there was a natural-log fold-change of 0.28 and -0.49 for *ct* and *vg*, respectively (FDR < 0.05 for both genes; positive natural-log fold-changes indicate higher expression in DIR_1, negative values indicate higher expression in all other clusters). (**F**) UMAPs of AMP single-cell data with cells colored by unsupervised clustering (as in **Figure 2.1N**) and by our classification into putative precursors of direct or indirect AFMs. UMAP color scales correspond to normalized (by total UMI) counts on a natural-log scale.

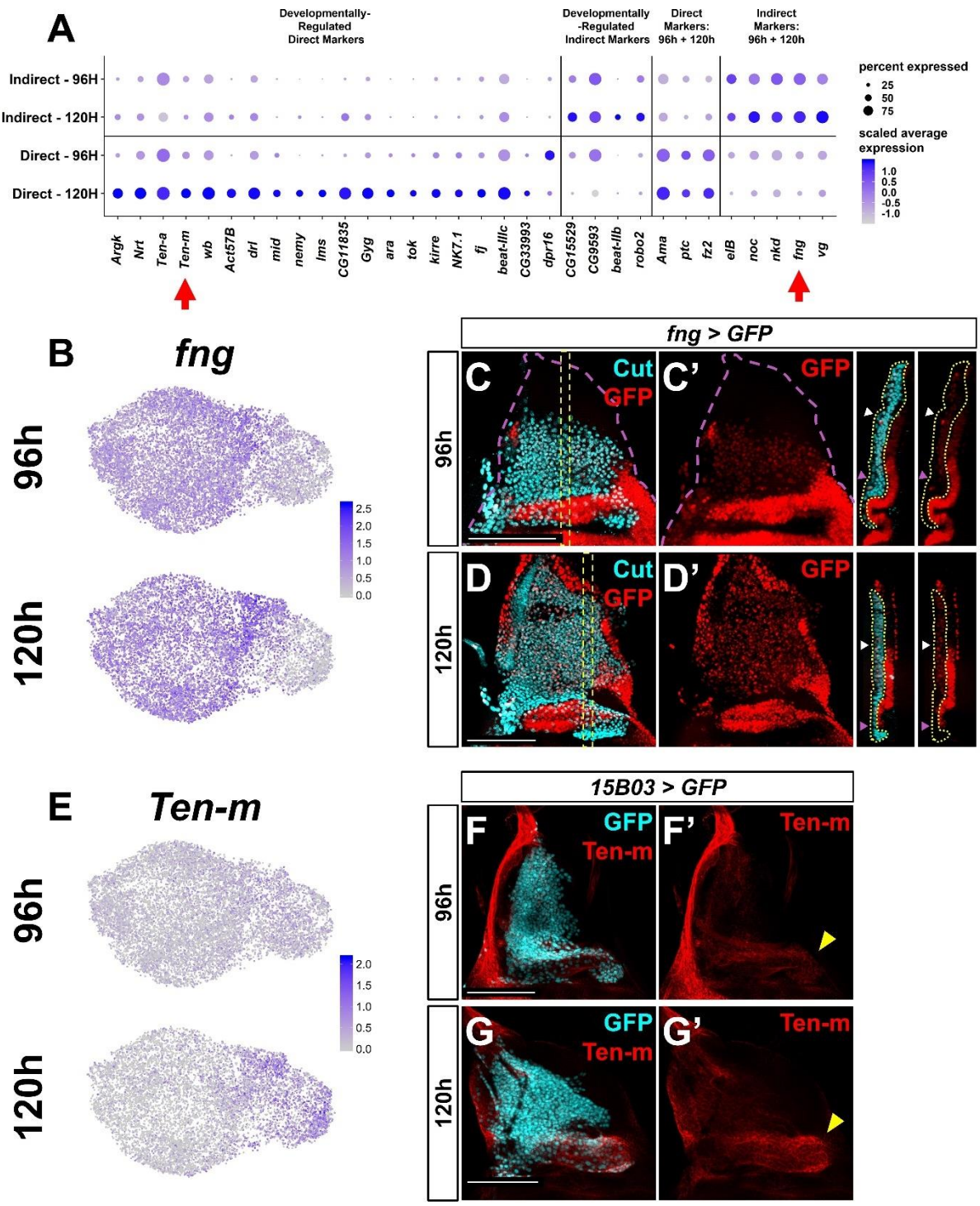

Figure 2.7. Temporal changes within the direct and indirect AMPs.
(**A**) Dot plot summarizing expression over time of genes that are differentially expressed within direct- or indirect-flight muscle precursors. Genes are grouped in the following manner: (1) genes that are differentially-expressed at higher levels within direct AMPs only at one time point ("Developmentally-Regulated Direct Markers"), (2) genes that are differentially-expressed at higher levels within indirect AMPs only at one time point ("Developmentally-Regulated Indirect Markers"), (3) genes that are differentially-expressed at higher levels within direct AMPs at both

time points ("Direct Markers: 96h + 120h"), and (4) genes that are differentially-expressed at higher levels within indirect AMPs at both time points ("Indirect Markers: 96h + 120h"). Arrows highlight *fng* and *Ten-m*, which are investigated in the following panels. (**B, E**) UMAPs of *fng* (**B**) and *Ten-m* (**E**) expression at 96h and 120h within AMPs. (**C-D**) Wing discs with GFP expression driven by *fng-Gal4* transgene at 96h (**C**) and 120h (**D**). AMPs are visualized with anti-Cut. Magenta dashed lines in **C** and **C'** provide an outline of the wing disc. Orthogonal sections correspond to orthogonal max projections within the yellow dashed boxes in **C** and **D**. Yellow dotted lines within the orthogonal sections (apical is left, basal is right) outline the AMPs. White and magenta arrowheads indicate regions of high and low GFP fluorescence, respectively. Note the higher levels of GFP in more dorsal AMPs (white arrowheads) compared to more ventral AMPs (magenta arrowheads) at both time points. (**F-G**) Wing discs stained for anti-Ten-m at 96h (**F, F'**) and 120h (**G, G'**). AMPs are visualized by expression of GFP via *15B03-Gal4* driver (see **Figure 2.10I**). Yellow arrowheads indicate regions of higher anti-Ten-m staining in direct AMPs. Note the increased levels of staining at 120h as compared to 96h. UMAP color scales correspond to normalized (by total UMI) counts on a natural-log scale. Images shown in **C** and **D** are single image slices containing AMPs, whereas images shown in **F** and **G** are max projections across image slices containing AMPs. Microscopy scale bars = 100 μm.

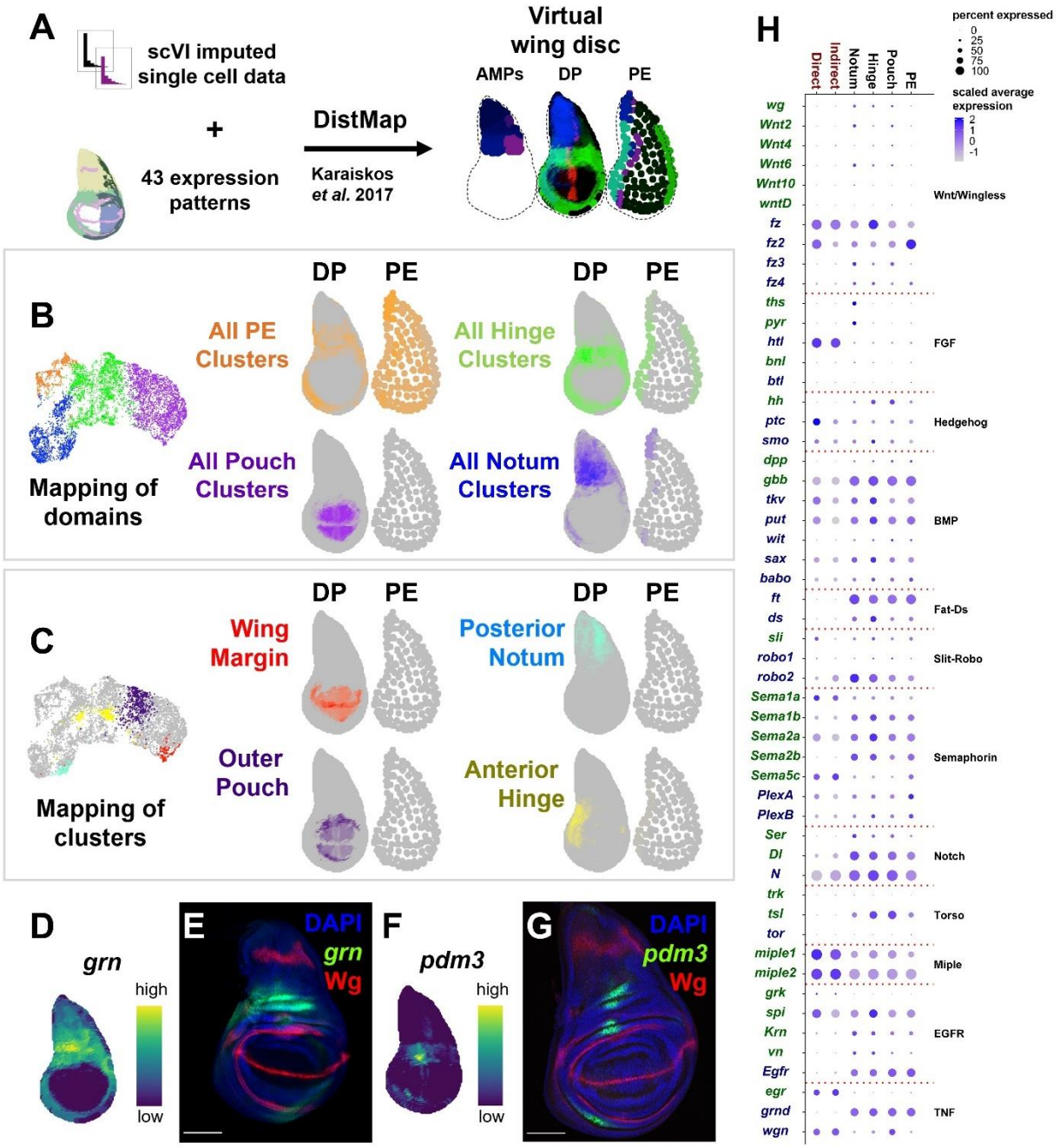

Figure 2.8. Spatial mapping of single-cell data to a virtual wing disc and analysis of receptor-ligand expression.

(**A**) Schematic describing the creation of a three-layered virtual wing disc (AMPs, disc proper (DP), and the peripodial epithelium (PE)) using DistMap (Karaiskos *et al.*, 2017) (see **Materials and Methods**). In contrast to the columnar cells of the DP, much of the PE is composed of squamous cells with flattened nuclei, and it is therefore represented as an outline which contains large dots. The virtual wing disc can be used to predict gene expression patterns (or virtual *in situs*), as shown for three example genes (*ptc* in red, *zfh2* in green, and *eyg* in blue). (**B, C**) Cells from the epithelial domains (**B**) and particular sub-regions (**C**) are shown both on the UMAP and mapped onto the DP and PE virtual wing disc. Stronger colors indicate higher predicted mapping;

gray indicates low predicted mapping for cells (see **Figure 2.9 A, B** for mappings of all cell clusters). (**D, F**) Predicted gene expression patterns of *grn* (**D**) and *pdm3* (**F**) in the epithelium disc proper, neither of which were used in building the virtual model. Yellow and dark purple correspond to high and low predicted expression, respectively. (**E, G**) Late 3rd instar wing-imaginal disc with transcriptional reporters for the genes *grn* (**E**) and *pdm3* (**G**). Note the relative similarity between the predicted expression and transcriptional reporters. (**H**) Dot plot summarizing the expression of genes encoding receptors and ligands from pathways that were differentially-expressed in at least one cell type. Dot size indicates the percent of cells that express the gene and the dot color indicates the relative gene expression level, within each of the cell groups. X-axis: Cell groups. Disc epithelium cell types are in black font, AMP cell types are in red font. Y-axis: Genes are either in blue or green font depending on their annotation as encoding for a receptor or ligand, respectively. Microscopy scale bars = 100 μm.

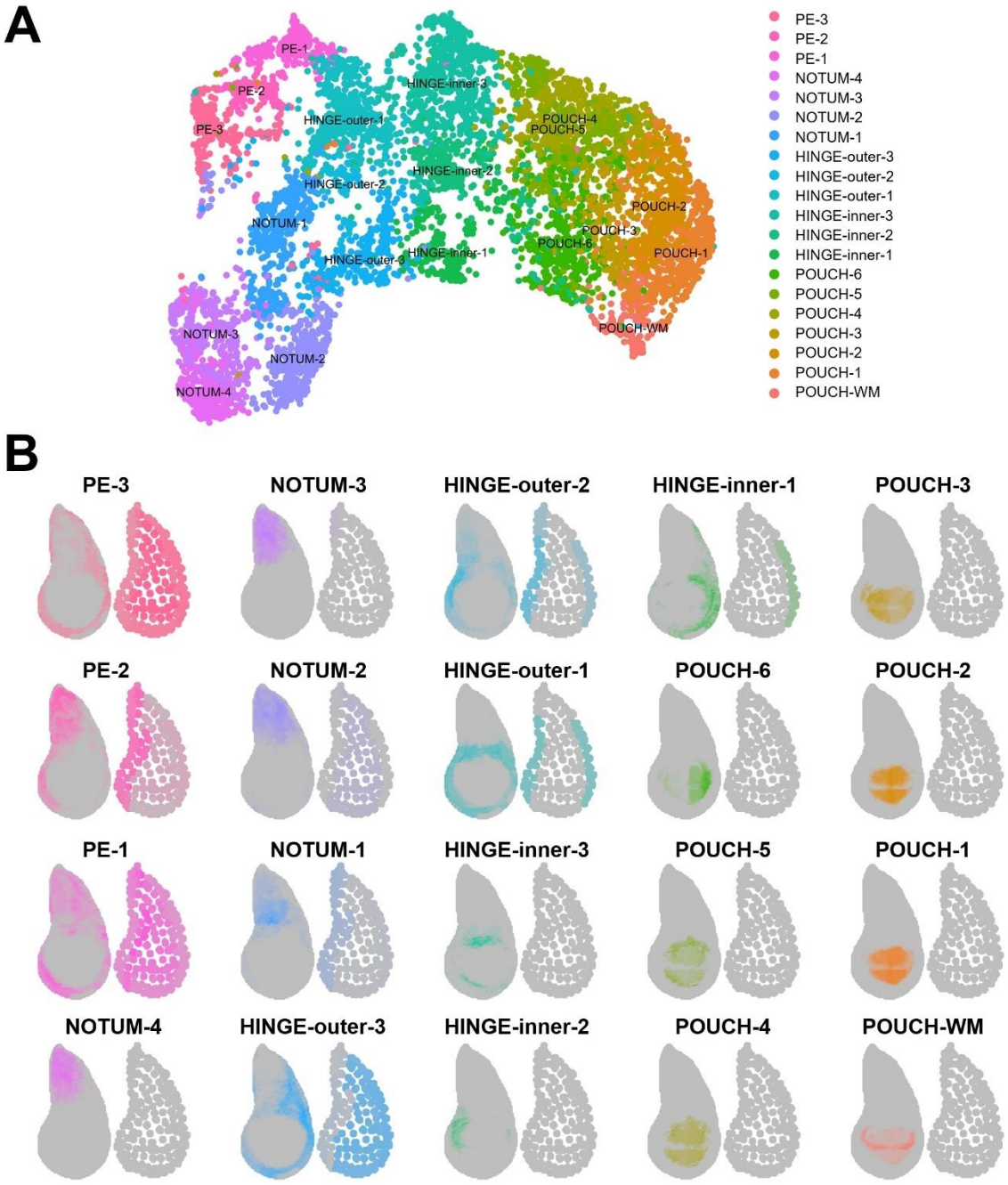

Figure 2.9. Mapping of epithelium cell clusters to the virtual wing disc.
(**A**) UMAP of epithelium cell clusters (same as the UMAP in **Figure 2.1H**). (**B**) Visualization of where epithelium cell clusters best map to the reference model for the disc proper layer (on left) and the peripodial epithelium layer (on right). Gray regions indicate low predicted mapping, whereas regions with darker color shades indicate higher predicted mapping. The mapping colors in **B** match the cluster colors in **A**.

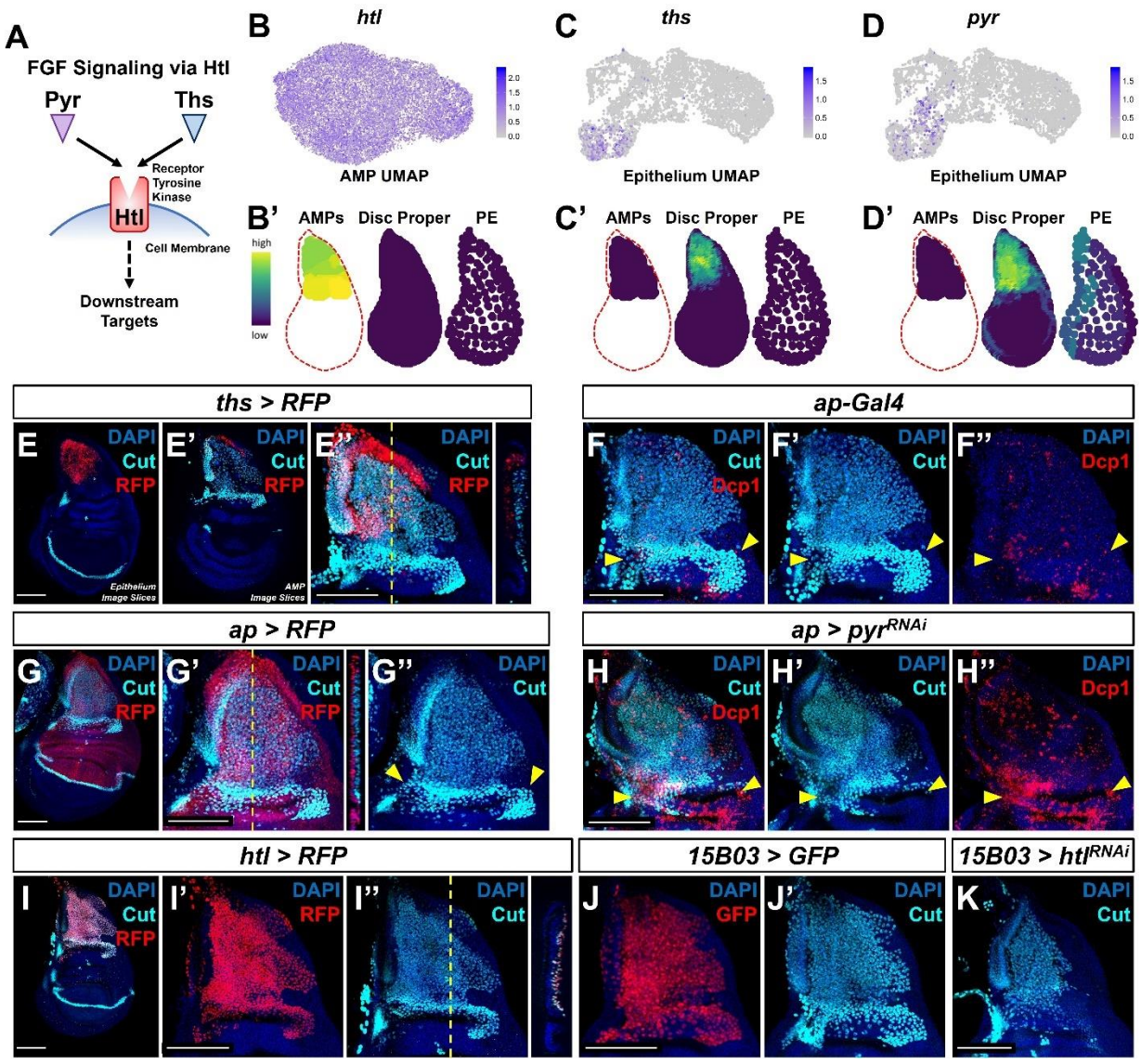

Figure 2.10. FGF signaling between AMPs and epithelium is critical for AMP viability and numbers.

(**A**) FGF signaling pathway diagram. FGF ligands (Pyr and Ths) signal through the FGFR (Htl). (**B-D**) Expression of *htl* (**B**), *ths* (**C**), and *pyr* (**D**) in the single-cell data. **B**, **C**, and **D** show UMAPs of these genes, in either the AMPs (for *htl*) or disc epithelium (for *ths* and *pyr*). **B'**, **C'**, and **D'** show virtual disc map predictions for the expression of these genes in the AMPs, disc proper, and peripodial epithelium. (**E**) *ths* expression domain (as shown by *ths-Gal4* driving the expression of RFP) in the wing disc. **E**, **E'**, and **E''** are max projections over image slices of the disc epithelium, images slices of the AMPs, and all image slices (both epithelium and AMPs), respectively. AMPs are visualized with anti-Cut (cyan). Orthogonal section (apical is left, basal is right) corresponds to the dashed yellow line in **E''**. Note that *ths* reporter expression is specific to the notum epithelium and absent from the AMPs. (**F-H**) Notum regions from wing discs with *ap-Gal4* transgene alone (**F**), or *ap-Gal4* transgene driving expression of either >*RFP* (**G**) or >pyr^{RNAi} (**H**). Discs shown in **F** and **H** are stained with anti-Dcp1 (red) to visualize cell death. Dashed yellow line in **G'** corresponds to the adjacent orthogonal section. Yellow arrowheads indicate expected location of direct AMPs. Note the loss of ventral- and posterior-localized AMPs following *pyr*

knock-down, with increased anti-Dcp1 staining (**H**). (**I**) *htl* expression domain (as shown *htl-Gal4* driving the expression of RFP) in the wing disc. Orthogonal section corresponds to the dashed yellow line in **I''** (contrast with the orthogonal section in **E**) . Note that the *htl* reporter is expressed by the AMPs which are visualized by anti-Cut staining (cyan). (**J, K**) The notum regions from wing discs with AMP-specific *15B03-Gal4* driving the expression of *>GFP* (**J**) and $>htl^{RNAi}$ (**K**). AMPs visualized with anti-Cut. Note the reduction of AMPs, especially of the direct AMPs. UMAP color scales correspond to normalized counts on a natural-log scale. All notum images are max projections across image slices. Microscopy scale bars = 100 μm.

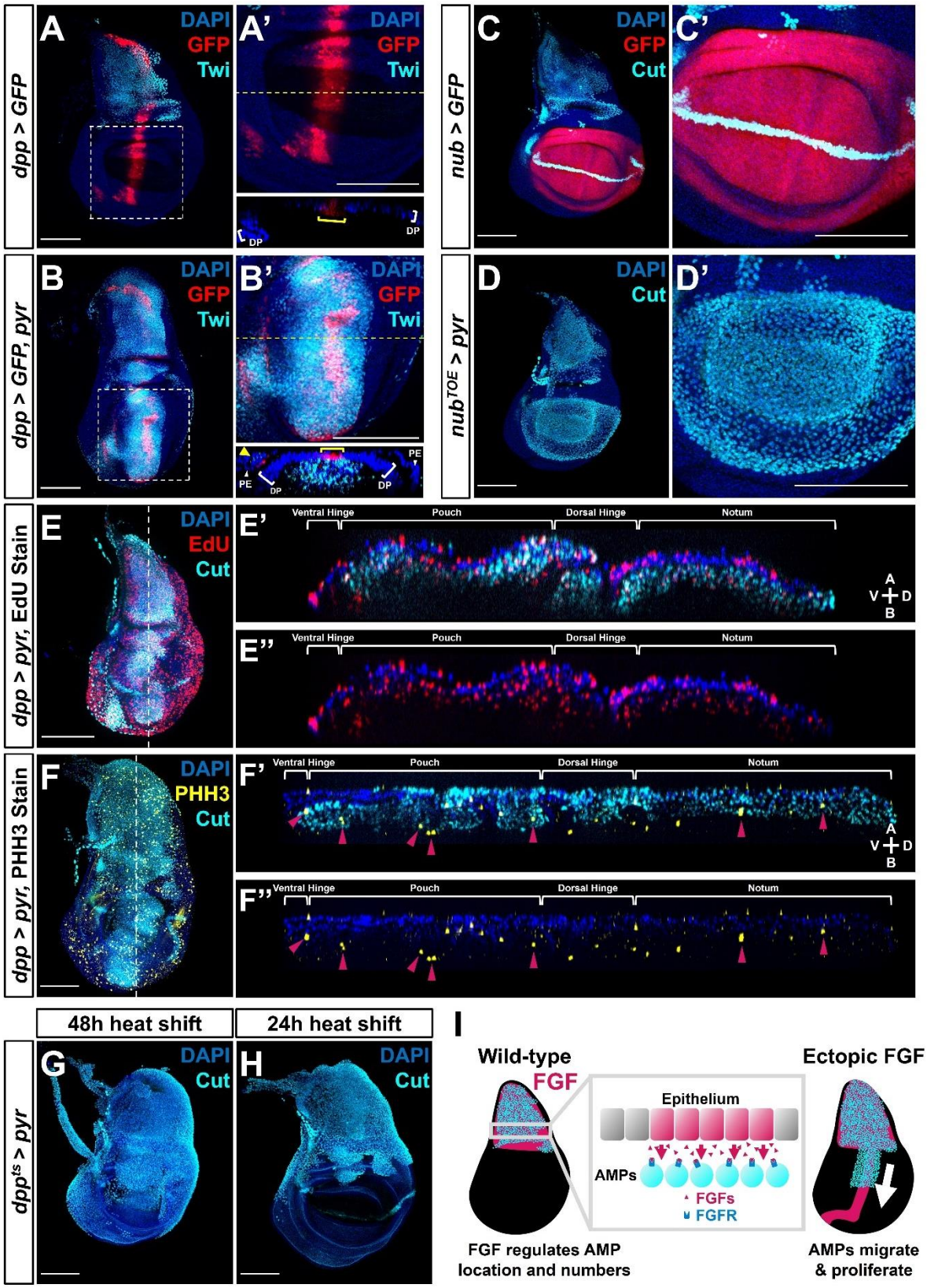

Figure 2.11. FGF from disc epithelium controls AMP localization.

