UC Irvine
UC Irvine Electronic Theses and Dissertations

Title
Quantifying variation In Drosophila immune response.

Permalink
https://escholarship.org/uc/item/6q73z8t0

Author
Ramirez-Corona, Bryan Alfonso

Publication Date
2021

Peer reviewed|Thesis/dissertation

University of California, Irvine

Quantifying variation In *Drosophila* immune response.

DISSERTATION

Submitted in partial satisfaction of the requirements for the degree of

DOCTOR OF PHILOSOPHY

In Developmental and Cell Biology

By

Bryan Alfonso Ramirez-Corona

Dissertation Committee:
Assistant Professor Zeba Wunderlich, Co-Chair
Professor Ali Mortazavi, Co-Chair
Associate Professor J.J. Emerson
Professor Kavita Arora
Professor Ken Cho

DEDICATION

Dedicado a Herlinda y Alfonso Ramirez,
quienes venieros a este paiz buscando una vida mejor para sus hijos,
y lo encontraron.

And to scientists that will come after,
may you always hold on to the belief,
that anything is possible.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

They say that it takes a village to raise a child. In a similar manner it takes a community to train a scientist, and I would be remiss not to thank those who helped me along the way.

Firstly, I would like to thank my mentor for the past 6 years Dr. Zeba Wunderlich. Dr. Wunderlich invited me into her lab and provided me with a model for conducting research that is both rigorous and compassionate. She provided me with the mentorship that has kept me excited to continue forward in science. I am proud to be forever of the Wunderlich scientific lineage.

Secondly, I would like to thank my thesis committee consisting of Dr. Emerson, Dr. Mortazavi, Dr. Arora and Dr. Cho. They have made committee meetings feel like an opportunity to grow and collaborate. Watching them tease apart the logic of my work and decisions has been an amazing opportunity to observe and participate in a microcosm of scientific community and discussion. I am absolutely indebted to them for taking the time to help train me.

A special thanks must also be given to my lab mates Dr. Lily Li, Dr. Rachel Waymack, Kevin Cabrera, Dr. Lianne Cohen, Flo Ramirez, Arash Abiri, Subhapradha Rangarajan, Mario Gad and Ariana Lee, among others. They have not only helped me grow as a scientist, but also as a person.

I would also like to thank Dr. Gabriela Balderrama Gutierrez, who has been one of my closest friends and an invaluable colleague.

Additionally, I would like to thank my colleagues and friends at the University of California Irvine. Particularly Dr. Lorrayne Serra, Dr. Camden Jansen, Dr. Nicholas Pervolakis whose friendship and advice helped carry me through the PhD process.

Lastly, I would like to thank my family who have been supportive of my decisions, albeit confused by the manner in which I go about making them. Their love in support has kept me going.

VITA

Bryan Alfonso Ramirez-Corona

EDUCATION

University of California, Irvine: PhD – Biology   2021

University of California, Irvine: M.S. – Biological Sciences   2018

University of California, Davis: B.S. - Genetics, Minor – Art Studio   2013

PUBLICATIONS

Ramirez-Corona, B. A., Love A. C., Yao, Z., Prescher, J. A., Wunderlich, Z. (2021) Longitudinal monitoring of individual infection progression in *Drosophila melanogaster*. (*Manuscript in Review)*

Ramirez-Corona, B. A., Fruth, S. M., Ofoegbu, O., Wunderlich, Z. (2021). The mode of expression divergence in *Drosophila* fat body is infection-specific. *Genome Res.* gr.269597.120 doi:10.1101/gr.269597.120

Arancibia, C., Riaz, S., Agüero, C., **Ramirez-Corona, B**., Alonso, R., Buscema, F., Martínez, L. and Walker, M. (2018), Grape phylloxera (*Daktulosphaira vitifoliae* Fitch) in Argentina: ecological associations to diversity, population structure and reproductive mode. Australian Journal of Grape and Wine Research, 24: 284-291. https://doi.org/10.1111/ajgw.12337

FELLOWSHIPS AND AWARDS

Broad Institute, Center of Excellence in Genomic Science Visiting Scientist Program, 2017.

Center for Complex Biological Systems Opportunity Award, 2017.

LSAMP, Bridge to the Doctorate Fellowship, 2015.

UC Irvine Diversity Recruitment Fellowship, 2015.

MBRS-IMSD Summer Research Fellowship, 2015.

ABSTRACT OF THE DISSERTATION
Quantifying variation in Drosophila immune response.

By

Bryan Alfonso Ramirez-Corona

Doctor of Philosophy in Biological Sciences

University of California, Irvine 2021

Assistant Professor Zeba Wunderlich, Co-Chair

Professor Ali Mortazavi, Co-Chair

A robust and specific immune response is critical for the maintenance of organismal health. The immune response's primary function is infection clearance; however, the outcome of an infection can vary greatly between individuals. This variation can be traced back to the combined effects of genetic background, environmental factors, and stochastic events. Understanding how these individual components influence the observed variation in immune response in *Drosophila* may help further our understanding of infection outcome in across metazoans, mammals, and plants.

Here, I investigate sequence variation that leads to variation in immune response and propose to novel methods that will help investigate additional sources of variation. In Chapter 2, I quantify the relative contributions of changes in *cis* and *trans* to expression divergence in the *Drosophila* immune response. I show that the proportions of these changes are condition specific. In Chapter 3, I describe a novel method for the longitudinal monitoring of infection progression using autobioluminscent bacteria. I demonstrate the methods utility for linking individual infection histories to observed outcomes. Lastly in Chapter 4, I adapt a method for the

cell specific chemical labeling of nascent RNA's using an orthogonal chemical-genetic labeling

system. I show preliminary results that suggest the method can specifically label RNA in a tissue

specific manner and provide next steps for the application of this method fully in *Drosophila.*

CHAPTER 1
Introduction

1.1 The importance of innate immunity

A robust and specific immune response is necessary for an organism's continued survival. The immune response is responsible for the clearance of pathogenic viruses, bacteria and other invading agents. In mammals this response consists of the evolutionarily ancient innate immune response and the relatively newer adaptive response (Janeway 2002). Invertebrates, including *Drosophila melanogaster*, do not possess an adaptive response and rely entirely on the innate immune response to fight off infection. The innate response system, however, is found across vertebrates, invertebrates, and plants; understanding innate immunity in one system has the potential to advance understanding of immunity across multiple kingdoms of life.

While the immune response's primary function is the clearance of infections, infection outcome depends on genetic background, environmental factors, and stochastic events. Genetics, diet, injury, and many other factors may all contribute to the immune response, and it is the specific combination of these factors that ultimately dictates infection outcome. This outcome can range from infection clearance, chronic infection, and in the worst-case scenario, death. One salient example can be seen in the recent global outbreak of Covid-19, where large amounts of variation in the severity of infection have been observed. This is true even after stratifying data by the largest contributing factors such as age, socioeconomic background, and comorbidities (Argenzian, 2020, Zhang 2021). As such there is a need to better understand the factors that contribute to infection variation.

In this thesis, I use the genetically amenable *Drosophila melanogaster* to investigate the genetic basis of immune response and describe methods for studying genetic and non-genetic sources of immune response and infection variation. In Chapter 2 I study the relative contributions of *cis*-acting and *trans*-acting sequence changes on expression of immune-responsive genes in the

Drosophila fat body under multiple conditions. In Chapter 3 I improve upon current methodologies for monitoring infection dynamics and describe a novel method to non-invasively monitor bacterial infection in flies. Finally, in Chapter 4 I describe a precise chemical-genetic method for labeling and isolating RNA in a tissue-specific manner. Together this work shows the most fine-grained image to date of sequence changes to immune response expression variation and provides methods to further investigate genetic and non-genetic variation in immune response.

Below I begin with an overview of the innate immune response of *Drosophila melanogaster*, focusing on the most relevant components for understanding the work presented in Chapter 2. In subsequent sections I give context for the following chapters by describing the state of the field specific to each chapter. I also describe current gaps in the knowledge and address how the work will contribute to our understanding of the field.

1.2   Specifics of the innate immune response

In *Drosophila* the innate immune response consists of a cellular response and a humoral or systemic response. In the cellular response, foreign agents are phagocytized or encapsulated by circulating plasmatocytes and lamellocytes. The humoral response to infection begins with the recognition of microbial ligands which set off signaling cascades that lead to the production of antimicrobial peptides (AMPs) in the fat body. This is the primary immune-responsive organ. These peptides are then released into the hemolymph where they can be transported to the sites of infection.

1.2.1 The fat body

In adult *Drosophila* the humoral response takes place primarily in the fat body. This tissue is found along the dorsal side of the adult fly and is in some ways comparable to the mammalian

liver. This tissue serves in the immune response as well as carrying out important metabolic roles. In *Drosophila* larvae, the fat body has been observed to be arranged into functionally distinct regions along the anterior-posterior axis, though in adults this has yet to be verified (Haunerland 1995). Unlike the larval fat body, the adult fat body has been largely understudied due to physical difficulties in working with this tissue. Whereas the larval fat body is relatively simple to dissect, the adult fat body is much more fragile and impossible to dissect intact. As such, the adult fat body remains a largely underexplored tissue with a great deal of promise for future discovery.

At the gene level, the production of AMPs is regulated by two conserved nuclear factor-κB (NF-κB) signaling cascades, the Toll pathway and the IMD pathway (Buchon 2014). A variety of pattern recognition receptors can recognize different classes of bacteria, activate the necessary cascades and initiate production of different AMPs (Lemaitre 2007, Lu 2019). The IMD response is generally turned on by gram-negative infections and the Toll response is generally turned on by gram-positive infections (Buchon 2014). Though historically these pathways have been thought of as distinct, research demonstrates that there is crosstalk between them (Troha 2018).

1.2.2 The Toll response

As the name suggests, the Toll pathway is regulated by the Toll-like receptor (TLR) genes of which there are nine in *Drosophila* that as act receptors to various ligands (Buchon 2019). Specifically, the Toll receptor and Spaetzle ligand are critical for activation of the Toll response pathway. This ligand is activated when extracellular recognition factors initiate protease cascades that lead to the cleavage of the Spatzle prodomain (Valanne 2011). Activation of the Toll pathway eventually leads to the nuclear localization of Dorsal and Dif, NF-κB homologs, via the degradation of Cactus, resulting in activation of AMP

genes (Valanne 2011). Although immune-related Toll signaling is conserved in mammals, in *Drosophila* it bears the additional role of being involved in development.

1.2.3 The IMD response

In response to gram-negative infections, the IMD pathway activates production of AMPs via nuclear localization of a third NF-κB factor Relish (Myllymäki 2014). This pathway is often compared to the mammalian Tumor necrosis factor receptor (TNFR) signaling pathway. Much like the mammalian TNFR pathway, this pathway bifurcates into the JNK pathway which is involved in wound healing (Myllymäki 2014). Direct microbial recognition by genes such as peptidoglycan recognition proteins (PGRPs) sets off protease cascades that result in the cleavage of the C-terminus of Relish. This allows for the nuclear localization of this factor and the activation of AMP production (Lu 2019).

1.3 *Cis* and *trans* variation in gene expression

In Chapter 2 of my thesis, I study the genetic underpinnings of expression divergence in *Drosophila melanogaster*. In eukaryotes, spatiotemporal gene expression patterns are generated by the interactions of *trans*-acting factors, such as transcription factors, and *cis*-regulatory elements (CREs), such as enhancers. Sequence variation in either protein-coding regions of *trans*-acting factors or CREs can change gene expression patterns. Changes in *trans* generally fall in coding regions and typically have widespread effects. This is because they affect all downstream targets of a gene carrying the change. On the other hand, changes in *cis* typically have more localized effects. Given that a CRE generally controls a single gene and only in a particular context, a change in *cis* is more likely to only affect the expression of a single gene at a time (Wittkopp 2012).

1.3.1 The importance of understanding *cis* and *trans* mutations

Understanding the relative contributions of *cis* and *trans* changes may help further our understanding of how organisms evolve. Changes in *cis* and *trans* underlie expression differences within and between species (Wittkopp 2012). Expression divergence in turn is important for processes such as adaptation and speciation. As such there is a need to understand the contribution of *cis* and *trans* mutations to better understand how organisms evolve (Emerson 2010, Wittkopp 2012). On one hand, a system can evolve through a small number of *trans* mutations that can bring about drastic changes but that are also more likely to have deleterious effects. On the other hand, a system can evolve through many fine-tuning mutations in *cis*, that allow for incremental changes that are less likely to have adverse effects (Wittkopp 2012). More than likely, it is some combination of both these strategies that result in observed expression divergence. And it is the relative proportions of these that may inform us of the evolutionary pressures on organisms and how they evolve and adapt in response to them (Wittkopp 2012).

In addition, identifying changes in *cis* and *trans* may serve as a stepping-off point for dissection of the regulatory logic of CREs. Common obstacles to studying the design principles of CREs is the difficulty in identifying functional CREs *in vivo*, pairing these to the genes they control, and then finding potentially informative sequence changes for understanding how CREs function. While the first two obstacles can be addressed using techniques such as DNase-seq and Hi-C, the last obstacle remains. However, one of the byproducts of quantifying *cis* and *trans* effects on a genome-wide scale is a list of genes with functional sequence changes in CREs. We can then combine this with other omics sets to identify potentially causative sequence changes for the observed differences in expression. This, in turn, may be informative to the understanding of how CREs function in controlling expression levels.

1.3.2 *Cis* and *trans* variation between species

Already, research has been conducted that investigates the relative proportions of *cis* and *trans* changes between species using a variety of organisms including yeast (Tirosh 2009, Metzger 2017), flies (Wittkopp 2004, Landry 2005, Wittkopp 2008, McManus 2010, Coolon 2014), vertebrates (Davidson 2016, Mattoli 2020) and plants (Shi 2012). When looking at expression divergence between species *cis* changes tend to dominate the system (Tirosh 2009, Metzger 2017, Wittkopp 2004, Wittkopp 2008, Mcmanus 2010, Coolon 2014, Mattoli 2020, Shi 2012). Greater evolutionary distances seem to result in a greater proportion of changes in *cis*, though notable exceptions to this trend exist (McManus 2010, Coolon 2014). This suggests that while changes in *cis* may be favored in expression divergence between species there are instances in which changes in *trans* may dominate.

1.3.3 *Cis* and *trans* variation within species

Given that changes in *cis* dominate between-species expression differences, we would expect this to remain true for the shorter evolutionary distances within species. This however is not always the case. Work examining this in yeast (Emerson 2010, Schaefke 2013, Metzger 2017), flies (Wittkopp 2008, Gonclaves 2012, Coolon 2014, Osada 2017, Benowitz 2020, Frochaux 2020) and plants (Bell 2013, Diaz-Valenzuela 2020) found a greater proportion of changes in *trans* to be responsible for within-species expression differences. Once more, exceptions to this trend exist, suggesting that while a type of change may be favored the contributions of each type of change may be specific to the evolutionary circumstances of the organisms involved (Gonclaves 2012, Osada 2017, Frochaux 2020). One potential explanation for the differences in relative contributions for changes in *cis* and *trans* within and between species is that changes in *trans* may

be relatively more common within a species but *cis* changes are more likely to become fixed in a population since they are less likely to have deleterious effects (Denver 2005).

1.3.3 *Cis* and *trans* variation in *Drosophila*

More specific to *Drosophila,* we see that changes in *cis* tend to dominate between species (Wittkopp 2004, Wittkopp 2008, Coolon 2014). The exception to this pattern is the expression divergence between *D. melanogaster* and *D. sechellia* (McManus 2010, Coolon 2014). It is worth noting that in the case of *D. sechellia* authors suggest that the break from the expected trend may be attributed to the specific evolutionary history of the species as *D. sechellia* an island species with very little intraspecific genetic variation (McManus 2010). As such the abundance of *trans*-acting changes may be a result of drift rather than selection.

With regards to intra-species expression variation, we observe more evidence of *trans*-acting changes being the dominant driver of expression divergence (Wittkopp 2008, Coolon 2014, Benowitz 2020). However, there are examples where this is not the case (Osada 2017, Frochaux 2020,). Given the large number of variations in experimental design between studies (genetic lines, conditions, treatments), it is difficult to determine the source of the observed differences in relative proportions of *cis* and *trans* effects between these works.

1.3.4 Compensatory effects

One trend that has emerged from previous studies is the prevalence of *cis* and *trans* effects that function in opposite directions to maintain gene expression levels, known as compensatory effects. This is a puzzling observation since it would suggest some sort of stabilizing pressure is maintaining gene expression levels. Nonetheless gene expression divergence emerges both within and between species. Thus, it is not clear how expression divergence is able to emerge when as much as 50% of *cis* and *trans* effects in a system are observed to be working antagonistically to

maintain gene expression. One explanation is that many *cis* effects may emerge to balance out the deleterious effects of a small number of *trans* effects. Additionally, the number of compensatory effects may be artificially inflated in some cases due to experimental design. As pointed out by Emerson *et. al.* 2010, estimates of *cis* and *trans* effects that rely on shared F1 hybrid samples will result in spurious inference of compensatory effects. This is because compensatory evolution is defined as a negative relationship between *cis* and *trans* variation and so in estimates that use a common F1 hybrid, these will be negatively correlated via error shared from the hybrid data (Zhang 2019). Thus, more controlled estimates of *cis* and *trans* effects spanning multiple conditions are necessary to better understand the true contributions of compensatory effects.

1.3.5 Limitations of current studies

Though previous works have been fundamental in our current understanding of *cis* and *trans* effects, they are often limited either by the technology of the time or the experimental design. Particularly in *Drosophila,* studies performed generally make use of RNA extracted from the whole body or only consider signal from a single control condition. This means that the signal is generally averaged across multiple tissues and in a single context. As such, these studies are unable to capture the contributions of *cis-* and *trans*-acting changes of specific biological processes. Capturing these contributions is important given that different processes and tissues may be under different selective pressures.

There is currently limited understanding of the broad rules regarding the mode in which gene expression differences arise. As such there is a need for a fine-grained examination of the contributions of *cis* and *trans* effects to expression divergence in specific biological processes across multiple conditions. In Chapter 2 I fill this particular gap in knowledge by measuring the contributions of *cis* and *trans* mutations in expression divergence of the humoral response. I

achieve this by using RNA from the fat body in both control and infected conditions. To date this study represents the most high-resolution pathway-specific examination of *cis* and *trans* effects in multicellular organisms.

1.4   Inter-individual variation in immune response

One common threat to the health of multicellular organisms is bacterial infection. For any given infection, the outcome can vary greatly between individuals. Bacteria are diverse, versatile, and extremely adaptable. In a host, the immune response is deployed to mitigate acute infection, but infection outcome is dependent on a combination of genetic background, environmental factors and stochastic events. The specific contributions of these factors can lead to variation in infection outcomes between individuals. There is a need to better understand the sources of outcome variation and the specific contributions of these sources therein. This need has been made particularly clear over the course of the ongoing Covid-19 global outbreak, where infection outcomes have been observed to vary drastically even among factors with the strongest effect sizes on outcome (Argenzian 2020, Zhang 2021). This knowledge not only has the potential to inform our understanding of variation in infection outcome not only in human health but also across insects and plant populations as well.

As mentioned before, infection outcome is dependent on the combination of genetic background, environmental factors, and stochastic events. In *Drosophila*, much of the research to date has addressed the contributions of genetic and environmental factors. This is done by either controlling for environmental conditions to test for differences in genetic background or by controlling for genetic background to test for different environmental conditions. While informative, these studies have given rise to the observation, that even among genetically identical

10

individuals there is a great deal of variation in infection outcome (Lazzaro 2004, Lazaro2006, Sackton 2010, Howick 2017, Duneau 2017). This variation is observed as differences in the bacterial load upon infection or the ability to survive infection within genetic lines. Historically, this variation has been attributed to stochastic differences such as variation in inoculation, time to immune initiation or other stochastic events (Duneau 2017, Ellner 2021). However, to our knowledge, detection of this interindividual variation has been limited in no small part due to limitations in existing methods.

1.4.1 Current methods and limitations

The most common method for measuring infection progression in *Drosophila,* and many other systems, is dilution plating. This method involves inoculating individuals with bacteria, then sacrificing individual flies at the time points of interest by grinding them up. A serial dilution of the homogenate is then plated on the appropriate media to quantify the colony-forming units (CFU) for each fly. In this manner, the mean CFU per fly can be calculated at a single time point and the average infection trajectory is determined by tracking how this mean changes over time. This method is labor-intensive, as it requires many individual measurements to account for biological variation in infection as well as variation introduced via the dilution plating process itself.

This method for measuring infection progression is also inadequate for determining sources of interindividual variation. This is because it requires animal sacrifice and as such can only provide a "snapshot" of an animal's physiological state. Additionally, this sacrifice means that all information of the infection progression up to and after the point of measurement is lost. Without access to individual infection trajectories, this method must average across many individuals to determine the average infection trajectory of a group. This in turn means that individual infection

outcomes cannot be directly linked to an individual infection trajectory. Thus, sources of variance between individuals can only be inferred but not directly observed.

1.4.2 Previous findings

Despite the caveats, the method of dilution plating has been critical for gaining insight into genetic variation in immune response infection progression between genetic lines (Lazzaro 2004, Lazaro2006, Sackton 2010, Howick 2017), as well as potential sources of non-genetic variation (Duneau 2017, Ellner 2021). Although it has never been directly measured, some work has been done in *Drosophila* on variation in infection outcome between genetically identical individuals, despite seemingly similar infection conditions. These works suggest that a potential explanation for this may be stochastic variation in the amount of time it takes the immune response to switch on in response to infection (Duneau 2017). Another explanation may be that mutual negative feedback loops between the pathogen and the host immune response amplify stochastic variation between individuals, resulting in differential infection outcomes (Ellner 2021). However, these works rely on models to explain observed variation in infection outcome, and while potentially informative they still require validation. As such, there is a need for a method that can longitudinally track infection in individual flies to directly link variation in infection outcome to infection trajectory and disentangle potential sources of variation in immune response.

To address this current need in the field, we developed a method to longitudinally monitor infection progression in individual *Drosophila melanogaster*. This method involves injecting autobioluminescent bacteria into flies and quantifying the amount of light emitted from bacteria through the fly cuticle. The light detected serves as a direct measure of the bacterial load over time without the need for sacrifice. Additionally, this method allows for the measurement of infection progression at the level of individuals while being high-throughput enough to allow between-

genotype measurements to also be feasible. In this way this new method improves on existing methods of monitoring infection by making them more high-throughput while allowing for the quantification of individual infection dynamics. In chapter 3 of my thesis, I describe this method in detail and demonstrate its potential for identifying sources of variation both between genetic lines and genetically identical individuals.

1.5   Tissue-specific labeling of nascent RNA

As mentioned above, differences in gene expression can drive larger scale phenotypic differences. Understanding what genes are differentially expressed between genotypes or conditions can inform us about the function and form of biological processes that are involved (Asyali 2006). However, gene expression is not homogeneous throughout an organism and expression programs may vary drastically between different tissues and cell types (Whitehead 2005, Asyali 2006). As such, gene expression studies that are interested in specific biological processes require expression measurements on the level of tissue or cell type to avoid confounding signals.

1.5.1 Current methods for tissue and cell-specific gene expression profiling

Several methods exist for acquiring expression data at the level of tissue or cell type, many of which require the physical isolation of tissues or cells of interest. To acquire tissue-specific measurements one of the most straightforward methods is to simply dissect or isolate the tissue for RNA extraction. This method works well for larger or highly structured tissues but becomes less feasible with smaller, more amorphous tissues or with rare cell types. For isolating cell-specific RNA, methods such as single-cell sequencing or cell sorting before sequencing do exist, but these methods can become expensive and they rely on the ability to dissociate a particular tissue into

individual cells (Otsuki 2014, Wang 2015). While these methods work well for sturdier cell types, they are not feasible for isolating cell types that cannot survive the mechanical stress of dissociation without rupturing. One common workaround for tissues and cells that are recalcitrant to standard tissue dissociation protocols is the sequencing of single nuclei. Research has shown that while single nuclei and single cells generally have highly correlated relative gene expression, notable differences do exist (Nguyen 2018). As such there is a need for methods of expression profiling that do not rely on mechanical means of isolation.

Currently, several systems exist for expression profiling that do not require tissue or cell isolation. In general, these methods either rely on the co-immunoprecipitation of RNA binding proteins or the labeling of nascent RNAs (Otsuki 2014). With the first method, one major concern is the level of background contamination introduced via biochemical procedures necessary for co-immunopurification which may result in the non-specific binding of proteins to RNA (Otsuki 2014). With the second method, however, this concern is mitigated since RNA is directly labeled and purified (Otsuki 2014). RNA labeling methods generally rely on inert nucleoside or nucleobase analogs carrying a chemical handle can be readily biotinylated (Miller 2009, Nguyen 2016, Beasley 2018). Tissue or cell-type specificity is conferred by the addition of an exogenous enzyme that in only expressed in predetermined tissues or cell types. This allows for the incorporation of the supplied analog into specific RNAs, and the chemical handle can then be used to specifically label and purify RNA from bulk extracted RNA.

1.5.2 Targeted RNA labeling in *Drosophila*

In *Drosophila* the established method for targeted labeling of nascent RNA was found to be lacking in specificity. This is because the system relies on the bacterial uracil phosphoribosyltransferase (UPRT), in combination with a uracil analog, to label RNA (Miller

2009, Nguyen 2016). Previously UPRT homologs in higher eukaryotes were found to be primarily inactive *in vitro* (Li 2007). This led to the use of microbial UPRT RNA labeling. However, *in vivo* work later found the *Drosophila* UPRT homolog to show activity that allowed for the nonspecific integration of uracil analogs into nascent RNA (Gosh 2015).

Consequently, there is a need for a method to specifically label RNA in *Drosophila* that does not rely on microbial UPRT. Fortunately, a separate system already exists that instead uses an inert protected nucleotide analog in combination with the bacterial-specific penicillin G amidase (PGA) (Nguyen 2016, Beasley 2019). Similar to the previous system, this inert analog can be activated in particular tissues or cell type via the specific expression of a PGA gene which leads to the removal of a PGA-specific protecting group and integration of the nucleotide analog into RNA (Nguyen 2016, Beasley 2019). Despite the promise of this system for specific labeling of RNA, it has yet to be adapted for use in *Drosophila*.

In Chapter 4 of my thesis, I describe how we adapted this PGA-dependent method for the labeling of nascent RNA in *Drosophila*. I demonstrate that this system can uptake and incorporate modified nucleotides administered orally. I also describe preliminary results that suggest that caged nucleotide analog coupled with tissue-specific expression of the PGA gene using a GAL4-UAS system appears to specifically label nascent RNAs. Though further work is required to verify and optimize the specificity of the methods' labeling, this work fills the current need for precise chemical tools for dissecting gene expression differences and may serve as a starting point for future work.

1.6   References

Argenzian, M. G., Bruc, S. L., Slate, C. L., Tia, J. R., Baldwi, M. R., Barr, R. G., … Chen, R. (2020). Characterization and clinical course of 1000 patients with coronavirus disease 2019 in New York: Retrospective case series. *The BMJ*, *369*(March). https://doi.org/10.1136/bmj.m1996

Asyali, M., Colak, D., Demirkaya, O., & Inan, M. (2008). Gene Expression Profile Classification: A Review. *Current Bioinformatics*, *1*(1), 55–73. https://doi.org/10.2174/157489306775330615

Bell, G. D. M., Kane, N. C., Rieseberg, L. H., & Adams, K. L. (2013). RNA-seq analysis of allele-specific expression, hybrid effects, and regulatory divergence in hybrids compared with their parents from natural populations. *Genome Biology and Evolution*, *5*(7), 1309–1323. https://doi.org/10.1093/gbe/evt072

Benowitz, K. M., Coleman, J. M., Allan, C. W., & Matzkin, L. M. (2020). Contributions of cis- And trans-Regulatory Evolution to Transcriptomic Divergence across Populations in the Drosophila mojavensis Larval Brain. *Genome Biology and Evolution*, *12*(8), 1407–1418. https://doi.org/10.1093/GBE/EVAA145

Buchon, N., Silverman, N., & Cherry, S. (2014). Immunity in Drosophila melanogaster-from microbial recognition to whole-organism physiology. *Nature Reviews Immunology*, *14*(12), 796–810. https://doi.org/10.1038/nri3763

Coolon, J. D., McManus, C. J., Stevenson, K. R., Graveley, B. R., & Wittkopp, P. J. (2014). Tempo and mode of regulatory evolution in Drosophila. *Genome Research*, *24*(5), 797–808. https://doi.org/10.1101/gr.163014.113

Daniel E Shumer, N. J. N. N. P. S. (2017). 乳鼠心肌提取 HHS Public Access. *Physiology & Behavior*, *176*(12), 139–148. https://doi.org/10.1016/j.tetlet.2018.09.040.Protected

Davidson, J. H., & Balakrishnan, C. N. (2016). Gene regulatory evolution during speciation in a songbird. *G3: Genes, Genomes, Genetics*, *6*(5), 1357–1364. https://doi.org/10.1534/g3.116.027946

Denver, D. R., Morris, K., Streelman, J. T., Kim, S. K., Lynch, M., & Thomas, W. K. (2005). The transcriptional consequences of mutation and natural selection in Caenorhabditis elegans. *Nature Genetics*, *37*(5), 544–548. https://doi.org/10.1038/ng1554

Diaz-Valenzuela, E., Sawers, R. H., & Cibrian-Jaramillo, A. (2020). Cis- And trans-regulatory variations in the domestication of the chili pepper fruit. *Molecular Biology and Evolution*, *37*(6), 1593–1603. https://doi.org/10.1093/molbev/msaa027

Ellner, S. P., Buchon, N., Dörr, T., & Lazzaro, B. P. (2021). Host-pathogen immune feedbacks can explain widely divergent outcomes from similar infections. *Proceedings of the Royal Society B: Biological Sciences*, *288*(1951). https://doi.org/10.1098/rspb.2021.0786

Emerson, J. J., Hsieh, L. C., Sung, H. M., Wang, T. Y., Huang, C. J., Lu, H. H. S., … Li, W. H. (2010). Natural selection on cis and trans regulation in yeasts. *Genome Research*, *20*(6), 826–836. https://doi.org/10.1101/gr.101576.109

Emerson, J. J., & Li, W. H. (2010). The genetic basis of evolutionary change in gene expression levels. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *365*(1552), 2581–2590. https://doi.org/10.1098/rstb.2010.0005

Frochaux, M. V., Bou Sleiman, M., Gardeux, V., Dainese, R., Hollis, B., Litovchenko, M., … Deplancke, B. (2020). cis-regulatory variation modulates susceptibility to enteric infection in the Drosophila genetic reference panel. *Genome Biology*, *21*(1), 6. https://doi.org/10.1186/s13059-019-1912-z

Goncalves, A., Leigh-Brown, S., Thybert, D., Stefflova, K., Turro, E., Flicek, P., … Marioni, J. C. (2012). Extensive compensatory cis-trans regulation in the evolution of mouse gene expression. *Genome Research*, *22*(12), 2376–2384. https://doi.org/10.1101/gr.142281.112

Haunerland, N. H., & Shirk, P. D. (1995). Regional and Functional Differentiation in the Insect Fat Body. *Annual Review of Entomology*, *40*(1), 121–145. https://doi.org/10.1146/annurev.ento.40.1.121

Howick, V. M., & Lazzaro, B. P. (2017). The genetic architecture of defence as resistance to and tolerance of bacterial infection in Drosophila melanogaster. *Molecular Ecology*, *26*(6), 1533–1546. https://doi.org/10.1111/mec.14017

Janeway, C. A., & Medzhitov, R. (2002). Innate immune recognition. *Annual Review of Immunology*, *20*(2), 197–216. https://doi.org/10.1146/annurev.immunol.20.083001.084359

Krukowski, R. A., West, D. S., Dicarlo, M., Cleves, M. A., Saylors, M. E., & Andres, A. (2017). Association of Gestational Weight Gain Expectations and Advice on Actual Weight Gain. *Obstetrics and Gynecology*, *129*(1), 76–82. https://doi.org/10.1097/AOG.0000000000001780

Landry, C. R., Wittkopp, P. J., Taubes, C. H., Ranz, J. M., Clark, A. G., & Hartl, D. L. (2005). Compensatory cis-trans evolution and the dysregulation of gene expression in interspecific hybrids of drosophila. *Genetics*, *171*(4), 1813–1822. https://doi.org/10.1534/genetics.105.047449

Lazzaro, B. P., Sackton, T. B., & Clark, A. G. (2006). Genetic Variation in Drosophila melanogaster Resistance to Infection : *Genetics*, *1554*(November), 1539–1554. https://doi.org/10.1534/genetics.105.054593

Lazzaro, B. P., Sceurman, B. K., & Clark, A. G. (2004). Genetic Basis of Natural Variation in D. melanogaster Antibacterial Immunity. *Science*, *303*(5665), 1873–1876. https://doi.org/10.1126/science.1092447

Lemaitre, B., & Hoffmann, J. (2007). The host defense of Drosophila melanogaster. *Annual Review of Immunology*, *25*, 697–743. https://doi.org/10.1146/annurev.immunol.25.022106.141615

Li, J., Huang, S., Chen, J., Yang, Z., Fei, X., Zheng, M., … Mao, Y. (2007). Identification and characterization of human uracil phosphoribosyltransferase (UPRTase). *Journal of Human Genetics*, *52*(5), 415–422. https://doi.org/10.1007/s10038-007-0129-2

Lu, Y., Su, F., Li, Q., Zhang, J., Li, Y., Tang, T., … Yu, X. Q. (2020). Pattern recognition receptors in Drosophila immune responses. *Developmental and Comparative Immunology*, *102*(July 2019), 103468. https://doi.org/10.1016/j.dci.2019.103468

Mattioli, K., Oliveros, W., Gerhardinger, C., Andergassen, D., Maass, P. G., Rinn, J. L., & Melé, M. (2020). Cis and trans effects differentially contribute to the evolution of promoters and enhancers. *Genome Biology*, *21*(1), 1–22. https://doi.org/10.1186/s13059-020-02110-3

McManus, C. J., Coolon, J. D., Duff, M. O., Eipper-Mains, J., Graveley, B. R., & Wittkopp, P. J. (2010). Regulatory divergence in Drosophila revealed by mRNA-seq. *Genome Research*, *20*(6), 816–825. https://doi.org/10.1101/gr.102491.109

Metzger, B. P. H., Wittkopp, P. J., & Coolon, J. D. (2017). Evolutionary Dynamics of Regulatory Changes Underlying Gene Expression Divergence among Saccharomyces Species. *Genome Biology and Evolution*, *9*(4), 843–854. https://doi.org/10.1093/gbe/evx035

Miller, M. R., Robinson, K. J., Cleary, M. D., & Doe, C. Q. (2009). TU-tagging: Cell type-specific RNA isolation from intact complex tissues. *Nature Methods*, *6*(6), 439–441. https://doi.org/10.1038/nmeth.1329

Myllymäki, H., Valanne, S., & Rämet, M. (2014). The Drosophila Imd Signaling Pathway . *The Journal of Immunology*, *192*(8), 3455–3462. https://doi.org/10.4049/jimmunol.1303309

Nguyen, K., Fazio, M., Kubota, M., Nainar, S., Feng, C., Li, X., … Spitale, R. C. (2017). Cell-Selective Bioorthogonal Metabolic Labeling of RNA. *Journal of the American Chemical Society*, *139*(6), 2148–2151. https://doi.org/10.1021/jacs.6b11401

Nguyen, Q. H., Pervolarakis, N., Nee, K., & Kessenbrock, K. (2018). Experimental considerations for single-cell RNA sequencing approaches. *Frontiers in Cell and Developmental Biology*, *6*(SEP), 1–7. https://doi.org/10.3389/fcell.2018.00108

Osada, N., Miyagi, R., & Takahashi, A. (2017). Cis- and trans- regulatory effects on gene expression. *Genetics*, *206*(August), 2139–2148. https://doi.org/10.1534/genetics.117.201459/-/DC1.1

Otsuki, L., Cheetham, S. W., & Brand, A. H. (2014). Freedom of expression: Cell-type-specific gene profiling. *Wiley Interdisciplinary Reviews: Developmental Biology*, *3*(6), 429–443. https://doi.org/10.1002/wdev.149

Sackton, T. B., Lazzaro, B. P., & Clark, A. G. (2010). Genotype and Gene Expression Associations with Immune Function in Drosophila. *PLoS Genetics*, *6*(1). https://doi.org/10.1371/journal.pgen.1000797

Schaefke, B., Emerson, J. J., Wang, T. Y., Lu, M. Y. J., Hsieh, L. C., & Li, W. H. (2013). Inheritance of gene expression level and selective constraints on trans-and cis-regulatory changes in yeast. *Molecular Biology and Evolution*, *30*(9), 2121–2133. https://doi.org/10.1093/molbev/mst114

Shi, X., Ng, D. W. K., Zhang, C., Comai, L., Ye, W., & Jeffrey Chen, Z. (2012). Cis- and trans-regulatory divergence between progenitor species determines gene-expression novelty in Arabidopsis allopolyploids. *Nature Communications*, *3*. https://doi.org/10.1038/ncomms1954

Tiroshauth, I., Reikhav, S., Levy, A. A., & Barkai, N. (2009). A yeast hybrid provides insight into the evolution of gene expression regulation. *Science*, *324*(5927), 659–662. https://doi.org/10.1126/science.1169766

Troha, K., Im, J. H., Revah, J., Lazzaro, B. P., & Buchon, N. (2018). *Comparative transcriptomics reveals CrebA as a novel regulator of infection tolerance in D. melanogaster*. *PLoS Pathogens* (Vol. 14). https://doi.org/10.1371/journal.ppat.1006847

Valanne, S., Wang, J.-H., & Rämet, M. (2011). The Drosophila Toll Signaling Pathway . *The Journal of Immunology*, *186*(2), 649–656. https://doi.org/10.4049/jimmunol.1002302

Wang, Y., & Navin, N. E. (2015). Advances and Applications of Single-Cell Sequencing Technologies. *Molecular Cell*, *58*(4), 598–609. https://doi.org/10.1016/j.molcel.2015.05.005

Whitehead, A., & Crawford, D. L. (2005). Variation in tissue-specific gene expression among natural populations. *Genome Biology*, *6*(2). https://doi.org/10.1186/gb-2005-6-2-r13

Wittkopp, P. J., Haerum, B. K., & Clark, A. G. (2004). Evolutionary changes in cis and trans gene regulation. *Nature*, *430*(6995), 85–88. https://doi.org/10.1038/nature02698

Wittkopp, P. J., Haerum, B. K., & Clark, A. G. (2008). Regulatory changes underlying expression differences within and between Drosophila species. *Nature Genetics*, *40*(3), 346–350. https://doi.org/10.1038/ng.77

Wittkopp, P. J., & Kalay, G. (2012). Cis-regulatory elements: Molecular mechanisms and evolutionary processes underlying divergence. *Nature Reviews Genetics*, *13*(1), 59–69. https://doi.org/10.1038/nrg3095

Zhang, X., & Emerson, J. J. (2019). Inferring Compensatory Evolution of *cis-* and *trans-*Regulatory Variation. *Trends in Genetics*, *35*(1), 1–3. https://doi.org/10.1016/j.tig.2018.11.003

Zhang, Y., Khullar, D., Wang, F., Steel, P., Wu, Y., Orlander, D., … Kaushal, R. (2021).
Socioeconomic variation in characteristics, outcomes, and healthcare utilization of COVID-19 patients in New York City. *PLoS ONE*, *16*(7 July), 1–17.
https://doi.org/10.1371/journal.pone.0255171

CHAPTER 2
The mode of expression divergence in *Drosophila* fat body is infection-specific

The contents of this chapter appear in the journal of *Genome Research* Ramirez-Corona 2021

2.1 Abstract

Transcription is controlled by interactions of *cis*-acting DNA elements with diffusible *trans*-acting factors. Changes in *cis* or *trans* factors can drive expression divergence within and between species, and their relative prevalence can reveal the evolutionary history and pressures that drive expression variation. Previous work delineating the mode of expression divergence in animals has largely used whole body expression measurements in one condition. Since *cis*-acting elements often drive expression in a subset of cell types or conditions, these measurements may not capture the complete contribution of *cis*-acting changes. Here, we quantify the mode of expression divergence in the *Drosophila* fat body, the primary immune organ, in several conditions, using two geographically distinct lines of *D. melanogaster* and their F1 hybrids. We measured expression in the absence of infection and in infections with Gram-negative *S. marcescens* or Gram-positive *E. faecalis* bacteria, which trigger the two primary signaling pathways in the *Drosophila* innate immune response. The mode of expression divergence strongly depends on the condition, with *trans*-acting effects dominating in response to Gram-positive infection and *cis*-acting effects dominating in Gram-negative and pre-infection conditions. Expression divergence in several receptor proteins may underlie the infection-specific *trans* effects. Before infection, when the fat body has a metabolic role, there are many compensatory effects, changes in *cis* and *trans* that counteract each other to maintain expression levels. This work demonstrates that within a single tissue, the mode of expression divergence varies between conditions and suggests that these differences reflect the diverse evolutionary histories of host-pathogen interactions.

2.2 Introduction

Differences in gene expression are drivers of phenotypic divergence in closely related species (King and Wilson 1975). These expression differences can arise through sequence changes in *cis*-regulatory elements, such as enhancers, or in the coding regions of *trans*-acting factors, such as transcription factors. These two types of changes differ in their impact. Changes in *cis* are local, typically affecting the expression of one gene at a time, whereas changes in *trans* can be broad, affecting all downstream targets of a gene. The relative prevalence of each of these types of changes may give insight into how expression divergence arises in a particular setting: through the accumulation many fine-tuning *cis*-acting changes, by a smaller number of large impact *trans*-acting changes, or both.

The prevalence and relative contributions of *cis* and *trans* changes are being explored in various model systems (Signor and Nuzhdin 2018). For example, within individual *Drosophila melanogaster* lines or between *Drosophila* species, the contributions of *cis*-acting changes generally increase with phylogenetic distance, and the precise balance of *cis* versus *trans* effects depends on the phylogenetic relationships and demographics of the genotypes being compared (Wittkopp et al., 2004, Wittkopp et al., 2008, McManus et al., 2010, Coolon et al., 2014, Osada et al., 2017). These studies have elucidated the mode and tempo of the changes driving expression divergence; however, most studies use whole body measurements of expression, thus averaging signal across multiple tissues. Therefore, these studies cannot examine the prevalence of *cis* and *trans* changes in specific biological processes, which may be subject to different types of selection pressure. In addition, given that many *cis*-regulatory elements act in a tissue-specific manner, studies that measure *cis* and *trans* effects with tissue-specific resolution may reveal effects undetectable in heterogenous samples.

Drosophila have an innate, but not adaptive, immune response, and this response is a powerful system for measuring the contributions of *cis* and *trans* changes for several reasons. First, the immune response is inducible, with active and inactive states. This allows for the clear delineation of the transcriptional response of the immune system from that of other processes. Second, the fat body within the immune system is an optimal tissue for study. Though other tissues participate in the immune system, the fat body is a primary driver of the humoral response (Buchon et al., 2014), and it is relatively easy to isolate. Lastly, there is ample variation in the resistance, survival, and transcriptional response to infection between individual *D. melanogaster* lines (Lazzaro et al., 2004, Lazzaro et al., 2006, Sackton et al., 2010, Hotson and Schneider 2015), suggesting there are many sequence changes driving these differences.

To quantify changes in *cis* and *trans* that drive transcriptional divergence in the immune response, we use allele-specific expression analysis (ASE) of RNA-seq data (Wittkopp et al. 2004, Signor and Nuzhdin 2018, Frochaux et al 2020). In this approach, we compare a gene's expression levels in two parental lines to the expression levels of each parental allele in the resulting F1 hybrids. Differences in expression due to changes in *cis*, e.g. a sequence change in a promoter or enhancer, will only affect the expression of the corresponding parental allele. Thus, changes in *cis* are independent of cellular environment and will be observed as allelic imbalance between the parents that is maintained in the hybrids. Differences in *trans*, e.g. a coding sequence change in a transcription factor, will affect the expression of both alleles in the F1 hybrids and thus will be observed as differential expression in the parental lines that is not maintained in the F1 hybrids. Combining ASE analysis with RNA-seq allows us to determine the prevalence of *cis* and *trans* changes genome-wide.

When comparing the innate immune response of different *D. melanogaster* lines, it is not clear whether *cis* or *trans* changes will dominate. Changes in *cis* generally affect a single gene's expression and thus may be easily tolerated, as they only introduce small amounts of phenotypic variation. Changes in *trans* can affect the expression of many genes at once and efficiently introduce a large amount of phenotypic variation, but changes in *trans* may be harder for the organism to tolerate, as they also increase the likelihood of deleterious effects. However, the specific biology of the innate immune response may temper this expectation. Antimicrobial peptides (AMPs) are among the most highly up-regulated genes in response to infection, but the deletion of individual AMP genes often has little to no measurable effect of infection survival (Hanson et al., 2019). This suggests that to get an appreciable phenotypic effect, synchronous changes in gene expression are required, which can result from a *trans*-acting change. In addition, within *D. melanogaster* lines, *trans* changes are typically more prevalent (Wittkopp et al., 2008, Coolon et al., 2014). In this setting, the observation of a large number of *cis*-acting changes would imply that immune-responsive expression divergence is achieved through the divergence of one gene at a time, suggesting a fine-tuning process. Conversely, a preponderance of *trans*-acting changes would imply that expression divergence is achieved through changes in upstream factors that can simultaneously modulate the expression of many target genes.

To measure the contributions of *cis* and *trans*-acting changes in the *Drosophila* innate immune response, we measured fat body gene expression in two sequenced inbred *D. melanogaster* lines and their F1 hybrids, in control and infection conditions. To find signaling pathway-specific effects, we separately infected the animals with either Gram-positive *Enterococcus faecalis* or the Gram-negative *Serratia marcescens*. These bacteria have different strengths of virulence and separately trigger the two primary immune signaling pathways in the

fly. We quantified the contribution of *cis* and *trans* effects in the control and in each infection condition. This approach enabled us to examine the evolutionary changes that drive expression divergence in response to a stimulus, while minimizing the confounding effects of multiple tissue types.