(**A, B**) Wing discs with *dpp-Gal4* driving the expression of *>GFP* alone (**A**) or *>GFP* together with *>pyr* (**B**), stained with anti-Twi (cyan) to visualize the AMPs. **A'** and **B'** correspond to white dashed boxes in **A** and **B**, respectively. Orthogonal sections (apical is top, basal is bottom, relative to DP) correspond to dashed yellow lines in **A'** and **B'**. Yellow brackets in orthogonal sections indicate *dpp-Gal4* expression in disc proper. Yellow arrowhead in the orthogonal of **B** indicates *dpp-Gal4* expression in the PE, which recruits AMP expansion to the PE surface when expressing *>pyr*. (**C, D**) Wing discs with *nub-Gal4* driving the expression of *>GFP* (**C**) or *>dCas9^{VPR}* (**D**), the latter being used in conjunction with a guide RNA targeting an upstream sequence of the *pyr* transcriptional start site (*pyr$^{TOE.GS00085}$*) to drive the overexpression of *pyr* in the wing pouch. Discs are stained with anti-Cut to visualize AMPs. Note that Cut is also expressed in the future wing margin of the disc epithelium (seen as a band through the wing pouch in **C** and **C'**). (**E, F**) Discs overexpressing *pyr* via *dpp > pyr*, stained for Cut (cyan) and either EdU incorporation (red) (**E**) or phosphohistone H3 (PHH3) (yellow) (**F**) to assess the ability of AMPs to replicate DNA and undergo mitosis. Magenta arrowheads in **F'** and **F''** highlight some AMPs that stain for PHH3. White brackets denote approximate domains of the disc proper. Note that ectopic AMPs stain for both EdU incorporation and PHH3, indicating that these cells are viable and proliferating outside of the endogenous AMP niche. A = apical; B = basal; V = ventral; D = dorsal, relative to disc proper. (**G, H**) Temperature-controlled expression of *>pyr* within the *dpp* domain, initiated at either mid L3 (**G**) or late L3 (**H**) (48 or 24 hours prior to pupariation, respectively). Note that even at these developmental stages, we observed ectopic AMPs that appear to be emigrating ventrally from the notum. (**I**) Model for the effects of FGF overexpression on AMP growth. FGF/FGFR interactions between the disc epithelium and adjacent AMPs are necessary for AMP viability, and ectopic expression of FGF ligands induces emigration of AMPs from the notum to a domain that broadly matches the pattern of FGF ligand expression. All wing disc images are max projections across all image slices. Microscopy scale bars = 100 μm.

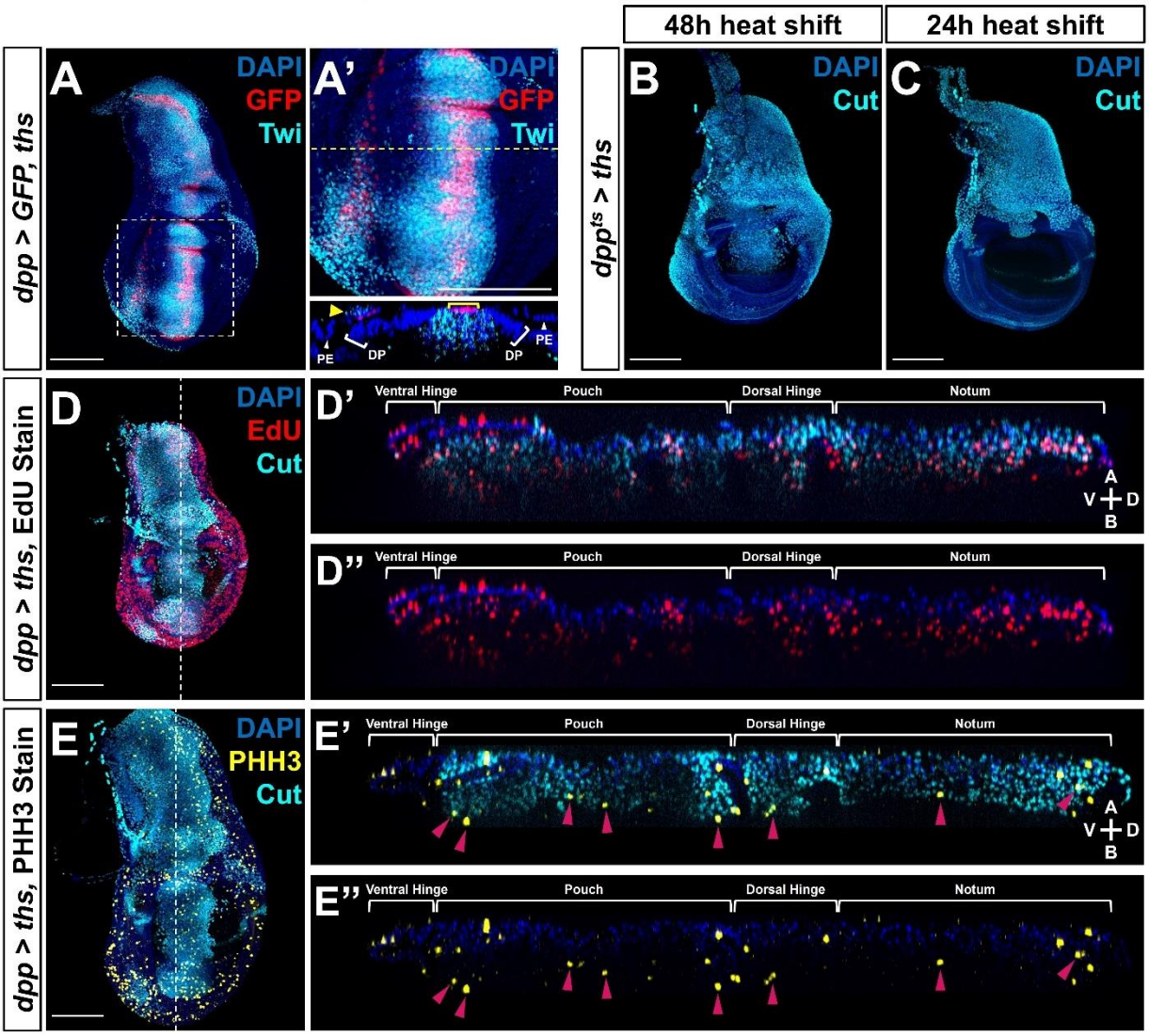

Figure 2.12. Ectopic Ths expression increases AMP number and migration.
(**A**) Wing discs with *dpp-Gal4* driving the expression of both >*GFP* and >*ths*, stained with anti-Twi to visualize the AMPs. **A'** corresponds to the white dashed boxes in **A**. Orthogonal section (apical is top, basal is bottom, relative to DP) corresponds to dashed yellow lines in **A'**. Yellow arrowhead in the orthogonal of **A'** indicates *dpp-Gal4* expression in the PE, which recruits AMP expansion to the PE surface when expressing >*ths*. (**B, C**) Temperature-controlled expression of >*ths* within the *dpp* domain, initiated at either mid L3 (**B**) or late L3 (**C**) (48 or 24 hours prior to pupariation, respectively). Note that even at these developmental stages, we observed ectopic AMPs that appear to be emigrating ventrally from the notum, similar to our results obtained with >*pyr* expression (see **Figure 2.11G, H**). (**D, E**) Discs overexpressing *ths* via *dpp* > *ths*, stained for Cut (cyan) and either EdU incorporation (red) (**D**) or phosphohistone H3 (PHH3) (yellow) (**E**) to assess the ability of AMPs to replicate DNA and undergo mitosis. Magenta arrowheads in **E'** and **E''** highlight some AMPs that stain for PHH3. White brackets denote approximate domains of the disc proper. Note that ectopic AMPs stain for both EdU incorporation and PHH3, indicating that these cells are viable and proliferating outside of the endogenous AMP niche. A = apical; B = basal; V = ventral; D = dorsal, relative to disc proper.

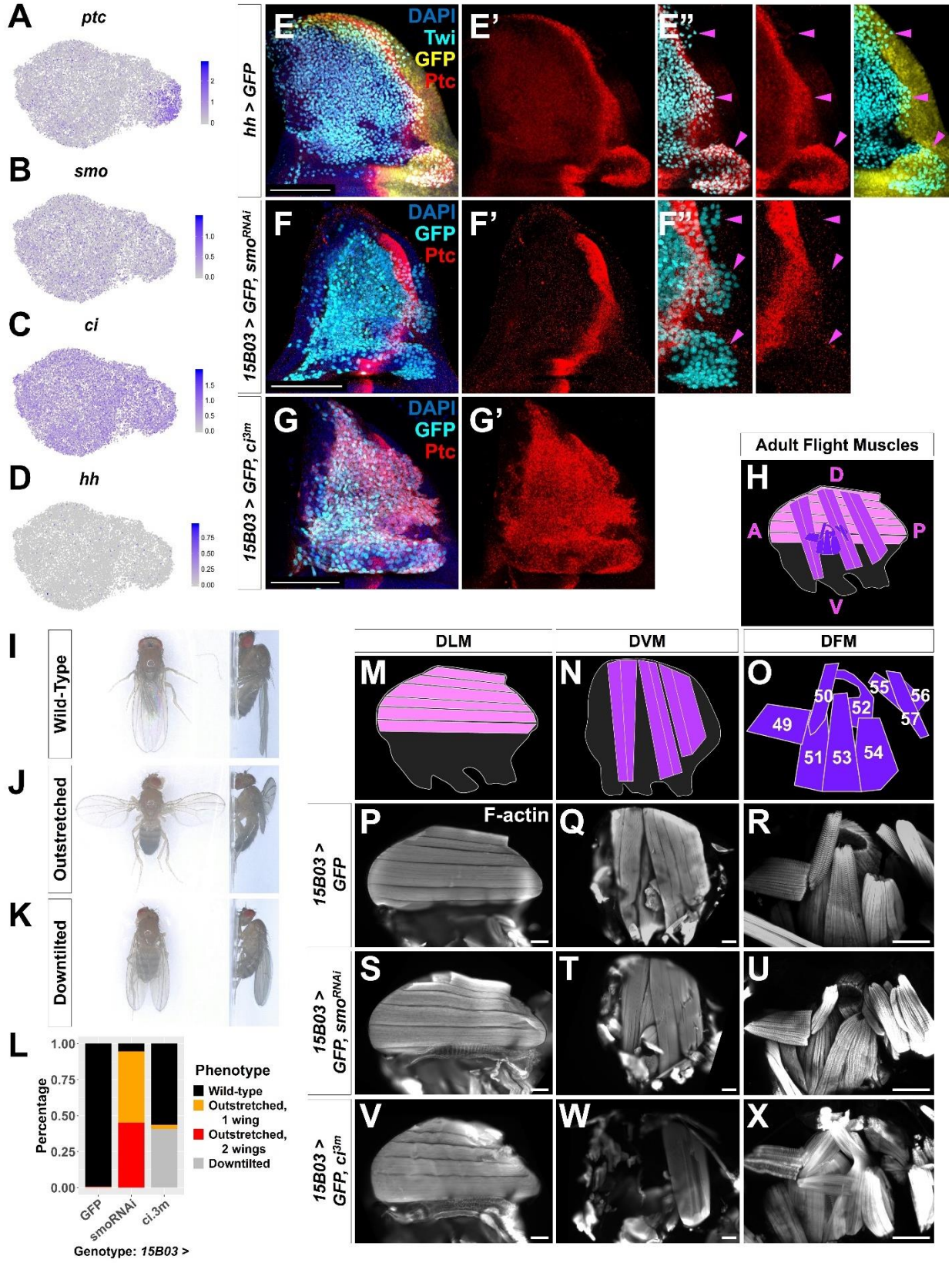

Figure 2.13. Hh signal from the disc epithelium patterns a subset of AMPs.

(**A-D**) UMAPs of *ptc* (**A**), *smo* (**B**), *ci* (**C**), and *hh* (**D**) expression in AMPs. (**E**) Notum of wing discs stained for anti-Ptc (red) with the *hh*-expressing epithelium cells marked by *hh-Gal4* driving *>GFP* (yellow). AMPs visualized by anti-Twi stain (cyan). Note Ptc expression in the posterior-localized AMPs, to the right of the epithelium Ptc stripe and in close proximity to the *hh*-expressing epithelium (magenta arrowheads) (**E''**) (see **Figure 2.14**). (**F, G**) Notum of wing discs stained for anti-Ptc (red) with *15B03-Gal4* driving *>GFP* together with either $>smo^{RNAi}$ (BL43134) to reduce Hh signaling within the AMPs (**F**) or $>ci^{3m}$ to mimic activated Hh signaling within the AMPs (**G**). Ptc expression is significantly reduced in the posterior-localised AMPs after *smo* knockdown (magenta arrowheads) without disrupting the stripe of Ptc expression in the epithelium (**F''**). Note that Ptc is expressed in all of the AMPs following $>ci^{3m}$ expression (**G'**). (**H**) Schematic of adult flight muscle fibers within the thorax where the muscle subtypes are differentially shaded: DLMs in pink, DVMs in purple, and DFMs in dark purple. (**I-K**) Wing posture phenotypes observed following Hh-signaling perturbations. Adults were imaged live, not anesthetized. (**I**) Wildtype posture, with wing blades folded along their dorsum. (**J**) Outstretched wing posture, where either one or both wings were always held perpendicular to the body axis. (**K**) Downtilted wing posture, with adults that hold their wings farther apart along their dorsum and tilted laterally downward. (**L**) Quantification of wing posture after Hh-signaling perturbation within AMPs (driven by *15B03-Gal4*). The number of adults assayed: *>GFP* = 341, $>smo^{RNAi}$ (BL43134) = 283, and $>ci^{3m}$ = 366. These smo^{RNAi} results were replicable with multiple RNAi lines (**Figure 2.15A**). (**M-O**) Separate schematics of expected DLMs (**M**), DVMs (**N**), and DFMs (**O**) morphology. Numbers on DFMs represent the canonical labels for the different fibers. (**P-X**) Adult flight muscles (visualized with F-actin staining) from animals with *15B03-Gal4* driving *>GFP* alone (**P-R**), or *>GFP* together with either $>smo^{RNAi}$ (**S-U**) or $>ci^{3m}$ (**V-X**). DLMs are shown in **P**, **S**, and **V**; DVMs are shown in **Q**, **T**, and **W**; DFMs are shown in **R**, **U**, and **X**. Adult flight muscles in *>GFP* flies had similar morphology in all adults examined (23 DLMs, 15 DVMs, and 12 DFMs). Adult flight muscles in $>smo^{RNAi}$ (BL43134) animals displayed abnormal DFMs (11/11 had muscles 53 and 54 misaligned, and 7/11 had muscles 55, 56, and 57 malformed) (**U**), while DLMs (n = 21/22) and DVMs (n=10/10) had relatively normal morphology (**S, T**). Adult flight muscles in $>ci^{3m}$ animals had normal DLMs morphology (n=7) (**V**), whereas the DVMs were either missing or severely disconnected (n=7) (**W**) and the DFMs appeared abnormal (n=4) (**X**). UMAP color scales correspond to normalized counts on a natural-log scale. All notum images are max projections across image slices containing AMPs. Microscopy scale bars = 100 μm.

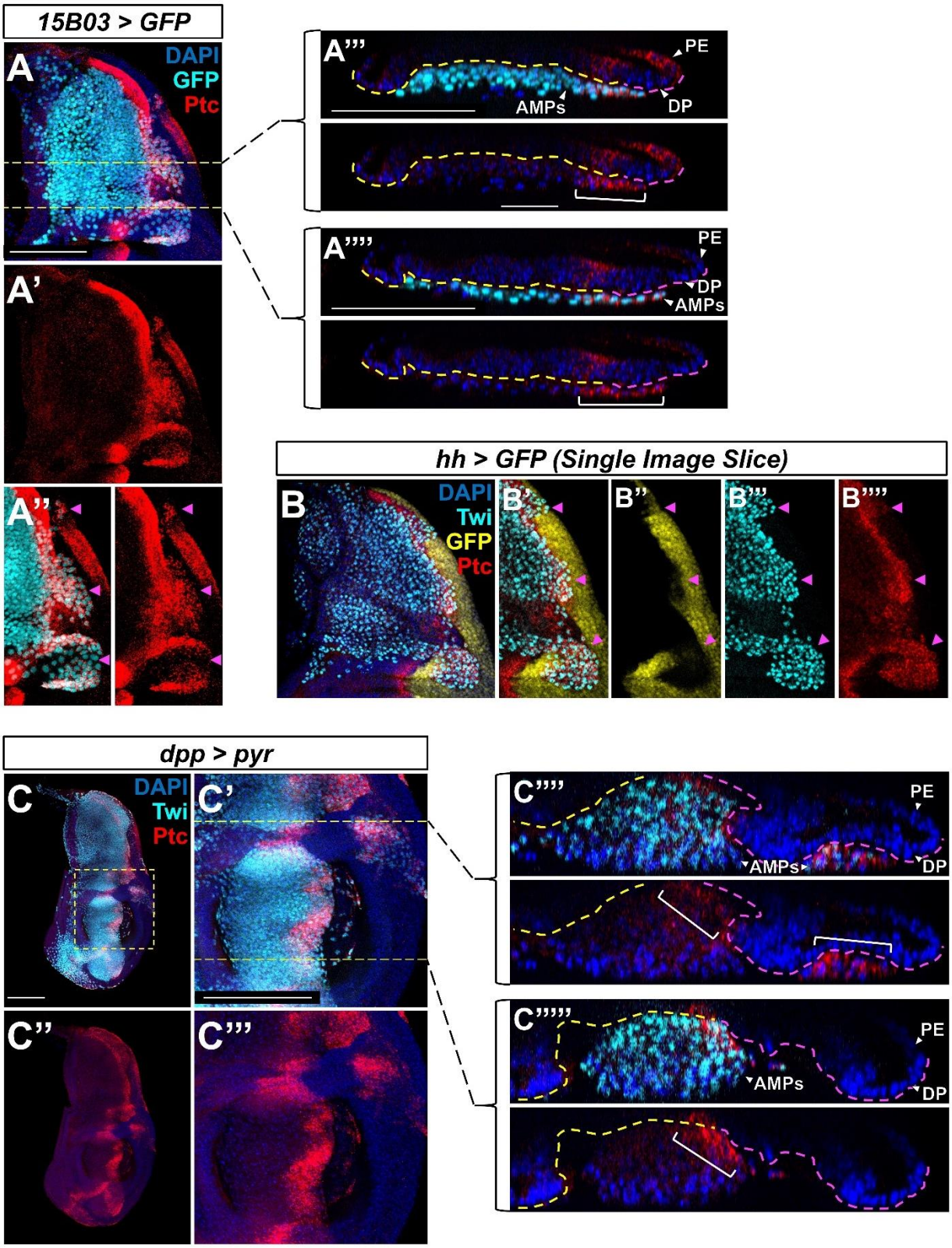

Figure 2.14. Ptc-expressing AMPs are neighboring the Hh-producing posterior compartment of the disc proper.

(**A**) Notum of wing discs stained for anti-Ptc (red) with *15B03-Gal4* driving *>GFP*. Magenta arrowheads in **A'' ** indicate posterior-localized AMPs that display high Ptc expression. Orthogonal sections **A'''** and **A''''** correspond to yellow dashed lines in (**A**). In these orthogonal sections, the interface between the AMPs and DP is outlined by a yellow and magenta dashed line (yellow = interface with anterior DP compartment, magenta = interface with posterior DP compartment). Note that in both **A'''** and **A''''**, we observe high Ptc expression in AMPs underneath the posterior compartment, extending a short distance underneath the anterior compartment (indicated by a white bracket). We estimate a distance of approximately 20-40 μm between the Hh-producing posterior compartment of the disc epithelium and the anterior-most AMP with high Ptc expression. (**B**) Notum of wing discs stained for anti-Ptc (red) and anti-Twi (cyan), with the *hh*-expressing epithelium cells marked by *hh-Gal4, >GFP* (yellow). A single image slice (rather than a max projection of all image slices, as in **Figure 2.13E**) is used to illustrate that the posterior-localized AMPs (indicated by magenta arrowheads) do not produce Hh ligand themselves, but are surrounded by Hh-producing epithelial cells. (**C**) Wing discs with *dpp-Gal4* driving the expression of *>pyr*, stained with both anti-Ptc (red) and anti-Twi (cyan). **C'** and **C'''** corresponds to the dashed yellow box in **C**, and orthogonal sections **C''''** and **C'''''** correspond to yellow dashed lines in **C'**. In **C''''** and **C'''''**, the interface between the AMPs and DP is outlined by a yellow and magenta dashed line (yellow = interface with anterior DP compartment, magenta = interface with posterior DP compartment). Note that even in these ectopic AMPs, we only observe high levels of Ptc in AMPs in close proximity to the Hh-secreting posterior compartment of the DP (such AMPs are indicated by white brackets).