2.3 Results

2.3.1 Two geographically distinct lines show genotype-specific immune response

To measure the relative contributions of *cis-* and *trans*-acting effects in the innate immune response, we needed two inbred, sequenced strains of *D. melanogaster* with abundant genetic variation and phenotypic differences in the immune response. The founder lines of the *Drosophila* Synthetic Population Resource fit these requirements, making them ideal candidates (King et al., 2012). To maximize the likelihood of finding variation in these lines, we selected two lines from different continents, the A4 line, also known as KSA2, collected from the Koriba Dam in South Africa, and the B6 line, collected from Ica, Peru. Using the available SNP data, we found 462,548 SNPs between A4 and B6, with about half of them falling into exonic regions, indicating that 0.9% of exonic bases varied between the genotypes, with an average of 25.25 variants per gene. The extensive variation in the coding regions allowed us to map, on average, 11.2% ($\pm$1.3%) of RNA-seq reads in an allele-specific manner.

To assess the divergence in the A4 and B6 immune responses, we measured gene expression pre- and post-infection in the abdominal fat body, the primary site of immune response. To do so, we performed RNA-seq on the dissected fat bodies of 4-day old males from both lines that had been infected with either Gram-positive *Enterococcus faecalis* (*Efae*) or Gram-negative

Serratia marcescens (*Smar*). We selected these bacteria because in *D. melanogaster*, Gram-positive infections generally stimulate the Toll pathway, and Gram-negative infections generally stimulate the IMD pathway, though there is additional nuance due to signaling crosstalk and the contributions of other signaling pathways (Buchon et al., 2014; Busse et al., 2007; Lemaitre and Hoffmann 2007; Tanji et al., 2010; Troha et al., 2018). We measured expression pre-infection and three hours post-infection, to capture the early transcriptional response prior to the complicating effects of feedback. As a control, we performed RNA-seq on the fat bodies of uninfected, unwounded animals from each genotype (see Methods). This choice means that, when compared to the control, the expression response observed in the infected samples includes both wound healing and infection responses.

In response to *Efae* infection, we found sizable genotype-specific effects in the immune response. To detect these effects, we performed two types of differential gene expression analysis: we compared control and infected samples to find *Efae*-responsive genes, and then within this group, we looked for genes differentially expressed between the A4 and B6 genotypes. We found 1165 differentially expressed genes between the control and infected samples regardless of genotype (Figure 2.1A). We categorized these *Efae*-responsive genes into four groups based on their differential expression between genotypes. Group 1 genes showed no genotype specific expression, Group 2 genes are differentially expressed only in the control samples, Group 3 genes are differentially expressed only in the infected samples, and Group 4 genes are differentially expressed in both control and infected samples. Of the 500 *Efae*-responsive genes showing genotype effects, 87% (433 genes) are in Group 3, while only 10 genes are in Group 1 and 57 genes in Group 4 (Figure 2.1B). This indicates that many *Efae*-responsive genes show genotype-specific expression, and these differences are typically only revealed in response to infection.

In response to the *Smar* infection, we found 1203 differentially expressed genes between the control and infected samples (Figure 2.1A). To look for genotype-specific expression, we categorized the 1203 *Smar*-responsive genes into the three previously mentioned groups. For this infection, we found roughly equal numbers of genes in Groups 2-4, with 88, 91, and 84 genes, respectively (Figure 2.1B). This indicates that a higher fraction of *Smar*-responsive genes show genotype effects prior to infection than *Efae*-responsive genes ($p = 1.7 \times 10^{-11}$, Chi-square test, Bonferroni corrected), while a higher fraction of *Efae*-responsive genes show genotype effects after infection ($p = 9.5 \times 10^{-67}$, Chi-square test, Bonferroni corrected).

To assess whether there is also phenotypic divergence on the organismal level, we performed the *Efae* and *Smar* infections and measured survival and bacterial load. In response to *Efae* infection, we found differences in the ability to survive infection between genotypes, with B6 surviving infection longer than A4 (Supplemental Figure S2.1A). In response to *Smar*, we found there were no significant differences in survival, but bacterial load was lower in A4 than in B6 (Supplemental Figure S2.1B, S2.1C). Together, these data demonstrate that there are differences between the two lines in their ability to resist or survive infection.

To compare our tissue-specific measurements to previous work, we intersected our *Efae*- and *Smar*-responsive genes to an existing list of immune-responsive genes. This list is an expanded version of the *Drosophila* immune responsive genes set (DIRGS) and constitutes the summation of more than two decades of work in *Drosophila* (De Gregorio et al., 2001; Lemaitre and Hoffman 2007; Troha et al., 2018). Of 538 genes on this list, we found more than half of these (297 genes) were identified as immune-responsive in our data (Figure 2.1C). Troha and colleagues identified a subset of immune-responsive genes as core, i.e., the gene that are differentially expressed regardless of the type of bacterial infection (Troha et al., 2018). Of these 252 core genes,

approximately 40% were found to be both *Smar-* and *Efae-*responsive in our data. Therefore, despite differences in the genetic background, tissue (previous studies were typically done with whole body sampling), and time points, our findings show concordance with previous studies of gene expression in response to infection. We also show that the A4 and B6 lines have divergence both immune-responsive expression, making them suitable for subsequent F1 hybrid experiments.

2.3.2 *Cis*-acting effects dominate expression variation in the uninfected fat body

To effectively quantify *cis* and *trans* effects, we needed to accurately analyze the allelic expression in F1 hybrids. Using the the Allele-Specific Alignment Pipeline (ASAP) (Krueger, https://www.bioinformatics.babraham.ac.uk/projects/ASAP/), we quantified allele-specific expression in our samples. Since we are working with males, we were able to use the fraction of misassigned X-chromosome reads as a metric of our pipeline's accuracy (Supplemental Methods). On average, 0.5% of X chromosome reads were mis-assigned (standard deviation = 3%; Supplemental Table S2.1). The consistent, low level of mis-assigned reads verifies our ability to accurately quantify allelic expression.

We next sought to quantify *cis* and *trans* effects in the control samples. We used the complete set of parental RNA-seq reads and the subset of the F1 hybrid reads that could be assigned to a specific allele. Using three separate generalized linear models, we tested for differential expression in the parents, allelic imbalance in the F1 hybrids, and *trans* effects between parents and F1 hybrids (see Methods) (Davidson and Balakrishnan, 2016; Osada et al., 2017; Takada et al., 2017). We then categorized each gene into one of six categories (Figure 2.2A). Genes showing no differential expression in the parents or F1 hybrids are ***conserved***. Genes showing differential expression in both the parents and F1 hybrids and no *trans* signal are ***cis-only***. Genes

showing differential expression in the parents and not the F1 hybrids are ***trans-only***. Some genes show evidence of both *cis* and *trans* effects and are either ***compensatory*** (if the changes on expression are in opposite directions) or ***cis + trans*** (if the changes on expression are coherent). Genes that do not fall into any of these categories are ***undetermined.***

Of the 4959 genes that were expressed in the pre-infection fat body that could be detected in an allele-specific manner, 77% were conserved (3802 genes; Figure 2.2B, F). We found 151genes showing unambiguous *cis* or *trans* effects. Among these 151 genes, *cis* effects dominated the signal: 90% of genes (135 genes) showed *cis* signal (including *cis-only*, *cis + trans* and compensatory genes), and 57% (87 genes) showed *cis-only* effects. 42% of genes (64 genes) showed *trans* signal and only 10% of genes (16 genes) showed *trans-only* effects. One-quarter of genes (37 genes) were compensatory, even when using an experimental design to avoid the artificial inflation of compensatory signal (Methods; Zhang and Emerson, 2019; Fraser et al., 2019). Additionally, to ensure that any differences in the quality of our in-house A4 and B6 transcriptomes do not affect our conclusions, we quantified *cis* and *trans* effects using sets of high confidence genes at multiple levels of stringency and found that this had negligible effects on the detected signal (Methods; Supplemental Figure S2.2; Supplemental Table S2.2). From these data, we can conclude that in the unstimulated state, most genes have conserved expression levels in the fat body, and among those genes that diverge, *cis* effects dominate, with a sizable number of genes showing compensatory *cis* and *trans* changes.

2.3.3 More *cis* than *trans* effects are found in *Efae*-infected fat body expression

We quantified *cis* and *trans* effects in *Efae*-infected samples following the same methodology. We found roughly 52% of genes (2580 genes) are conserved and 379 genes showed unambiguous *cis*

or *trans* effects (Figure 2.2C). To identify genes whose expression divergence is specific to the immune response, we eliminated genes that show *cis* or *trans* signal in the control sample. After this filtering, roughly 69% of the genes showing *cis* or *trans* effects (263 genes) remained; 66% of these genes (174 genes) show *cis-only* signal, and 28% (75 genes) show *trans-only* signal. Only 8 genes (3%) show concordant *cis + trans* effects, and only 6 genes show compensatory effects. Of the genes that show *cis-only* signal, roughly even numbers of genes show higher expression in each genotype, consistent with the idea that *cis*-acting changes affect a single gene at a time. In contrast, of the genes showing *trans-only* signal, nearly twice as many were expressed more highly in the B6 genotype (48 genes) than in the A4 genotype (27 genes) ($p = 0.0105$, Chi-square test). This suggests that one or a few changes in upstream regulatory factors are responsible for this observation, and below, we identify candidate genes. Since we do not observe this trend towards higher B6 expression in the control samples and have removed genes that showed any evidence of mapping bias (Methods), we are confident this trend reflects true biological differences in the immune response. In sum, we find both *cis* and *trans* effects drive *Efae*-responsive expression divergence, with *cis* effects dominating.

2.3.4 *Trans* effects dominate expression variation in the *Smar*-infected fat body

Lastly, we quantified *cis* and *trans* effects in response to *Smar* infection. We found roughly 82% of genes (4106 genes) are conserved, and 355 genes showed unambiguous *cis* or *trans* signal (Figure 2.2D). We again filtered out genes that show *cis* or *trans* effects in the control samples and were left with 251 genes that have immune-specific signal. Of these, 31% (79 genes) showed *cis-only* signal, and roughly equal numbers of *cis-only* genes showed higher expression in each genotype. Seven genes showed *cis + trans* effects, and 16 genes had compensatory signal. 59% of

genes (149 genes) showed *trans-only* signal. Within *trans-only* genes, we found that 71% (106 genes) showed greater expression in B6. In summary, in response to *Smar* infection, *trans* effects drive the majority of expression divergence between the two genotypes and few genes show compensatory effects.

2.3.5 Comparisons of *cis* and *trans* signals in different conditions reveal both infection-specific and shared divergence

To systematically assess modes of expression variance under different conditions, we compared the proportion of genes falling into the different categories. The control and *Efae*-infected samples had a greater proportion of *cis-only* genes than the *Smar* samples (control vs. *Smar* $p = 3.96 \times 10^{-6}$, *Efae* vs. *Smar* $p = 6.83 \times 10^{-14}$, Chi-square test, Bonferroni-corrected). All three groups differ in the proportion of *trans-only* genes, with *Smar*-infected samples showing more than twice the proportion of genes with *trans*-only signal, followed by *Efae,* and then the control samples (control vs. *Efae* $p = 3.48 \times 10^{-4}$, control vs. *Smar* $p < 1.54e-16$, *Efae* vs. *Smar* $p = 3.05 \times 10^{-11}$, Chi-square test, Bonferroni-corrected). We also found that the uninfected fat body showed significantly more compensatory signal than either infected sample (control vs. *Efae* $p < 1.54e-16$, control vs. *Smar* $p = 1.8 \times 10^{-6}$, Chi-square test, Bonferroni-corrected). Taken together, this suggests one of two possibilities. One possibility is that before infection, when the fat body is carrying out its metabolic functions, there is less pressure for expression divergence. An alternative interpretation is that immune-responsive genes are more tolerant of expression divergence and subject to less pressure to maintain expression levels. In response to infection, there is ample expression divergence, which is driven by both *cis* and *trans* effects. The extent to which each type of effect contributes is

dependent on the particular pathogen, suggesting that the relative importance of local and pleiotropic changes is specific to different infection pressures.

Though we generally expect the two infections to regulate gene expression via distinct signaling pathways, we also anticipated some genes would be regulated in both infections, either due to crosstalk between the IMD and Toll pathways (Busse et al., 2007; Tanji et al., 2010) or via more general infection and wound responses. We found 86 genes with unambiguous *cis* and/or *trans* signal in response to both *Efae* and *Smar* infection (Supplemental Data S2.1). Of these genes, 71 showed concordant classification. Therefore, in the majority of genes shared between these two infections, the same genetic differences are likely driving the expression divergence in both infection conditions.

2.3.6 Differential expression of detection genes is a likely source for genotype expression bias in observed *trans* effects

Since we observed that genes with *trans-only* effects tended to be more highly expressed in B6 than in A4 in both infection conditions, we hypothesized that changes in a handful of upstream immune factors are responsible for this phenomenon. The changes in upstream regulators could either be infection-specific or shared. Out of 202 genes showing *trans-only* signal in either infection, only 17 genes were shared, indicating that the bulk of *trans*-acting changes are likely infection-specific.

Immune detection genes, signaling genes, or transcription factors differentially expressed between genotypes are likely sources of *trans*-acting changes, since these genes have the ability to affect the expression of many downstream targets. We posited that these genotype-specific differences had to be present in the control to have the effects at the 3-hour post-infection

timepoint. Of the 295 genes that are differentially expressed between genotypes in the control samples, we found 22 genes that are prime candidates, which we will refer to as *trans*-source candidates (Table 1).

Five peptidoglycan recognition proteins (PGRP) genes are potential mediators of the large number of *trans* effects observed in the *Smar* infection. Four of these PGRPs (PGRP-SC1a, PGRP-SC1b, PGRP-SC2, PGRP-LB) are negative regulators of the IMD response, and the last gene, PGRP-SD is positive Toll and IMD regulator (Bischoff et al., 2006**;** Zaidman-Rémy et al., 2006; Iatsenko et al., 2016; Charroux et al., 2018; Lu et al., 2020). Three of the negative regulators, PGRP-SC1a, PGRP-SC1b, PGRP-SC2, are more highly expressed in A4. Given that these are negative regulators of the IMD pathway, this finding is congruent with the observation that genes showing *trans-only* signal tend to show greater expression in B6. PGRP-SD is more highly expressed in B6, and, given its role as a positive regulator of the IMD response, it is also consistent with the trend of higher B6 expression of genes showing *trans-only* signal. The last negative regulator of IMD response, PGRP-LB, has higher expression in B6. Since three of the four negative regulators are more highly expressed in A4, it is possible this balance can account for the expression trend in *Smar trans-only* genes. It is also possible that the greater expression of PGRP-SD is enough to account for the differences observed.

Though there were fewer *trans* effects in the *Efae*-infected samples than in the *Smar*-infected samples, the pattern wherein most *trans*-only genes showed greater expression in B6 than A4 was maintained. Of the 22 *trans*-source candidates, we found two Toll-specific genes: *Spatzle-Processing Enzyme* (*SPE*) and *spatzle* (*spz*), which are both more highly expressed in B6. *Spatzle* is the Toll receptor ligand, and *SPE* is required to generate the active form of *spz*, so differential expression of these genes can drive a large number of downstream changes. In addition, PGRP-

SD can act as a positive regulator of both the Toll and IMD responses and is also found to have higher expression in the B6 line.

In addition to differences in expression between genotypes, function-altering differences in the coding sequences of immune genes may also be the source of *trans*-acting changes. As a first approach, we analyzed the coding sequence differences between A4 and B6 in the 22 *trans* source candidates identified above using the Ensembl Variant Effect Predictor (McLaren et al., 2016). There are a number of non-synonymous changes, some of which fall into functional domains (Supplemental Figure S2.3, Supplemental Table S2.3 and S2.4). Predicting the effect of these mutations on individual protein function, however, remains a challenge.

As an alternative approach, we analyzed the proportions of synonymous to nonsynonymous coding changes between A4 and B6 in several larger gene sets. Previous work has demonstrated that immune-related genes have a higher average rate of adaptive evolution than other gene classes (Sackton, et al. 2007; Obbard, et al. 2009). We wanted to see if, for our particular genotypes and genes of interest, the same held true. We considered all genes expressed in the fat body above a threshold of 1 count per million (CPM), and then sorted them into two groups: genes that are differentially expressed in response to either or both infections (*DE infection*) and those that are not (*fat body detected*). We then intersected each of these gene lists with our curated immune-responsive gene set to generate both a list of differentially and non differentially-expressed immune genes (*DE immune* and *non-DE immune* respectively; Figure 2.3A). We posited that, given the large number of *trans* effects in response to infection, differentially-expressed immune-related genes may have a greater proportion of nonsynonymous changes compared to the fat body detected gene set. We found that DE immune genes have a significantly higher fraction of nonsynonymous sequence changes (24%) compared to the fat body detected genes (21%) ($p =$

0.01, Chi-square test, Bonferroni-corrected), suggesting that some of these changes may be under selection and possibly the source of our *trans*-acting signal (Figure 2.3A-B). By comparison, the non-DE immune genes had a lower proportion of nonsynonymous changes (19 %, p=1.6×10^{-4}, Chi-square test, Bonferroni-corrected), suggesting that the elevated rate of nonsynonymous changes in DE immune genes is not simply reflective of their immune status. In summary, we find that differentially-expressed immune genes have a larger proportion of non-synonymous changes between our genomes of interest than fat body detected or non-differentially expressed immune genes. Some of these non-synonymous changes may be capable of altering the function of these proteins and therefore drive expression divergence of downstream genes in a *trans*-acting fashion.

2.3.7 Genes with *cis* effects have greater transcription factor binding site divergence than to genes with *trans effects*

The above analysis sought to identify changes in expression or protein sequence that may drive the observed *trans* effects *Cis*-acting changes also drive expression divergence of a large number of genes. These changes encompass mutations in several types of DNA features, including promoters, enhancers, and untranslated regions. We analyzed the patterns of divergence in immune-responsive transcription factor binding sites (TFBS) to see if they were consistent with our delineation of *cis* and *trans*-acting effects. We hypothesized that genes whose divergence was due to *cis*-acting effects would show more divergence in the associated TFBS than those without them.

We scanned potential regulatory regions of our genes of interest for TFBS in the A4 and B6 genomes. There are relatively few characterized immune-responsive enhancers in the fat body, so instead we extracted 1kb regions upstream of the transcription start site of genes showing any

36

cis or *trans*-acting changes in infected conditions. We searched these regions for binding sites corresponding to four known immune-responsive transcription factors Dorsal (Dl), Relish (Rel), Serpent (Srp) and CrebA (Shazman et al., 2014). CrebA modulates transcription in response to both Gram-positive and Gram-negative bacteria (Troha et al., 2018). Srp binding sites have been previously used to identify immune-responsive enhancers (Senger et al., 2004). Relish is a NF-kB transcription factor downstream of the IMD pathway, and Dl and its paralog Dorsal-related immunity factor (Dif) are downstream of the Toll signaling pathway. For this analysis however, only Dl was considered since Dif homodimers have less specific binding preferences than Dl and Dif/Rel heterodimers bind sequences similar to Rel homodimers (Senger et al., 2004). Given the cross-talk between the Toll and IMD pathways, we searched both *Efae-* and *Smar*-responsive genes for both Dl and Rel binding sites. For each gene, we calculated the difference in the total number of TFBS in the A4 and B6 genomes. We then compared the genotype differences between genes showing any *cis* effects and genes showing exclusively *trans* effects (see Methods). We hypothesized that genes showing *cis* effects would have more differences in TFBS than the *trans* effected genes, which would be observed as a broader distribution in TFBS differences.

For all transcription factors except Dl (Figure 2.4A-E), the genes with *cis* effects did indeed show a broader distribution of difference than those with *trans* effects (all TFs: $p = 8.8 \times 10^{-13}$, Rel: $p = 2.9 \times 10^{-2}$, Srp: $p = 7.1 \times 10^{-10}$, CrebA: $p = 1.5 \times 10^{-7}$, F-test to compare distribution variances, Bonferroni corrected). While most genes do not differ in TFBS numbers, 22% of genes with *cis* changes differed, as opposed to only 18% of *trans* affected genes, though this difference was not significant (Figure 2.4F). As the number of characterized immune-responsive enhancers and transcription factors increases, we will be able to refine this analysis to more accurately identify potential causative mutations of *cis*-effects.

2.4 Discussion

Here, we quantified the mode and extent of expression divergence in the *Drosophila* abdominal fat body, both in an uninfected control condition, where it carries out a variety of metabolic roles, and in response to two types of infection. We found that two geographically isolated lines of *D. melanogaster* are phenotypically distinct in their immune responses, differing both on the organismal and transcriptional levels. By comparing gene expression in the fat body between these lines and their F1 hybrids, we quantified the contributions of *cis* and *trans* effects to expression divergence in the uninfected control, *Efae*-infected and *Smar*-infected conditions. Both the control and *Efae* infection conditions were dominated by *cis* effects, while the *Smar* infection condition had an abundance of *trans* effects. The uninfected control also showed a greater proportion of compensatory effects, suggesting that there is stabilizing selection to maintain fat body expression levels of certain genes in the absence of an infection. Among the genes showing changes in *trans*, we found that expression of the B6 allele is typically higher, and we identified expression divergence in a group of proteins that may drive these *trans* effects. By analyzing the TFBS content of upstream regions of genes, we found that genes with *cis* effects show evidence of more TFBS divergence than genes with *trans* effects. Overall, we find that the mode of evolution in expression divergence can vary between conditions in a single tissue and likely represents condition-specific selection pressures.

Our unique approach to measuring the mode of expression divergence gave rise to several novel observations about the relative contributions of *cis* and *trans* effects on expression variation. While there have been a number of studies aimed at disentangling the contribution of *cis* and *trans* changes to gene expression in *Drosophila,* few have sought to answer this question using a single organ or with different physiological stimuli (Wittkopp et al., 2004, Wittkopp et al., 2008,

McManus et al., 2010, Coolon et al., 2014, Osada et al., 2017). Our approach allows us to examine evolutionary changes in response to perturbation while minimizing the confounding effects of multiple tissue types. A previous study by Juneja, et al. (2016) found, among geographically distinct flies, a large number of *cis*-acting changes that cause whole body expression divergence in response to an infection with mixture of bacteria. This is concordant with our finding of a large number of *cis*-acting changes in both infection conditions, but this study did not quantify *trans*-acting changes or distinguish between Toll- and IMD-specific responses. By measuring expression in the heads and abdomens of multiple *D. melanogaster* lines, another group reported the predominance of changes in *cis* over those in *trans* but did not measure these differences in different physiological states or attempt to dissect individual tissues in the head or abdomen (Osada et al., 2017). Most recently, two studies sought to uncover the underlying genetics of resistance to either *P. entomophila* or *E. faecalis* infection, and each identified novel drivers of phenotypic divergence (Chapman et al., 2020; Frochaux et al., 2020). Here, we sought to directly assess the contribution of *cis* and *trans* sequence changes in a single tissue in the context of multiple treatment conditions, giving a uniquely high-resolution view of the evolutionary sequence changes underlying expression divergence.

With our approach we were able to uncover two notable trends. First, we found that compensatory mutations were more frequent in the control samples than in either of the infected conditions. Previous studies in several organisms had suggested that compensatory effects were very prevalent (McManus et al., 2010, Gonclaves et al., 2012, Schaefke et al., 2013, Coolon et al., 2014). However, certain choices in experimental design can inflate estimates of compensatory effects (Zhang and Emerson 2019; Fraser et al., 2019). Our study avoids this artifact, and therefore yields a more accurate estimate of compensatory effects across multiple conditions. Additionally,

a large proportion of studies addressing *cis* and *trans* effects in animals do so in "control" conditions, which may not reveal the full extent of selection forces that act on gene expression (Gonclaves et al., 2012, Osada et al., 2017, Davidson and Balakrishnan 2016, Signor and Nuzhdin 2018). We find evidence that the genes involved in the maintenance of basic metabolic functions of the uninfected fat body are under different selective pressures than those involved in immune response. Unlike the immune-responsive genes, which must contend with a continuously evolving pathogen landscape, the genes carrying out metabolic functions may be subject to stabilizing selection, given relatively unchanging nutritional availability. In future studies, it will be interesting to further probe which systems and conditions show enrichment for these different patterns of expression divergence.

Secondly, we observe that the relative contribution of *cis*- and *trans*-acting changes are perturbation-specific. In response to *Efae* infection, *cis* effects dominate expression changes, while in the *Smar* infection, *trans* changes are predominant. The prevalence of either *cis* or *trans* effects can be reasonably justified in our system, but we did not anticipate that the proportion of these effects would be infection specific. Because changes in *trans* factors have pleiotropic effects, it has been suggested that changes to these factors are under more selective constraint than *cis*-acting elements, and, thus, *cis* effects can more readily introduce small-scale variation into a system (Schaefke et al., 2013). In some cases, however, arriving at a more fit phenotype may require the coordinated alteration of expression of many genes, which may be more readily achieved by changes to *trans*-acting factors. In our *D. melanogaster* lines, *S. marcescens* is more virulent than *E. faecalis* – a higher dose of *E. faecalis* is needed to achieve similar levels of mortality to that of *S. marcescens* (Supplemental Figure S2.1). It is possible that adaptation to highly virulent pathogens or rapidly evolving pathogens requires large-scale, synchronous changes to expression,

whereas adaptation to less virulent pathogens is possible with smaller, localized mutations. Experiments with a wider range of pathogens, particularly those that trigger the same signaling pathway, will further illuminate the relationship between the mode of expression divergence and the host-pathogen relationship. In addition, expansion of the study to more *D. melanogaster* genotypes or to other time points will yield a more complete picture of the modes of expression divergence in the immune response.

In summary, we find that the mode of expression divergence, as represented by the proportion of *cis* and *trans* effects in a system, is condition-specific in the *Drosophila melanogaster* abdominal fat body. This specificity may be a result of the distinct selective pressures that different host-pathogen interactions exert on the *D. melanogaster* immune system. In the course of our study, we found several candidate genes that may be the sources of the observed *trans* effects, which are most prominent in *Smar* infection. In the future, we can combine the data sets presented here with other types of functional genomics experiments to identify the specific sequence changes that drive *cis*-acting divergence. Taken together, these studies will provide a more comprehensive view of how regulation of expression in this rapidly changing system is wired and evolves.

2.5 Methods

2.5.1 Animal genotypes, infection protocols, and survival analysis

The A4 and B6 *D. melanogaster* lines, SNP tables, and genomic reads were received from the *Drosophila* Synthetic Population Resource (King et al., 2012). Flies were reared at 25°C on standard cornmeal fly food (Brent and Oster 1974). For all RNA-seq experiments four-day-old males were infected with approximately 15 nL of $A_{600} = 0.5$ OD solution of either *Enterococcus*

faecalis or *Serratia marcescens* via microinjection, yielding an infection of ~10,000 CFUs/fly (Khalil et al., 2015). Survival and bacterial load experiments were performed using a modified infection protocol (Supplemental Methods). Uninfected controls were placed on a carbon dioxide pad for 6 minutes to mimic the effects of anesthesia used for microinjection. Bacteria were grown in liquid culture on a shaker at 37˚C overnight and then diluted 1:1000 in fresh media in the morning. Cultures were grown until exponential phase then pelleted down and resuspended in PBS for OD measurement and injection. Injections took place between 3:00 and 5:00 pm to account for the impact of circadian rhythm on immune response (Scheiermann et al., 2013).

To determine the number of unique SNPs between A4 and B6, we downloaded published SNP tables from the DSPR website (King et al., 2012). We selected SNPs that were not shared between lines and that also showed a reference allele frequency of < 0.05 . We then calculated total SNP differences for exonic and non-exonic regions using exon coordinates from FlyBase (dm6/iso-1: FB2019_01) (Thurmond et al., 2019).

2.5.2 Preparation and sequencing of RNA-seq libraries

For sequencing experiments abdominal filets with the attached fat bodies were prepared as in (Krupp and Levine et al., 2010) 3 hours post infection. Three fat bodies per sample were suspended in TRIzol on ice (Life Technologies) and immediately stored at -80˚C for later extraction (Kono et al., 2016). To mitigate the impact of batch effects, injections and RNA extractions were done in groupings of 6-8 samples, with at least two treatment conditions and two genotypes (A4, B6, A4B6 or B6A4) represented in each batch. A minimum of three biological replicates were collected for each treatment condition/genotype combination. Both the order of treatment and the order of RNA extraction was randomized for each batch. RNA was extracted using Zymo Research Direct-zol RNA Extraction Kits. Library construction was completed protocol outlined in (Serra

et al., 2018). Samples were then sequenced on Illumina NextSeq Platform with NextSeq 500/550 High Output Kit v2.5 to generate 43bp paired end reads. Data was imported to the UCI High Performance Computational Cluster for trimming and mapping of sequenced reads.

2.5.3 Differential expression analysis

Reads were trimmed and filtered using Trimmomatic 0.35 (Bolger et al., 2014), specifying the parameters ILLUMINACLIP:TruSeq3-PE.fa:2:30:10 LEADING:6 SLIDINGWINDOW:4:15 MINLEN:30. Count and TPM data for each sample was then calculated using Salmon 0.12.0 aligner (Patro et al., 2017) using the dm6/iso-1 transcriptome and the parameters -l A --validateMappings. Count matrices of gene-level data were then constructed in R using the Tximport 1.12.3 package (Soneson et al., 2015). To find genes either differentially expressed in response to each infection, compared to control, or differentially expressed between genotypes, we used the edgeR 3.26.5 package (Robinson et al., 2010, McCarthy et al., 2012). For this analysis we excluded lowly expressed genes (CPM<1), accounted for extraction batch in our model, and corrected p-values with false discovery rate (Benjamini and Yekutieli et al., 2001). Genes with an FDR < 0.05 were considered differentially expressed. Additionally, we assessed the potential effect of absolute expression on our ability to call genotype effects, and we did not find any significant sources of bias (Supplemental Figure S2.5). Code and accompanying files related to this section are in Supplemental Code as both R-notebooks and HTML documents (Script1_fig1).

2.5.4 Generation of A4 and B6 transcriptome annotations

To map RNA-seq reads in an allele-specific manner, we created two reference transcriptomes by lifting over iso-1 genome annotations to sequenced A4 and B6 genomes. Using the UCSC liftOver suite, custom chain files were created by mapping iso-1 homologous sequences to the A4 or B6 genome using BLAT (parameters -tileSize=12 -minScore=100 -minIdentity=98)(Salinas et al.,

2016). A subset of 7654 high confidence genes were used for the subsequent analysis (Supplemental Methods)

2.5.5 Allele-specific expression analysis

RNA reads were assigned parental alleles using Allele Specific Alignment Pipeline (Krueger, https://www.bioinformatics.babraham.ac.uk/projects/ASAP/) using the A4 and B6 genomes and allowing for no mismatches. Non-uniquely assignable reads were discarded. Count and TPM data were then generated by aligning allelic reads to the corresponding transcriptome. Count matrices of gene-level data were then constructed in R using the Tximport 1.12.3 package (Soneson et al., 2015).

To characterize expression divergence into *cis* and *trans* categories, differential expression was determined with unparsed parental reads and allele-specific reads from the F1 hybrids, using edgeR and three distinct GLM structures. Lowly expressed genes (CPM<1) and X Chromosome genes were excluded from the analysis. For each condition, we first tested for differential gene expression between parental samples (Murad et al., 2019). Next, we tested for allelic imbalance, taking into account parent of origin and maternal genotype effects as outlined in (Osada et al., 2017; Takada et al., 2017). For this test we used half of the F1 hybrid samples. Finally, we tested for *trans* effects using parental samples and the remaining F1 hybrid samples (J. Coolon pers. comm., Supplemental_Code: Script2_fig2.rmd Section 4). In all three tests, we assigned significance after adjusted *p*-values for multiple comparisons using the False Discovery Rate method (Benjamini and Yekutieli et al., 2001). Using the results from each test, we categorized each gene into one of five classes using the logic outlined in Table 2, which is based on previous studies (Emerson and Li 2010, McManus et al., 2010). Any genes that did not fit into the described patterns were categorized as "undetermined" and were excluded from further analysis. A complete

list of genes and their categories for each condition is available in the Supplemental Data S2.1. Code and accompanying files related to this section are available Supplemental Code as both R-notebook and HTML document form (Script2_fig2).

2.5.6 Identification of sources of trans effects

To investigate potential sources of observed *trans* effects, we looked for genes differentially expressed in uninfected samples. We selected genes that show differential expression between A4 and B6 in uninfected samples. These genes were then intersected with a list of known *Drosophila* transcription factors as well as known immune genes (De Gregorio et al., 2001; Lemaitre and Hoffman 2007; Hammonds et al., 2013, Troha et al., 2018). Only genes that were transcription factors, immune detection genes, or immune signaling genes were considered to be candidates.

2.5.7 Analysis of SNPs in coding sequences

To better understand the effects of sequence changes on coding regions between our lines we used the Ensemble Variant Effect Predictor Tool (VEP) to predict the effects of SNPs on the resulting amino acid sequence (McLaren et al., 2016). The *fat body expressed* gene set consists of genes expressed in the unstimulated fat body above a CPM of 1 and excludes genes differentially expressed in response to infection. *DE infeciton* genes are those differentially expressed in response to infection with either *Efae* or *Smar. DE immune* genes are those differentially expressed genes that are also previously verified immune response genes, and *non-DE immune* genes are previously-verified immune genes in the fat body expression gene set. Unless otherwise stated, figures were generated using ggplot2 3.3.2 package in R 3.6.0 (Wickham 2016, R Core Team 2019). Code and accompanying files related to this section are available in Supplemental Code as both R-notebook and HTML document form (Script3_figure3).

2.5.8 Analysis of transcription factor binding site variation

To investigate the effects of noncoding sequence changes on observed expression divergence, we identified differences in TFBS in potential *cis* elements of genes showing evidence of expression divergence. We selected 1kb regions upstream of the transcription start site of genes showing *cis* or *trans* effects in response to infection (421 genes). TFBS for these regions were passed through the tool for Finding Individual Motif Occurrences (FIMO) from the MEME suite (v 5.1.0) at p-value thresholds of either p =0.001, or p =0.0001 and using default parameters (Bailey et al., 2009). MEME motif files were generated using the sites2meme utility and TFBS sequences from OnTheFly (Shazman et al., 2014). Binding site data was downloaded into R 3.6.0 for analysis and plotting (R core Team 2019). Binding sites with a p-value < .001 were considered in downstream analysis. This threshold was selected based on the ability to call a majority of previously identified Rel and Srp binding sites in four immune responsive enhancers (Senger et al 2004, Supplemental Table S5). For comparison, we categorized genes into two groups *cis* genes or *trans* genes. *Cis* genes were defined as genes showing any *cis* effect (*cis*-only, *cis* + *trans* and compensatory categories) in response to either infection (219 genes). T*rans* genes were defined and genes that showed *trans*-only effects and no other effects in response to either infection (199 genes). Genes showing any combination of *trans*-only and any *cis* effects were excluded from this analysis (3 genes). Differences in the number of TFBS were calculated by subtracting the number of TFBS for each gene's upstream region in B6 from A4, for all TFs combined as well as for each TF separately. We tested for significance in the distribution of these TFBS differences between the *cis* and *trans* affected genes using an F test for variance with the R 3.6.0 function var.test. We also repeated this analysis using the TFBS score instead of number, and the results mirrored those found for the TFBS number (Supplemental Figure S2.4). Code and accompanying files related to this

section can be found in Supplemental Code as both R-notebooks and HTML document form (Script4_fig4).

2.5.9 Description of statistical tests

p-values for all single and multiple proportion comparisons were calculated using the R 3.6.0 prop.test function which performs a Chi-square test with Yate's continuity correction. For data where more than one statistical test was performed on the same set of data, *p*-values were Bonferroni corrected for familywise type I error by multiplying the p-value by the number of tests performed.

2.6 Data Access

All raw and processed sequencing data generated in this study have been submitted to the NCBI Gene Expression Omnibus (GEO; https://www.ncbi.nlm.nih.gov/geo/) under accession number GSE155033.

2.7 Acknowledgements

2.8 Disclosure Declarations

The authors have no conflicts of interest to declare.

2.9 References

Bailey, T. L., Boden, M., Buske, F. A., Frith, M., Grant, C. E., Clementi, L., … Noble, W. S. (2009). MEME Suite: tools for motif discovery and searching. *Nucleic Acids Research*, *37*(suppl_2), W202–W208. https://doi.org/10.1093/nar/gkp335

Benjamini, Y., & Yekutieli, D. (2001). The control of the false discovery rate in multiple testing under dependency. *Annals of Statistics*, *29*(4), 1165–1188. https://doi.org/10.1214/aos/1013699998

Bischoff, V., Vignal, C., Duvic, B., Boneca, I. G., Hoffmann, J. A., & Royet, J. (2006). Downregulation of the Drosophila immune response by peptidoglycan- recognition proteins SC1 and SC2. *PLoS Pathogens*. https://doi.org/10.1371/journal.ppat.0020014

Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics*, *30*(15), 2114–2120. https://doi.org/10.1093/bioinformatics/btu170

Brent, M. M., & Oster, I. I. (1974). Nutritional substitution -- a new approach to microbial control for Drosophila cultures. *Drosophila Information Service*, (52), 155–157.

Buchon, N., Silverman, N., & Cherry, S. (2014). Immunity in Drosophila melanogaster — from microbial recognition to whole- organism physiology. *Nature Publishing Group*, *14*(12), 796–810. https://doi.org/10.1038/nri3763

Busse, M. S., Arnold, C. P., Towb, P., Katrivesis, J., & Wasserman, A. (2007). A κB sequence code for pathway-specific innate immune responses, *26*(16), 3826–3835. https://doi.org/10.1038/sj.emboj.7601798

Chapman, J. R., Dowell, M. A., Chan, R., & Unckless, R. L. (2020). The Genetic Basis of Natural Variation in Drosophila melanogaster Immune Defense against Enterococcus faecalis. *Genes* . https://doi.org/10.3390/genes11020234

Charroux, B., Capo, F., Kurz, C. L., Peslier, S., Chaduli, D., Viallat-lieutaud, A., & Royet, J. (2018). Cytosolic and Secreted Peptidoglycan-Degrading Enzymes in Drosophila Respectively Control Local and Systemic Immune Responses to Microbiota. *Cell Host and Microbe*, *23*(2), 215-228.e4. https://doi.org/10.1016/j.chom.2017.12.007

Coolon, J. D., McManus, C. J., Stevenson, K. R., Graveley, B. R., & Wittkopp, P. J. (2014). Tempo and mode of regulatory evolution in Drosophila. *Genome Research*, *24*(5), 797–808. https://doi.org/10.1101/gr.163014.113

Davidson, J. H., & Balakrishnan, C. N. (2016). Gene Regulatory Evolution During Speciation in a Songbird. *G3&#58; Genes|Genomes|Genetics*, *6*(5), 1357–1364. https://doi.org/10.1534/g3.116.027946

De Gregorio, E., Spellman, P. T., Rubin, G. M., & Lemaitre, B. (2001). Genome-wide analysis of the Drosophila immune response by using oligonucleotide microarrays. *Proceedings of the National Academy of Sciences of the United States of America*, *98*(22), 12590–12595. https://doi.org/10.1073/pnas.221458698

Delignette-Muller, M. L., & Dutang, C. (2015). fitdistrplus: An R Package for Fitting Distributions. *Journal of Statistical Software*, *64*(4), 1–34.

Fraser, H. B. (2019). Improving Estimates of Compensatory cis–trans Regulatory Divergence. *Trends in Genetics*, *35*(1), 3–5. https://doi.org/10.1016/j.tig.2018.09.003

Frochaux, M. V., Bou Sleiman, M., Gardeux, V., Dainese, R., Hollis, B., Litovchenko, M., … Deplancke, B. (2020). cis-regulatory variation modulates susceptibility to enteric infection in the Drosophila genetic reference panel. *Genome Biology*, *21*(1), 6. https://doi.org/10.1186/s13059-019-1912-z

Goncalves, A., Leigh-Brown, S., Thybert, D., Stefflova, K., Turro, E., Flicek, P., … Marioni, J. C. (2012). Extensive compensatory cis-trans regulation in the evolution of mouse gene expression. *Genome Research*, *22*(12), 2376–2384. https://doi.org/10.1101/gr.142281.112

Hammonds, A. S., Bristow, C. A., Fisher, W. W., Weiszmann, R., Wu, S., Hartenstein, V., … Celniker, S. E. (2013). Spatial expression of transcription factors in Drosophila embryonic organ development, 1–15.

Hanson, M. A., Dostálová, A., Ceroni, C., Poidevin, M., Kondo, S., & Lemaître, B. (2019). Synergy and remarkable specificity of antimicrobial peptides in vivo using a systematic knockout approach. *ELife*, *8*, 1–24. https://doi.org/10.7554/eLife.48778

Hotson, A. G., & Schneider, D. S. (2015). Drosophila melanogaster natural variation affects growth dynamics of infecting Listeria monocytogenes. *G3: Genes, Genomes, Genetics*, *5*(12), 2593–2600. https://doi.org/10.1534/g3.115.022558

Iatsenko, I., Kondo, S., Mengin-Lecreulx, D., & Lemaitre, B. (2016). PGRP-SD, an Extracellular Pattern-Recognition Receptor, Enhances Peptidoglycan-Mediated Activation of the Drosophila Imd Pathway. *Immunity*, *45*(5), 1013–1023. https://doi.org/10.1016/j.immuni.2016.10.029

Juneja, P., Quinn, A., & Jiggins, F. M. (2016). Latitudinal clines in gene expression and cis - regulatory element variation in Drosophila melanogaster. *BMC Genomics*, 1–11. https://doi.org/10.1186/s12864-016-3333-7

Kassambara, A., & Kosinski, M. (2019). Survminer: Drawing Survival Curves using "ggplot2." Retrieved from https://cran.r-project.org/package=survminer

Khalil, S., Jacobson, E., Chambers, M. C., & Lazzaro, B. P. (2015). Systemic bacterial infection and immune defense phenotypes in Drosophila melanogaster. *Journal of Visualized Experiments*, *2015*(99), 1–9. https://doi.org/10.3791/52613

King, E. G., Macdonald, S. J., & Long, A. D. (2012). Properties and power of the Drosophila synthetic population resource for the routine dissection of complex traits. *Genetics*. https://doi.org/10.1534/genetics.112.138537

King, M.-C., & Wilson, A. C. (1975). Evolution ay Two Levels in humans and Chimpanzees. *Science*, *188*(4184), 107–118. Retrieved from http://www.jstor.org/stable/1739875

Kono, N., Nakamura, H., Ito, Y., Tomita, M., & Arakawa, K. (2016). Evaluation of the impact of RNA preservation methods of spiders for de novo transcriptome assembly. *Molecular Ecology Resources*, *16*(3), 662–672. https://doi.org/10.1111/1755-0998.12485

Krueger, F. (n.d.). ASAP-Allele-sp[ecific alignment pipeline. Retrieved from https://www.bioinformatics.babraham.ac.uk/projects/ASAP/

Krupp, J. J., & Levine, J. D. (2010). Dissection of oenocytes from adult Drosophila melanogaster. *Journal of Visualized Experiments : JoVE*, (41), 2242. https://doi.org/10.3791/2242

Lazzaro, B. P., Sceurman, B. K., Clark, A. G., & Sciences, B. (2004). Genetic Basis of Natural Variation in D . melanogaster. *Science*, *303*(March), 1873–1877.

Lazzaro, B. P., Sackton, T. B., & Clark, A. G. (2006). Genetic Variation in Drosophila melanogaster Resistance to Infection : *Genetics*, *1554*(November), 1539–1554. https://doi.org/10.1534/genetics.105.054593

Lemaitre, B., & Hoffmann, J. (2007). The Host Defense of Drosophila melanogaster. https://doi.org/10.1146/annurev.immunol.25.022106.141615

Lu, Y., Su, F., Li, Q., Zhang, J., Li, Y., Tang, T., … Yu, X. Q. (2020). Pattern recognition receptors in Drosophila immune responses. *Developmental and Comparative Immunology*. https://doi.org/10.1016/j.dci.2019.103468

McCarthy, D. J., Chen, Y., & Smyth, G. K. (2012). Differential expression analysis of multifactor RNA-Seq experiments with respect to biological variation. *Nucleic Acids Research*, *40*(10), 4288–4297. https://doi.org/10.1093/nar/gks042

McLaren, W., Gil, L., Hunt, S. E., Riat, H. S., Ritchie, G. R. S., Thormann, A., … Cunningham, F. (2016). The Ensembl Variant Effect Predictor. *Genome Biology*, *17*(1), 1–14. https://doi.org/10.1186/s13059-016-0974-4

McManus, C. J., Coolon, J. D., Duff, M. O., Eipper-Mains, J., Graveley, B. R., & Wittkopp, P. J. (2010). Regulatory divergence in Drosophila revealed by mRNA-seq. *Genome Research*, *20*(6), 816–825. https://doi.org/10.1101/gr.102491.109

Murad, R., Macias-Muñoz, A., Wong, A., Ma, X., & Mortazavi, A. (2019). Integrative analysis of *Hydra* head regeneration reveals activation of distal enhancer-like elements. *BioRxiv*, 544049. https://doi.org/10.1101/544049

Obbard, D. J., Welch, J. J., Kim, K. W., & Jiggins, F. M. (2009). Quantifying adaptive evolution in the Drosophila immune system. *PLoS Genetics*, *5*(10). https://doi.org/10.1371/journal.pgen.1000698

Osada, N., Miyagi, R., & Takahashi, A. (2017). Cis - and trans -regulatory effects on gene expression in a natural population of Drosophila melanogaster. *Genetics*, *XXX*(June), 1–12. https://doi.org/10.1534/genetics.XXX.XXXXXX

Patro, R., Duggal, G., Love, M. I., Irizarry, R. A., & Kingsford, C. (2017). Salmon provides fast and bias-aware quantification of transcript expression. *Nature Methods*, *14*(4), 417–419. https://doi.org/10.1038/nmeth.4197

Pearson, W. R. (2013). Selecting the Right EMR Vendor Selecting the Right EMR Vendor. *Curr Protoc Bioinformatics*, (43), 3.5.1–3.5.9. https://doi.org/10.1002/0471250953.bi0305s43.Selecting

Pérez-Portela, R., & Riesgo, A. (2013). Optimizing preservation protocols to extract high-quality RNA from different tissues of echinoderms for next-generation sequencing. *Molecular Ecology Resources*, *13*(5), 884–889. https://doi.org/10.1111/1755-0998.12122

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

Sackton, T. B., Lazzaro, B. P., & Clark, A. G. (2010). Genotype and Gene Expression Associations with Immune Function in Drosophila. *PLoS Genetics*, *6*(1). https://doi.org/10.1371/journal.pgen.1000797

Sackton, T. B., Lazzaro, B. P., Schlenke, T. A., Evans, J. D., Hultmark, D., & Clark, A. G. (2007). Dynamic evolution of the innate immune system in Drosophila. *Nature Genetics*, *39*(12), 1461–1468. https://doi.org/10.1038/ng.2007.60

Salinas, F., De Boer, C. G., Abarca, V., García, V., Cuevas, M., Araos, S., … Cubillos, F. A. (2016). Natural variation in non-coding regions underlying phenotypic diversity in budding yeast. *Scientific Reports*, *6*, 1–13. https://doi.org/10.1038/srep21849

Schaefke, B., Emerson, J. J., Wang, T. Y., Lu, M. Y. J., Hsieh, L. C., & Li, W. H. (2013). Inheritance of gene expression level and selective constraints on trans-and cis-regulatory changes in yeast. *Molecular Biology and Evolution*, *30*(9), 2121–2133. https://doi.org/10.1093/molbev/mst114

Scheiermann, C., Kunisaki, Y., & Frenette, P. S. (2013). Circadian control of the immune system. *Nature Reviews Immunology*, *13*(3), 190–198. https://doi.org/10.1038/nri3386

Senger, K., Armstrong, G. W., Rowell, W. J., Kwan, J. M., Markstein, M., & Levine, M. (2004). Immunity Regulatory DNAs Share Common Organizational Features in Drosophila, *13*, 19–32.