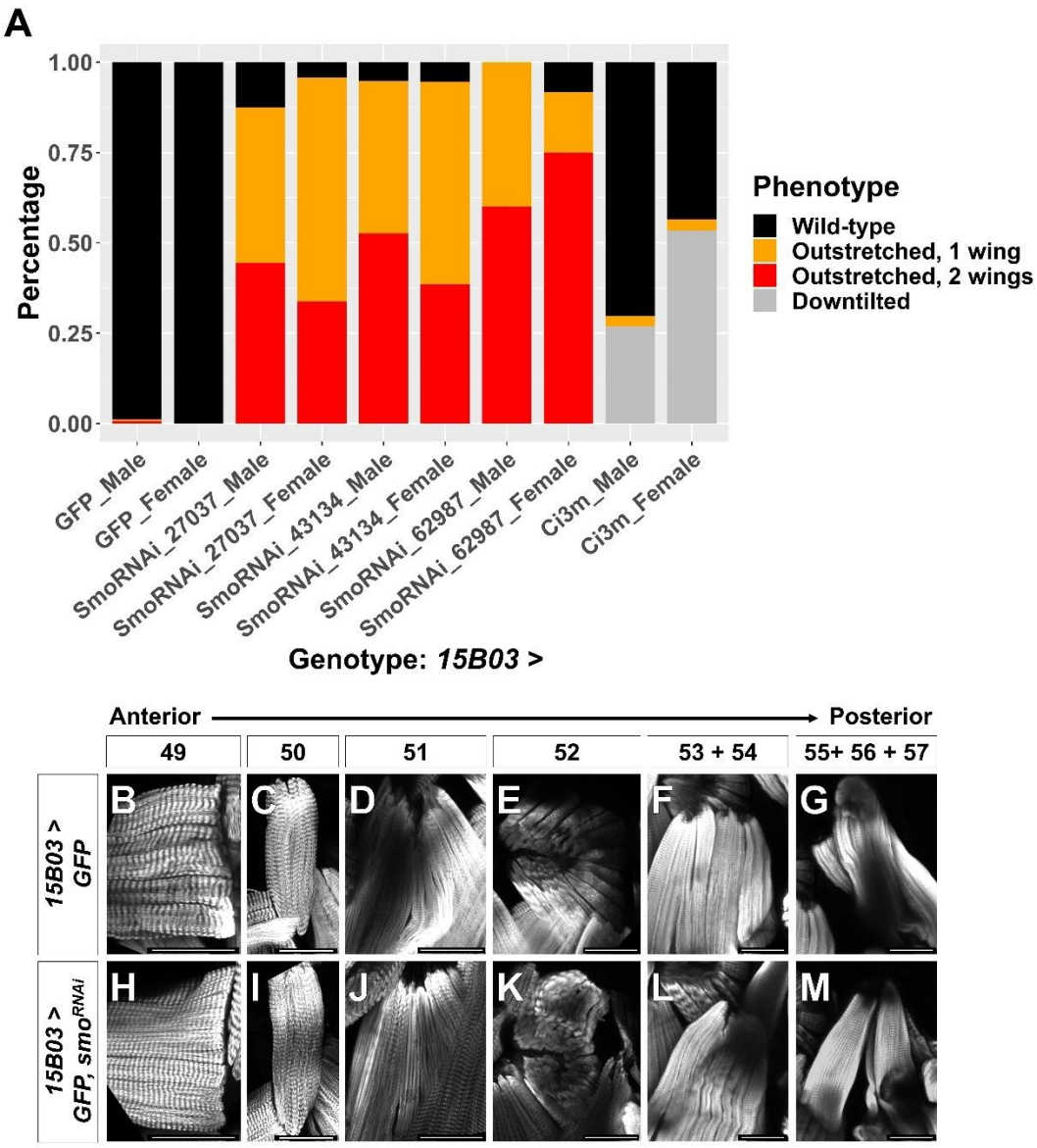

Figure 2.15. Adult wing-posture phenotypes and morphology of individual muscle fibers after Hh-signaling perturbation.

Quantification of observed wing posture phenotypes under conditions of Hh-signaling perturbation within AMPs (for description of wing posture phenotypes, see **Figure 2.13**). All *UAS* lines are driven by *15B03-Gal4*. Data are presented for three different *UAS-smo*^{RNAi} lines and separated by sex. Wing posture phenotypes were fairly consistent between males and females. Number of flies examined: GFP: 164 (male) and 179 (female), *smo*^{RNAi} BL27037: 72 (male) and 71 (female), *smo*^{RNAi} BL43134: 135 (male) and 148 (female), *smo*^{RNAi} BL62987: 20 (male) and 12 (female), *ci*^{3m}: 175 (male) and 191 (female) (**B-M**) DFMs in adults where *15B03-Gal4* drives >*GFP* alone (**B-G**) or >*GFP* together with >*smo*^{RNAi} (**H-M**). Muscle fibers are shown in the order of relative anterior-posterior positioning within the thorax, with numbering nomenclature as described in **Figure 2.13O**. Note that in >*smo*^{RNAi} flies, the posterior edge of muscle 52 appears to be truncated (compare **K** with **E**); muscles 53 and 54 are indistinguishable and both project to the dorsal attachment site of muscle 54 (compare **L** with **F**); and muscles 55, 56, and 57 are disorganized (compare **M** with **G**). Microscopy scale bars = 50 μm.

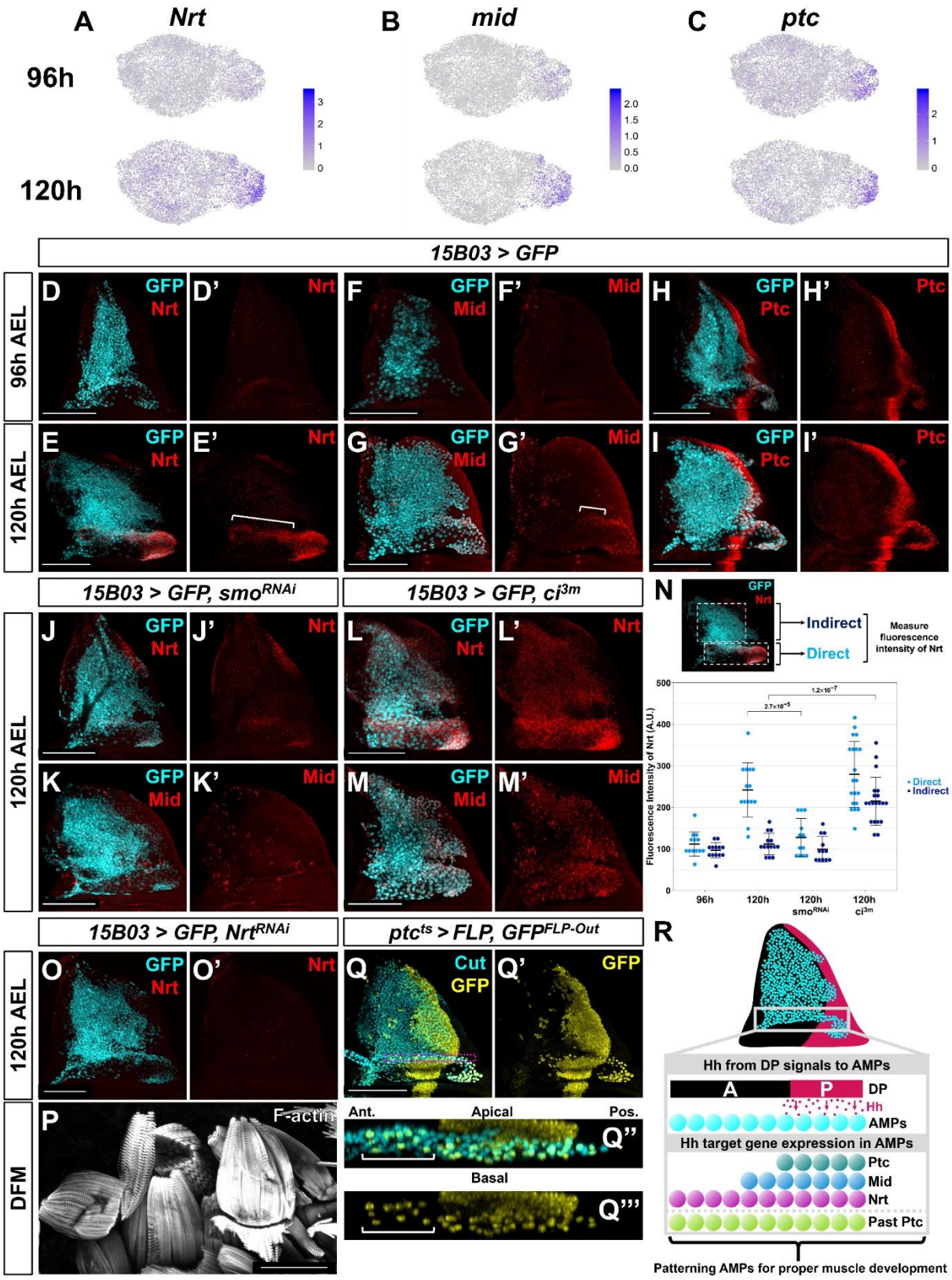

Figure 2.16. Nrt and Mid are novel downstream Hh-pathway targets in the AMPs.

(**A-C**) UMAPs of *Nrt* (**A**), *mid* (**B**), and *ptc* (**C**) expression at 96h and 120h in AMPs. Note the increase in expression for *Nrt* and *mid* from 96h to 120h, whereas *ptc* expression is relatively unchanged. (**D-I**) Wing discs stained with anti-Nrt at 96h (**D**) and 120h (**E**), anti-Mid at 96h (**F**) and 120h (**G**), and anti-Ptc at 96h (**H**) and at 120h (**I**). AMPs are visualized by expression of GFP (cyan) via *15B03-Gal4* driver. Note the negligible staining of anti-Nrt and anti-Mid in AMPs at 96h, matching the scRNAseq expression data. (**J-M**) Wing discs expressing either $>smo^{RNAi}$ to reduce Hh signaling within the AMPs (**J, K**) or $>ci^{3m}$ to activate Hh signaling within the AMPs via *15B03-Gal4* driver, along with *>GFP* to visualize AMPs (cyan). Discs are either stained with anti-Nrt (**J, L**) or anti-Mid (**K, M**). Note that the knockdown of *smo* prevents the expression of Nrt and Mid in the direct AMPs and that the overexpression of ci^{3m} leads to ectopic expression in the indirect AMPs. (**N**) Quantification of anti-Nrt staining within direct and indirect AMPs. The graph shows binned values of average fluorescent intensity. P-values were calculated from unpaired t-tests and error bars indicate standard deviation. Note that *smo* knockdown prevents Nrt expression within the direct AMPs at 120h, and the expression of activated *ci* increases Nrt expression in both the direct and indirect AMPs. (**O**) Wing disc expressing $>Nrt^{RNAi}$ in AMPs via *15B03-Gal4* driver, stained with anti-Nrt at 120h. Note that the knockdown of *Nrt* eliminates Nrt staining in the AMPs (**O'**). (**P**) DFMs in adults where *15B03-Gal4* drives $>Nrt^{RNAi}$. Note the enlarged posterior DFMs, specifically muscles 55, 56, and 57 (similar phenotypes were observed in all 5 flies examined) (compare to the control in **Figure 2.13R**). (**Q**) Lineage tracing of AMPs that have previously expressed *ptc* earlier in development; genotype: *ptc-Gal4, tub-Gal80*^{ts}*, >FLP, Ubi-FRT-stop-FRT-GFP* (ptc^{ts}*>FLP, GFP*$^{FLP\text{-}Out}$). Cells that expressed *ptc-Gal4* when larva are shifted to higher temperature (non-permissive temperature for Gal80^{ts}) and their progeny will be permanently labeled by GFP expression. The temperature shift from 18 ^{o}C to 30 ^{o}C was done at 5 days AEL for 24h, larvae were dissected at late 3rd instar. AMPs were visualized with anti-Cut staining. Orthogonal max projection is shown in **Q''** and **Q'''**, corresponding to the dashed purple box in **Q**. Note that a subset of the anterior labeled AMPs (indicated by white bracket) expresses GFP. (**R**) Model explaining the protein expression of Hh-signaling targets in AMPs. Posterior-localized AMPs receive Hh from the DP, activating expression of Hh-signaling targets Ptc, Nrt, and Mid (the latter two being AMP-specific targets). These AMPs migrate anteriorly, either a result of active cell movement or due displacement caused by proliferation. Anterior-localized AMPs quickly degrade Ptc protein, but Nrt and Mid perdure longer.

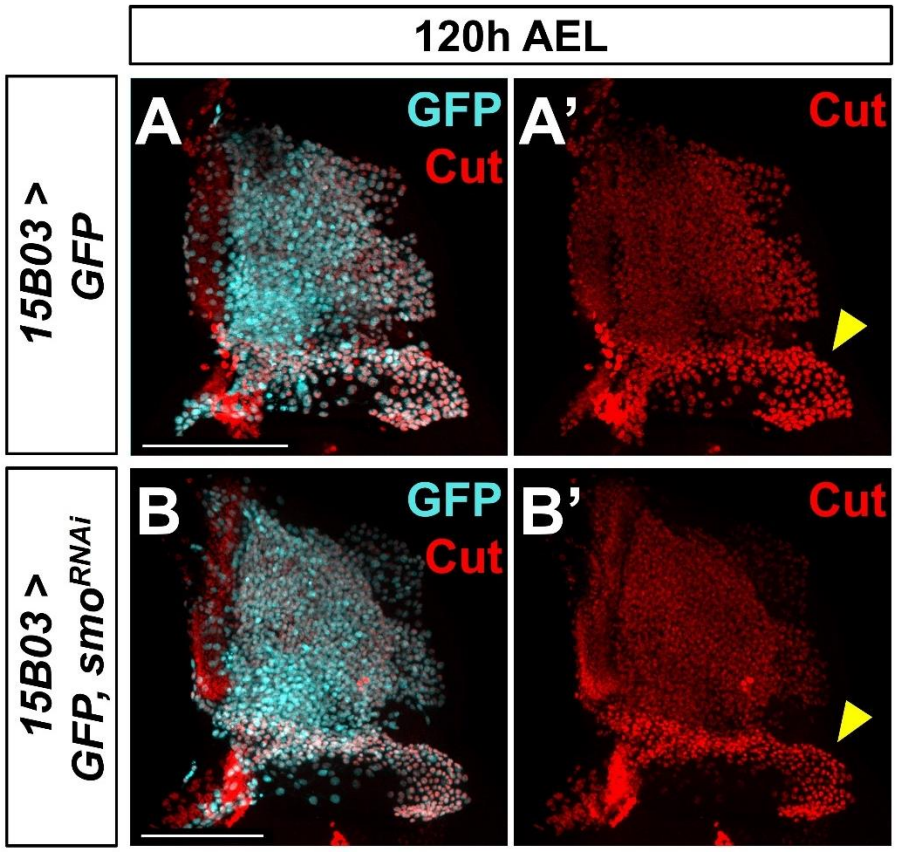

Figure 2.17. Knockdown of *smo* does not affect Ct protein levels in AMPs.
(**A, B**) Notum of wing discs stained for anti-Cut (red) with *15B03-Gal4* driving *>GFP* alone (**A**) or *>GFP* with *>smo^{RNAi}* (**B**). Yellow arrowheads indicate higher levels of Ct staining in direct AMPs. Note that Ct staining in direct AMPs is unaffected by *smo^{RNAi}* expression. Microscopy scale bars = 100 μm.

SUPPLEMENTARY FILES

Supplementary Files can be found with the online published article:

Everetts, N. J.*, Worley, M. I.*, Yasutomi, R., Yosef, N., & Hariharan, I. K. (2021). Single-cell transcriptomics of the Drosophila wing disc reveals instructive epithelium-to-myoblast interactions. *eLife, 10*, e61276.

Supplementary File 2.1. Genes with differential expression between 96h and 120h within the epithelium and AMPs. Genes were selected based on being significantly and consistently upregulated or downregulated between the two time points in either the disc epithelium and/or the AMPs. The average gene expression within cells (natural-log scale), fraction of cells expressing a given gene, fold-change between time points (natural-log scale), and FDR for differential-expression significance are reported. These gene expression, detection, and fold-change calculations are averaged across each of the pairwise comparisons performed, and the max FDR value is shown (see **Materials and Methods** for details on differential expression between time points). Negative fold-change values indicate higher expression at 96h and are colored magenta. Positive fold-change values indicate higher expression at 120h and are colored green. N.R. = not replicable; calculations in which the fold-change direction differed between pairwise comparisons.

Supplementary File 2.2. Genes with differential expression between 96h and 120h within the epithelial cell clusters. Genes were selected based on being significantly and consistently upregulated or downregulated between the two time points in at least one epithelial cluster. The natural-log of the fold change between 96h and 120h is reported, averaged across each of the pairwise comparisons performed (see **Materials and Methods** for details on differential expression between time points). Negative values indicate higher expression at 96h and are colored magenta. Positive values indicate higher expression at 120h and are colored green. Values that were not significant (based on max FDR) are reported with a "-".

Supplementary File 2.3. Genes with differential expression between 96h and 120h within the direct and indirect AMP. Genes were selected based on being significantly and consistently upregulated or downregulated between the two time points in either the direct and/or the indirect AMPs. The average gene expression within cells (natural-log scale), fraction of cells expressing a given gene, fold-change between time points (natural-log scale), and FDR for differential-expression significance are reported. These gene expression, detection, and fold-change calculations are averaged across each of the pairwise comparisons performed, and the max FDR value is shown (see **Materials and Methods** for details on differential expression between time points). Negative fold-change values indicate higher expression at 96h and are colored magenta. Positive fold-change values indicate higher expression at 120h and are colored green. N.R. = not replicable; calculations in which the fold-change direction differed between pairwise comparisons; calculations in which the fold-change direction differed between pairwise comparisons.

Supplementary File 2.4. Geometry of disc model. CSV file of the X, Y, Z geometry used in reference gene expression patterns (**Supplementary file 5**). Formatted as used in DistMap to generate virtual wing-disc.

Supplementary File 2.5. Reference gene expression patterns. CSV file of the binarized reference gene expression patterns (along with geometry in **Supplementary file 4**). Formatted as used in DistMap to generate virtual wing-disc.

REFERENCES

Ariss, M. M., Islam, A. B. M. M. K., Critcher, M., Zappia, M. P., & Frolov, M. V. (2018). Single cell RNA-sequencing identifies a metabolic aspect of apoptosis in Rbf mutant. Nature Communications, 9(1), 5024. https://doi.org/10.1038/s41467-018-07540-z

Ayala-Camargo, A., Anderson, A. M., Amoyel, M., Rodrigues, A. B., Flaherty, M. S., & Bach, E. A. (2013). JAK/STAT signaling is required for hinge growth and patterning in the Drosophila wing disc. Developmental Biology, 382(2), 413–426. https://doi.org/10.1016/j.ydbio.2013.08.016

Bageritz, J., Willnow, P., Valentini, E., Leible, S., Boutros, M., & Teleman, A. A. (2019). Gene expression atlas of a developing tissue by single cell expression correlation analysis. Nature Methods, 16(8), 750–756. https://doi.org/10.1038/s41592-019-0492-x

Bate, M. (1993). The Mesoderm and its Derivatives. In M. Bate & A. Martinez Arias (Eds.), The Development of Drosophila melanogaster (pp. 1013–1090). Cold Spring Harbor Laboratory Press.

Bate, M., Rushton, E., & Currie, D. A. (1991). Cells with persistent twist expression are the embryonic precursors of adult muscles in Drosophila. Development (Cambridge, England), 113(1), 79–89. http://www.ncbi.nlm.nih.gov/pubmed/1765010

Beiman, M., Shilo, B. Z., & Volk, T. (1996). Heartless, a Drosophila FGF receptor homolog, is essential for cell migration and establishment of several mesodermal lineages. Genes and Development, 10(23), 2993–3002. https://doi.org/10.1101/gad.10.23.2993

Bothe, I., & Baylies, M. K. (2016). Drosophila myogenesis. Current Biology, 26(17), R786–R791. https://doi.org/10.1016/j.cub.2016.07.062

Browaeys, R., Saelens, W., & Saeys, Y. (2020). NicheNet: modeling intercellular communication by linking ligands to target genes. Nature Methods, 17(2), 159–162. https://doi.org/10.1038/s41592-019-0667-5

Buescher, M., Svendsen, P. C., Tio, M., Miskolczi-McCallum, C., Tear, G., Brook, W. J., & Chia, W. (2004). Drosophila T Box Proteins Break the Symmetry of Hedgehog-Dependent Activation of wingless. Current Biology, 14(19), 1694–1702. https://doi.org/10.1016/j.cub.2004.09.048

Butler, M. J., Jacobsen, T. L., Cain, D. M., Jarman, M. G., Hubank, M., Whittle, J. R. S., Phillips, R., & Simcox, A. (2003). Discovery of genes with highly restricted expression patterns in the Drosophila wing disc using DNA oligonucleotide microarrays. In Development (Vol. 130, Issue 4, pp. 659–670). https://doi.org/10.1242/dev.00293

Cao, J., Spielmann, M., Qiu, X., Huang, X., Ibrahim, D. M., Hill, A. J., Zhang, F., Mundlos, S., Christiansen, L., Steemers, F. J., Trapnell, C., & Shendure, J. (2019). The single-cell transcriptional landscape of mammalian organogenesis. Nature, 566(7745), 496–502. https://doi.org/10.1038/s41586-019-0969-x

Cohen, S. M. (1993). Imaginal disc development. In M. Bate & A. Martinez Arias (Eds.), The Development of Drosophila melanogaster (pp. 747–841). Cold Spring Harbor Laboratory Press.

Delgado, I., & Torres, M. (2017). Coordination of limb development by crosstalk among axial patterning pathways. Developmental Biology, 429(2), 382–386. https://doi.org/10.1016/j.ydbio.2017.03.006

Deng, M., Wang, Y., Zhang, L., Yang, Y., Huang, S., Wang, J., Ge, H., Ishibashi, T., & Yan, Y. (2019). Single cell transcriptomic landscapes of pattern formation, proliferation and growth in Drosophila wing imaginal discs. Development (Cambridge), 146(18). https://doi.org/10.1242/dev.179754

Edgar, B. A., & O'Farrell, P. H. (1990). The three postblastoderm cell cycles of Drosophila embryogenesis are regulated in G2 by string. Cell, 62(3), 469–480. https://doi.org/10.1016/0092-8674(90)90012-4

Evans, C. J., Olson, J. M., Ngo, K. T., Kim, E., Lee, N. E., Kuoy, E., Patananan, A. N., Sitz, D., Tran, P., Do, M.-T., Yackle, K., Cespedes, A., Hartenstein, V., Call, G. B., & Banerjee, U. (2009). G-TRACE: rapid Gal4-based cell lineage analysis in Drosophila. Nature Methods, 6(8), 603–605. https://doi.org/10.1038/nmeth.1356

Fabre, P. J., Leleu, M., Mascrez, B., Lo Giudice, Q., Cobb, J., & Duboule, D. (2018). Heterogeneous combinatorial expression of Hoxd genes in single cells during limb development. BMC Biology, 16(1), 101. https://doi.org/10.1186/s12915-018-0570-z

Feregrino, C., Sacher, F., Parnas, O., & Tschopp, P. (2019). A single-cell transcriptomic atlas of the developing chicken limb. BMC Genomics, 20(1), 401. https://doi.org/10.1186/s12864-019-5802-2

Fernandes, J., Bate, M., & Vijayraghavan, K. (1991). Development of the indirect flight muscles of Drosophila. Development (Cambridge, England), 113(1), 67–77. http://www.ncbi.nlm.nih.gov/pubmed/1765009

Franke, A., & Baker, B. S. (1999). The rox1 and rox2 RNAs Are Essential Components of the Compensasome, which Mediates Dosage Compensation in Drosophila. In Molecular Cell (Vol. 4).

Frémion, F., Darboux, I., Diano, M., Hipeau-Jacquotte, R., Seeger, M. A., & Piovant, M. (2000). Amalgam is a ligand for the transmembrane receptor neurotactin and is required for neurotactin-mediated cell adhesion and axon fasciculation in Drosophila. The EMBO Journal, 19(17), 4463–4472. https://doi.org/10.1093/emboj/19.17.4463

Garcia-Bellido, A., Ripoll, P., & Morata, G. (1973). Developmental Compartmentalisation of the Wing Disk of Drosophila. Nature New Biology, 245(147), 251–253. https://doi.org/10.1038/newbio245251a0

Ghazi, A., Anant, S., & VijayRaghavan, K. (2000). Apterous mediates development of direct flight muscles autonomously and indirect flight muscles through epidermal cues. Development (Cambridge, England), 127(24), 5309–5318. http://www.ncbi.nlm.nih.gov/pubmed/11076753

Gunage, R. D., Dhanyasi, N., Reichert, H., & VijayRaghavan, K. (2017). Drosophila adult muscle development and regeneration. Seminars in Cell & Developmental Biology, 72, 56–66. https://doi.org/10.1016/j.semcdb.2017.11.017

Gunage, R. D., Reichert, H., & VijayRaghavan, K. (2014). Identification of a new stem cell population that generates Drosophila flight muscles. ELife, 3(August2014), 1–25. https://doi.org/10.7554/eLife.03126

Hatori, R., & Kornberg, T. B. (2020). Hedgehog produced by the Drosophila wing imaginal disc induces distinct responses in three target tissues. Development (Cambridge, England), 147(22). https://doi.org/10.1242/dev.195974

Held Jr, L. I. (2002). Imaginal Discs. Cambridge University Press. https://doi.org/10.1017/CBO9780511529733

Hong, W., Mosca, T. J., & Luo, L. (2012). Teneurins instruct synaptic partner matching in an olfactory map. Nature, 484(7393), 201–207. https://doi.org/10.1038/nature10926

Hortsch, M., Patel, N. H., Bieber, A. J., Traquina, Z. R., & Goodman, C. S. (1990). Drosophila neurotactin, a surface glycoprotein with homology to serine esterases, is dynamically expressed during embryogenesis. Development (Cambridge, England), 110(4), 1327–1340. http://www.ncbi.nlm.nih.gov/pubmed/2100266

Kadam, S., McMahon, A., Tzou, P., & Stathopoulos, A. (2009). FGF ligands in Drosophila have distinct activities required to support cell migration and differentiation. Development, 136(5), 739–747. https://doi.org/10.1242/dev.027904

Karaiskos, N., Wahle, P., Alles, J., Boltengagen, A., Ayoub, S., Kipar, C., Kocks, C., Rajewsky, N., & Zinzen, R. P. (2017). The Drosophila embryo at single-cell transcriptome resolution. Science (New York, N.Y.), 358(6360), 194–199. https://doi.org/10.1126/science.aan3235

Kidd, T., Bland, K. S., & Goodman, C. S. (1999). Slit Is the Midline Repellent for the Robo Receptor in Drosophila. Cell, 96(6), 785–794. https://doi.org/10.1016/S0092-8674(00)80589-9

Laurichesse, Q., & Soler, C. (2020). Muscle development : a view from adult myogenesis in Drosophila. Seminars in Cell & Developmental Biology, 104, 39–50. https://doi.org/10.1016/j.semcdb.2020.02.009

Lawrence, P. A. (1982). Cell Lineage of the Thoracic Muscles of Drosophila. In Cell (Vol. 29).

Leal, S. M., Qian, L., Lacin, H., Bodmer, R., & Skeath, J. B. (2009). Neuromancer1 and Neuromancer2 regulate cell fate specification in the developing embryonic CNS of Drosophila melanogaster. Developmental Biology, 325(1), 138–150. https://doi.org/10.1016/j.ydbio.2008.10.006

Lehner, C. F., & O'Farrell, P. H. (1990). The roles of Drosophila cyclins A and B in mitotic control. Cell, 61(3), 535–547. https://doi.org/10.1016/0092-8674(90)90535-M

Lin, S., Ewen-Campen, B., Ni, X., Housden, B. E., & Perrimon, N. (2015). In Vivo Transcriptional Activation Using CRISPR/Cas9 in Drosophila. Genetics, 201(2), 433–442. https://doi.org/10.1534/genetics.115.181065

Lopez, R., Regier, J., Cole, M., Jordan, M., & Yosef, N. (2018). Bayesian Inference for a Generative Model of Transcriptome Profiles from Single-cell RNA Sequencing. 292037. https://doi.org/10.1101/292037

Madhavan, M., & Schneiderman, H. A. (1977). Histological analysis of the dynamics of growth of imaginal discs and histoblast nests during the larval development of Drosophila melanogaster. Wilhelm Roux's Archives of Developmental Biology, 183(4), 269–305. https://doi.org/10.1007/BF00848459

McInnes, L., Healy, J., & Melville, J. (2018). UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction. ArXiv E-Prints. http://arxiv.org/abs/1802.03426

Meller, V. H., & Rattner, B. P. (2002). The roX genes encode redundant male-specific lethal transcripts required for targeting of the MSL complex. The EMBO Journal, 21(5), 1084–1091. https://doi.org/10.1093/emboj/21.5.1084

Miller, A. (1950). The internal anatomy and histology of the imago of Drosophila melanogaster. In M. Demerec (Ed.), The Biology of Drosophila (pp. 420–534). John Wiley and Sons.

Neto-Silva, R. M., Wells, B. S., & Johnston, L. A. (2009). Mechanisms of Growth and Homeostasis in the Drosophila Wing. Annual Review of Cell and Developmental Biology, 25(1), 197–220. https://doi.org/10.1146/annurev.cellbio.24.110707.175242

Pau, G., Fuchs, F., Sklyar, O., Boutros, M., & Huber, W. (2010). EBImage--an R package for image processing with applications to cellular phenotypes. Bioinformatics, 26(7), 979–981. https://doi.org/10.1093/bioinformatics/btq046

Poodry, C. A., & Schneiderman, H. A. (1970). The ultrastructure of the developing leg of Drosophila melanogaster. Wilhelm Roux' Archiv Für Entwicklungsmechanik Der Organismen, 166(1), 1–44. https://doi.org/10.1007/BF00576805

Price, M. A., & Kalderon, D. (1999). Proteolysis of cubitus interruptus in Drosophila requires phosphorylation by protein kinase A. Development (Cambridge, England), 126(19), 4331–4339. http://www.ncbi.nlm.nih.gov/pubmed/10477300

Requena, D., Álvarez, J. A., Gabilondo, H., Loker, R., Mann, R. S., & Estella, C. (2017). Origins and Specification of the Drosophila Wing. Current Biology, 27(24), 3826-3836.e5. https://doi.org/10.1016/j.cub.2017.11.023

Ribatti, D., & Santoiemma, M. (2014). Epithelial-mesenchymal interactions: a fundamental Developmental Biology mechanism. The International Journal of Developmental Biology, 58(5), 303–306. https://doi.org/10.1387/ijdb.140143dr

Roy, S., & VijayRaghavan, K. (1997). Homeotic genes and the regulation of myoblast migration, fusion, and fibre-specific gene expression during adult myogenesis in Drosophila. Development, 124(17), 3333–3341.

Roy, S., & VijayRaghavan, K. (1999). Muscle pattern diversification in Drosophila: the story of imaginal myogenesis. BioEssays, 21(6), 486–498. https://doi.org/10.1002/(SICI)1521-1878(199906)21:6<486::AID-BIES5>3.0.CO;2-M

Satija, R., Farrell, J. A., Gennert, D., Schier, A. F., & Regev, A. (2015). Spatial reconstruction of single-cell gene expression data. Nature Biotechnology, 33(5), 495–502. https://doi.org/10.1038/nbt.3192

Schier, A. F. (2020). Single-cell biology: beyond the sum of its parts. In Nature Methods (Vol. 17, Issue 1, pp. 17–20). Nature Research. https://doi.org/10.1038/s41592-019-0693-3

Stathopoulos, A., Tam, B., Ronshaugen, M., Frasch, M., & Levine, M. (2004). Pyramus and thisbe: FGF genes that pattern the mesoderm of Drosophila embryos. Genes and Development, 18(6), 687–699. https://doi.org/10.1101/gad.1166404

Stuart, T., Butler, A., Hoffman, P., Hafemeister, C., Papalexi, E., Mauck, W. M., Hao, Y., Stoeckius, M., Smibert, P., & Satija, R. (2019). Comprehensive Integration of Single-Cell Data. Cell, 177(7), 1888-1902.e21. https://doi.org/10.1016/j.cell.2019.05.031

Sudarsan, V., Anant, S., Guptan, P., VijayRaghavan, K., & Skaer, H. (2001). Myoblast diversification and ectodermal signaling in Drosophila. Developmental Cell, 1(6), 829–839. https://doi.org/10.1016/s1534-5807(01)00089-2

Tanimoto, H., Itoh, S., Ten Dijke, P., & Tabata, T. (2000). Hedgehog creates a gradient of DPP activity in Drosophila wing imaginal discs. Molecular Cell, 5(1), 59–71. https://doi.org/10.1016/S1097-2765(00)80403-7

Terriente, J., Perea, D., Suzanne, M., & Díaz-Benjumea, F. J. (2008). The Drosophila gene zfh2 is required to establish proximal-distal domains in the wing disc. Developmental Biology, 320(1), 102–112. https://doi.org/10.1016/j.ydbio.2008.04.028

Tran, K. D., Miller, M. R., & Doe, C. Q. (2010). Recombineering Hunchback identifies two conserved domains required to maintain neuroblast competence and specify early-born neuronal identity. Development, 137(9), 1421–1430. https://doi.org/10.1242/dev.048678

Vento-Tormo, R., Efremova, M., Botting, R. A., Turco, M. Y., Vento-Tormo, M., Meyer, K. B., Park, J.-E., Stephenson, E., Polański, K., Goncalves, A., Gardner, L., Holmqvist, S., Henriksson, J., Zou, A., Sharkey, A. M., Millar, B., Innes, B., Wood, L., Wilbrey-Clark, A., … Teichmann, S. A. (2018). Single-cell reconstruction of the early maternal–fetal interface in humans. Nature, 563(7731), 347–353. https://doi.org/10.1038/s41586-018-0698-6

Vishal, K., Lovato, T. L., Bragg, C., Chechenova, M. B., & Cripps, R. M. (2020). FGF signaling promotes myoblast proliferation through activation of wingless signaling. Developmental Biology, 464(1), 1–10. https://doi.org/10.1016/j.ydbio.2020.05.009

Waddington, C. H. (1940). Organisers and genes. Organisers and Genes.

Worley, M. I., Setiawan, L., & Hariharan, I. K. (2013). TIE-DYE: a combinatorial marking system to visualize and genetically manipulate clones during development in Drosophila melanogaster. Development, 140(15), 3275–3284. https://doi.org/10.1242/dev.096057

Yamaguchi, M., Nishida, Y., Moriuchi, T., Hirose, F., Hui, C. C., Suzuki, Y., & Matsukage, A. (1990). Drosophila Proliferating Cell Nuclear Antigen (Cyclin) Gene: Structure, Expression during Development, and Specific Binding of Homeodomain Proteins to Its 5'-Flanking Region. Molecular and Cellular Biology, 10(3), 872–879. https://doi.org/10.1128/MCB.10.3.872

Zappia, M. P., Castro, L., Ariss, M. M., Jefferson, H., Islam, A. B., & Frolov, M. V. (2020). A cell atlas of adult muscle precursors uncovers early events in fibre-type divergence in Drosophila. EMBO Reports, 21(10), e49555. https://doi.org/10.15252/embr.201949555

Zeev-Ben-Mordehai, T., Paz, A., Peleg, Y., Toker, L., Wolf, S. G., Rydberg, E. H., Sussman, J. L., & Silman, I. (2009). Amalgam, an axon guidance Drosophila adhesion protein belonging to the immunoglobulin superfamily: Over-expression, purification and biophysical characterization. Protein Expression and Purification, 63(2), 147–157. https://doi.org/10.1016/j.pep.2008.09.019

Zirin, J. D., & Mann, R. S. (2007). Nubbin and Teashirt mark barriers to clonal growth along the proximal–distal axis of the Drosophila wing. Developmental Biology, 304(2), 745–758. https://doi.org/10.1016/j.ydbio.2007.01.025

Portions of the following chapter were originally published as an article on
bioRxiv 2021.07.08.451678; doi: https://doi.org/10.1101/2021.07.08.451678

Chapter 3

Critical genetic program for Drosophila imaginal disc regeneration revealed by single-cell analysis

ABSTRACT

Whether regeneration is primarily accomplished by re-activating gene regulatory networks used previously during development or by activating novel regeneration-specific transcriptional programs remains a longstanding question. Currently, most genes implicated in regeneration also function during development. Using single-cell transcriptomics in regenerating *Drosophila* wing discs, we identified two regeneration-specific cell populations within the blastema. They are each composed of cells that upregulate multiple genes encoding secreted proteins that promote regeneration. In this regenerative secretory zone, the transcription factor Ets21C controls the expression of multiple regeneration-promoting genes. While eliminating *Ets21C* function has no discernible effect on development, it severely compromises regeneration. This Ets21C-dependent gene regulatory network is also activated in blastema-like cells in tumorous discs, suggesting that pro-regenerative mechanisms can be co-opted by tumors to promote aberrant growth.

RESULTS AND DISCUSSION

A long-standing question in the field of regenerative biology is whether regeneration is mainly accomplished by reactivation of gene regulatory networks (GRNs) used during earlier stages of development or, alternatively, by GRNs that are specifically activated during regeneration. Studies of regenerating tissues have provided evidence for cellular states that are not observed during normal development and for patterns of gene expression that seem specific for regeneration (for example Gerber *et al.*, 2018; Aztekin *et al.*, 2019). However, so far, there is little evidence for genes that are needed for regeneration but not for normal development.

To identify transcriptional programs initiated during regeneration, we examined regeneration of *Drosophila* larval wing imaginal discs, the epithelial tissues that differentiate into the adult wings and thorax. Imaginal discs are capable of regenerating after damage through the formation of a blastema, defined by localized proliferation and increased cellular plasticity (reviewed by Worley *et al.*, 2012). To search for regeneration-specific GRNs, we compared regenerating and developing wing discs using single-cell transcriptomics. Tissue damage was induced by temporarily expressing the pro-apoptotic TNF ortholog *eiger* within the wing pouch, the portion of the disc that generates the wing blade (Smith-Bolton *et al.*, 2009) (**Figure 3.1**). Subsequent regeneration occurs by localized cell proliferation and cell-fate re-specification. We collected wing discs after 24 hours of regeneration, approximately one third of the way through the regenerative process, and sequenced a total of 14,320 cells from two biological replicates, with an average of >3,000 genes detected per cell. Three major cell types were identified: epithelial cells, myoblasts, and hemocytes (**Figure 3.2**). Since imaginal disc regeneration is driven by epithelial cell proliferation (Smith-Bolton *et al.*, 2009; Worley *et al.*, 2012), we focused further analysis on these cells.

To identify potential regeneration-specific GRNs, we harmonized data from epithelial cells from regenerating discs with our previously collected data from undamaged discs (Everetts *et al.*, 2021) using scVI (Gayoso *et al.*, 2021) (see **Materials and Methods**) (**Figure 3.3A, B**). We assigned cell clusters to specific subregions of the wing disc epithelium based on the expression of known marker genes (Bageritz *et al.*, 2019; Deng *et al.*, 2019; Zappia *et al.*, 2020; Everetts *et al.*, 2021) (**Figure 3.3B, C; Figure 3.4**). As expected, cell clusters with pouch identity were underrepresented in the regenerating sample, as this portion of the tissue was ablated (**Figure 3.4**). From our single-cell analysis, we observed two clusters, denoted Blastema1 and Blastema2, that were almost exclusively composed of cells from the regenerating sample (181/186 and 519/564 cells, respectively) (**Figure 3.3B; Figure 3.4**). Within these two regeneration-specific clusters, we observed the upregulation of genes known to be induced around the site of damage, including the Wnt ligands *wingless (wg)* and *Wnt6* (Harris *et al.*, 2016), *Matrix metalloproteinase 1 (Mmp1)*, and *Insulin-like peptide 8 (Ilp8)* (**Figure 3.3C, D**).

Both Blastema1 and Blastema2 clusters express *Ilp8*, which is strongly upregulated around the site of damage in the regenerating disc (**Figure 3.3F, G**). However, Blastema2 showed a higher expression of hinge-identity markers, such as transcription factor *Zn finger homeodomain 2* (*zfh2)*, than Blastema1 (**Figure 3.3C**), suggesting that these cells might occupy a more proximal (outer) position. Indeed, in regenerating tissue, we observed higher Zfh2 expression in the outer ring of *Ilp8*-expressing cells (**Figure 3.3G**). In contrast, Blastema1 cells expressed higher levels of the *unpaired (upd1, upd2, upd3)* ligands, *asperous (aspr)*, and *PDGF- and VEGF-related factor 1 (Pvf1)* (**Figure 3.3C, E**). The Upd ligands activate the JAK/STAT pathway, which is important for cellular plasticity and regeneration (Katsuyama *et al.*, 2015; Santabarbara-Ruiz *et al.*, 2015; La

Fortezza *et al.*, 2016; Worley *et al.*, 2018). The gene *aspr* encodes a secreted protein with multiple EGF-repeats important for regeneration (Harris *et al.*, 2020). Pvf1 binds to its receptor Pvr and the resulting signaling is known to contribute to wound healing (Wu *et al.*, 2009), and homologs are involved in regeneration in other systems (Currie *et al.*, 2016; Johnson *et al.*, 2020). We determined that *Pvf1*, *upd3*, and *Ilp8* were all expressed at the center of the blastema **(Figure 3.3H, I; Figure 3.5**), which is surrounded by cells that express *Ilp8* but not *Pvf1* or *upd3*. Thus, the Blastema1 cells are located at the center of the blastema and are surrounded by Blastema2 cells; cells in both regions secrete ligands, some of which are known to promote regeneration, and are likely acting on the surrounding tissue. We refer to these regions together as the regenerative secretory zone.

From our single-cell analysis, we used gene signatures to determine that the cells within the regenerative secretory zone were in an intermediate state between hinge and pouch identities (**Figure 3.3J; Figure 3.6**). This finding suggested that these cells were derived from the surrounding inner-hinge region and were in the process of acquiring more distal pouch fates. To investigate this process, we examined the location of proliferating cells and found high levels of EdU incorporation surrounding the regenerative secretory zone (**Figure 3.3K**) (Cosolo *et al.*, 2019). As regeneration proceeded, the EdU incorporation extended more centrally to occur within the regenerative secretory zone (**Figure 3.3L**). To determine if these proliferating cells are reprogrammed during regeneration to replace the ablated pouch, we performed a lineage-tracing experiment with an enhancer that is normally only expressed in a ring of cells of the inner-hinge. In the absence of pouch ablation, these cells and their progeny remain confined to the hinge (**Figure 3.3M, N**). However, after regeneration following pouch ablation, most of the regenerated pouch was derived from cells that once expressed this enhancer (**Figure 3.3O**). Thus, the ablated pouch is regenerated by the proliferation and reprogramming of more proximally fated inner hinge cells, likely driven by the ligands secreted by the regenerative secretory zone (**Figure 3.3P**).

To search for a regulator of these regeneration-specific transcriptional changes, we analyzed our single-cell data for a transcription factor that was specifically expressed within the blastema cells. We found that *Ets at 21C* (*Ets21C*) was specifically expressed during regeneration, primarily within the cells of the regenerative secretory zone, and not in cells from developing wing discs which were undamaged (**Figure 3.7A-C**). *Ets21C* was also upregulated after physically wounding of the wing disc (**Figure 3.7D**), implying that *Ets21C* is involved in a general regeneration response. *Ets21C* had previously been shown to be upregulated during disc regeneration by bulk sequencing of blastema-enriched cells (Khan *et al.*, 2017). Our single-cell data indicates that *Ets21C* expression is highly correlated with *Ilp8* and *Mmp1* expression during regeneration (**Figure 3.8**). *Ets21C* expression was induced during the genetic ablation period and was maintained throughout regeneration (**Figure 3.9**), suggesting that Ets21C could function at multiple stages of regeneration.

To determine if *Ets21C* was important either for normal development or for regeneration, we turned to mutant analysis. First, we observed that homozygous $Ets21C^{-/-}$ null mutants generate viable and fertile adults, as previously noted (Mundorf *et al.*, 2019), whose wings were of normal size and shape (**Figure 3.7E-G**). By generating mosaic eyes composed of marked wild-type cells and $Ets21C^{-/-}$ mutant cells, we found that mutant cells did not display defects in cell proliferation even in a competition scenario with wild-type cells (**Figure 3.7H, I**). Thus, *Ets21C* is dispensable for normal development and its absence does not impair cell proliferation.

Next, we tested if *Ets21C* is essential for imaginal disc regeneration. Following our genetic ablation assay, homozygous null $Ets21C^{-/-}$ mutants showed a dramatic defect in the extent of wing regeneration when compared to either $Ets21C^{+/-}$ heterozygotes or wild-type controls (**Figure 3.7J**).

This effect was observed with the null mutation in trans to a chromosomal deletion (**Figure 3.7J, K**), indicating that the effect was indeed due to the loss of *Ets21C* function. Thus, *Ets21C* is required for effective regeneration.

Ets21C is part of the Ets-family of DNA binding transcription factors that are broadly conserved in animals. The *Ets21C* mammalian orthologs are *Ets-related gene (ERG)* and *Friend Leukemia Integration 1 Transcription Factor (FLI1)*, both of which can act as proto-oncogenes (Kar and Gutierrez-Hartmann, 2013). In *Drosophila*, although *Ets21C* is not expressed in undamaged third instar wing discs, its expression is upregulated in tumorous imaginal discs (Kulshammer *et al.*, 2015; Toggweiler *et al.*, 2016) and it has been shown to be involved in adult midgut homeostasis (Jin *et al.*, 2015; Mundorf *et al.*, 2019). *Ets21C* is a downstream target of JNK/AP1 signaling in these contexts (Kulshammer *et al.*, 2015; Toggweiler *et al.*, 2016; Mundorf *et al.*, 2019). Similarly, we found that even in undamaged discs, activation of the JNK pathway induces *Ets21C* expression (**Figure 3.10**). The JNK/AP1 pathway is known to be critical for regeneration (Bosch *et al.*, 2005; Mattila *et al.*, 2005; Katsuyama *et al.*, 2015; Harris *et al.*, 2016; Harris *et al.*, 2020). Thus, we hypothesized that Ets21C could be functioning downstream of JNK/AP1 signaling to activate a regeneration-specific GRN in the blastema.

We investigated if *Ets21C* mutant tissues would fail to upregulate secreted molecules during regeneration. At 24 hours of regeneration, *Pvf1* and *upd3* are expressed in the inner region of the regenerative secretory zone (**Figures 3.3 and 3.5**), and we further examined their expression patterns earlier in regeneration in both wild-type and mutant *Ets21C* tissue. In wild-type regenerating discs at the start of regeneration, Pvf1 antibody and *upd3-lacZ* (Bunker *et al.*, 2015) expression was detected within the center of the blastema and also within cellular debris (**Figure 3.11A, C**). In contrast, 0-hour regenerating *Ets21C*$^{-/-}$ mutant discs showed a substantial decrease both in Pvf1 and *upd3-lacZ* expression (**Figu re 3.11B, D**), suggesting that *Ets21C* is required to initiate expression of both ligands in response to damage. We also investigated the expression of *Pvf1* and *upd3* using hybridization chain reaction (HCR), a sensitive and robust mRNA *in situ* technique (Choi et al., 2018). Interestingly, in contrast to the results obtained from the Pvf1 antibody and *upd-lacZ* reporter, HCR detected noticeable transcript expression of both genes within the 0-hour regeneration blastema of *Ets21C* mutant discs albeit at much lower levels compared to wild-type *Ets21C* discs (**Figure 3.12A, C**). When assayed by HCR at 24 hours into regeneration, expression of both genes had attenuated in the *Ets21C* mutant but persisted in the *Ets21C* wild-type tissue (**Figure 3.12B, D**). These results indicate that *Ets21C* is more important for maintaining *Pvf1* and *upd3* over the course of regeneration, as all assays indicated loss of both *Pvf1* and *upd3* expression at 24 hours into regeneration. However, while *Ets21C* is not completely necessary for the activation of *Pvf1* and *upd3* (as indicated via HCR), loss of *Ets21C* may still compromise the expression of these genes such that antibody and genetic reporters (likely less sensitive than HCR) show only negligible levels at the start of regeneration. We examined whether compromised expression of *upd3* and possibly its paralogs in *Ets21C* mutant tissue impacted downstream JAK/STAT signaling. In control tissues, the STAT activity reporter was expressed in the center of the blastema at the start of regeneration (**Figure 3.11E**) as well as in the surrounding hinge regions where the JAK/STAT pathway is active during development (Ayala-Camargo *et al.*, 2013; La Fortezza *et al.*, 2016). In contrast, *Ets21C*$^{-/-}$ mutant tissues failed to activate the JAK/STAT reporter within the cells at the center of the blastema (**Figure 3.11F**), while JAK/STAT signaling in the hinge remained unaffected. Thus, *Ets21C* is required for maintained expression of Pvf1 and Upd ligands within the inner regenerative secretory zone, even though it is not the sole activator of these genes. The disruption of the inner regenerative secretory zone is also observed

when we assess the pattern of cell proliferation in regenerating *Ets21C* mutant discs, which did not maintain a central non-proliferating zone for the same duration as wild-type *Ets21C* discs (**Figure 3.13**). This observation suggests that Ets21C is required for proper function of the inner regenerative secretory zone and potentially for its maintained establishment.

We next examined if Ets21C regulated the expression of genes expressed in the outer regenerative secretory zone. Wg expression during regeneration was unaffected in *Ets21C* mutants (**Figure 3.14**). While Mmp1 expression was observed at early stages of regeneration, it was prematurely absent by 24 hours in *Ets21C* mutants (**Figure 3.11G-J**), demonstrating that Ets21C is required to maintain Mmp1 expression during regeneration. Mmp1 is important for proper blastema formation (McClure *et al.*, 2008) and effective regeneration (Harris *et al.*, 2020). Similarly, *Ilp8* expression appeared normal at earlier time points of regeneration, but showed a slight decrease as regeneration progressed (**Figure 3.11K, L**). Regenerating discs, when compared to normal development, have a relatively immature transcriptional state (**Figure 3.15**), marked by the delay in the expression of the transcription factor Broad (Narbonne-Reveau and Maurange, 2019) (**Figure 3.15**). *Ets21C*$^{-/-}$ mutants, however, showed premature expression of Broad during regeneration (**Figure 3.15**). In addition, a regeneration-induced developmental delay that occurs in more distant tissues, specifically the eye disc, was also reduced in *Ets21C*$^{-/-}$ larvae (**Figure 3.15**). *Ilp8* is crucial for delaying pupariation (Colombani *et al.*, 2012; Garelli *et al.*, 2012), and this delay is correlated with regeneration outcomes (Smith-Bolton *et al.*, 2009; Halme *et al.*, 2010; Katsuyama *et al.*, 2015; Harris *et al.*, 2016). Indeed, *Ets21C* animals ended the larval phase of development approximately 30h before regenerating controls (**Figure 3.16**), which is likely the result of a decrease in Ilp8 levels and other signaling molecules. Thus, *Ets21C* mutants have both local and systemic defects in their regenerative response that collectively contribute to the reduced regeneration.