Serra, L., Chang, D., Macchietto, M., Williams, K., Murad, R., Lu, D., … Mortazavi, A. (2018). Adapting the Smart-seq2 Protocol for Robust Single Worm RNA-seq. *Bio-Protocol*, *8*(4), 1–16. https://doi.org/10.21769/bioprotoc.2729

Shazman, Shula ; Lee, Hunjoong ; Socol, Yakov ; Mann, Richard ; Honig, B. (2014). OnTheFly: a database of Drosophila melanogaster transcription factors and their binding sites (NAR 42, D167-171). *Nucleic Acids Research*, *42*(D167–D171). https://doi.org/doi:10.1093/nar/gkt1165

Signor, S. A., & Nuzhdin, S. V. (2018). The Evolution of Gene Expression in cis and trans. *Trends in Genetics*, *34*(7), 532–544. https://doi.org/10.1016/j.tig.2018.03.007

Soneson, C., Love, M. I., & Robinson, M. D. (2015). Differential analyses for RNA-seq: transcript-level estimates improve gene-level inferences. *F1000Research*, *4*(1521). https://doi.org/10.12688/f1000research.7563.1

Takada, Y., Miyagi, R., Takahashi, A., Endo, T., & Osada, N. (2017). A Generalized Linear Model for Decomposing Cis -regulatory, Parent-of-Origin, and Maternal Effects on Allele-Specific Gene Expression. *G3 (Bethesda, Md.) Genes|Genomes|Genetics Genes|Genomes|Genetics*, *7*(7), 2227–2234. https://doi.org/10.1534/g3.117.042895

Tanji, T., Yun, E. Y., & Ip, Y. T. (2010). Heterodimers of NF-κB transcription factors DIF and Relish regulate antimicrobial peptide genes in Drosophila. *Proceedings of the National Academy of Sciences of the United States of America*. https://doi.org/10.1073/pnas.1009473107

Team, R. C. (2019). R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from https://www.r-project.org/

Therneau, T. M. (2020). A Package for Survival Analysis in R. Retrieved from https://cran.r-project.org/package=survival

Therneau, T. M. (2000). *Modeling Survival Data: Extending the {C}ox Model*. New York: Springer. Retrieved from https://cran.r-project.org/package=survival

Thurmond, J., Goodman, J. L., Strelets, V. B., Attrill, H., Gramates, L. S., Marygold, S. J., … Baker, P. (2019). FlyBase 2.0: The next generation. *Nucleic Acids Research*, *47*(D1), D759–D765. https://doi.org/10.1093/nar/gky1003

Troha, K., Im, J. H., Revah, J., Lazzaro, B. P., & Buchon, N. (2018). *Comparative transcriptomics reveals CrebA as a novel regulator of infection tolerance in D. melanogaster*. *PLoS Pathogens* (Vol. 14). https://doi.org/10.1371/journal.ppat.1006847

Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York.

Wittkopp, P. J., Haerum, B. K., & Clark, A. G. (2004). Evolutionary changes in cis and trans gene regulation. *Nature Publishing Group*, *430*(1), 85–88.

Wittkopp, P. J., Haerum, B. K., & Clark, A. G. (2008). Regulatory changes underlying expression differences within and between Drosophila species. *Nature Genetics*, *40*(3), 346–350. https://doi.org/10.1038/ng.77

Zaidman-Rémy, A., Hervé, M., Poidevin, M., Pili-Floury, S., Kim, M. S., Blanot, D., … Lemaitre, B. (2006). The Drosophila Amidase PGRP-LB Modulates the Immune Response to Bacterial Infection. *Immunity*, *24*(4), 463–473. https://doi.org/10.1016/j.immuni.2006.02.012

Zhang, X., & Emerson, J. J. (2019). Inferring Compensatory Evolution of cis- and trans-Regulatory Variation. *Trends in Genetics*, *35*(1), 1–3. https://doi.org/10.1016/j.tig.2018.11.003

2.10 Figures
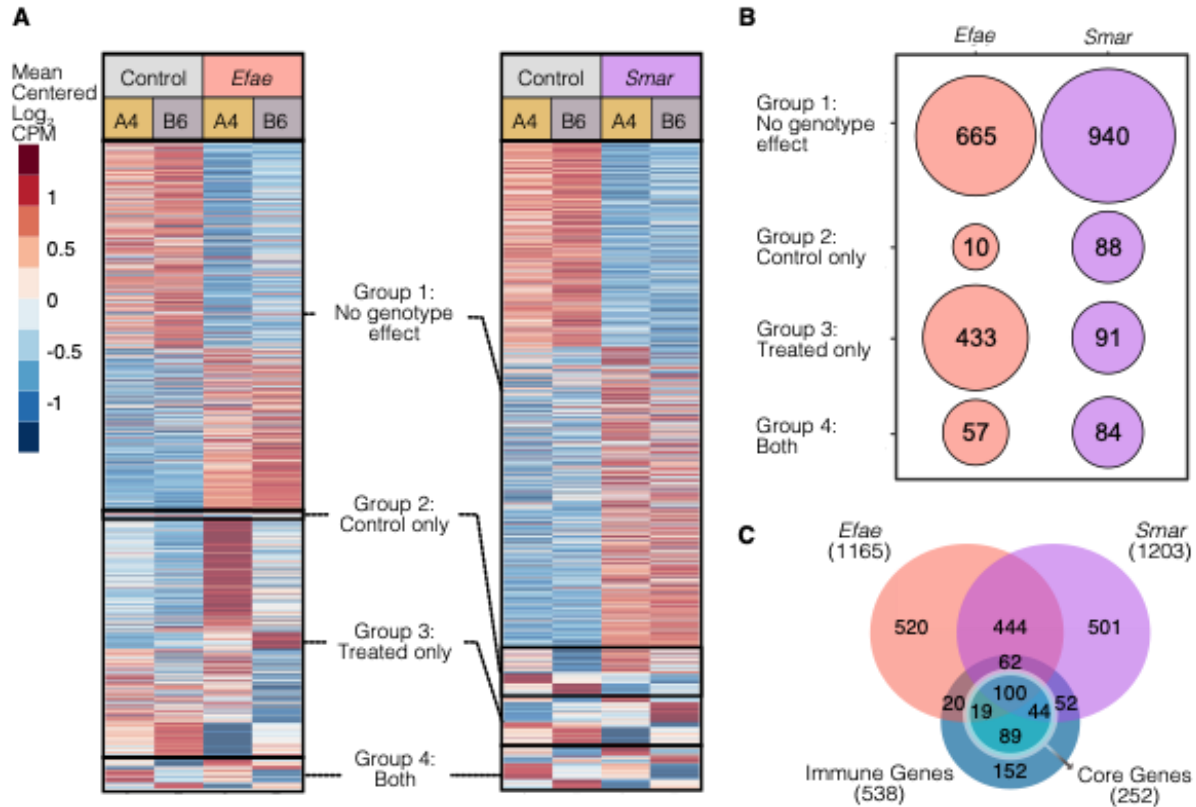
Figure 2.1. The A4 and B6 *D. melanogaster* **lines have variation in their response to Gram-positive** *E. faecalis* **infection and Gram-negative** *S. marcescens*
A) We measured expression in the fat bodies of the A4 and B6 lines infected with Gram-positive *Enterococcus faecalis* (*Efae*) or with Gram-negative *Serratia marcescens* (*Smar*), 3 hours post-infection. We found 1165 and 1205 differentially expressed in response to infection to *Efae* and *Smar* respectively, relative to control samples. Mean centered log$_2$ average CPM values for each condition are displayed. We categorized the infection responsive genes into four groups, based on their differential expression between the two fly genotypes: genes showing no genotype-specific expression (Group 1), genes showing genotype-specific expression only in the control condition (Group 2), genes showing genotype-specific expression only in the infected condition (Group 3) and genes showing genotype-specific expression in both control and infected conditions (Group 4). B) Among genes showing genotype effects, the majority of genes in *Efae* fell into the Group 2 classification, indicating a large amount of genotype-specific expression variation is revealed upon infection with *Efae*. Among *Smar*-responsive genes, roughly equal numbers show expression differences between the genotypes before (Group 1), after (Group 2), and both before and after infection (Group 3). C) We intersected the genes we identified as differentially expressed in response to infection and a list of previously published immune responsive genes. More than half of the verified immune genes were identified as differentially expressed in the abdominal fat body, with half of these immune genes being shared between conditions. Among these previously identified immune genes, core genes are differentially expressed across all infections. We detected

roughly 40% of the core set as differentially expressed in both our infection conditions, despite differences in the genetic background, tissue type, and time point used in our study versus previous work.

A

$\log_2 \left(\frac{B6_{F1}}{A4_{F1}} \right)$ (y-axis)

$\log_2 \left(\frac{B6_{Parents}}{A4_{Parents}} \right)$ (x-axis)

cis + trans — compensatory — *cis-only*

trans-only — *trans-only*

cis-only — compensatory — *cis + trans*

B Uninfected Control

C *Enterococcus faecalis*

D *Serratia marcescens*

E

Percent

cis only | *cis + trans* | compensatory | *trans*

Treatment
- Control
- *Efae*
- *Smar*

*p<3 x 10⁻³
Bonferonni corrected

F

Category	Control	*Efae*	*Smar*
Conserved	3802	2580	4106
cis-only	87	174	79
cis + trans	11	8	7
compensatory	37	6	16
trans-only	16	75	149

56

Figure 2.2. The relative contributions of *cis* and *trans* effects to expression divergence are condition specific.

A) Here we show a schematic of the expected locations for genes falling into four classifications of causes of expression divergence in plots that show the expression ratio of a gene in the parental lines (x-axis) against the allele expression ratio in the F1 hybrids (y-axis). B) In the uninfected control condition, of 4960 genes that could be detected in an allele-specific manner, 153 genes showed *cis* or *trans* signal. Of these 153 genes, most showed *cis*-acting effects. Panel F) displays the precise numbers of genes in each category. C) In response to *Efae* infection, expression divergence is driven predominantly by changes in *cis*. D) In response to *Smar* infection, expression divergence is dominated by changes in *trans*. E) We compared the fraction of genes categorized into each divergence class in the three conditions and found that the modes of expression divergence were condition-specific.

A

Category	Fat Body Detected	DE Infection	DE Immune	Non-DE Immune
Synonymous SNPs	74144	14194	1957	1844
Synonymous %	0.789	0.797	0.764	0.813
Nonsynonymous SNPs	19819	3610	604	424
Nonsynonymous %	0.211	0.203	0.236	0.187

B
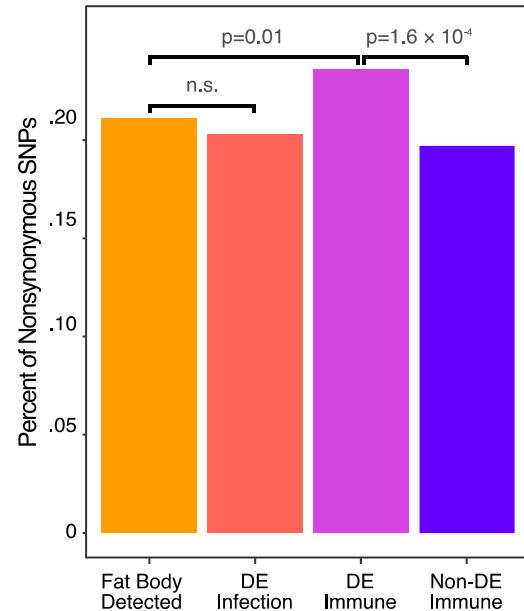
Figure 2.3: There is a greater proportion of non-synonymous SNPs in previously identified immune-responsive genes.
A) To look for the prevalence of non-synonymous SNPs in our genotypes and genes of interest, we defined four gene sets. Among genes detected in the fat body samples, we separated genes into those that were differentially expressed in response to either infection (DE infection) and those that were not (fat body detected). Within the fat body detected genes, we defined previously identified immune genes showing no differential expression in response to infection (Non-DE Immune), and among the DE infection genes, we refined the gene list to include previously identified immune genes (DE immune). The numbers indicate the total number of SNPs found in each gene set and the percentages of synonymous and non-synonymous SNPs. B) DE immune genes have a higher proportion of non-synonymous SNPs than the fat body expressed genes, which suggests they may carry function-altering SNPs at a higher rate than the fat body expressed genes. *p*-values are Bonferroni-corrected Chi-square test with the proportion of non-synonymous SNPS relative to the fat body expressed gene set.
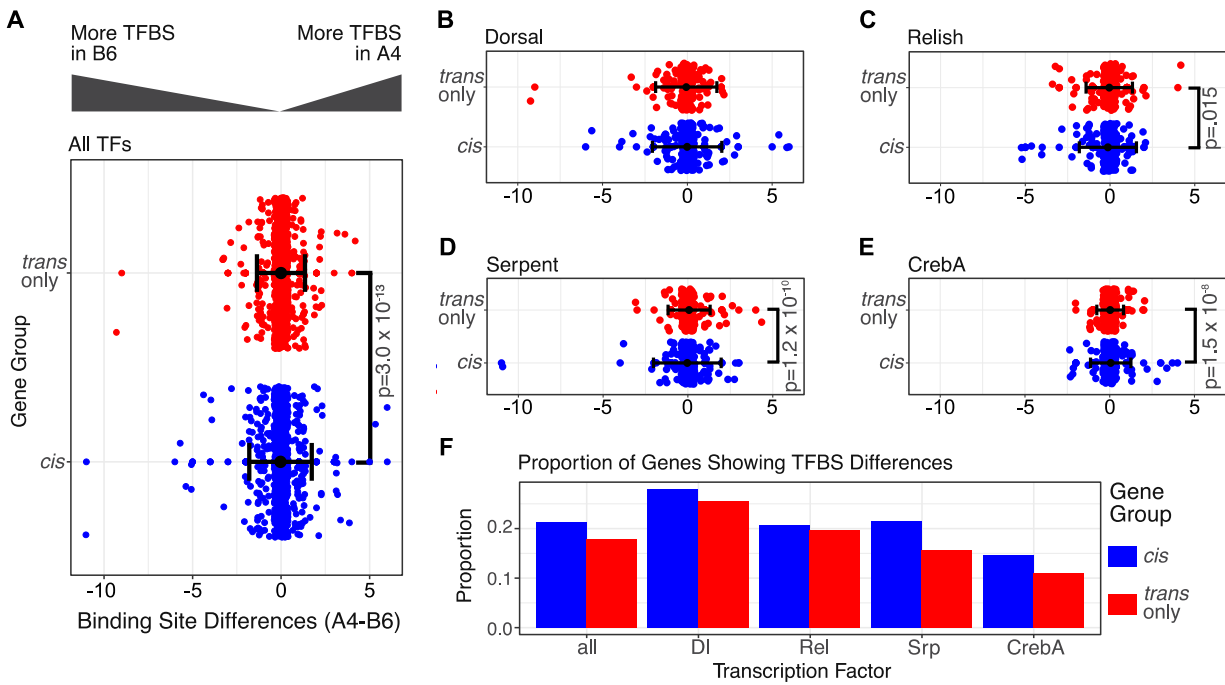
Figure 2.4: There are greater differences in TFBS in *cis* affected genes than *trans* affected genes.
A) We identified TFBS for 4 immune responsive transcription factors Dl, Rel, Srp and CrebA in 1kb region upstream of 219 *cis*-affected genes and 199 *trans*-affected genes. Differences in total TFBS numbers between genotypes were calculated for each gene and plotted. We find that variance in the distribution of these differences is significantly greater in genes showing *cis* effects (F-test to compare distribution variances, Bonferoni corrected). B) For Dl TFBS, there was not a a significant difference in the width of the TFBS distribution between genes showing *cis* effects and *trans* effects. C-E) For Rel, Srperpent and CrebA TFBS, there was a broder distribution of TFBS differences in genes with *cis* effects than genes with *trans* effects. F) A larger proportion of genes showing *cis* effects had a difference in total TFBS than genes showing *trans* effects, though the differences in these proportions were not significant.

59

FB Gene ID	Gene Symbol	Type	Log$_2$ Fold Change (B6/A4)	More Highly Expressed in:	A4 Average CPM	B6 Average CPM	Immune involvement
FBgn0029822	*CG12236*	TF	-3.13	A4	28	2.7	Unclear
FBgn0039075	*CG4393*	Signaling	2.06	B6	8.8	39	Unclear
FBgn0038978	*tHMG1*	TF	3.06	B6	7	57	Unclear
FBgn0287768	*esg*	TF	-3.27	A4	11	1	Unclear
FBgn0039932	*fuss*	TF	2.50	B6	1.1	6.3	Wound healing
FBgn0250732	*gfzf*	TF	8.24	B6	0	2	Unclear
FBgn0000448	*Hr3*	TF	-4.12	A4	10	0.7	Unclear
FBgn0016675	*Lectin-galC1*	Detection	2.72	B6	79	570	Binding and agglutination
FBgn0035993	*Nf-YA*	TF	-10.16	A4	10	0	Unclear
FBgn0028542	*NimB4*	Detection	-1.08	A4	40	22	Phagocytosis and microbial pattern recognition
FBgn0259896	*NimC1*	Detection	-3.06	A4	97	27	Phagocytosis and microbial

							pattern recognition
FBgn0003130	*Poxn*	TF	-4.38	A4	1.2	0.07	Unclear
FBgn0014033	*Sr-CI*	Detection	-2.39	A4	84	38.6	Phagocytosis and microbial pattern recognition
FBgn0004606	*zfh1*	Signaling / TF	1.77	B6	50	17	Haematopoiesis
FBgn0031973	*Spn28Dc*	Signaling	2.56	B6	9.4	2.4	Negative regulator of melanization
FBgn0037906	*PGRP-LB*	Detection	4.536	B6	111.9	75.2	Negative regulator of IMD pathway
FBgn0043576	*PGRP-SC1a*	Detection	-5.57	A4	4.3	6.8	Negative regulator of IMD pathway
FBgn0033327	*PGRP-SC1b*	Detection	-5.24	A4	3.9	0.2	Negative regulator of IMD pathway
FBgn0043575	*PGRP-SC2*	Detection	-4.02	A4	15	1.3	Negative regulator of IMD pathway
FBgn0035806	*PGRP-SD*	Detection	4.25	B6	97.6	19.2	Positive regulator of IMD pathway
FBgn0039102	*SPE*	Signaling	2.41	B6	491.9	255.2	Positive regulator of Toll pathway
FBgn00	*spz*	Signaling	0.68	B6	72.8	45.5	Positive

							regulator of Toll pathway
03495							

Table 2.1: Transcription factors and immune genes identified as potential sources of trans effects in infection.
List of genes potentially driving the *trans* effects for *Efae* and *Smar* infection. Candidate genes were identified by finding genes that had genotype-specific expression differences in the uninfected control conditions and that were classified as either a transcription factor, immune signaling gene, or immune detection gene.

Category	Differential gene expression in parents	F1 allelic imbalance	*Trans* test
cis only	True	True	False
trans only	True	False	True
cis + trans	True	True	True
Compensatory	False	True	True
Conserved	False	False	False

Table 2.2: Logic for *cis* and *trans* effect gene categories.
Genes were designated into categories based on the results of three statistical tests. Here, 'True' indicates a significant test result at FDR<.05 and 'False' indicates an insignificant test result.

2.11 Supplemental Materials

Supplemental Methods

Survival and Bacterial load tracking of A4 and B6 lines.

To more effectively ascertain differences in survival, we used lower doses of bacteria for the survival analysis than for the RNA-seq analysis (5,000 CFUs of *E. faecalis* or 1,000 CFUs of *S. marcescens*). Once per day following infection, the survival status of the flies was recorded and the bacterial load was measured via dilution plating of a live flies as in (Khalil et al., 2015; Supplemental Figure S2.1). Kaplan-Meier estimates of survival were calculated using the survival 3.2-3 package in R (Therneau et al., 2000; Therneau et al.,2020), and log-rank tests and plotting were performed using the survminer 0.4.4 package (Kassambara and Kosinski 2019).

Filtering low confidence annotations from A4 and B6 transcriptome annotations.

To assess the quality of our annotations and remove genes with poor annotations, genomic sequencing reads for A4 and B6 from the DSPR website were downloaded and aligned to our transcriptome files using Salmon 0.12.0 aligner (Thurmond et al., 2019). We hypothesized that well-annotated genes would show similar coverage of genomic reads in both the A4 and B6 transcriptomes. We then filtered genes using two methods for outlier calling: a Poisson distribution-based method and a negative binomial generalized linear model (GLM) method, similar to that used for differential gene expression in RNA-seq experiments. For the Poisson method, we fitted a Poisson distribution to gene count data for the A4 and B6 transcriptomes separately, using the fitdistributionplus 1.0-14 package in R and called outlier genes using three thresholds of increasing stringency p = 0.001, 0.01 and 0.025 (Delignette-Muller & Dutang 2015). For the GLM-based approach, we looked for gene counts that were significantly different between the A4 and B6 transcriptomes and filtered genes using FDR thresholds of 0.01, 0.05 and 0.09. As our threshold for significance became more stringent, we filtered out an increasing number of genes but the differences between the final filtered sets show about a 3% difference in terms of total genes and less than 1% difference in genes shown to be differentially expressed (Supplemental Figure S2.2). Genes found not to be outliers in either the Poisson or GLM method were then combined into gene sets based on the stringency of filtering. These gene sets were then used to quantify *cis* and *trans* effects for all three conditions. We found that the stringency of filtering did not significantly impact the total number or proportions of *cis* and *trans* effects between conditions. For the allele-specific expression analysis presented in Figure 2.3, we used a set of genes filtered using a combination of both methods at medium stringency.