To test whether *Ets21C* regulates *Ilp8* cell-autonomously, we generated mosaic discs containing patches of both *Ets21C*$^{-/-}$ mutant and wild-type cells prior to tissue damage (see **Materials and Methods**). In regenerating discs, clones of *Ets21C* cells (marked by the absence of RFP) are comparable in size to wild-type clones, indicating that *Ets21C* does not have a cell-autonomous function in regulating cell proliferation during regeneration (**Figure 3.11M**). However, the *Ets21C*$^{-/-}$ mutant cells showed a cell-autonomous decrease in *Ilp8-GFP* expression after 48h of regeneration compared to wild-type cells (**Figure 3.11M''**), indicating that as with *Mmp1*, Ets21C is required to sustain *Ilp8* expression.

While *Ets21C* function is required for regeneration and not necessary for normal development, it is known to be expressed in tumorous imaginal discs that have mutations that disrupt apicobasal polarity (Kulshammer *et al.*, 2015; Toggweiler *et al.*, 2016). Moreover, in one study, reducing *Ets21C* function was shown to reduce overall tumor size (Toggweiler *et al.*, 2016). Recent single-cell studies (Deng *et al.*, 2019)(Ji *et al.*, 2019) of tumorous imaginal discs caused by mutations in the apicobasal polarity regulator *scribble* (*scrib*) (Bilder *et al.*, 2000) have demonstrated considerable cellular heterogeneity. By harmonizing our data with published single-cell RNAseq data derived from tumorous *scrib* discs (Deng *et al.*, 2019) (**Figure 3.17A, B**), we found that a subset of cell clusters have similar transcriptomes to the regenerative secretory zone (Blastema1 or Blastema2 cell clusters) (**Figure 3.17C-H**). Notably, these cell clusters express *Ets21C* along with *upd3, Pvf1, Mmp1,* and *Ilp8* (**Figure 3.17D-F**) and the cells are more prevalent at earlier stages of disc overgrowth (**Figure 3.17H**). Thus, while most cells in the disc have defects in apicobasal polarity, only a small subset of cells appear to activate this pro-regenerative GRN featuring *Ets21C.* Since reducing *Ets21C* function reduces growth in similar tumor models

(Toggweiler *et al.*, 2016), the presence of blastema-like cells could be critical for promoting the overgrowth of tumorous discs.

In conclusion, we have discovered a GRN that is dispensable for normal development yet essential for regeneration. Regeneration-specific GRNs may also exist in vertebrates and their reactivation could be valuable for regenerative medicine. Finally, the role of pro-regenerative GRNs in oncogenesis merits further exploration.

MATERIALS AND METHODS

Single-cell data collection

For each sample, approximately 300 regenerating wing-imaginal discs were collected after 24 hours of regeneration. Regenerating discs were dissected within 1 hour in Supplemented Schneider's Medium. The samples were then processed according to the protocol outlined in (Everetts *et al.*, 2021). Briefly, we used a mixture of trypsin and collagenase to enzymatically dissociate the tissues. Then we used FACS to eliminate both apoptotic cells and cellular debris. Because our tissue dissociation protocol enriched for myoblasts, we decided to specifically sort out myoblasts during the collection of our second regeneration sample. This was done with a *Holes in muscle (Him)-GFP* construct that specifically labeled the myoblasts (Rebeiz *et al.*, 2002). The myoblasts represented 75.67 % of the cells from the sample without *Him-GFP* and 12.71 % cells of the sample with *Him-GFP* (as determined by single-cell analysis). Single-cell suspensions were barcoded for single-cell RNA sequencing with the 10X Chromium Single Cell platform (v2 chemistry). Barcoded samples were sequenced on an Illumina NovaSeq (S2 flow cell) to over 60% saturation.

Single-cell data analysis

Single-cell sequencing reads were aligned with the 10x Genomics CellRanger pipeline (v.2.2.0) to the *Drosophila melanogaster* transcriptome (version 6.24, FlyBase). Analysis of the single-cell data was conducted in the R and Python programming languages, primarily using the packages scvi-tools v0.9.1 (Gayoso *et al.*, 2021) and Seurat v3 (Stuart *et al.*, 2019).

Before cell filtering, we used scvi-tools to harmonize our single-cell data from regenerating wing discs with the single-cell data from developing wild-type wing discs presented in our previous study (accession number GSE155543) (Everetts *et al.*, 2021). We used Seurat's variance-stabilizing transformation method to select 1000 variable genes for each batch, and the scVI VAE model was trained on the union of these genes with the following parameters: n_latent = 15, n_layers = 2, gene_likelihood = "nb", max_epochs = 400, and train_size = 0.8. The scVI latent space was used as the input for Seurat's clustering algorithm (parameters: clustering k.param = 35, clustering resolution = 2.0; default parameters otherwise). Known transcriptional markers were used to classify cell clusters: *SPARC* and *twist* for AMPs, *Fasciclin 3(Fas3)* and *narrow* for the disc epithelium, and *regucalcin* and *Hemese (He)* for hemocytes (**Figure 3.2**). We also identified an unusual cell cluster that expressed both AMP and epithelium markers, with slightly elevated average nGene and nUMI counts, which suggested that these cells were actually doublets. When we applied the tool DoubletFinder (McGinnis *et al.*, 2019) (parameters: 30 PCs, pN = 0.25, expected doublets (nExp) = 7.5% of total cells in each batch, and pK determined by the recommended BCmvn method) to each individual batch, the majority of cells within this cluster had been classified as potential doublets within their respective batches. We determined this cluster to represent AMP-epithelium doublets and removed it from subsequent analysis. We isolated the remaining disc epithelium and AMP clusters and filtered these cell types separately.

When filtering the disc epithelium cells, we first processed each batch using the standard Seurat pipeline (parameters: nfeatures = 1000, npcs = 30, k.param = 20, clustering resolution = 2.0; default parameters otherwise) and removed low-quality clusters. We classified low-quality clusters as having: 1) an average nGene less than 1 standard deviation below the average nGene of all cells, 2) an average percent.mito greater than 1 standard deviation above the average percent.mito of all cells, and 3) an abundance of negative marker genes compared to positive

markers genes (as calculated by a Wilcoxon test). After removing low-quality clusters, we marked potential doublets within the epithelium cells by applying DoubletFinder (same parameters as described initially) to the epithelium cells in each batch. Batches were harmonized with scVI (same parameters as described initially), trained on the union of the top 1000 variable genes within the epithelium cells for each batch as determined by Seurat. After harmonization, we used the scVI latent space as a basis for Seurat clustering (parameters: k.param = 35, resolution = 2.0; default parameters otherwise). We removed a cluster that we determined to be epithelium-epithelium doublets, based on the following characteristics: (1) a noticeably higher average nGene compared to all other clusters (the only cluster with an average nGene > 1 standard deviation above the average nGene of all cells), (2) an extreme abundance of potential doublets as classified by DoubletFinder from each batch (~70% of all potential doublets classified were contained within this cluster), and (3) a lack of marker genes (both positive and negative) when compared to other clusters. We also removed a cluster that we determined to represent a small number of trachea cells, based on the unique expression of marker genes tracheal-prostasin and waterproof. We re-ran our variable gene selection, scVI harmonization, and Seurat clustering. Data was visualized in 2 dimensions with UMAP (parameters: min.dist = 0.1; default parameters otherwise).

Gene signature analysis of the blastema

For each identity combination (hinge-pouch, pouch-notum, and notum-hinge), gene signatures were constructed as follows: First, differential expression was performed between wild-type (non-regenerating) cells of each identity pair (e.g., for the hinge-pouch signature, differential expression was performed between cells from (Everetts *et al.*, 2021) classified as hinge vs. cells classified as pouch). This was conducted using a Wilcoxon test via Seurat's FindMarkers function, selecting genes with a natural-log fold-change of greater than 0.25 (logfc.threshold = 0.25) and a Bonferroni-corrected p-value of < 0.05. This provided three gene sets that differentiated hinge-pouch, pouch-notum, and notum-hinge identities. Second, principal component analysis was performed on all cells using each gene set. The first principal components from each analysis were defined as the gene signatures, as they best separated cells of the different identities. The signature scores of cell clusters were visualized in 2-dimensions using Seurat's VlnPlot function (**Figure 3.3J; Figure 3.6A, B**) and in 3-dimensions using the R package Plotly (**Figure 3.6C**).

Gene signature of cellular maturity

To determine the relative cellular maturity (or developmental progression) of individual cells within the regenerating tissue we generate a gene signature score based on genes with differential expression during normal development. First, we selected genes with consistent differential expression between epithelial cells from mid (younger) and late (older) 3rd instar imaginal discs (with a threshold of greater than 0.25 natural-log fold-change). Second, this gene set was then used to perform principal component analysis to derive a cellular maturity score. The relative cellular maturity score of cells from normal developing and regenerating discs were visualized on the UMAP (**Figure 3.14B**).

Single-cell comparison of regenerating and scrib tissues

The expression matrices for the *scrib* single-cell data were downloaded from GEO, accession number GSE130566 (Deng *et al.*, 2019). Gene names were updated to match those within our regeneration and wild-type datasets. All *scrib* datasets (4d, 5d, 8d, and 14d) were harmonized with scVI (n_latent = 15, n_layers = 2, gene_likelihood = "nb", max_epochs = 400, and train_size =

0.8), trained on the union of the top 1000 variable genes for each batch as determined by Seurat. Clustering was performed using Seurat, and we isolated the *scrib* epithelium clusters (identifiable by high expression of *Fasciclin 3* and *narrow*) for subsequent comparison with the regeneration and wild-type epithelium data. No *scrib* epithelium cells were filtered during this comparative analysis.

The epithelium data from regeneration, wild-type, and *scrib* samples was initially harmonized with scVI (n_latent = 15, n_layers = 2, gene_likelihood = "nb", max_epochs = 400, and train_size = 0.8), trained on the union of the top 1000 variable genes for each batch as determined by Seurat. The weights from this scVI model were used to initialize a scANVI model (using the from_scvi_model function) for semi-supervised training and label transfer. The cluster identities from our regeneration analysis (**Figure 3.3B**) were supplied as input labels (via setup_anndata), with all *scrib* cells marked as "Unknown". The scANVI model was trained for 50 epochs (max_epochs = 50) to predict the probability of "Unknown" cells belonging to each of the input labels. Input labels were subsampled during training (n_samples_per_label = 150) to prevent the loss of labels with relatively few cells (e.g., Blastema1) during label transfer. After training, the scANVI latent space was used as a basis for UMAP (**Figure 3.16B**), and the transferred labels corresponded to the highest predicted identity for each cell (**Figure 3.16C**).

Immunohistochemistry and microscopy

The following antibodies were from the Developmental Studies Hybridoma Bank (DSHB): mouse anti-Wg (1:100, 4D4), mouse anti-Mmp1 (1:100, a combination of 14A3D2, 3A6B4 and 5H7B11), mouse anti-Broad-Z1 (BrZ1) (1:100, Z1.3C11.OA1), and rat anti-Elav (1:50, Elav-7E8A10). The following antibodies were gifts: rat anti-Zfh2 (1:100, Chris Doe (Tran *et al.*, 2010)), rat anti-Twist (1:1000, Eric Wieschaus), rat anti-Pvf1 (1:500, Ben-Zion Shilo (Rosin *et al.*, 2004)), and pan-hemocyte anti-H2 (1:100) (Kurucz *et al.*, 2003). The following antibodies are from commercial sources: rabbit anti-cleaved Death caspase-1 (Dcp-1) (1:250, Cell Signaling); chicken anti-GFP (1:500, ab13970 Abcam, Cambridge, UK); rabbit anti-PHH3 (1:500, Millipore-Sigma). Secondary antibodies were from Cell Signaling. Nuclear staining with DAPI (1:1000). Tissues were imaged on a Zeiss Axioplan microscope with Apotome attachment, using 10x and 20x objectives. Image files were processed with ImageJ software.

EdU assay and quantification

For EdU staining, live discs were incubated in Schneider's medium (ThermoFisher 21720024) with EdU for 30 minutes, following the protocol for the Click-iT EdU Cell Proliferation Kit, Alexa Fluor 555 (ThermoFisher C10338) and Alexa Fluor 647 (ThermoFisher C10340). After the incubation, discs were fixed in 4% PFA for 15 min, before proceeding with standard antibody staining, as detailed above.

EdU intensity was quantified using the ImageJ software. For each regenerating disc, a square box was drawn, centered around the blastema. The length of the box was 140 microns for the 0h R discs, 160 microns for the 24h and 36h R discs, and 200 microns for the 48h R discs. The EdU intensity was measured at every pixel along the two diagonals of each box using ImageJ's "Plot Profile" function. Subsequent analysis was done using R software. The measured EdU intensities were first z-normalized (i.e., for all values in a measured profile, subtract the mean and divide by the standard deviation) and then averaged across all diagonals from all processed discs at each regenerating time point. The average normalized (scaled) EdU intensity was plotted with

the package ggplot2, and smoothed curves were added using the stat_smooth function with method = "gam".

Drosophila stocks and husbandry

The stocks that were used in this study include: *Ets21C$^{\Delta 10}$* (Mundorf *et al.*, 2019); *upd3-lacZ* (Bunker *et al.*, 2015); *eyFLP; arm-lacZ FRT40A ; hsFLP; FRT40A ; hsFLP; FRT40A ubi-RFP; rn-GAL4, tub-GAL80^{ts}, UAS-rpr (Smith-Bolton et al., 2009)*; *UAS-his2A::RFP ;* and *Him-GFP* (Rebeiz *et al.*, 2002). Stocks obtained from the Bloomington Stock Center include: *Ilp8-GFP (Ilp8^{MI00727}*, Bl33079); *10XSTAT-DGFP (10XSTAT92E-DGFP*, Bl26199, Bl26200) (Bach *et al.*, 2007); *Ets21C-GFP* (*Pbac-Ets21C-GFP.FLAG^{VK00033}*, Bl38639)*; hh-Gal4; rn-Gal4* (Bl7405) ; *rn-GAL4, tub-GAL80^{ts}, UAS-egr* (Bl51280)(*4*); *Df(2L)BSC456* (Bl24960); *UAS-hepWt* (Bl9308)*; Ubi-FRT-stop-FRT-GFP^{nls}* (BL32251) (Evans *et al.*, 2009) *; lexAOp-FLP* (Bl55819); and *GMR26E03-lexA* (Bl54354) (Pfeiffer *et al.*, 2010).

Regeneration experiments

Unless otherwise noted, the genetic ablation system used to study regeneration was *rn-GAL4, tub-GAL80^{ts}, UAS-eiger* (Smith-Bolton *et al.*, 2009). Genetic ablation experiments were conducted by synchronizing development by collecting eggs on grape plates and picking 55 L1 larvae into vials with yeast paste. Temperature shifts to induce ablation (from 18℃ to 30℃) were conducted on day 7 after egg lay (AEL) for 40 hours. The extent of adult wing regeneration was scored by binning the resulting wings into 5 categories (0%, 25%, 50%, 75%, and 100%) (*4*). The resulting regeneration scores were calculated per population. Experimental replicates were done on separate days with a minimum of 2 vials per genotype and three replicates per genotype. Statistical comparison performed on regeneration scores using ANOVA followed by Tukey test for significance.

Mitotic clones during regeneration

Mosaic tissues were generated by recombinase-driven (FLP/FRT) mitotic recombination within the genetic background of the ablation system. The expression of *hsFLP* was induced by an 1h heat-shock at 37℃ on day 3 AEL, which generated clones throughout the imaginal discs prior to genetic ablation and regeneration. Mutant cells were labeled by the absence of RFP and wild-type cells were marked by 2X RFP. The genotype of the experimental larvae used to generate *Ets21C* mutant clones during regeneration: *hsFLP; Ets21C$^{\Delta 10}$, FRT40A / ubi-RFPnls, FRT40A; rn-GAL4, tub-GAL80^{ts}, UAS-eiger / Ilp8-GFP* (**Figure 3.11O**).

Lineage-tracing experiments

We identify an enhancer for the gene *grain (grn)* that was primarily expressed in the inner-hinge, G*MR26E03-lexA* (Pfeiffer *et al.*, 2010) during normal development (**Figure 3.3N**). Lineage-tracing was performed by permanently labelling the cells that expressed *GMR26E03-lexA* by driving the expression of the recombinase FLP (*lexAop-FLP*) to induce the removal of a stop-cassette (*Ubi-FRT-stop-FRT-GFP^{nls})* (**Figure 3.3M-O**).

Pupariation timing experiments

Images were taken every 20 minutes of vials that contained animals as they transitioned between larva to pupa. This was performed at 18℃ with a wide-angle camera (Arducam). Pupariation was scored by observing when the animals stopped moving and darken in color.

Physical wounding assay

Wing discs were physically wounded *in situ* as described in (Yoo *et al.*, 2016). Briefly, L3 larvae with the wing pouch fluorescently labeled (*rn-GAL4, UAS-his2A::RFP)* were visualized using a fluorescence microscope. The right wing pouch was wounded by carefully applying pressure on the larval cuticle using a thin gauge insulin needle without penetrating the larval cuticle. Larvae were then returned to vials containing Bloomington food and dissected 6 hours or 24 hours later.

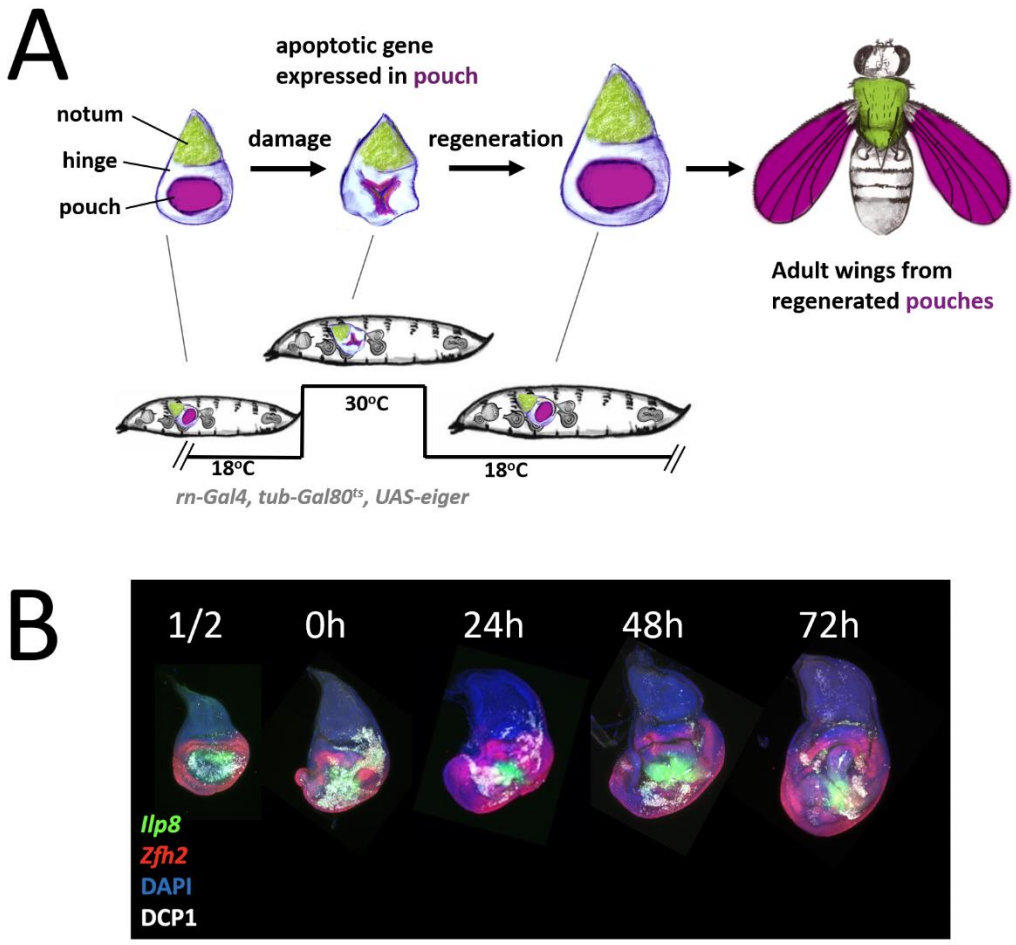

Figure 3.1. Schematic of genetic ablation system to study imaginal disc regeneration.
(**A**) Schematic of genetic ablation system. The canonical domains of the wing disc and the adult structures to which they give rise are colored in green (notum), white (hinge), and purple (pouch). Expression of the pro-apoptotic gene *eiger* is targeted to the wing pouch using *rn-GAL4* and *UAS-eiger*. Gal4 function is inhibited at 18° C and permitted at 30° C by a ubiquitously-expressed temperature-sensitive Gal80 (*tub-Gal80^{ts}*). (**B**) Imaginal discs can be dissected and analyzed during and after the ablation period. The half-way point through the 40h ablation is indicated by "½". Other times refer to the time after the downshift to 18° C - the phase when regeneration occurs.

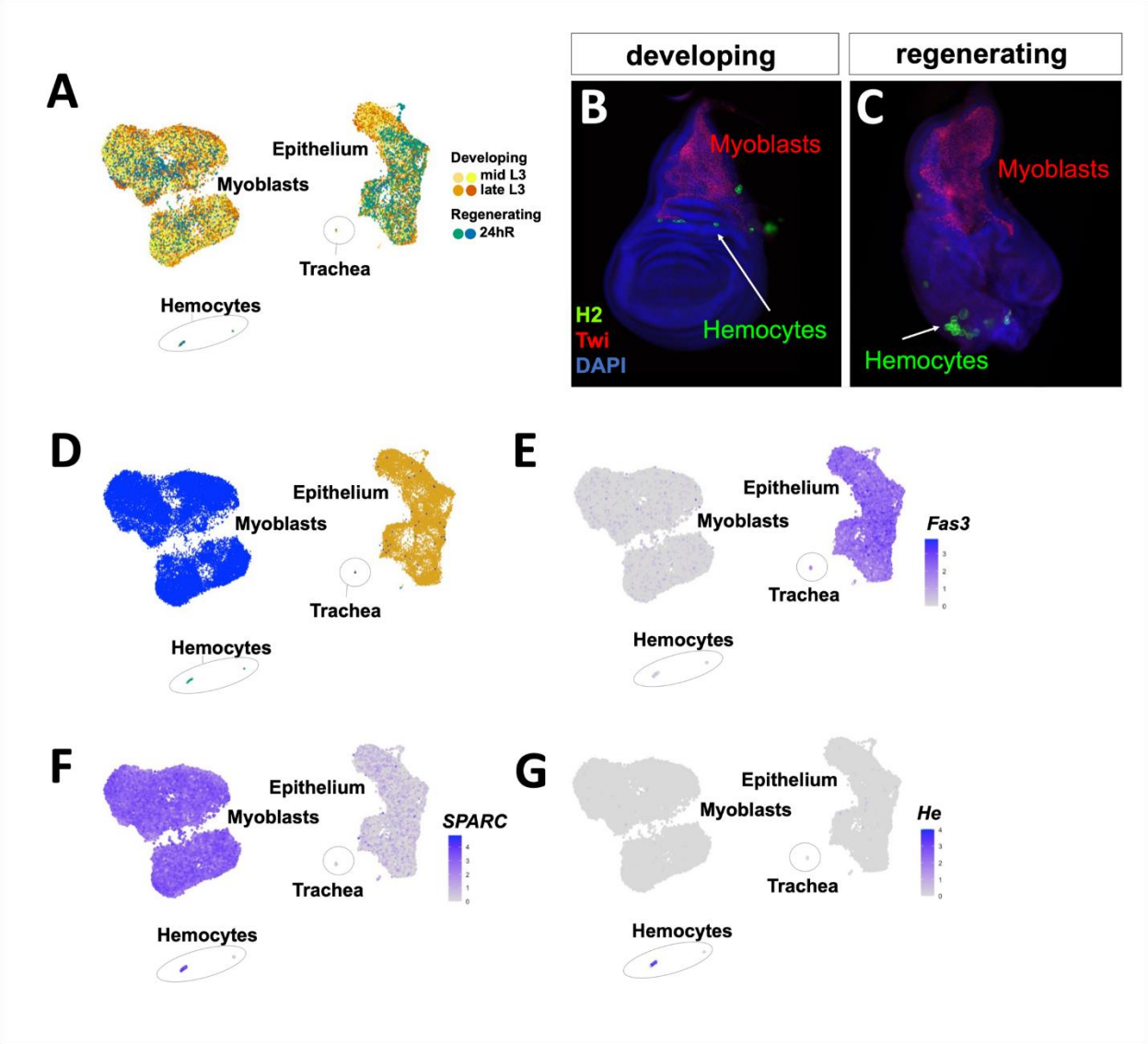

Figure 3.2. Three major cell types were identified from scRNAseq of wing imaginal discs. (**A**) Harmonized UMAP of scRNA data from wing imaginal discs. Data colored by sample of origin and labeled. Samples were derived from developing discs at the middle at late stages of the third larval instar (L3), as described previously (Everetts *et al.*, 2021), and from regenerating discs 24h after the downshift to 18^{o} C. Two biological replicates were obtained for each sample (see **Materials and Methods**). The three major cell types identified were epithelial cells, myoblasts and hemocytes. In addition, a few trachea cells were also identified. The cell counts from the regenerating discs were: 6,613 epithelial cells, 7,466 myoblasts, 224 hemocytes and 17 trachea cells. (**B, C**) Wing-imaginal discs stained with anti-H2 to label the hemocytes and anti-Twist to label the myoblasts in developing (**B**) and regenerating (**C**) wing discs. (**D**) UMAP colored by major cell types: myoblasts, epithelial cells, and hemocytes are shown in different colors (**E-G**) Expression of marker genes for the three major cell types: (**E**) *Fasciclin 3* (*Fas3)* expression marks the epithelium; (**F**) *Secreted protein, acidic, cysteine-rich* (*SPARC)* expression marks the myoblasts; and (**G**) *Hemese* (*He)* marks the hemocytes.

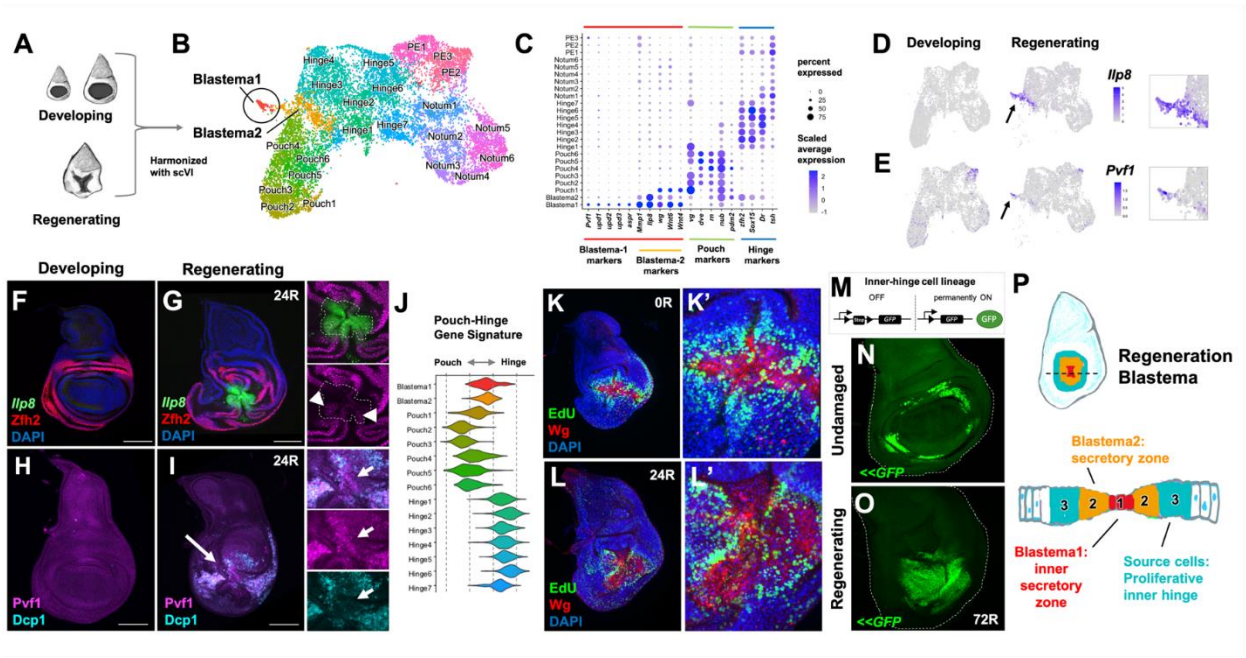

Figure 3.3. Single-cell analysis reveals two distinct cell states in the regeneration blastema. (**A**) Diagram of imaginal disc samples compared by scRNAseq. (**B**) UMAP of harmonized epithelial cells. (**C**) Dot plot summarizing gene expression for cluster marker genes. (**D, E**) Expression of *Ilp8* (**D**) and *Pvf1* (**E**) as visualized on UMAP. (**F-I**) Developing and regenerating wing discs, after 24 h of regeneration (24R), (**F, G)** with an *Ilp8-GFP* reporter stained with anti-Zfh2 (hinge marker), and (**H, I**) stained with anti-Pvf1 and anti-cleaved Death caspase-1 (Dcp-1) (detects apoptotic cells and debris). (**J**) Pouch-Hinge gene signature analysis of blastema cell clusters. (**K, L**) Regenerating wing discs at 0R and 24R with cells in S-phase visualized by EdU incorporation. (**M**) Schematic of lineage-tracing technique with an inner-hinge enhancer. Lineage tracing in (**N**) normal development and (**O**) following 72h or regeneration. (**P**) Schematic of distinct cell types of the blastema.

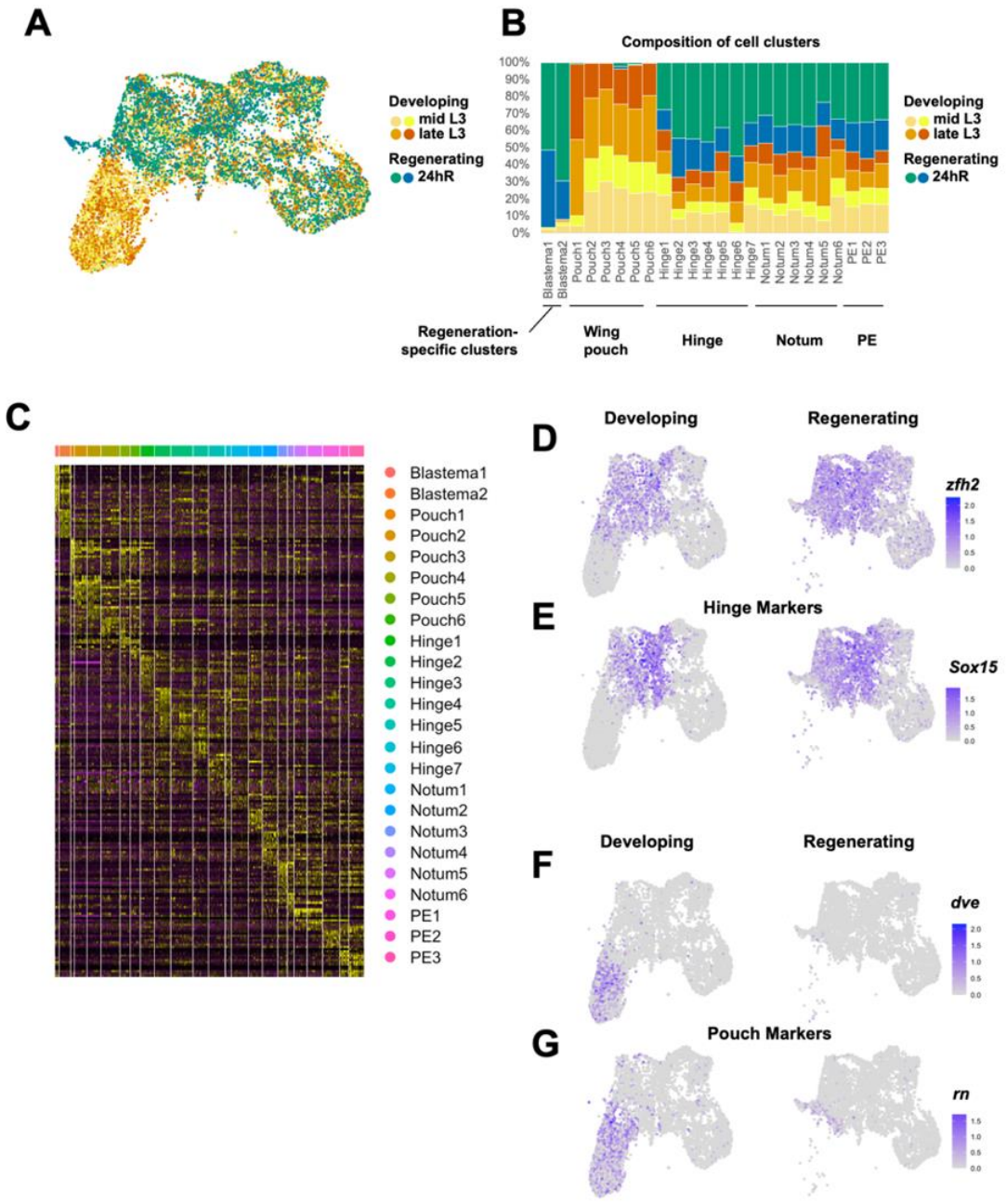

Figure 3.4. Composition of cell clusters in developing and regenerating discs.
(**A**) UMAP of harmonized data from epithelial cells from regenerating and developing (from two time points) samples. Each dataset, including replicates, is represented in a distinct color. (**B**) Composition of cell clusters, as shown in **Figure 3.3B**. Note the underrepresentation of cells assigned to pouch clusters in regenerating discs and the near absence of cells assigned to the Blastema1 and Blastema2 clusters in developing discs. (**C**) Heatmap showing differential expression of marker genes through the different cell clusters of the harmonized epithelial cell object. Gene expression with individual cells from the single-cell data for the hinge makers *Zn finger homeodomain 2* (*zfh2)* (**D**) and *Sox box protein 15* (*Sox15)* (**E**) and for the pouch markers *defective proventriculus* (*dve)* (**F**) and *rotund (rn)* (**G**).

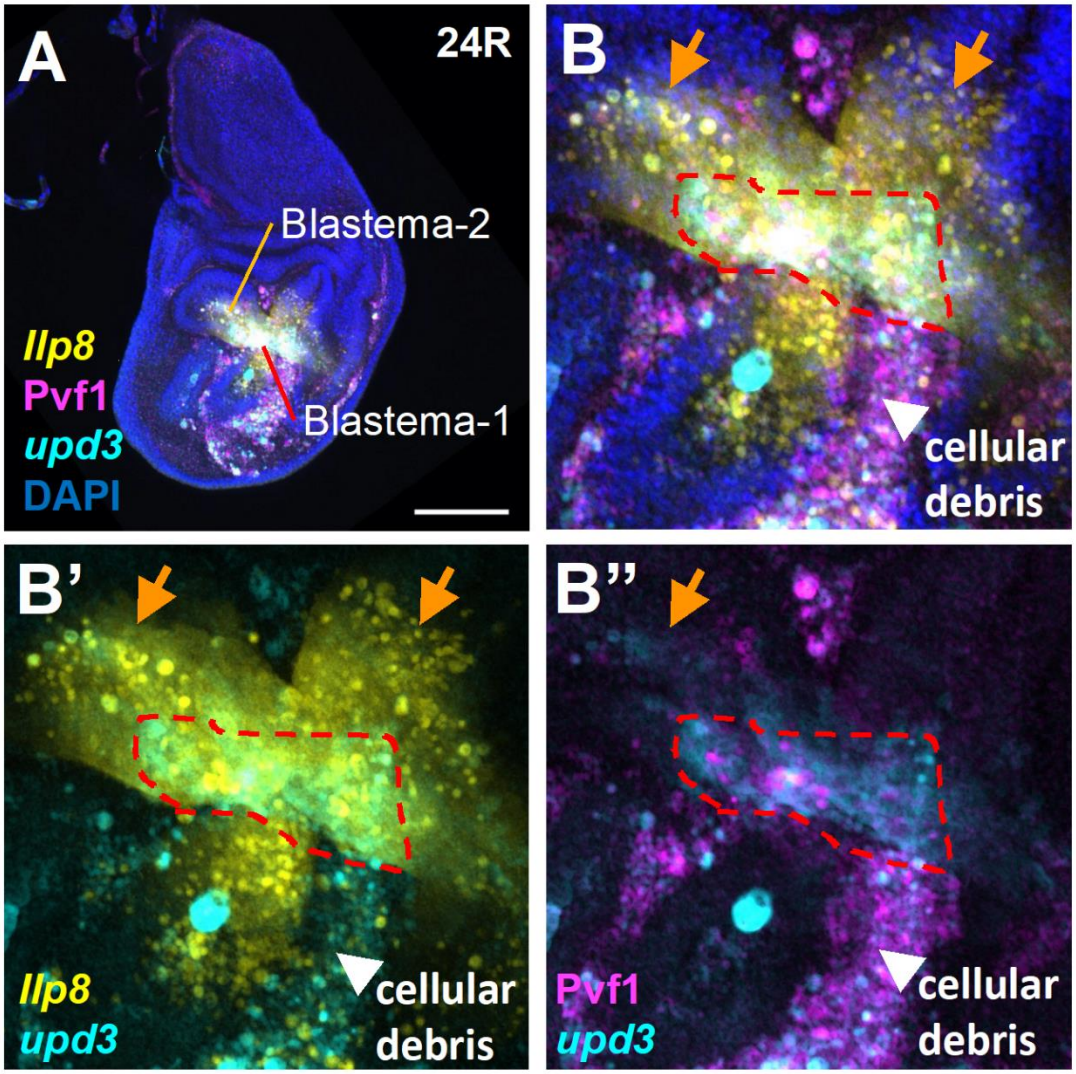

Figure 3.5. Co-expression of Blastema-1 and Blastema-2 marker genes within the regenerating epithelium.

(**A, B**) Regenerating wing-imaginal disc, after 24h of regeneration, with transcriptional reporters and antibody staining to highlight the nested position of Blastema-1 and Blastema-2 cells within the regenerating epithelium. *Ilp8-GFP* expression is shown in yellow, anti-Pvf1 stain shown in magenta, and *upd3-lacZ (Bunker et al., 2015)* shown in cyan. (**B**) Magnification of blastema. Orange arrow highlights the region of *Ilp8* expression and the red dotted line highlights the region of higher Pvf1 and *upd3-lacZ* expression that is in the surviving epithelial cells. Note that the cellular debris shows evidence of expression of all three of these marker genes (white arrowhead). Microscopy scale bars = 100 μm.

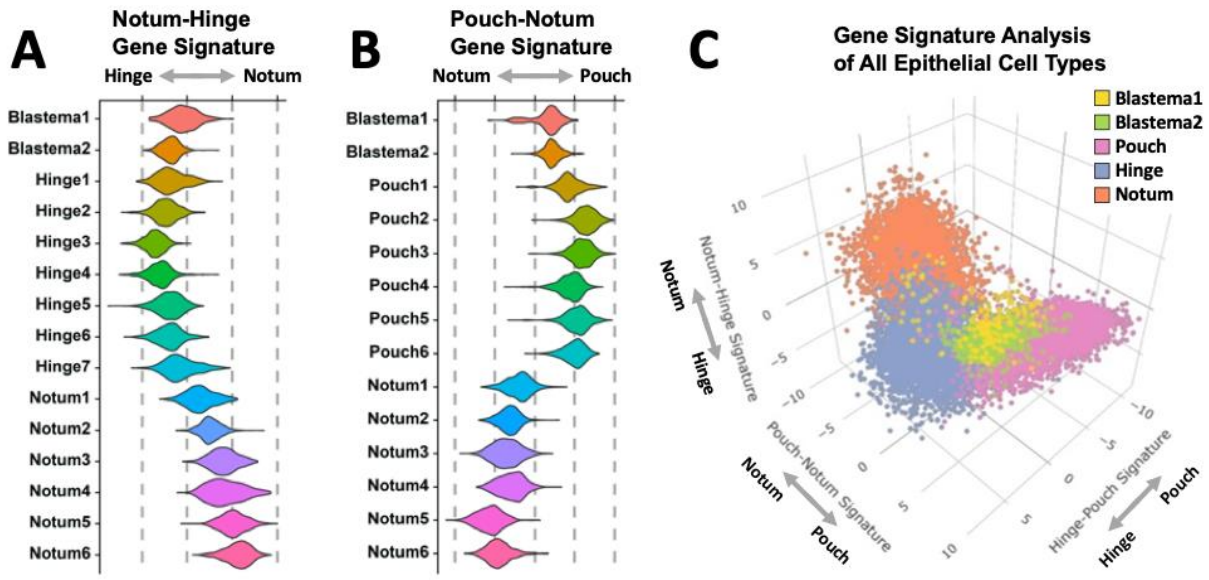

Figure 3.6. Comparative gene signature analysis of major epithelial domains for the blastema clusters.

(**A, B**) Clusters within the scRNAseq dataset scored with a notum-hinge (**A**) or pouch-notum (**B**) gene signature. Note that scores for the blastema clusters are more closely aligned with the hinge (**A**) or pouch (**B**) rather than the notum. The hinge-pouch signature is shown in **Figure 3.3J.** (**C**) 3D signature plot of scRNAseq clusters scored by notum-hinge, pouch-notum, and pouch-hinge gene signatures. Axes correspond to values shown in **A**, **B**, and **Figure 3.3J**. Note that both Blastema1 and Blastema2 cells are centered between hinge and pouch fates.

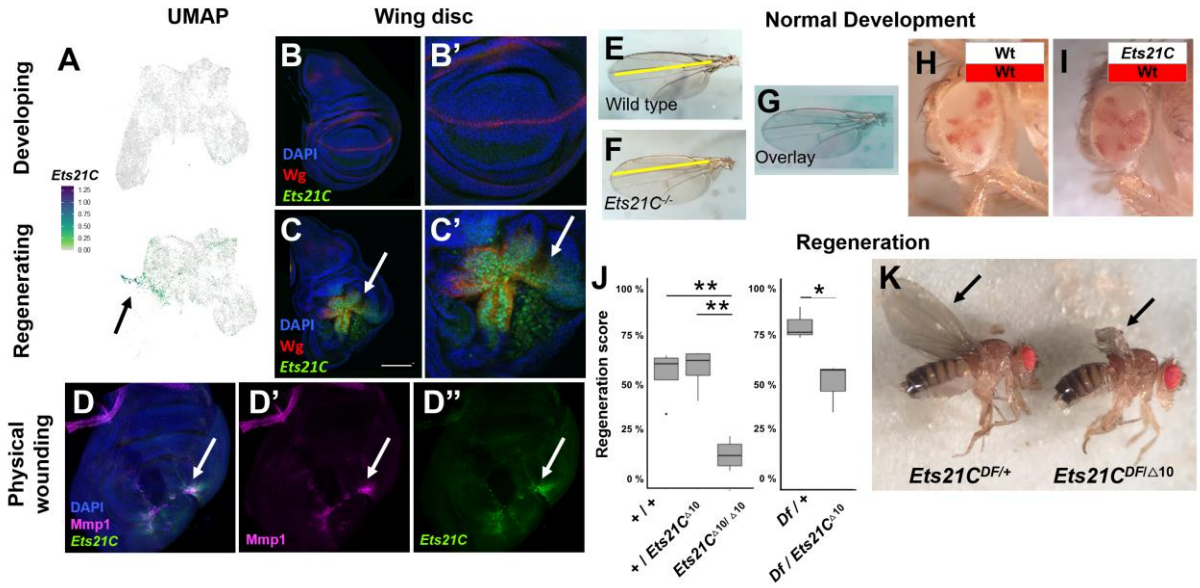

Figure 3.7. Transcription factor Ets21C is specifically required for regeneration.
(**A**) *Ets21C* expression in developing and regenerating scRNAseq data. (**B, C**) *Ets21C-GFP* expression in (**B**) developing and (**C**) regenerating wing discs. (**D**) *Ets21C-GFP* expression following physically wounding disc through larval cuticle. (**E-G**) Wing blades from wild-type and *Ets21C* mutant animals, raised in standard conditions, and shown overlaid. (**H, I**) Mosaic adult eyes with control (**H**) and *Ets21C* (**I**) cells marked by the absence of red pigment. Note that *Ets21C* mutant cells (**I**) contribute to tissue at a similar proportion as control cells (**H**). (**J, K**) The extent of regeneration, as scored by the size of the resulting wing blades, p values: * <0.05, ** <0.005 (see **Materials and Methods**).

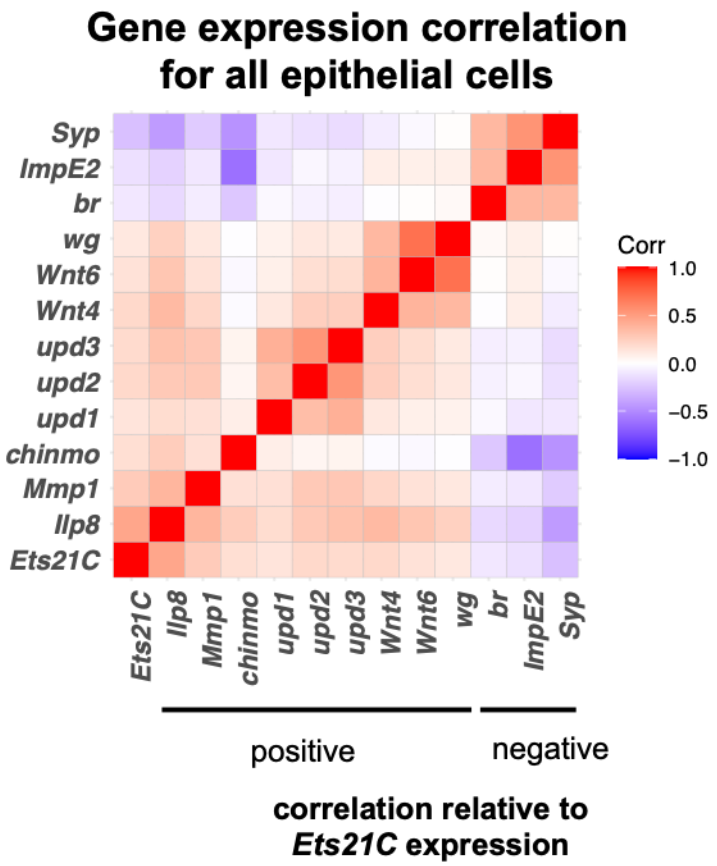

Figure 3.8. Genes that positively or negatively correlate with *Ets21C* expression.
Pearson correlation of normalized gene expression data within all epithelial cells (both developing and regenerating datasets). The gene with the most correlated expression to *Ets21C* expression is *Ilp8*. *Ets21C* expression is also positively correlated with the expression of *Mmp1*, *chinmo*, *upd1*, *upd2*, *upd3*, *Wnt4*, *Wnt6* and *wg*. Examples of genes that show negative gene expression correlated with *Ets21C* include *broad (br)*, *Ecdysone-inducible gene E2* (*ImpE2)*, and *Syncrip* (*Syp)*.