Assessing accuracy of ASAP allele calling using X Chromosome reads.
To verify the accuracy of our quantification allelic expression in F1 hybrids, we used the RNA-seq data from the A4 and B6 parental lines and data from the F1 hybrids (A4♂x B6♀) and reciprocal crosses (B6♂x A4♀), in the control, *Efae*-infected, and *Smar*-infected conditions. Since we are using males, if our allele-specific expression analysis is correct, none of the X Chromosome reads should map to the paternal genotype. Using the published A4 and B6 genomes and the Allele-Specific Alignment Pipeline (ASAP) (Krueger, https://www.bioinformatics.babraham.ac.uk/projects/ASAP/), we quantified the fraction of X Chromosome reads that incorrectly map to the paternal genotype. On average, samples had 0.5% mis-assigned reads (standard deviation = 3%), with the highest fraction being 1.2% (Supplemental Table S1). The consistent, low level of mis-assigned reads verifies our ability to accurately quantify allelic expression. Given that all the flies are male, any reads aligning to the paternal X Chromosome can definitively be classified as mis-assigned.
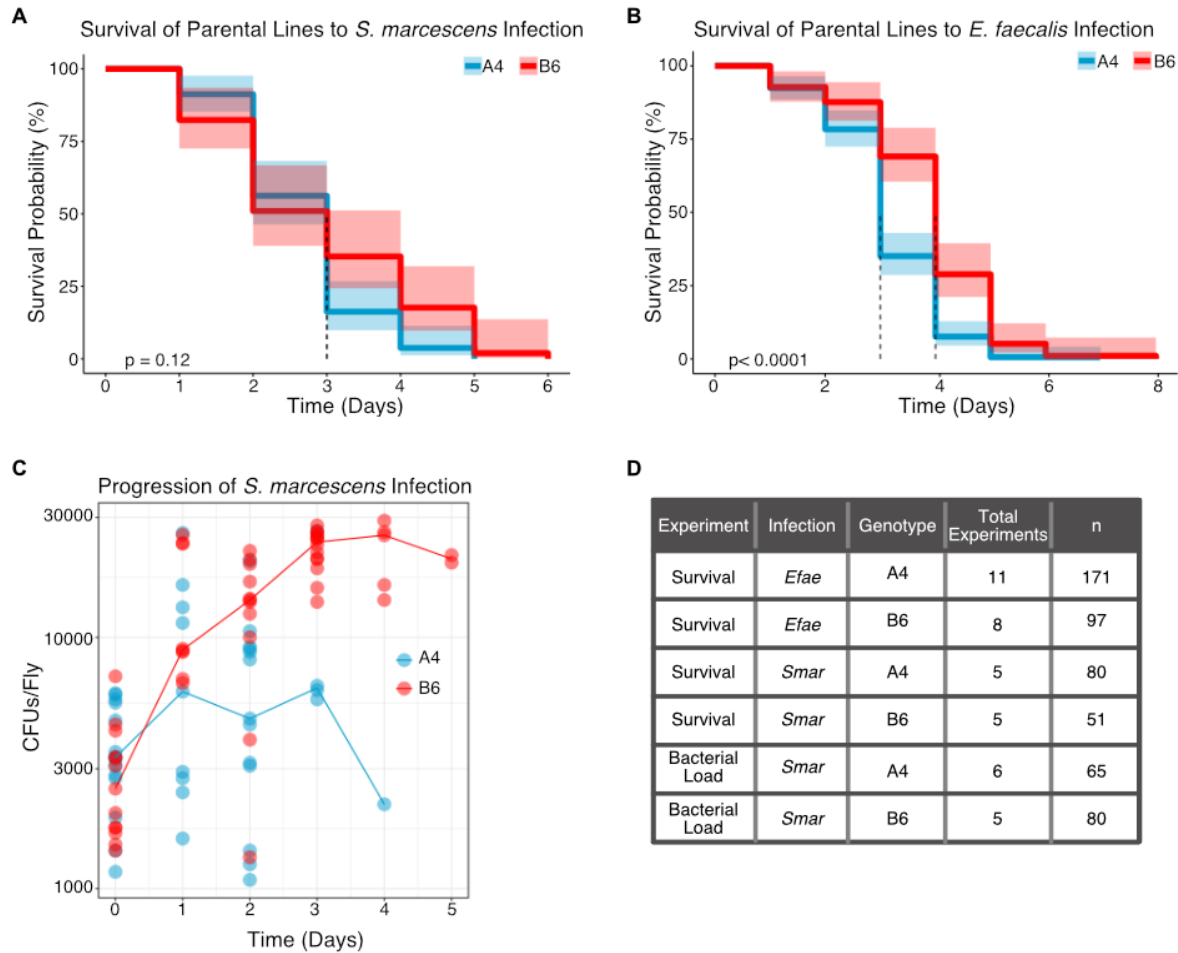
A Survival of Parental Lines to *S. marcescens* Infection

B Survival of Parental Lines to *E. faecalis* Infection

C Progression of *S. marcescens* Infection

D

Experiment	Infection	Genotype	Total Experiments	n
Survival	*Efae*	A4	11	171
Survival	*Efae*	B6	8	97
Survival	*Smar*	A4	5	80
Survival	*Smar*	B6	5	51
Bacterial Load	*Smar*	A4	6	65
Bacterial Load	*Smar*	B6	5	80

Figure S2.1: A4 and B6 lines show differences in survival in response to Gram-positive but not Gram-negative infection.

A) Survival curves and confidence intervals for flies infected with an average 1000 CFUs of *S. marcescens*, observed once per day. P-value was calculated using a log-rank test. B6 flies survive *Efae* infection for longer than the A4 flies. B) Survival curves and confidence intervals for flies infected with approximately 5000 CFUs of *E. faecalis*, observed once per day. P-value was calculated using a log-rank test. There is no significant difference in infection survival between the two genotypes. C) Bacterial load of A4 and B6 lines in response to *S. marcescens* infection, assessed by dilution plating of homogenized infected flies. Points represent a single animal's bacterial load measurement (an average of three technical replicates per animal), and solid lines indicate the median values of bacterial load for each day. Though the flies do not show a significant difference in survival, it appears that A4 shows greater resistance to *Smar*, while B6 shows greater tolerance of the infection. D) Table showing sample sizes for the results depicted in this figure.
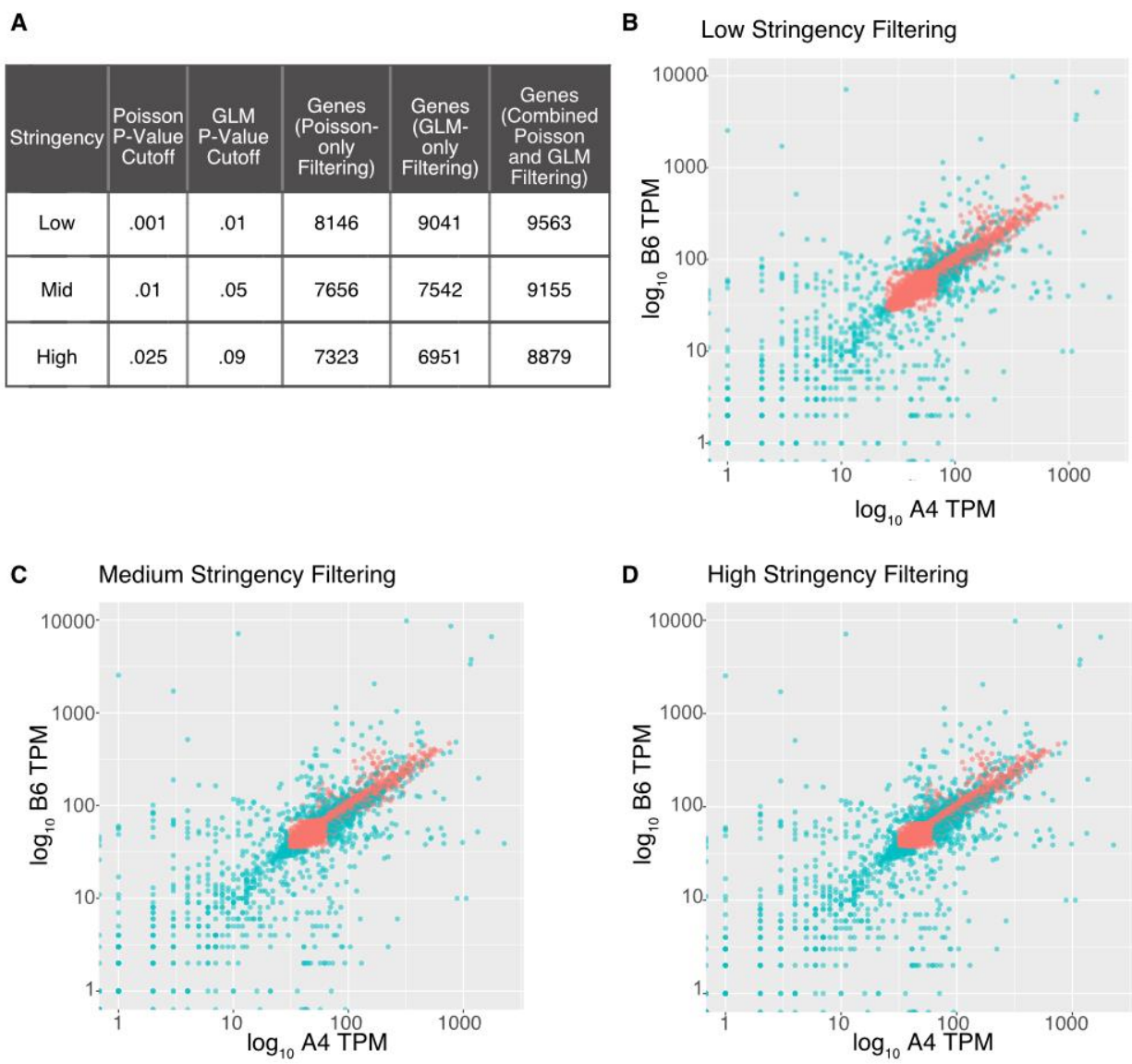
A

Stringency	Poisson P-Value Cutoff	GLM P-Value Cutoff	Genes (Poisson-only Filtering)	Genes (GLM-only Filtering)	Genes (Combined Poisson and GLM Filtering)
Low	.001	.01	8146	9041	9563
Mid	.01	.05	7656	7542	9155
High	.025	.09	7323	6951	8879

B Low Stringency Filtering

C Medium Stringency Filtering

D High Stringency Filtering

Figure S2.2: Differences in assembly quality minimally affect *cis* and *trans* effects in downstream analysis.
To quantify the effects of assembly quality of on *cis* and *trans* effects, we filtered out genes with poor annotations at increasingly restrictive thresholds and quantified differences in *cis* and *trans* effects (Supplemental Table S2). We identified potentially problematic genes by aligning A4 and B6 genomic reads to their respective transcriptomes. We posited that each gene of the lifted over transcriptome should receive roughly the same amount of coverage once normalized for gene length and that genes deviating from this coverage were poorly annotated. We used two methods for calling outlier genes: a Poisson distribution-based method and a GLM based method (see Methods for details). A) Here we report the non-outlier (retained) gene numbers for different methods and degrees of stringency. The gene numbers do not decrease rapidly with increasing stringency. B-D) These graphs plot the gene counts in transcripts per million (TPM) using the A4

and B6 genomic reads mapped to their respective transcriptomes. Outlier genes are shown in teal and retained genes are shown in pink. The quantification of *cis* and *trans* effects for these different gene sets are shown in Supplemental Table S2.
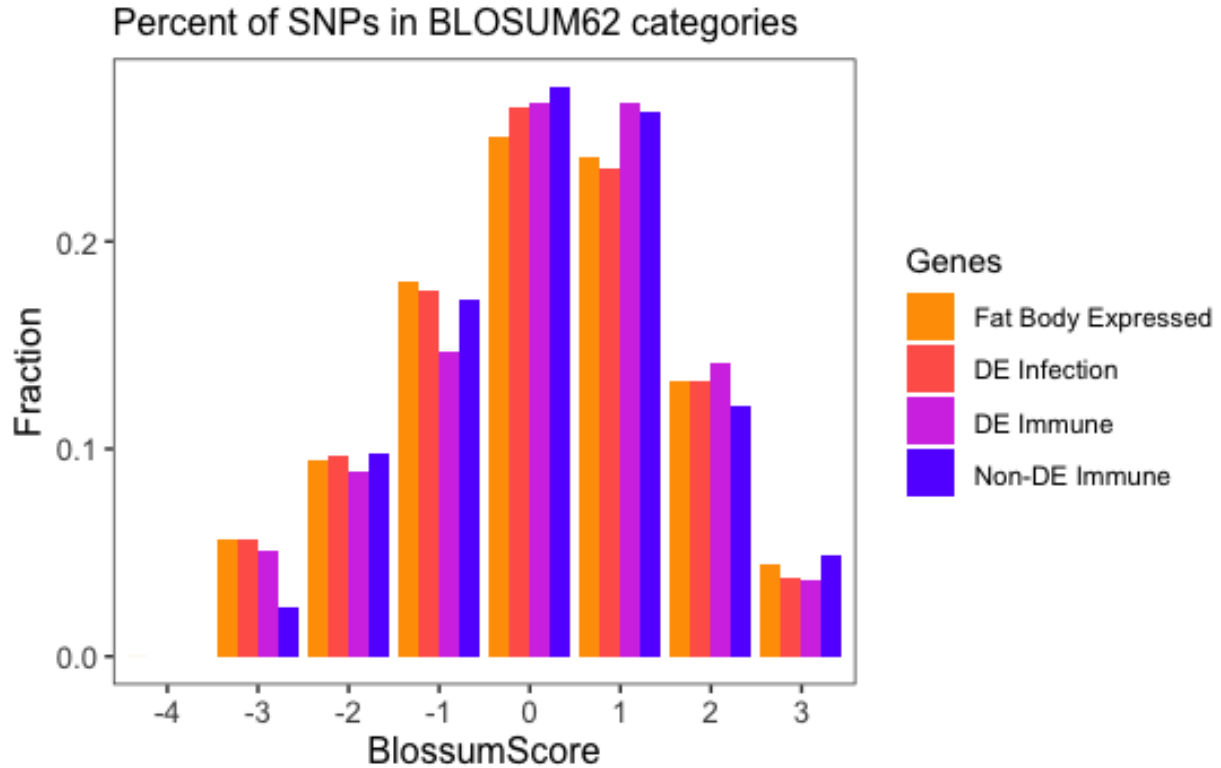

Figure S2.3: Most nonsynonymous mutations have non-negative BLOSUM62 scores.
As a coarse-grained approximation of the effects of non-synonymous changes on protein function, we analyzed the distribution of BLOSUM62 scores for the four gene sets described. The BLOSUM62 score is a homology-based metric that describes the likelihood of a particular residue change, positive numbers indicate frequently observed changes, while negative numbers indicate rare amino acid substitution (Pearson 2013). For all gene sets, non-negative scores dominate, with 67% for *fat body detected*, 67% for *DE infected*, and 71% for *DE immune,* 55% Non-DE immune. This suggests that there are some nonsynonymous mutations that may alter protein function, but the fraction of these disruptive mutations does not significantly differ between gene sets.
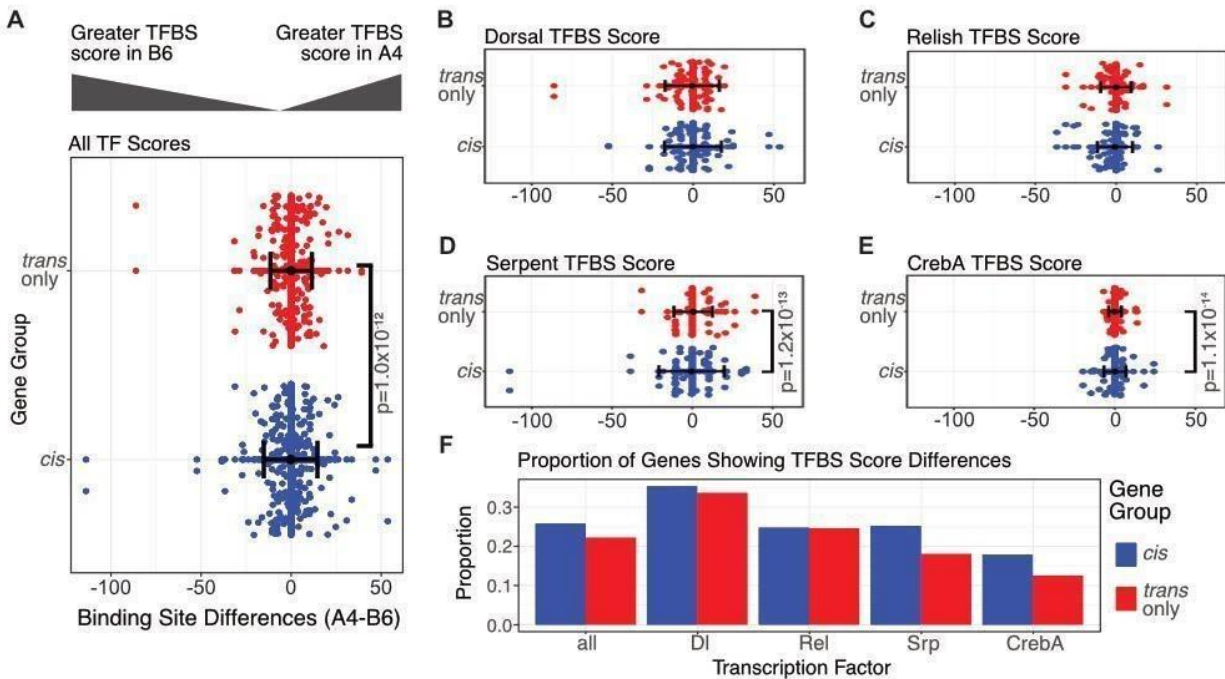
Figure S2.4: There are greater differences in TFBS score in *cis* affected genes than *trans* affected genes.
A) We quantified TFBS score for four immune responsive transcription factors: Dl, Rel, Srp and CrebA in 1kb region upstream of 219 *cis* affected genes and 199 *trans* affected genes. The differences in total TFBS score in the B6 and A4 upstream regions were calculated for each gene. We find that variance in the distribution of these differences is greater in genes showing *cis* effects (F-test to compare distribution variances, Bonferoni corrected). We then looked at the distribution of these differences for each of the four transcription factors separately. B-C) The variances in the score difference distributions for Dl and Rel were not significantly different between genes showing *cis* effects and *trans* effects. D-E) The variances of the score differnt distributions for Srp and CrebA are significantly differet between genes showing *cis* effects and *trans* effects. F) A higher fraction of genes showing *cis* effects had differences in total TFBS score than genes showing *trans* effects, though these fractions were not significantly different.
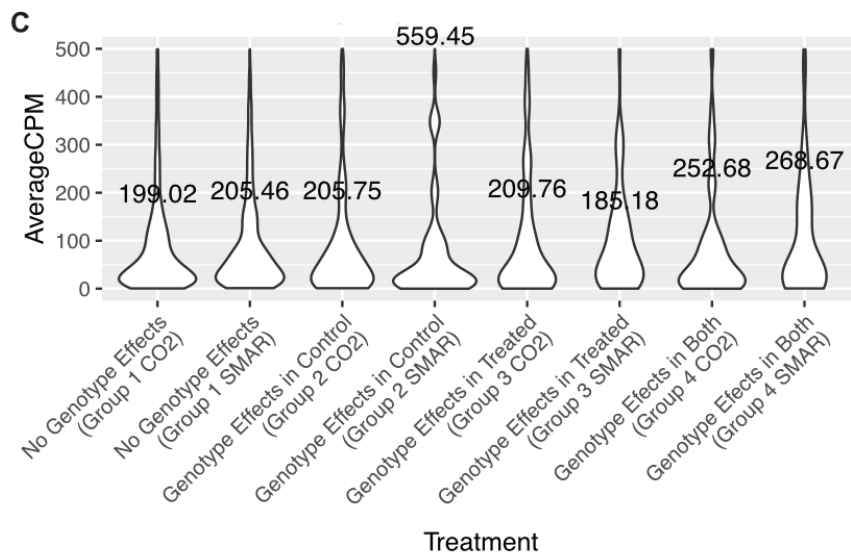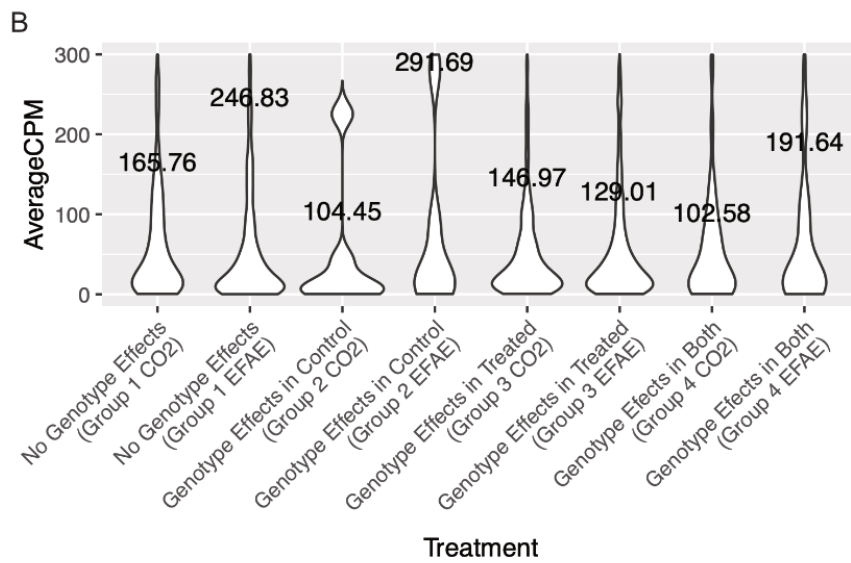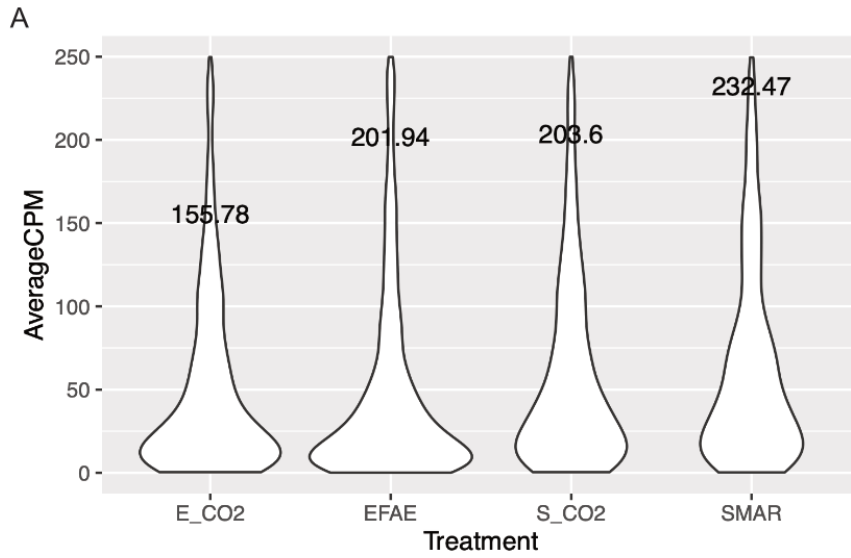
68

Figure S2.5: Average CPM for genes across different identified gene groups.
A) To determine if absolute expression between immune stimulated and control samples may be biasing our ability to detect genotype-specific effects, we looked at average CPM values for differentially expressed genes in both infection conditions. Specifically, we wanted to ensure that we were not finding more expression divergence effects in the infected samples because genes have higher expression in response to infection than in the control condition. The expression levels of differentially expressed genes in *Efae* (1165 genes) and *Smar* (1203 genes) conditions were compared to the corresponding genes in the control samples, *E_CO2* and *S_CO2* respectively (Supplemental Code, Script1_fig1). We performed two-sided Wilcoxon rank-sum tests to compare average CPM values from immune stimulated samples to average cpm values of the corresponding genes in the control conditions. Using a p-value threshold of *p*=0.05 we found no significant differences in average CPM of infection-responsive genes between treated samples and untreated controls (two-sample Wilcoxon rank sum test, Bonferroni corrected). B) Here we show average CPM values for each of the four gene groups showing different genotype effects in response to *Efae* infection (as shown in Figure 2.1). We do not observe a significantly different average CPM between treated and control conditions (two-sample Wilcoxon rank sum test, Bonferroni corrected). C) Here we show average CPM values for the four gene groups showing different genotype effects in response to *Smar* infections (as shown in Figure 2.1). No group shows significance in average CPM value between the treated and control samples (two sample Wilcoxon rank sum test, Bonferroni corrected).