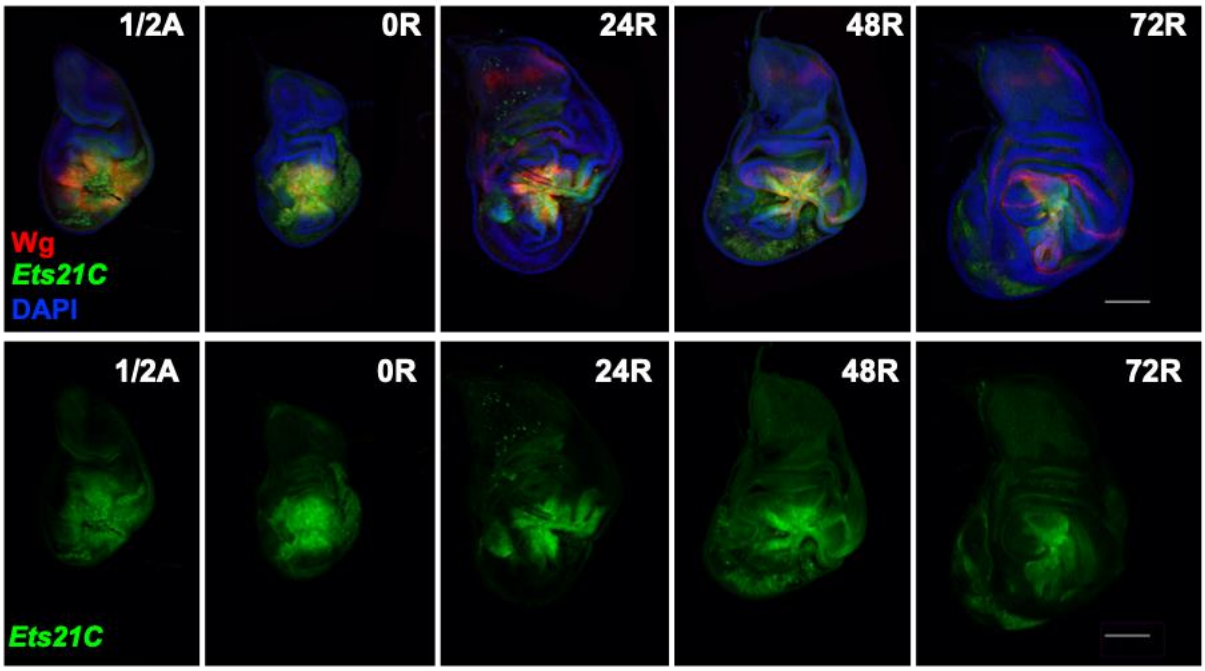

Figure 3.9. Expression of *Ets21C-GFP* over the course of regeneration.
Wing imaginal discs dissected half-way (20 h) through the ablation period (1/2A) and time points (indicated in hours) during regeneration at 18C (0R, 24R, 48R, 72R). Regenerating wing discs are also stained with anti-Wg (in red). Regeneration is near complete by 72 h. Microscopy scale bars = 100 μm.

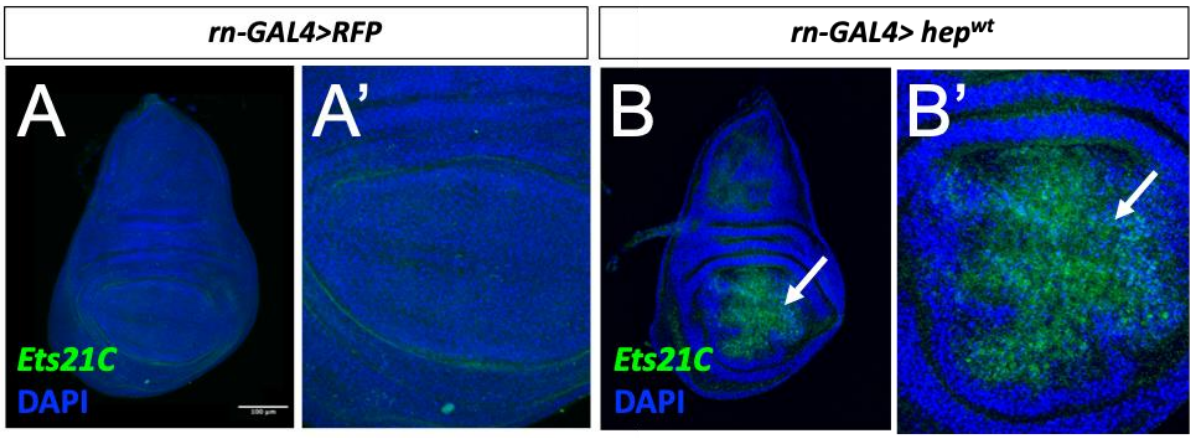

Figure 3.10. Activation of the JNK/AP1 pathway in non-ablated tissue induces the expression of *Ets21C-GFP*.
(**A, B**) Expression of *Ets21C-GFP* in a wing disc with *rn-GAL4* driving the expression of *UAS-RFP* (**A**) or *UAS-hep*^{wt} (**B**). Note that driving the expression of the wild-type version of the JNK-kinase *hemipterous* (*hep*) within the wing pouch leads to the expression of *Ets21C*. Microscopy scale bars = 100 μm.

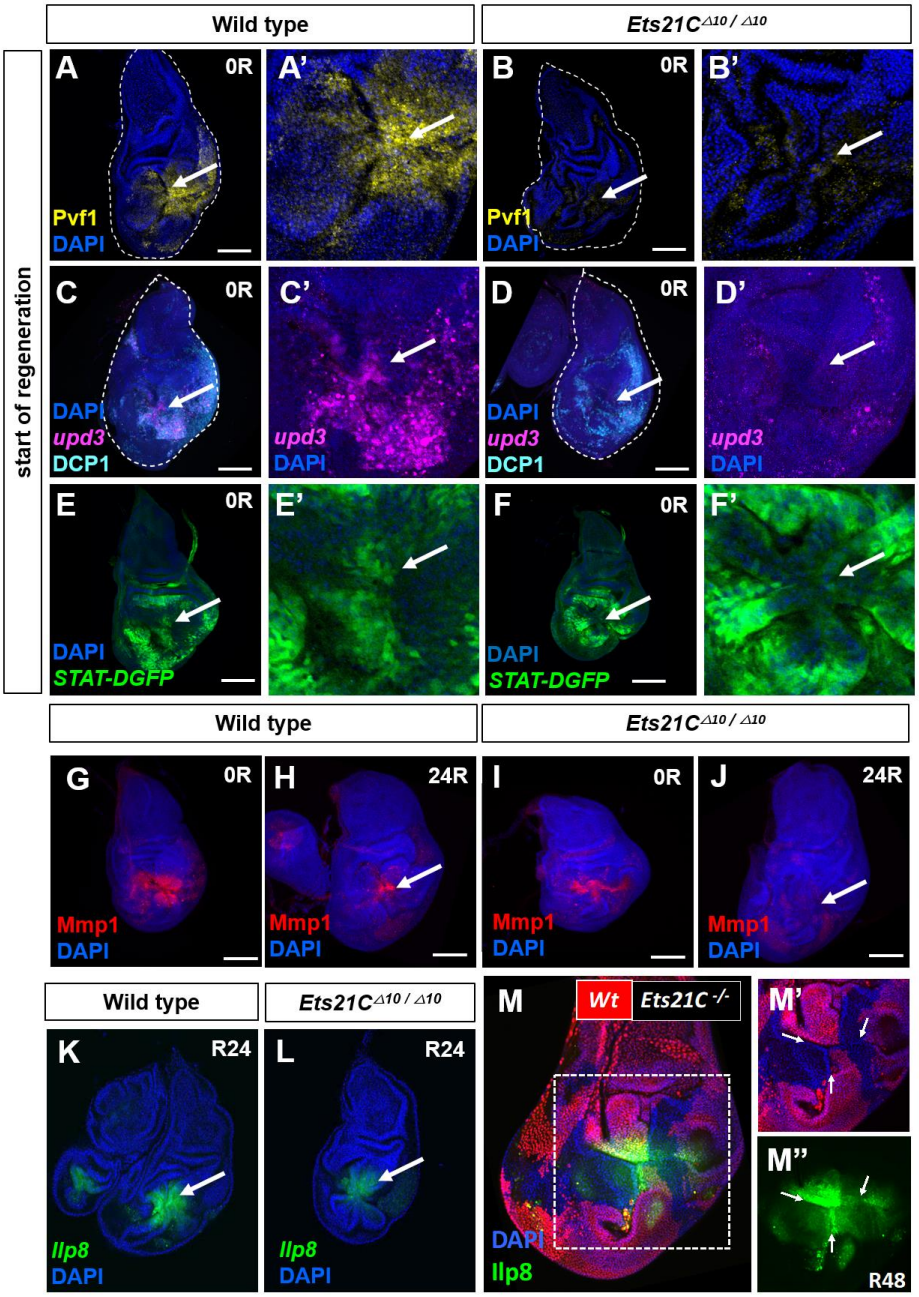

Figure 3.11. *Ets21C* regulates a pro-regenerative program.
(**A-L**) Wild-type or homozygous *Ets21C* mutant regenerating imaginal discs. (**A-F**) Imaginal discs at the start of the regeneration period (0R), (**A, B**) showing Pvf1 expression, (**C, D**) *upd3-lacZ* expression and apoptosis (cleaved Dcp-1), and (**E, F**) JAK/STAT activity (*STAT-DGFP* reporter). Note the decreased activation in *Ets21C* mutant tissues. (**G-J**) Mmp1 expression at the start (0R) and after 24h of regeneration (24R). (**K, L**) *Ilp8-GFP* reporter expression at 24R. (**O**) Mosaic tissues created by mitotic recombination. *Ets21C* mutant cells marked by the absence of RFP. Note that *Ets21C* mutant cells are able to proliferate during regeneration, but show a cell-autonomous decrease in *Ilp8-GFP* expression at 48h of regeneration.

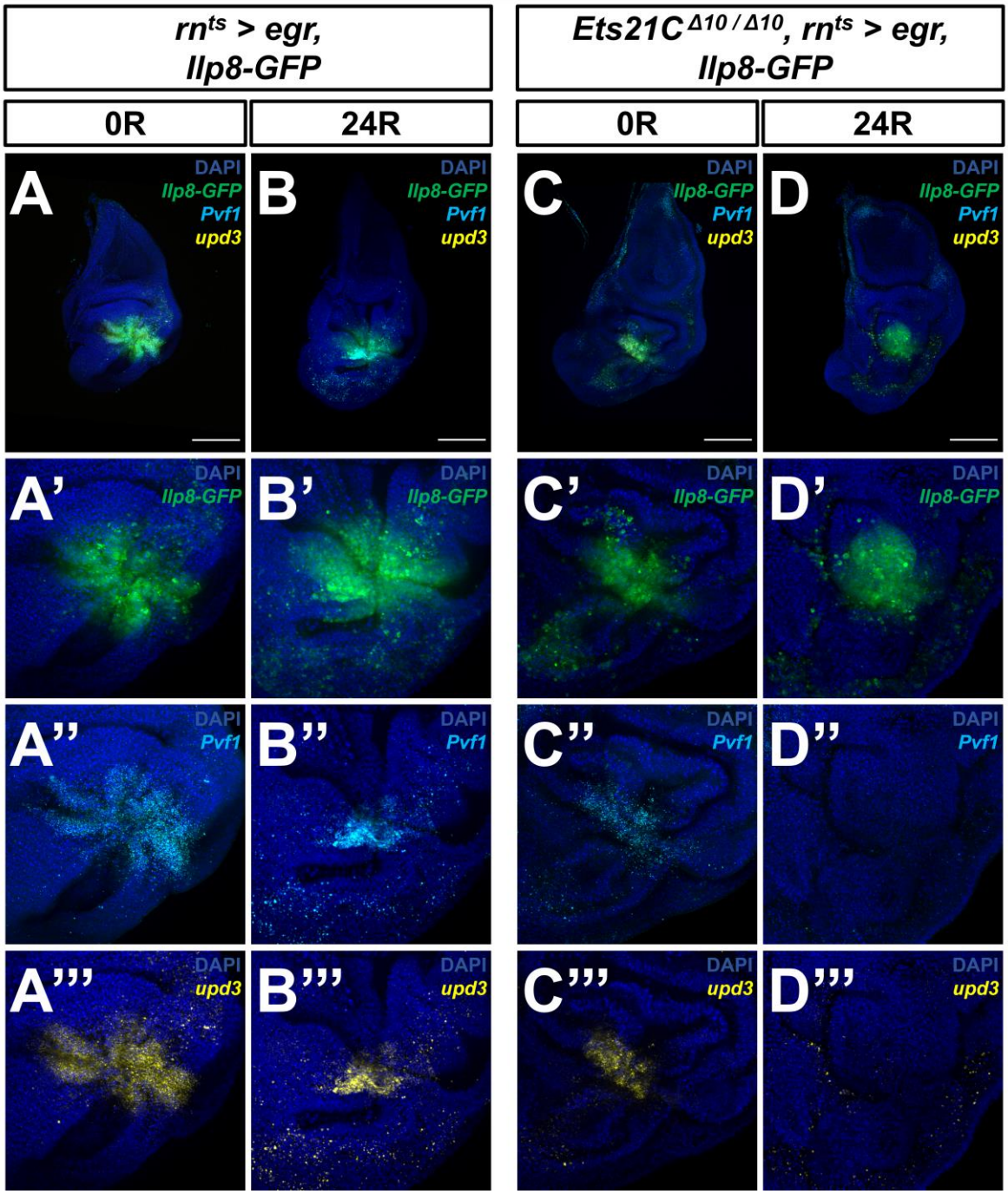

Figure 3.12. HCR reveals *Ets21C*-dependent maintenance of *Pvf1* and *upd3* transcription. (**A, B**) Wild-type or (**C, D**) *Ets21C* mutant wing discs at either the start (0R) (**A, C**) or after 24h of regeneration (24R) (**B, D**). Discs are all stained for *Pvf1* and *upd3* transcripts using hybridization chain reaction (HCR) (Choi et al., 2018). Note that in wild-type *Ets21C* discs, *Pvf1* and *upd3* are expressed at appreciable levels at both 0R and 24R, albeit with higher expression at 0R. In *Ets21C* mutant discs, *Pvf1* and *upd3* are only noticeable at 0R; by 24R, expression of these genes has attenuated.

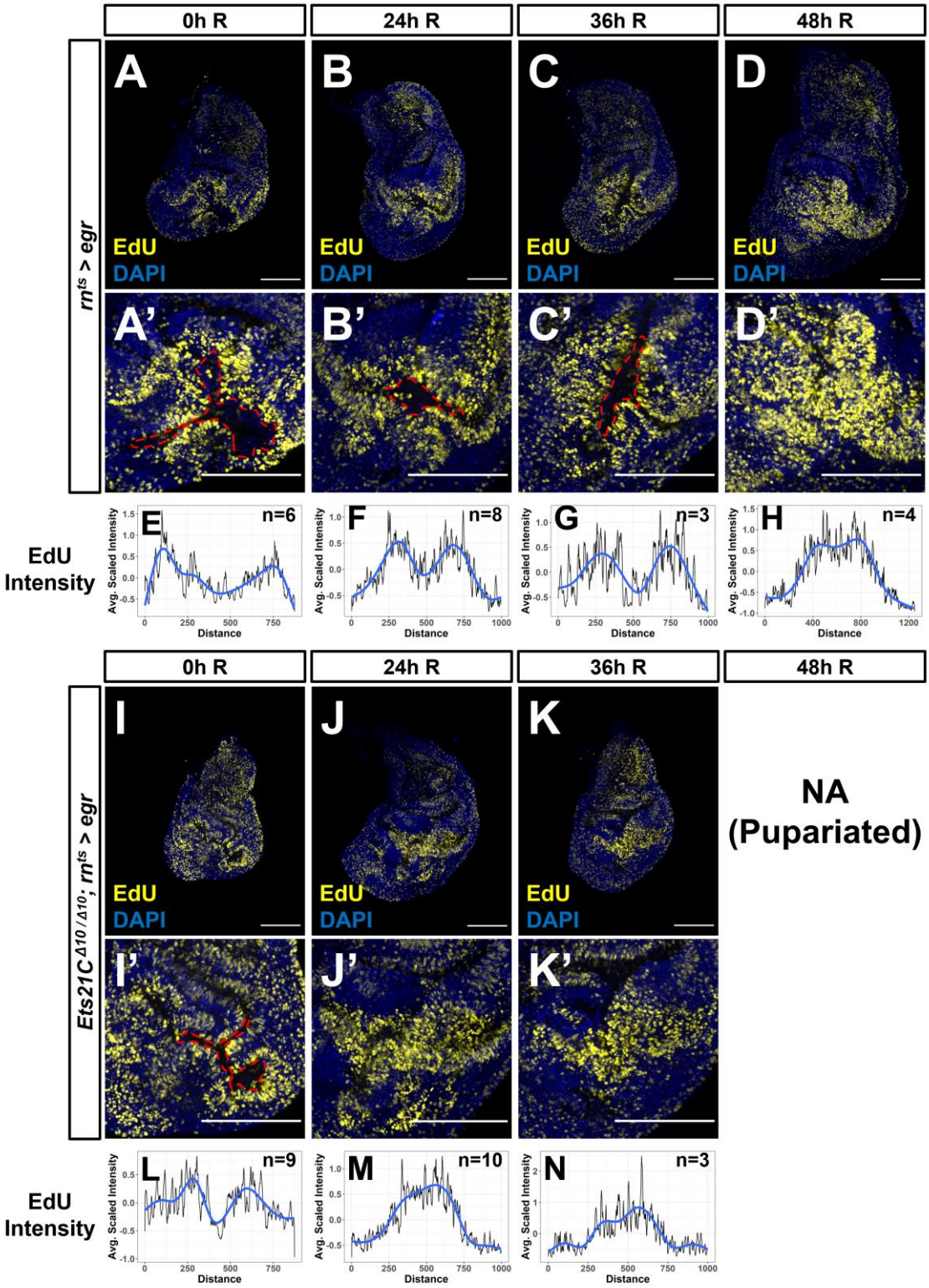

Figure 3.13. Loss of *Ets21C* disrupts the non-proliferative zone at the blastema center.

(**A-D, I-K**) Imaginal discs at different time points during regeneration (0R, 24R, 36R, 48R) for wild-type (**A-D**) or *Ets21C* mutant (**I-K**) tissues. Cell proliferation is marked by the incorporation of the thymidine analog EdU. Note that for the *Ets21C* mutants, the regeneration period is ended prematurely by pupariation (**Figure 3.15**). (**E-H, L-N**) Profiles of average EdU intensity within the blastema at each regeneration time point for wild-type (**E-H**) or *Ets21C* mutant (**L-N**) discs. The y-axis corresponds to the average of z-normalized EdU intensity (see **Materials and Methods**). The x-axis corresponds to measured pixels centered around the blastema (i.e., the center of the distance axis is the center of the blastema). Note that in the wild-type tissues there is reduced EdU intensity within the center of the blastema, highlighted by the areas within the red dashed lines and by the bimodality of the EdU intensity profiles (most notable at 0R, 24R, and 36R). This pattern of proliferation is less robust within *Ets21C* mutant tissue, as evidenced by the unimodality of the EdU intensity profiles at 24R and 36R.

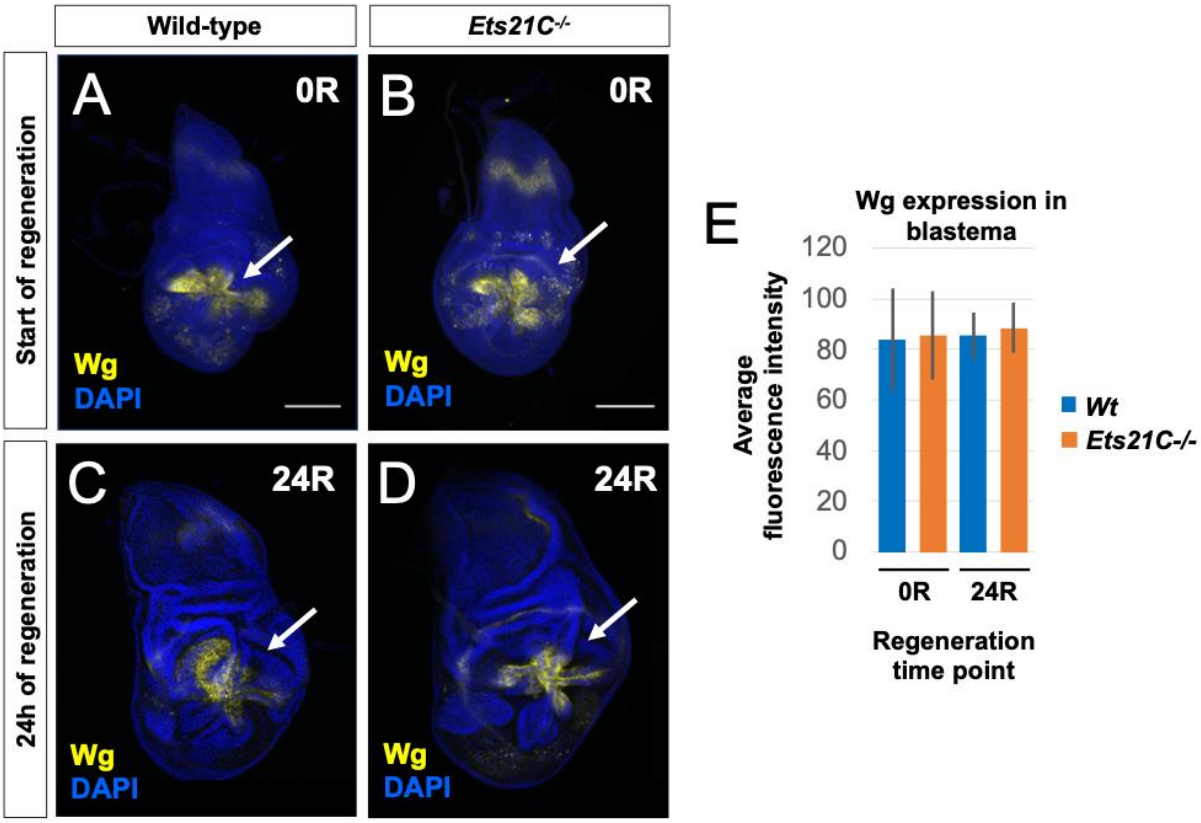

Figure 3.14. Wnt ligand Wg expression is unaffected during regeneration in *Ets21C* mutants. (**A-D**) Regenerating wing imaginal discs at the start (**A, B**) and after 24 h of regeneration (**C**, **D**) for wild type and *Ets21C* mutant tissues. Discs stained with anti-Wg and DAPI. Arrows are pointing to the area of the blastema. (**E**) Average fluorescent intensity of the anti-Wg stain within the blastema region for both wild type (Wt) and *Ets21C* mutant tissues. No change in the amount of Wg expression was detected at either regeneration time point.