Sample	Treatment	Genotype ♂/♀	Total Reads	Mapped Reads	% Uniquely Aligned to A4	% Uniquely Aligned to B6	% Mis-assigned to X Chromosome
1	control	A4	26369673	24364366	9.4	0.1	0.3
2	control	A4	14870917	13878016	9.5	0.1	0.3
3	control	A4	18732323	17558251	9.4	0.1	0.3
4	control	A4	34580046	32442180	10.3	0.1	0.4
5	control	A4B6	41318671	19649962	5.4	6.4	0.6
6	control	A4B6	41205951	19378946	5.6	6.6	0.6
7	control	A4B6	53666239	50178799	4.6	5.5	0.8
8	control	A4B6	82417525	65605513	5.7	6.4	0.6
9	control	B6	17980721	16879587	0.4	10.2	0.6
10	control	B6	19997738	18798790	0.4	8.7	0.8
11	control	B6	19129651	17946593	0.4	10.4	0.8
12	control	B6	24984658	23547941	0.3	9.1	1.0
13	control	B6A4	53543764	10893030	6.0	4.5	0.3
14	control	B6A4	47079732	24895491	6.7	5.0	0.3
15	control	B6A4	47509119	21329979	6.3	4.7	0.3
16	control	B6A4	49562943	46726476	6.0	4.6	0.3
17	*Efae*	A4	11521847	10597039	10.9	0.0	0.4
18	*Efae*	A4	26211400	24598530	12.2	0.1	0.4
19	*Efae*	A4	16272150	15204121	12.0	0.0	0.3
20	*Efae*	A4	24759445	23361494	11.0	0.1	0.3
21	*Efae*	A4B6	36234287	33302637	5.4	6.0	0.9
22	*Efae*	A4B6	54770680	51649242	6.0	6.7	0.5
23	*Efae*	A4B6	37724992	35152256	5.5	6.0	0.7
24	*Efae*	A4B6	52373459	49185996	7.4	8.0	0.4

25	*Efae*	B6	20269632	18651459	0.2	10.4	1.1
26	*Efae*	B6	22075327	20668129	0.3	11.9	0.5
27	*Efae*	B6	28118298	26565158	0.3	9.2	1.2
28	*Efae*	B6	28488360	26831112	0.3	12.7	0.6
29	*Efae*	B6A4	43346696	39878989	5.9	4.8	0.4
30	*Efae*	B6A4	50841666	47062579	6.6	5.2	0.3
31	*Efae*	B6A4	45437286	42562754	6.2	4.9	0.3
32	*Efae*	B6A4	62113778	57926378	6.6	5.2	0.3
33	*Smar*	A4	20932070	19569646	10.7	0.1	0.3
34	*Smar*	A4	22220731	20314035	7.7	0.1	0.3
35	*Smar*	A4	13096294	12215786	11.5	0.1	0.3
36	*Smar*	A4	19474264	17939316	10.8	0.0	0.3
37	*Smar*	A4B6	46702136	13363924	5.2	6.3	0.6
38	*Smar*	A4B6	42722535	19500222	6.1	6.7	0.5
39	*Smar*	A4B6	70196188	17932361	6.0	6.8	0.5
40	*Smar*	A4B6	49839957	22491147	5.7	6.5	0.5
41	*Smar*	A4B6	48904532	45834532	6.3	6.7	0.5
42	*Smar*	B6	9730094	9132793	0.4	10.4	0.5
43	*Smar*	B6	11254219	10569215	0.4	11.7	0.4
44	*Smar*	B6	16858117	15638969	0.2	10.3	0.8
45	*Smar*	B6A4	45215266	8235284	6.4	4.8	0.2
46	*Smar*	B6A4	70994061	11427623	6.0	4.7	0.3
47	*Smar*	B6A4	54223062	50351817	6.8	5.4	0.4

Table S2.1: Sample read numbers and alignment statistic.
Sample treatment categories are uninfected (control), *E. faecalis*-infected (*Efae*) and *S. marcescens*-infected (*Smar*). Genotype of samples are listed to indicate hybrid cross order: male genotype is listed first and female genotype second. We also show counts of 43bp paired end reads for each sample before and after alignment, percentages for A4 and B6 uniquely mapping reads, and percentages of mis-assigned X Chromosome reads (total mis-assigned X Chromosome reads over total X Chromosome genotype-specific reads).

Stringency	Treatment	*Cis*-only genes	*Trans*-only genes	*Cis* + *Trans* genes	Compensatory genes	Conserved genes	Undetermined genes
Poisson Med	Control	86	16	11	38	3808	1001
Combined Low	Control	89	16	10	46	3989	1046
Combined Med	Control	86	16	11	38	3808	1001
Combined High	Control	86	15	13	35	3688	962
Poisson Med	*Efae*	169	73	8	6	2586	1993
Combined Low	*Efae*	177	75	8	5	2734	2064
Combined Med	*Efae*	169	73	8	6	2586	1993
Combined High	*Efae*	165	77	8	8	2488	1929
Poisson Med	*Smar*	72	144	6	18	4107	496
Combined Low	*Smar*	77	153	6	15	4319	500
Combined Med	*Smar*	72	144	6	18	4107	496
Combined High	*Smar*	69	139	7	19	3965	485

Table S2.2: Increased stringency of problematic gene filtering minimally impacts overall number of *cis* and *trans* effects.
For each treatment, using sets of genes filtered at various levels of stringency, we quantified the number of genes falling into each of the *cis* and *trans* categories. We found that within treatment conditions the number and proportions of genes did not greatly differ as we increased the stringency of filtering.

Location	Allele	Gene Symbol	Gene	Feature	CDS position	Protein position	Amino acids	Codons	BLOSUM62
2L:4122351	T	*Sr-CI*	FBgn0014033	FBtr0346582	526	176	H/Y	Cac/Tac	2
2L:4122897	C	*Sr-CI*	FBgn0014033	FBtr0077467	947	316	S/T	aGc/aCc	1
2L:4123356	T	*Sr-CI*	FBgn0014033	FBtr0346582	1406	469	K/M	aAg/aTg	-1
2L:8005499	A	*Spn28Dc*	FBgn0031973	FBtr0079549	763	255	A/S	Gcg/Tcg	1
2L:8005523	G	*Spn28Dc*	FBgn0031973	FBtr0079549	739	247	I/L	Att/Ctt	2
2L:8005549	G	*Spn28Dc*	FBgn0031973	FBtr0079549	713	238	V/A	gTc/gCc	0
2L:8006451	A	*Spn28Dc*	FBgn0031973	FBtr0079549	682	228	T/S	Aca/Tca	1
2L:8006864	C	*Spn28Dc*	FBgn0031973	FBtr0079549	269	90	N/S	aAc/aGc	1
2L:13968919	C	*NimB4*	FBgn0028542	FBtr0080617	832	278	T/A	Acc/Gcc	0
2L:13974306	G	*NimC1*	FBgn0259896	FBtr0080615	1787	596	I/T	aTa/aCa	-1
2L:13974690	T	*NimC1*	FBgn0259896	FBtr0343644	1409	470	P/H	cCt/cAt	-2
2L:13974703	G	*NimC1*	FBgn0259896	FBtr0080615	1390	464	S/P	Tca/Cca	-1
2L:13975363	T	*NimC1*	FBgn0259896	FBtr0080615	730	244	V/M	Gtg/Atg	1
2L:13975380	T	*NimC1*	FBgn0259896	FBtr0080615	713	238	G/D	gGc/gAc	-1

2L:13975 515	G	*NimC 1*	FBgn0259 896	FBtr0080 615	578	193	V/A	gTc/g Cc	0
2L:13975 735	T	*NimC 1*	FBgn0259 896	FBtr0080 615	358	120	G/S	Ggc/A gc	0
2L:13976 157	C	*NimC 1*	FBgn0259 896	FBtr0343 644	40	14	S/A	Tca/G ca	1
2R:87170 36	G	*PGRP -SC2*	FBgn0043 575	FBtr0088 709	70	24	I/V	Atc/G tc	3
2R:10207 902	C	*Hr3*	FBgn0000 448	FBtr0330 609	1570	524	P/A	Cca/G ca	-1
2R:10232 873	T	*Hr3*	FBgn0000 448	FBtr0452 140	439	147	S/T	Tcg/A cg	1
2R:10237 018	G	*Hr3*	FBgn0000 448	FBtr0112 799	23	8	N/T	aAc/a Cc	0
3L:76517 52	T	*PGRP -SD*	FBgn0035 806	FBtr0076 807	548	183	S/F	tCc/tT c	-2
3L:94418 76	A	*Nf-YA*	FBgn0035 993	FBtr0076 504	17	6	S/I	aGc/a Tc	-2
3R:71486 18	C	*gfzf*	FBgn0250 732	FBtr0334 671	1480	494	H/D	Cac/G ac	-1
3R:71506 21	A	*gfzf*	FBgn0250 732	FBtr0091 512	10	4	P/S	Ccc/T cc	-1
3R:23378 558	T	*CG43 93*	FBgn0039 075	FBtr0339 617	3322	1108	L/I	Tta/At a	2
3R:23378 567	C	*CG43 93*	FBgn0039 075	FBtr0301 085	3313	1105	P/A	Cca/G ca	-1
3R:23378 571	A	*CG43 93*	FBgn0039 075	FBtr0339 616	3309	1103	E/D	gaG/g aT	2
3R:23379 640	G	*CG43 93*	FBgn0039 075	FBtr0339 617	2375	792	Q/P	cAa/c Ca	-1
3R:23379 641	T	*CG43 93*	FBgn0039 075	FBtr0301 085	2374	792	Q/K	Caa/A aa	1

3R:23381 986	T	*CG43 93*	FBgn0039 075	FBtr0301 085	548	183	T/N	aCc/a Ac	0
3R:27066 830	G	*spz*	FBgn0003 495	FBtr0085 137	199	67	T/P	Acc/C cc	-1
3R:30773 707	A	*zfh1*	FBgn0004 606	FBtr0331 180	232	78	Q/K	Cag/A ag	1
3R:30774 111	T	*zfh1*	FBgn0004 606	FBtr0085 701	386	129	K/M	aAg/a Tg	-1
3R:30774 123	T	*zfh1*	FBgn0004 606	FBtr0331 180	398	133	A/V	gCc/g Tc	0
3R:30774 165	C	*zfh1*	FBgn0004 606	FBtr0085 701	440	147	S/T	aGc/a Cc	1
3R:30785 831	T	*zfh1*	FBgn0004 606	FBtr0085 701	2861	954	A/V	gCg/g Tg	0

Table S2.3: Sequence changes in the list of candidate genes identified as being potential sources of *trans* effects.

Of the 46 SNPs falling into the coding regions of 22 genes identified as potential *trans* sources, 37 SNPs resulted in amino acid substitutions in 12 genes. Roughly 20% (8 SNPs) of these SNPs fell into the phagocytic gene *NimC1* alone. In all cases, the majority of affected protein domains were in unnamed domains. Of the 5 PGRPs, only 2 (*SC2* and *SD*) were found to carry mutations that resulted in coding region substitutions. These mutations fell into a transmembrane helix domain for *PGRP-SC2* but in an unknown domain for *PGRP-SD*. Additionally we found 5 missense mutations in *Spaetzle processing enzyme* and a single mutation in *Spaetzle,* though in both cases these mutations fell on unnamed protein domains. This underscores the large gap in our understanding of many of the domains important in the function of innate immunity genes and may serve as potential points of interest for future investigation.

Locati on	All lel e	Gen e Sy mbo l	Domains
2L:41 22351	T	*Sr-CI*	Gene3D:2.60.120.200,Pfam:PF00629,PROSITE_profiles:PS50060,PANT HER:PTHR23282,SMART:SM00137,Superfamily:SSF49899,CDD:cd062 63
2L:41 22897	C	*Sr-CI*	Gene3D:2.60.120.200,PANTHER:PTHR23282
2L:41 23356	T	*Sr-CI*	PANTHER:PTHR23282,MobiDB_lite:mobidb-lite,Low_complexity_(Seg):seg
2L:80 05499	A	*Spn 28D c*	Gene3D:3.30.497.10,Pfam:PF00079,PANTHER:PTHR11461,PANTHER: PTHR11461:SF281,SMART:SM00093,Superfamily:SSF56574,CDD:cd0 0172
2L:80 05523	G	*Spn 28D c*	Gene3D:3.30.497.10,Pfam:PF00079,PANTHER:PTHR11461,PANTHER: PTHR11461:SF281,SMART:SM00093,Superfamily:SSF56574,CDD:cd0 0172
2L:80 05549	G	*Spn 28D c*	Gene3D:3.30.497.10,Pfam:PF00079,PANTHER:PTHR11461,PANTHER: PTHR11461:SF281,SMART:SM00093,Superfamily:SSF56574,CDD:cd0 0172
2L:80 06451	A	*Spn 28D c*	Gene3D:3.30.497.10,Pfam:PF00079,PANTHER:PTHR11461,PANTHER: PTHR11461:SF281,SMART:SM00093,Superfamily:SSF56574,CDD:cd0 0172
2L:80 06864	C	*Spn 28D c*	PANTHER:PTHR11461,PANTHER:PTHR11461:SF281,Superfamily:SS F56574
2L:13 96891 9	C	*Nim B4*	PANTHER:PTHR24047,Gene3D:2.10.25.10,SMART:SM00181
2L:13 97430 6	G	*Nim C1*	PANTHER:PTHR24047,PANTHER:PTHR24047:SF29,Transmembrane_ helices:TMhelix
2L:13 97469 0	T	*Nim C1*	PANTHER:PTHR24047,PANTHER:PTHR24047:SF29

2L:13 97470 3	G	*Nim C1*	PANTHER:PTHR24047,PANTHER:PTHR24047:SF29
2L:13 97536 3	T	*Nim C1*	Gene3D:2.10.25.10,PANTHER:PTHR24047,PANTHER:PTHR24047:SF 29,SMART:SM00181,Superfamily:SSF57184
2L:13 97538 0	T	*Nim C1*	Gene3D:2.10.25.10,PANTHER:PTHR24047,PANTHER:PTHR24047:SF 29,SMART:SM00181,Superfamily:SSF57184
2L:13 97551 5	G	*Nim C1*	Gene3D:2.10.25.10,PANTHER:PTHR24047,PANTHER:PTHR24047:SF 29,SMART:SM00181
2L:13 97573 5	T	*Nim C1*	Gene3D:2.10.25.10,PROSITE_patterns:PS00022,PANTHER:PTHR24047 ,PANTHER:PTHR24047:SF29,SMART:SM00181
2L:13 97615 7	C	*Nim C1*	Cleavage_site_(Signalp):SignalP-noTM
2R:87 17036	G	*PG RP-SC2*	Gene3D:3.40.80.10,PIRSF:PIRSF037945,PANTHER:PTHR11022,SMA RT:SM00701,Superfamily:SSF55846,Transmembrane_helices:TMhelix
2R:10 20790 2	C	*Hr3*	Low_complexity_(Seg):seg
2R:10 23287 3	T	*Hr3*	-
2R:10 23701 8	G	*Hr3*	PANTHER:PTHR45805,PANTHER:PTHR45805:SF2
3L:76 51752	T	*PG RP-SD*	Gene3D:3.40.80.10,PIRSF:PIRSF037945,PANTHER:PTHR11022,PANT HER:PTHR11022:SF67,Superfamily:SSF55846
3L:94 41876	A	*Nf-YA*	-
3R:71 48618	C	*gfzf*	PANTHER:PTHR43969,PANTHER:PTHR43969:SF7

3R:71 50621	A	*gfzf*	PANTHER:PTHR43969,PANTHER:PTHR43969:SF7
3R:23 37855 8	T	*CG 439 3*	PANTHER:PTHR24174,PANTHER:PTHR24174:SF1,MobiDB_lite:mobi db-lite
3R:23 37856 7	C	*CG 439 3*	PANTHER:PTHR24174,PANTHER:PTHR24174:SF1,MobiDB_lite:mobi db-lite
3R:23 37857 1	A	*CG 439 3*	PANTHER:PTHR24174,PANTHER:PTHR24174:SF1,MobiDB_lite:mobi db-lite
3R:23 37964 0	G	*CG 439 3*	PANTHER:PTHR24174,PANTHER:PTHR24174:SF1,MobiDB_lite:mobi db-lite
3R:23 37964 1	T	*CG 439 3*	PANTHER:PTHR24174,PANTHER:PTHR24174:SF1,MobiDB_lite:mobi db-lite
3R:23 38198 6	T	*CG 439 3*	Gene3D:1.25.40.20,PROSITE_profiles:PS50297,PANTHER:PTHR24174 ,PANTHER:PTHR24174:SF1,Superfamily:SSF48403
3R:27 06683 0	G	*spz*	PANTHER:PTHR23199,PANTHER:PTHR23199:SF4
3R:30 77370 7	A	*zfh1*	Pfam:PF13912,PROSITE_patterns:PS00028,PROSITE_profiles:PS50157, PANTHER:PTHR24391,PANTHER:PTHR24391:SF27,SMART:SM0035 5
3R:30 77411 1	T	*zfh1*	PANTHER:PTHR24391,PANTHER:PTHR24391:SF27,MobiDB_lite:mo bidb-lite,MobiDB_lite:mobidb-lite
3R:30 77412 3	T	*zfh1*	PANTHER:PTHR24391,PANTHER:PTHR24391:SF27,MobiDB_lite:mo bidb-lite,MobiDB_lite:mobidb-lite,Low_complexity_(Seg):seg
3R:30 77416 5	C	*zfh1*	PANTHER:PTHR24391,PANTHER:PTHR24391:SF27,MobiDB_lite:mo bidb-lite,MobiDB_lite:mobidb-lite,Low_complexity_(Seg):seg

3R:30 78583 1	T	*zfh1*	PANTHER:PTHR24391,PANTHER:PTHR24391:SF27

Table S2.4: Domains associated with sequence changes in the list of candidate genes identified as being potential sources of *trans* effects.
List of protein domains affected by sequence changes in exonic regions from Supplemental Table S3.

Genotype	P-value threshold	Identified Rel sites	Matched Rel sites	Missing Rel sites	Identified Srp sites	Matched Srp sites	Missing Srp sites
A4	.001	26	10	1	13	7	0
A4	.0001	12	4	7	7	0	7
B6	.001	29	11	0	13	7	0
B6	.0001	12	4	7	7	0	7

Table S2.5: Determination of ap-value threshold for transcription factor binding site analysis
To determine an appropriate p-value threshold for identifying transcription factor binding sites (TFBS), we tested FIMO's ability to detect previously identified Rel and Srp binding sites in the upstream regions of four immune responsive genes. The Identified Rel sites and Identified Srp sites columns give the total identified binding sites for the selected TF by the FIMO utility. The Matched Rel sites and Matched Srp sites columns give the number of identified sites that match the previously described binding sites (Senger et al., 2004). The Missing Rel sites and Missing Srp sites columns give the number of previously identified sites that were not able to be detected by a given threshold. Based on this analysis, we used a p-value threshold of 0.001 for our TFBS analysis.

CHAPTER 3
Longitudinal monitoring of individual infection progression in *Drosophila melanogaster*

3.1 Abstract

The innate immune system is critical for host survival of infection. Infection models in organisms like *Drosophila melanogaster* are key for understanding evolution and dynamics of innate immunity. However, current toolsets for fly infection studies are limited in their ability to resolve changes in pathogen load on the hours time-scale, along with stochastic responses to infection in individuals. Here we report a novel bioluminescent imaging strategy enabling non-invasive characterization of pathogen load over time. We demonstrate that photon flux from autobioluminescent reporter bacteria can be used to estimate pathogen count. *Escherichia coli* expressing the *ilux* operon were imaged in whole, living flies at relevant concentrations for immune study. Because animal sacrifice was not necessary to estimate pathogen load, stochastic responses to infection were characterized in individuals for the first time. The high temporal resolution of bioluminescence imaging also enabled visualization of the fine dynamics of microbial clearance on the hours time-scale. Overall, this non-invasive imaging strategy provides a simple and scalable platform to observe changes in pathogen load *in vivo* over time.

3.2 Introduction

Bacteria are widespread and can cause severe disease in animals, including humans (Hunter et al., 2010). Hosts mitigate acute illness through deployment of the immune system. Infection progression and outcome is ultimately determined by a combination of genetics, environment, and stochastic events (Carruthers et al., 2020; Duneau et al., 2017). Determining the relative contribution of each of these factors to individual prognosis will enable the identification of genetic markers and early predictors of infection outcome. Such discoveries will contribute to targeted treatment of bacterial infections.

Determination of genetic and stochastic contributions to infection outcome requires a host organism amenable to genetic manipulation and high-throughput experimentation. *Drosophila melanogaster* fulfills both of these criteria. There are thousands of inbred and sequenced *D. melanogaster* lines, and flies are tractable for high-throughput experimentation in 96-well plates, unlike common mammalian model organisms (Lack, Lange, Tang, Corbett-Detig, & Pool, 2016). Additionally, *Drosophila* possess an innate immune system composed of signaling pathways that are highly conserved in mammals (Lemaitre & Hoffmann, 2007). Flies use both the Toll and IMD signaling pathways in their immune response. The Toll pathway was initially discovered in flies and is analogous to the Toll-like receptor signaling pathway found in mammals (Lemaitre & Hoffmann, 2007). In both flies and mammals, the pathway depends on molecular recognition through pattern recognition receptors (PRRs) to initiate downstream immune response (Moy & Cherry, 2013). The IMD pathway in flies is orthologous to the TNF receptor family signaling cascade in mammals (Buchon, Silverman, & Cherry, 2014; Lemaitre & Hoffmann, 2007). Beyond their use as a platform for discovered conserved immune genes, studying immunity in insects like *Drosophila* provides insight into how insect vectored diseases spread, and how they may be contained (D. Schneider, 2000).