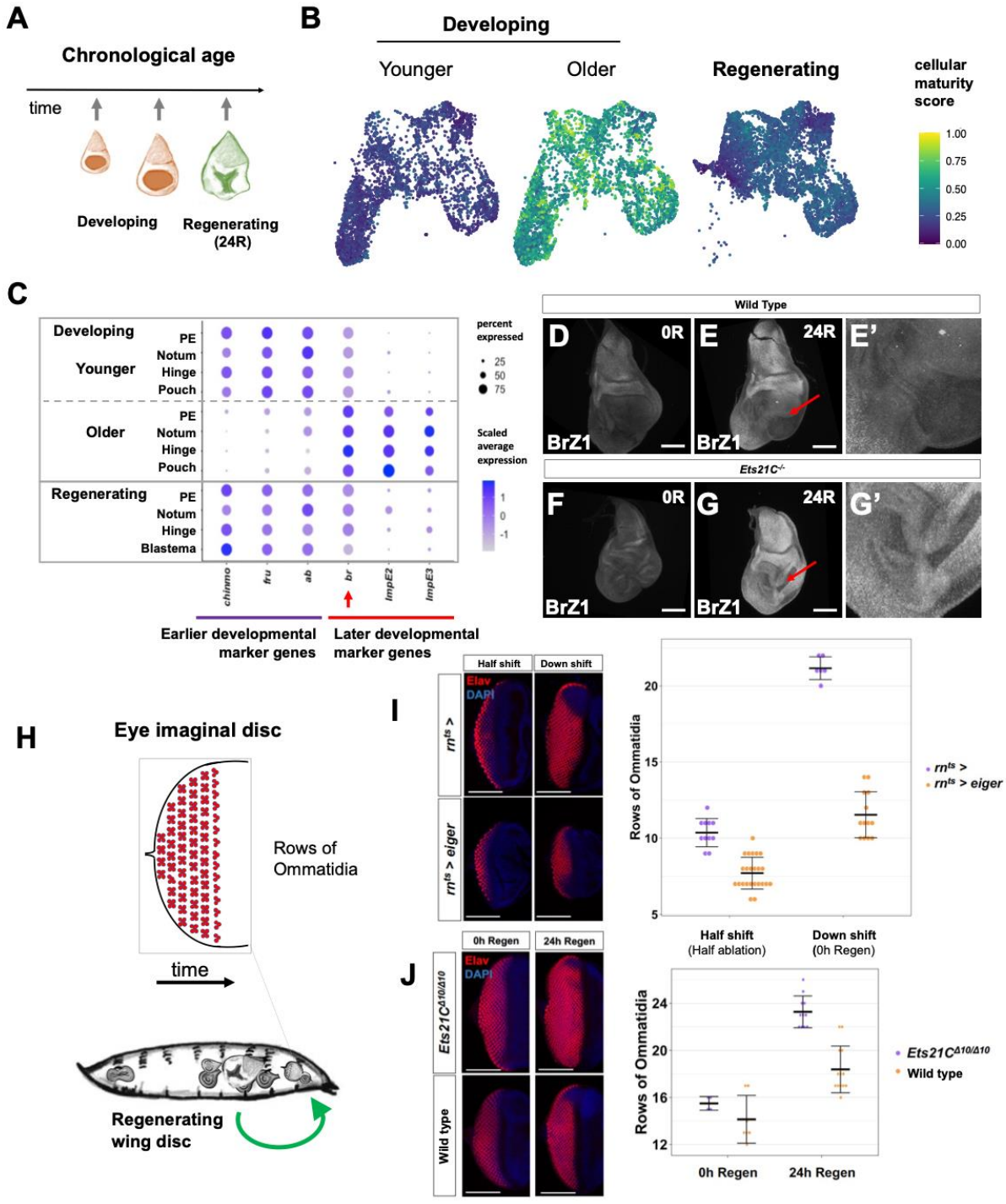

Figure 3.15. Developmental progression and cellular maturity during regeneration.
(**A**) Diagram of the relative chronological age of our single-cell datasets. Note that the regenerating imaginal discs are chronologically older than both developmental time points, which are from mid and late 3rd instar, because tissue damage results in an extended larval phase during with regeneration occurs (Smith-Bolton *et al.*, 2009; Halme *et al.*, 2010; Katsuyama *et al.*, 2015; Harris *et al.*, 2016). (**B**) Based on genes with high differential expression during normal development, we calculated a cellular maturity score to determine the relative developmental maturity of individual

cells from the single-cell data. Note that the cells within the regenerating sample show a lower or intermediate cellular maturity score, indicating that the developmental progression of the cells within the epithelium have been either paused or rejuvenated to an earlier time point in development. (**C**) Dot plot showing the relative expression levels for several genes that change over the course of development, including three genes expressed earlier development: *Chronologically inappropriate morphogenesis* (*chinmo), fruitless (fru)* and *abrupt (ab)*; and three gene expressed in older imaginal discs during normal development: *broad (br), Ecdysone-inducible gene E2 (ImpE2),* and *Ecdysone-inducible gene E3 (ImpE3).* Note that the regenerating sample shows similar expression patterns to the younger sample. (**D-G**) Regenerating imaginal discs stained with an antibody that recognizes the Z1 isoform of the Broad transcription factor for wild-type (**D, E**) and *Ets21C* mutant (**F, G**) discs. Note that *Ets21C* expression was found to negatively correlate with *broad (br)* expression (**Figure 3.8**). Discs were dissected at the start of the regeneration period (0R) (**D, F**) and after 24h of regeneration (24R) (**E, G**). Note that BrZ1 levels increase from 0R to 24R and are at higher levels in the non-regenerating regions of the tissue and lower in the regenerating pouch region. Also note that *Ets21C* mutant discs have higher BrZ1 levels at 24R, especially in the region of the regenerating pouch. (**H-J**) To investigate the developmental progression of other tissues with the animal undergoing regeneration, we have counted the number of ommatidial rows within the eye-imaginal disc. Previous work has suggested that the growth of other imaginal discs may pause during the process of wing-imaginal disc regeneration, based on the overall size of the imaginal discs (Boulan *et al.*, 2019). (**H**) Diagram of how the "eye-clock" can be used to compare the organismal-wide developmental progression. (**I**) The number of ommatidial rows for undamaged controls as compared to regenerating larvae. Note that the rate of addition of new rows of ommatidia added has slowed in the regenerating sample. (**J**) Comparison of the number of ommatidial rows for wild-type and *Ets21C* mutants during regeneration. Note that the *Ets21C* mutant animals show an increased number of ommatidial rows by 24h of regeneration, indicating that there is a reduction in the organism-wide developmental delay that is observed in wild-type regenerating larvae.

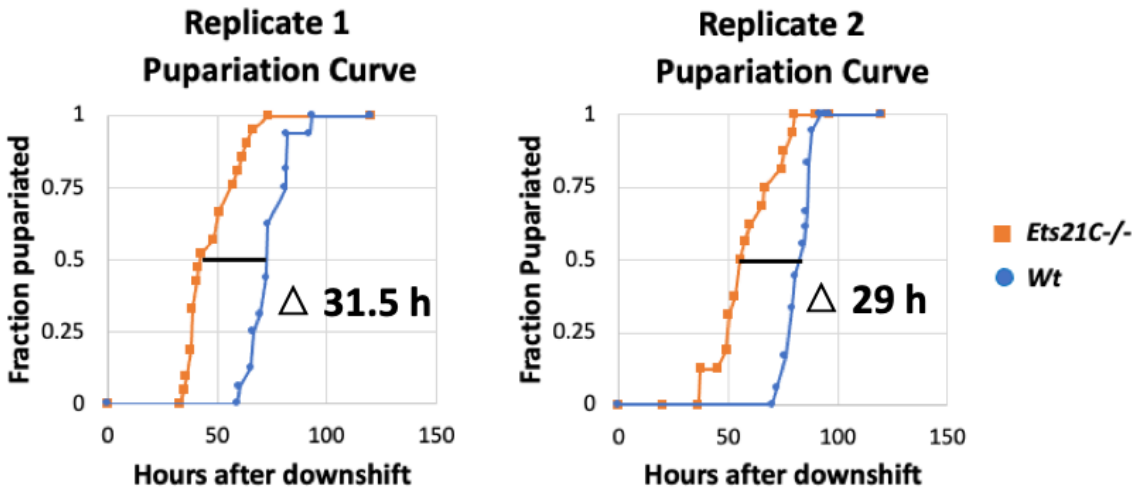

Figure 3.16. Relative pupariation timing for wild type (Wt) and *Ets21C* mutant animals following damage and regeneration.
Images of vials were taken at 10-minute intervals and then scored based on coloration changes during pupariation. Replicates were biological replicates conducted on separate days. The relative difference in pupariation timing was calculated based on the point when one-half of the animals scored had pupariated. Following ablation and regeneration, *Ets21C* mutant larvae formed pupa 31.5 hours and 29 hours earlier than the wildtype controls.

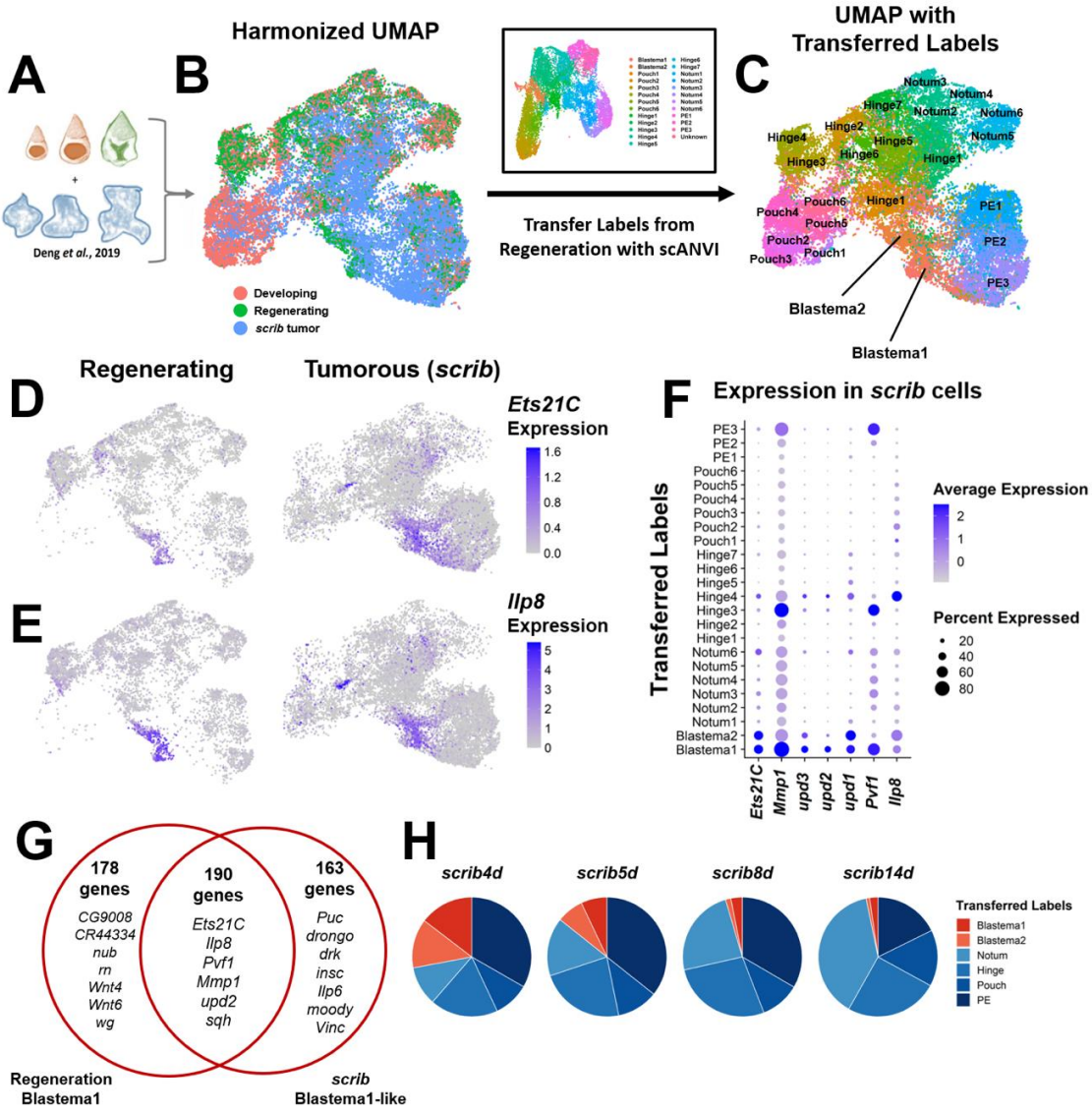

Figure 3.17. Single-cell comparison of regenerating and *scrib* tumorigenic discs.
(**A**) Diagram of combined single-cell data. (**B**) Harmonized UMAP with cells colored by sample of origin (as also shown in **Figure 3.11Q**). (**C**) Harmonized UMAP colored by cell identity as generated by transferring labels from the regeneration cell atlas onto the *scrib* dataset using scANVI (see Materials and Methods). (**D, E**) Expression UMAPs of *Ets21C* (**D**) and *Ilp8* (**E**) in regenerating and *scrib* tumor datasets (note that expression for *scrib* tumors alone was also shown in **Figure 3.11R**). (**F**) Dot plot summarizing the expression of blastema markers within the *scrib* data. Note the co-expression of these genes in *scrib* clusters with Blastema1 and Blastema2 transferred labels. (**G**) Venn diagram comparing the overlap of markers for the Blastema1 cells within the regenerating data and the Blastema1-like cells within the *scrib* data. (**H**) Quantification of transferred labels within the *scrib* datasets, over the course of tumor development as collected by *Deng et al.* 2019 (*8*) (4d, 5d, 8d, and 14d = 4, 5, 8, and 14 days of tumor development, respectively). Note that there are more blastema-like cells on day 4 and 5 than on day 8 and 14 of within the *scrib* tumorous discs. This data is also summarized as "*scrib* early*"* and *"scrib* late" in **Figure 3.11S**.

REFERENCES

Alon, U., 2007. Network motifs: theory and experimental approaches. Nat Rev Genet 8, 450-461.

Ayala-Camargo, A., Anderson, A.M., Amoyel, M., Rodrigues, A.B., Flaherty, M.S., Bach, E.A., 2013. JAK/STAT signaling is required for hinge growth and patterning in the Drosophila wing disc. Dev Biol 382, 413-426.

Aztekin, C., Hiscock, T.W., Marioni, J.C., Gurdon, J.B., Simons, B.D., Jullien, J., 2019. Identification of a regeneration-organizing cell in the Xenopus tail. Science 364, 653-658.

Bach, E.A., Ekas, L.A., Ayala-Camargo, A., Flaherty, M.S., Lee, H., Perrimon, N., Baeg, G.H., 2007. GFP reporters detect the activation of the Drosophila JAK/STAT pathway in vivo. Gene Expr Patterns 7, 323-331.

Bageritz, J., Willnow, P., Valentini, E., Leible, S., Boutros, M., Teleman, A.A., 2019. Gene expression atlas of a developing tissue by single cell expression correlation analysis. Nat Methods 16, 750-756.

Bilder, D., Li, M., Perrimon, N., 2000. Cooperative regulation of cell polarity and growth by Drosophila tumor suppressors. Science 289, 113-116.

Bosch, M., Serras, F., Martin-Blanco, E., Baguna, J., 2005. JNK signaling pathway required for wound healing in regenerating Drosophila wing imaginal discs. Dev Biol 280, 73-86.

Boulan, L., Andersen, D., Colombani, J., Boone, E., Leopold, P., 2019. Inter-Organ Growth Coordination Is Mediated by the Xrp1-Dilp8 Axis in Drosophila. Dev Cell 49, 811-818 e814.

Bunker, B.D., Nellimoottil, T.T., Boileau, R.M., Classen, A.K., Bilder, D., 2015. The transcriptional response to tumorigenic polarity loss in Drosophila. Elife 4.

Choi, H. M. T., Schwarzkopf, M., Fornace, M. E., Acharya, A., Artavanis, G., Stegmaier, J., Cunha, A., & Pierce, N. A. (2018). Third-generation in situ hybridization chain reaction: Multiplexed, quantitative, sensitive, versatile, robust. Development (Cambridge), 145(12). https://doi.org/10.1242/DEV.165753/48466

Colombani, J., Andersen, D.S., Leopold, P., 2012. Secreted peptide Dilp8 coordinates Drosophila tissue growth with developmental timing. Science 336, 582-585.

Cosolo, A., Jaiswal, J., Csordás, G., Grass, I., Uhlirova, M., Classen, A.K., 2019. JNK-dependent cell cycle stalling in G2 promotes survival and senescence-like phenotypes in tissue stress. Elife 8.

Currie, J.D., Kawaguchi, A., Traspas, R.M., Schuez, M., Chara, O., Tanaka, E.M., 2016. Live Imaging of Axolotl Digit Regeneration Reveals Spatiotemporal Choreography of Diverse Connective Tissue Progenitor Pools. Dev Cell 39, 411-423.

Deng, M., Wang, Y., Zhang, L., Yang, Y., Huang, S., Wang, J., Ge, H., Ishibashi, T., Yan, Y., 2019. Single cell transcriptomic landscapes of pattern formation, proliferation and growth in Drosophila wing imaginal discs. Development 146.

Evans, C.J., Olson, J.M., Ngo, K.T., Kim, E., Lee, N.E., Kuoy, E., Patananan, A.N., Sitz, D., Tran, P., Do, M.T., Yackle, K., Cespedes, A., Hartenstein, V., Call, G.B., Banerjee, U., 2009. G-TRACE: rapid Gal4-based cell lineage analysis in Drosophila. Nat Methods 6, 603-605.

Everetts, N.J., Worley, M.I., Yasutomi, R., Yosef, N., Hariharan, I.K., 2021. Single-cell transcriptomics of the Drosophila wing disc reveals instructive epithelium-to-myoblast interactions. Elife 10.

Garelli, A., Gontijo, A.M., Miguela, V., Caparros, E., Dominguez, M., 2012. Imaginal discs secrete insulin-like peptide 8 to mediate plasticity of growth and maturation. Science 336, 579-582.

Gayoso, A., Lopez, R., Xing, G., Boyeau, P., Wu, K., Jayasuriya, M., Melhman, E., Langevin, M., Liu, Y., Samaran, J., Misrachi, G., Nazaret, A., Clivio, O., Xu, X., Ashuach, T., Lotfollahi, M., Svensson, V., da Veiga Beltrame, E., Talavera-López, C., Pachter, L., Theis, F.J., Streets, A., Jordan, M.I., Regier, J., Yosef, N., 2021. scvi-tools: a library for deep probabilistic analysis of single-cell omics data. bioRxiv. (doi: https://doi.org/10.1101/2021.04.28.441833)

Gerber, T., Murawala, P., Knapp, D., Masselink, W., Schuez, M., Hermann, S., Gac-Santel, M., Nowoshilow, S., Kageyama, J., Khattak, S., Currie, J.D., Camp, J.G., Tanaka, E.M., Treutlein, B., 2018. Single-cell analysis uncovers convergence of cell identities during axolotl limb regeneration. Science 362.

Halme, A., Cheng, M., Hariharan, I.K., 2010. Retinoids regulate a developmental checkpoint for tissue regeneration in Drosophila. Curr Biol 20, 458-463.

Harris, R.E., Setiawan, L., Saul, J., Hariharan, I.K., 2016. Localized epigenetic silencing of a damage-activated WNT enhancer limits regeneration in mature Drosophila imaginal discs. Elife 5.

Harris, R.E., Stinchfield, M.J., Nystrom, S.L., McKay, D.J., Hariharan, I.K., 2020. Damage-responsive, maturity-silenced enhancers regulate multiple genes that direct regeneration in Drosophila. Elife 9.

Ji, T., Zhang, L., Deng, M., Huang, S., Wang, Y., Pham, T.T., Smith, A.A., Sridhar, V., Cabernard, C., Wang, J., Yan, Y., 2019. Dynamic MAPK signaling activity underlies a transition from growth arrest to proliferation in Drosophila scribble mutant tumors. Dis Model Mech 12.

Jin, Y., Ha, N., Fores, M., Xiang, J., Glasser, C., Maldera, J., Jimenez, G., Edgar, B.A., 2015. EGFR/Ras Signaling Controls Drosophila Intestinal Stem Cell Proliferation via Capicua-Regulated Genes. PLoS Genet 11, e1005634.

Johnson, G.L., Masias, E.J., Lehoczky, J.A., 2020. Cellular Heterogeneity and Lineage Restriction during Mouse Digit Tip Regeneration at Single-Cell Resolution. Dev Cell 52, 525-540 e525.

Kar, A., Gutierrez-Hartmann, A., 2013. Molecular mechanisms of ETS transcription factor-mediated tumorigenesis. Crit Rev Biochem Mol Biol 48, 522-543.

Katsuyama, T., Comoglio, F., Seimiya, M., Cabuy, E., Paro, R., 2015. During Drosophila disc regeneration, JAK/STAT coordinates cell proliferation with Dilp8-mediated developmental delay. Proc Natl Acad Sci U S A 112, E2327-2336.

Khan, S.J., Abidi, S.N.F., Skinner, A., Tian, Y., Smith-Bolton, R.K., 2017. The Drosophila Duox maturation factor is a key component of a positive feedback loop that sustains regeneration signaling. PLoS Genet 13, e1006937.

Kulshammer, E., Mundorf, J., Kilinc, M., Frommolt, P., Wagle, P., Uhlirova, M., 2015. Interplay among Drosophila transcription factors Ets21c, Fos and Ftz-F1 drives JNK-mediated tumor malignancy. Dis Model Mech 8, 1279-1293.

Kurucz, E., Zettervall, C.J., Sinka, R., Vilmos, P., Pivarcsi, A., Ekengren, S., Hegedus, Z., Ando, I., Hultmark, D., 2003. Hemese, a hemocyte-specific transmembrane protein, affects the cellular immune response in Drosophila. Proc Natl Acad Sci U S A 100, 2622-2627.

La Fortezza, M., Schenk, M., Cosolo, A., Kolybaba, A., Grass, I., Classen, A.K., 2016. JAK/STAT signalling mediates cell survival in response to tissue stress. Development 143, 2907-2919.

Mattila, J., Omelyanchuk, L., Kyttala, S., Turunen, H., Nokkala, S., 2005. Role of Jun N-terminal Kinase (JNK) signaling in the wound healing and regeneration of a Drosophila melanogaster wing imaginal disc. Int J Dev Biol 49, 391-399.

McClure, K.D., Sustar, A., Schubiger, G., 2008. Three genes control the timing, the site and the size of blastema formation in Drosophila. Dev Biol 319, 68-77.

McGinnis, C.S., Murrow, L.M., Gartner, Z.J., 2019. DoubletFinder: Doublet Detection in Single-Cell RNA Sequencing Data Using Artificial Nearest Neighbors. Cell Syst 8, 329-337 e324.

Mundorf, J., Donohoe, C.D., McClure, C.D., Southall, T.D., Uhlirova, M., 2019. Ets21c Governs Tissue Renewal, Stress Tolerance, and Aging in the Drosophila Intestine. Cell Rep 27, 3019-3033 e3015.

Narbonne-Reveau, K., Maurange, C., 2019. Developmental regulation of regenerative potential in Drosophila by ecdysone through a bistable loop of ZBTB transcription factors. PLoS Biol 17, e3000149.

Pfeiffer, B.D., Ngo, T.T., Hibbard, K.L., Murphy, C., Jenett, A., Truman, J.W., Rubin, G.M., 2010. Refinement of tools for targeted gene expression in Drosophila. Genetics 186, 735-755.

Rebeiz, M., Reeves, N.L., Posakony, J.W., 2002. SCORE: a computational approach to the identification of cis-regulatory modules and target genes in whole-genome sequence data. Site clustering over random expectation. Proc Natl Acad Sci U S A 99, 9888-9893.

Rosin, D., Schejter, E., Volk, T., Shilo, B.Z., 2004. Apical accumulation of the Drosophila PDGF/VEGF receptor ligands provides a mechanism for triggering localized actin polymerization. Development 131, 1939-1948.

Santabarbara-Ruiz, P., Lopez-Santillan, M., Martinez-Rodriguez, I., Binagui-Casas, A., Perez, L., Milan, M., Corominas, M., Serras, F., 2015. ROS-Induced JNK and p38 Signaling Is Required for Unpaired Cytokine Activation during Drosophila Regeneration. PLoS Genet 11, e1005595.

Smith-Bolton, R.K., Worley, M.I., Kanda, H., Hariharan, I.K., 2009. Regenerative growth in Drosophila imaginal discs is regulated by Wingless and Myc. Dev Cell 16, 797-809.

Stuart, T., Butler, A., Hoffman, P., Hafemeister, C., Papalexi, E., Mauck, W.M., 3rd, Hao, Y., Stoeckius, M., Smibert, P., Satija, R., 2019. Comprehensive Integration of Single-Cell Data. Cell 177, 1888-1902 e1821.

Toggweiler, J., Willecke, M., Basler, K., 2016. The transcription factor Ets21C drives tumor growth by cooperating with AP-1. Sci Rep 6, 34725.

Tran, K.D., Miller, M.R., Doe, C.Q., 2010. Recombineering Hunchback identifies two conserved domains required to maintain neuroblast competence and specify early-born neuronal identity. Development 137, 1421-1430.

Worley, M.I., Alexander, L.A., Hariharan, I.K., 2018. CtBP impedes JNK- and Upd/STAT-driven cell fate misspecifications in regenerating Drosophila imaginal discs. Elife 7.

Worley, M.I., Setiawan, L., Hariharan, I.K., 2012. Regeneration and transdetermination in Drosophila imaginal discs. Annu Rev Genet 46, 289-310.

Wu, Y., Brock, A.R., Wang, Y., Fujitani, K., Ueda, R., Galko, M.J., 2009. A blood-borne PDGF/VEGF-like ligand initiates wound-induced epidermal cell migration in Drosophila larvae. Curr Biol 19, 1473-1477.

Yoo, S.K., Pascoe, H.G., Pereira, T., Kondo, S., Jacinto, A., Zhang, X., Hariharan, I.K., 2016. Plexins function in epithelial repair in both Drosophila and zebrafish. Nat Commun 7, 12282.

Zappia, M.P., de Castro, L., Ariss, M.M., Jefferson, H., Islam, A.B., Frolov, M.V., 2020. A cell atlas of adult muscle precursors uncovers early events in fibre-type divergence in Drosophila. EMBO Rep 21, e49555.