Variation in infection outcome between *D. melanogaster* lines is in part determined by the genetic backgrounds of the host (Duneau et al., 2017; Hotson & Schneider, 2015; Lazzaro, Sackton, & Clark, 2006; Sackton, Lazzaro, & Clark, 2010). Using genetically distinct *D. melanogaster* lines, previous work has identified how different loci affect either the ability of an animal to reduce bacterial load or to induce gene expression upon infection (Frochaux et al., 2020; Lazzaro et al., 2006). Less studied both in its mechanism of action and its variation between lines is the fly's ability to endure infection, i.e tolerance. Though studies have shown that the genes

involved in infection tolerance somewhat overlap with those involved in resistance, this has yet to be comprehensively determined for different pathogens and across genetically diverse lines (Ayres, Freitag, & Schneider, 2008; Ayres & Schneider, 2008; D. S. Schneider & Ayres, 2008; Troha, Im, Revah, Lazzaro, & Buchon, 2018).

D. melanogaster lines also display variability in infection response within genetically identical individuals. This variation has been indirectly linked to stochastic differences in bacterial growth within the colonized host early in the infection process and variation in the onset of the animal's immune response (Duneau et al., 2017; Ellner, Buchon, Dörr, & Lazzaro, 2021).The exact source of stochasticity has yet to be directly observed, largely due to a lack of tools capable of providing information on bacterial load noninvasively over time. Characterization of infection progression in flies has typically relied on destructive methods to establish bacterial loads at static time points, e.g. dilution plating. Dilution plating involves sacrificing individuals and quantifying bacterial load using colony counts from serially diluted fly homogenates. While these averaged "snapshot" analyses can provide biological insight into mechanisms of infection progression and clearance, they are unable to capture the stochastic variability that occurs between individuals (Chambers, Jacobson, Khalil, & Lazzaro, 2019; Kutzer & Armitage, 2016). Furthermore, fine resolution of early immune dynamics has historically been difficult to determine, owing to low temporal sensitivity of existing methods. New strategies that enable high resolution, non-invasive measurement of pathogen load are necessary for understanding infection dynamics and outcomes of genetically similar and diverse populations.

Historically, a "go-to" method for non-invasive imaging in rodent models is bioluminescence (Love & Prescher, 2020; Zambito, Chawda, & Mezzanotte, 2021). Bioluminescence employs luciferase enzymes that oxidize luciferin substrates, producing photons

of light (Kaskova, Tsarkova, & Yampolsky, 2016). These photons can be detected through tissues in whole organisms, enabling sensitive and non-invasive readouts (James & Gambhir, 2012). Because no external excitation source is necessary, background signals are very low compared to other optical (e.g., fluorescent) readouts (Contag & Bachmann, 2002). Despite these advantages, bioluminescence has only been sporadically used in *D. melanogaster* (Brandes et al., 1996; Stanewsky, Jamison, Plautz, Kay, & Hall, 1997; Stempfl et al., 2002). One potential reason for its limited use is that uniform delivery of the luciferin substrate by feeding is difficult, as feeding patterns can vary from fly to fly (Ja et al., 2007). One solution is to use autobioluminescent systems, which produce light without the need for exogenous substrate delivery. There are operons that produce both the luciferase and luciferin, allowing transgenic organisms to continuously glow (Kaskova et al., 2016). One popular autobioluminescent system derived from bacteria is the *ilux* system (Gregor, Gwosch, Sahl, & Hell, 2018). Engineered from the bacterial *lux* system, *ilux* emits blue light (490 nm) and exhibits enhanced brightness and thermal stability. While autobioluminescent systems have been used for decades to illuminate the spread of various pathogens *in vivo* (Cronin et al., 2012; Massey et al., 2011; Morrissey, Hill, & Begley, 2013), they have yet to be applied for studying bacterial clearance in *D. melanogaster*.

Here, we report a novel method employing the *ilux* system for longitudinally monitoring bacterial load in *D. melanogaster*. By expressing the *ilux* operon in requisite bacteria and using photon count as a reporter for relative microbial load, we can non-invasively monitor infection progression and clearance over time. With this method, we are able to observe distinct infection dynamics between lines of D. melanogaster as well as between genetically identical individuals. This method generated real-time, high-throughput and longitudinal measures of infection in *D. melanogaster*, a feat that has not yet been accomplished in the field.

3.3 Results

Drosophila infection models have historically required animal sacrifice to determine pathogen load, which limits measurements to static time points (Figure 3.1A). While this method has provided key insights into fly innate immunity and disease progression (Chambers et al., 2019; Chambers Moria, Jacobson, Khalil, Lazzaro Brian, & Bäumler, 2014; Duneau et al., 2017; Kutzer & Armitage, 2016; Lazzaro et al., 2006), the fine dynamics of infection progression and tolerance among individuals remains difficult to measure. To address these limitations, we designed a non-invasive imaging strategy to monitor pathogen load over time (Figure 3.1B). We employed bioluminescent *E. coli* that constitutively expressed the *ilux* reporter (*ilux-Ecoli*) as a proof-of-concept platform (Gregor et al., 2018). The load of these autoluminescent bacteria could then be tracked post-injection, with photon flux reporting on pathogen count. Thus, we hypothesized that this method could be used to track differences in immune response between genetically distinct individuals and identify stochastic differences in infection progression within groups of genetically identical individuals.

To employ autobioluminscence as a reporter for pathogen count, we first measured how photon flux correlated with bacterial optical density (OD) in liquid culture (Figure S3.1). To do so, we serially diluted *ilux-Ecoli* in liquid culture and measured bioluminescent output. Photon counts correlated exponentially with bacterial OD. Since OD can be converted to bacteria colony forming units (CFUs), we can determine relative pathogen load from total flux (Figure 3.2A).

We then sought to determine how flux correlates with bacterial count when injected into living flies. The *ilux* system emits blue light, with an emission maximum of 490 nm (Figure S3.2). While blue light is difficult to detect in the thicker tissues of mammalian model organisms, the fly cuticle is thin, and thus we hypothesized that the blue emission would be readily detectable in

infected flies. Indeed, upon injection and imaging *ilux-Ecoli* in wild-type, male Oregon-R flies, we were able to reliably detect as few as 1000 CFUs. The thermal noise on the imaging instrument used (IVIS Lumina II) is ~10^2 photons/sec. Thus, we could not reliably image <1000 CFU. A positive correlation was observed between CFU and photon flux indicating that flux can report on relative microbial count *in vivo* (Figure 3.2B). To determine whether sexually dimorphic pigmentation of the cuticle affected this relationship, we also compared the best fit line of radiance/CFU of male and female flies. We found no differences between sexes (Figure S3.3) Thus, we were confident bioluminescence could be employed to determine pathogen count in both male and female flies. When we compared the relationship between CFU and total flux in liquid culture versus in flies, we found no difference in relationship (Figure S3.4). This suggests that the fly cuticle does not interfere with photon penetrance, and that we can use standard curves of flies injected with known amounts of bacteria to calculate relative load at a given point in time.

To determine whether bioluminescence could be employed for longitudinal tracking of pathogen load over time, we monitored flies for infection progression over several days. Given the duration of the experiment, the flies required housing both compatible with imaging and including food to prevent starvation. To this end, flies were housed in black 96-well plates by preparing small aliquots of food for each well and placing a glass sheet overtop of the plate (Figure S3.5). The glass sheet was secured with black electrical tape to mitigate aberrant photon scattering. With the housing in hand, we infected 48 flies with increasing concentrations of *ilux-Ecoli* and transferred them to individual wells in the prepared housing. The flies were then imaged for four days, and photon fluxes recorded (Figure 3.3A-B). Since the flies were freely moving during the 3–5-minute image acquisition, they created a donut shaped signal by traveling around the sides of

the well (Figure 3.3A). We can account for fly activity in quantification by summing the photon count of the entire well.

Since wild-type flies mount a robust immune response against *E. coli* infection, none of the flies died during the course of the experiment. For lowest doses of bacteria administered, photon flux values indicated that the infection had been cleared to levels <1000 CFU, i.e., below the detection limit. At higher doses, detectable amounts of bacteria remained on day 4. Further monitoring would distinguish whether flies exposed to higher initial doses of *E. coli* maintain a systemic, but non-lethal, infection (Chambers et al., 2019; Chambers Moria et al., 2014) or whether more time is needed to clear the infection. To ensure the decreasing photon counts as a function of time were due to infection clearance rather than loss of the *ilux* plasmid, we performed the same time course and performed bioluminescence imaging of plated fly homogenate for each day (Figure S3.6, S3.7). We observed minimal plasmid loss. Clearance patterns varied among flies receiving the same dose of pathogen (Figure 3.3C). For example, when looking at the trajectories from the two lowest doses, there are individual flies that experience resurgent infections, which are unobservable with dilution plating assays. Some flies receiving the highest dose show markedly faster clearance dynamics than others. Our longitudinal measurements allow us to determine that this is true variation in the infection response, as opposed to differences in initial dose. Together, these results highlight how bacterial bioluminescence can be used to image stochastic responses to infection.

Historically, stochasticity of immune response in the first 12 hours of infection has been difficult to observe due to the need for fly sacrifice to obtain pathogen load information. We hypothesized our bioluminescent imaging strategy would be useful for studying differences in pathogen clearance during these critical first hours of infection. For example, we can measure

87

when the immune system is activated, which may vary between individuals, genotypes or in response to different pathogens. To test this capability, we used two fly lines, Oregon-R wild-type flies, and immunodeficient imd[10191] flies (Pham, Dionne, Shirasu-Hiza, & Schneider, 2007). Wild-type flies are resilient to *E. coli* infection and clear the Gram-negative microbes easily, as shown in Figure 3.3. Imd[10191] flies, by contrast, more easily succumb to infection. These organisms bear a frameshift mutation in the IMD protein, effectively eliminating the immune response to Gram-negative bacteria. Therefore, we anticipated imd[10191] flies would have different pathogen loads compared to wild-type files. Indeed, upon injection of a large quantity (1,000,000 CFU) of *ilux-Ecoli*, imd[10191] flies sustained high pathogen levels over time, while Oregon-R flies steadily cleared the infection as evidenced by reduced emission levels (Figure 3.4B). When handling highly concentrated pathogens for injection, variance in initial dose can occur. In this experiment, the imd[10191] flies received a slightly lower initial dose of bacteria than the wild-type flies. However, by the two-hour time point, the imd[10191] flies carried a higher load than the wild-type flies. Although we aim to deliver consistent initial doses, this result highlights a feature of this method: we can censor individual animals that receive aberrant initial doses. This quality control step is not possible with dilution plating-based methods.

Beyond quality control, we can actually use variation in initial dose to answer biological questions. For example, it has been shown that within a genotype variation in initial dose contributes to the differences in bacterial load of chronic infections observed, though this study relies on group averages and is unable to assess the impact of initial infection load (Chambers et al., 2019). To address this question, we plotted the infection dynamics of individual flies (Figure 3.4C). In the imd[10191] line, one individual received a low inoculation of bacteria. To determine if the initial load of infection contributed to the differences in bacterial load at the end of the time

course, we divided samples in each genotype as having received an initial dose above the mean (high dose) or below the mean (low dose) and tested for differences in the mean of the final bacterial loads. We found the initial dose of pathogen does not correlate with differences in the final load (Oregon-R high vs low: p=0.18, imd[10191] high vs low: p=0.78, Welch two sample t-test, Figure S3.8). The bioluminescent method thus reports on initial inoculation differences and can provide insight into alternative hypotheses for variance in infection dynamics among genetically identical populations.

In the experiment above, the high initial dose of bacteria killed the immune-deficient flies in a small-time window, which made it difficult to assess potential drivers of death. We posited that a lower dose of bacteria would kill immunodeficient flies more slowly and with greater variation in the time to death. To test this hypothesis, we injected wild-type and imd[10191] flies with 10,000 CFUs of *ilux-Ecoli*. We then imaged flies every hour for 2 days post-infection. While both strains of flies were injected with the same concentration of bacteria, within the first hour, pathogen load differed between the two populations (Figure 3.5A-B). Over time, the wild-type flies cleared the infection to low pathogen load and survived to the end of the experiment. Conversely, the imd[10191] flies showed a gradual increase in pathogen load over time, with all flies succumbing to the infection by the end of the experiment. While initial loads varied between individuals, inoculation load did not correlate with final bacterial load upon death for imd[10191] individuals, or upon end of experiment for wild type individuals (S8, S9). This suggests variance in initial load does not fully explain variation in bacterial load over time. In line with our hypotheses, we did observe substantial variance in time to death among the imd[10191] individuals (Figure 3.5C).

To further explore the variation in the infection progression in immunodeficient flies, we plotted their individual dynamics (Figure 3.6A). To identify groups of individuals showing distinct

infection profiles, samples were hierarchically clustered using Euclidean dissimilarity (Figure 3.6B, Figure S3.10, (Montero & Vilar, 2014)). This clustering requires that each individual have a measurement from all time points; therefore, we included bacterial load data post-mortem for flies that died prior to the end of the experiment. We separated the trajectories into four clusters, with the majority of flies falling into cluster 2 (blue, n=29) or cluster 3 (yellow, n=13), and clusters 1 and 4 having three flies each (magenta and green, respectively). Clusters 2 and 3 appear to separate based on the lag time to unchecked bacterial growth, with cluster 2 having a lag time of 5-10 hours and cluster 3 having a lag time of 10-15 hours after infection. Clusters 1 and 4 appear to cluster primarily based on receiving a below-average initial dose (cluster 4) or on bacterial dynamics post-mortem (cluster 1). No cluster appeared to correlate with time of death (Welch two sample t-test on every cluster combination, Figure S3.11). While all samples ultimately succumbed to the infection, these distinct dynamics would have been missed using previously established methods.

We further investigated potential sources of the variability observed in time to death of the imd[10191] flies. Several possibilities exist to explain these differences in dynamics, including variation in the injury upon infection, variation in the initial pathogen load, or differences in the physiological state of the fly. To examine if this variance in time to death can be explained by differences in the initial load of infection, we performed Spearman rank correlations between time of death and the initial load. We also measured the correlations between time of death and load at 20 h post injection and the area under curve (AUC) of pathogen load at 20 h. We found initial load was negatively correlated with time of death (ρ = -0.30, 95% bootstrapped confidence interval = (-0.56, 0.00)) (Figure 3.6C). Intuitively, this aligns with expectations: flies receiving higher inoculations will die more quickly. The strongest correlation was found between time to death and

observed load at 20 h post injection (ρ = -0.61, 95% bootstrapped confidence interval = (-0.78, -0.39)) (Figure 3.6D). A similarly strong correlation was observed between time to death and area under the flux curve at 20 h post injection (ρ = -0.59, 95% bootstrapped credible interval =(-0.75, -0.36)) (Figure 3.6E). These correlations suggest that a threshold may exist where host colonization can no longer be contained by the immune system, and death becomes inevitable. This highlights how factors beyond initial load contribute to the time of death observed in individual flies, and the importance of stochasticity in infection outcome. Indeed, the variable paths flies take toward a potential "point-of-no-return" in bacterial colonization can now be investigated in fine detail using the temporal resolution afforded by our bioluminescent method. Ongoing work involving integration of additional luminescent reporters to label immune system components would enable even more thorough investigation into how different variables can contribute to infection outcome.

Unlike the immunodeficient line, all of the wild-type flies survived well past two days, with the overall trend of clearing the infection (Figure 3.7A). Despite the average drop in pathogen load, we did observe large amounts of variance, particularly between hours 10 and 20 of infection. Hierarchical clustering sorted the individual traces into groups that corresponded to early pathogen load (Figure S3.12). In studying this data, we noticed that some individuals experienced a resurgence in infection, and these individuals fell into more than one cluster. Therefore, we highlighted profiles based on the presence of a secondary distinct increase of pathogen load after the initial infection (magenta, n = 10) rather than a sustained gradual decrease (blue, n = 38, Figure 3.7B). Again, we found the initial dose of infection did not correlate with the presence of a secondary peak (Oregon-R C1 vs C2: p=0.41, Welch two sample t-test, Figure S3.13). Despite the presence of an increase in pathogen load in some flies, the population ultimately converges towards

clearing the infection. We expect that this method's ability to report fine differences in pathogen load over time will enable future research to uncover the origins of such stochasticity and better characterize the paths hosts may take to clear an infection.

3.4 Discussion

Here we show that employing an autobioluminescent bacterial reporter enables simple, non-invasive pathogen load determination in *Drosophila melanogaster*. We show that total flux can report on changes in bacterial load over time. The non-invasive feature of this method fundamentally changes our view into infection dynamics by enabling longitudinal tracking of infections in individual animals. Traditional dilution plating approaches only allow for the measurement of an average infection trajectory in a population of genetically identical individuals, while longitudinal measurements allow us to reveal individual variation in dynamics and features such as resurgent infections. Further, these traces can be used to test hypotheses about what features (e.g. initial dose, time to immune system engagement, or colonization progress) drive the ultimate outcome of infection. The simple housing requirements and rapid acquisition times allow for the efficient measurement of large numbers of animals or genotypes at fine time resolution. Because each animal is repeatedly sampled throughout the experiment, the longitudinal measurements effectively reduce the sample size needed to identify differences in distinct genotypes of flies. The decrease in sample size, coupled with the increase in throughput, will make previously laborious genetic screens to identify new components of the immune response more accessible.

The flexibility and simplicity of this method should enable its use in a wide range of settings. For example, the *ilux* cassette can be manipulated via molecular cloning, allowing expression in a wide variety of pathogens. The fly housing requires only simple components found

in most laboratories. We also examined the feasibility of using a plate reader instead of an IVIS imager for luminescence measurements (Figure S3.15). Using flies injected with bacterial doses spanning 6 orders of magnitude, we found that while the IVIS has better sensitivity in detecting low pathogen loads, the plate reader performed comparably to the IVIS at higher doses. This indicates that the plate reader may be useful for examining infections leading to binary outcomes such as death.

The longitudinal measurements and temporal resolution achieved using bioluminescence enables novel observations of infection dynamics. For example, although we found wild type flies cleared *ilux-Ecoli* in all experiments, low initial doses prompted a resurgence in infection among certain individuals (Figure 3.7). Using previous dilution plating methods, resurgence cannot be definitely identified. It would appear to be variation between sacrificed individuals at discrete time points. Using our bioluminescent method, resurgence can be easily visualized as an increase in total flux observed in certain individuals of a population and future studies may investigate the causes and predictors of a resurgent infection.

New insights were also gleaned from experiments with immunodeficient flies. We found a moderate correlation between the initial load and time to death for immunodeficient flies, and a stronger correlation between time to death and bacterial load at 20 h post injection for individuals receiving a low initial dose of *ilux-Ecoli* (Figure 3.6D). This correlation suggests that flies may have a bacteria "threshold" where colonization proceeds unchecked, and death becomes inevitable (Duneau et al., 2017). Taken together, it seems both variation in initial load and infection progression combine to determine the ultime time of death. Thus, bacterial loads at earlier timepoints may be able to predict whether survival or death may be expected.

In summary, non-invasive tracking of pathogen load in *Drosophila melanogaster* over time offers many advantages when compared to traditional methods. Dilution plating requires animal sacrifice to determine average pathogen load at static time points, limiting investigations into individual infection dynamics. Fine differences in pathogen load, infection progression, and immune activation are unable to be determined. The bioluminescent method presented herein offers a facile approach to non-invasively monitor pathogen load on the individual level. Because images may be acquired on the minute-time scale, bioluminescence enables infection progression to be monitored with exceptional resolution. Using this method, we can begin to quantify the contributing factors that result in stochasticity, resurgence, and ultimate infection outcome.

3.5 Materials and Methods

3.5.1 Preparing ilux E. coli

An *E. coli* strain harboring a plasmid with the *ilux* operon (*ilux* pGEX(-)) was obtained from Addgene (plasmid # 107897, deposited by Stefan Hell) and streaked on an LB agar plate with ampicillin (100 µg/mL) to afford single colonies. A single colony was picked and grown in 5 mL LB broth containing ampicillin (100 µg/mL, LB-AMP). The culture was miniprepped according to the manufacturer's instructions (kit purchased from Zymo Research). The concentration of the plasmid was determined using a NanoDrop 3000 (Thermo Fisher). Plasmid (10 ng) was transformed into chemically competent TOP10 *E. coli* (20 µL). The transformant was recovered with SOC (50 µL) for 30 mins at 37 °C and 25 µL plated on an agar plate containing ampicillin. A single colony was picked and expanded in LB-AMP, and a glycerol stock was made for long term storage at -80 °C (500 µL culture with 500 µL 50% v/v glycerol). This glycerol stock is referred to as *ilux-Ecoli*.

3.5.2 Drosophila lines and rearing

Oregon R and imd-10191 were used for this study (Pham et al., 2007). Both lines were reared on standard cornmeal media at 20°C (Brent and Oster 1974). Four-day old male and female flies were collected for injections to ensure full replacement of the larval fat body by the adult fat body (Johnson & Butterworth, 1985).

3.5.3 Drosophila infection induction

Prior to infection, *ilux-Ecoli* was cultured in liquid LB-AMP on a shaker at 37°C for 8 hours. Bacteria were then pelleted using a table-top micro-centrifuge at 5000 rpm/g and resuspended in 200μl of 1X phosphate buffered saline. Optical density was then measured using a NanoDrop 2000 (Thermo Fisher). Injection solutions were prepared at the appropriate OD by dilution in additional PBS. Flies were injected with 34 nL of bacterial solution using Narishige IM 300 Microinjector along the scutescutullar suture and immediately placed into black 96-well plates (Grenier Bio One). For time course experiments, 96-well plates for imaging were prepared by punching out circles of standard cornmeal media and placing these at the bottom of wells before placing flies into the plate. A 4-inch by 6-inch glass cover was placed on the plate during the duration of the time course to prevent individual escape. For single time point measurements, flies were placed in 96-well plates lacking food.

3.5.4 Dilution plating: ilux-Ecoli

To determine the concentration of *ilux-Ecoli* at per OD measurement, *ilux-Ecoli* was grown in 10 mL LB-AMP on a shaker at 37°C for 6-7 hours while the bacteria was still in exponential growth phase. Bacteria was then pelleted down and resuspended in 200 uL 1X phosphate buffered saline and OD of the solution was determined using a NanoDrop 2000 (Thermo Fisher). A stock solution of OD 1 was prepared and serially diluted using 1x PBS by 6million fold. CFUs where then quantified in two ways. The first manner was using 5μl of solution for each dilution which was

spot plated in triplicate on LB plates supplemented with 100µg/ml ampicillin. The second manner was using 90µl of solution which was plated on LB plates supplemented with 100µg/ml ampicillin. Colonies were then counted for each dilution step to determine the concentration of CFUs at OD1 for both methods (Supplementary Figure 3.1). To determine the relationship between *ilux-Ecoli* concentration and total flux, a solution of bacteria was prepared as described above, 90µl of each dilution was then placed into black 96 well plates for imaging (see "imaging parameters" section below).

3.5.5 Dilution plating: infected Drosophila

To determine the concentration of bacteria injected into individual flies, flies were suspended in 250µl of 1X phosphate buffered saline and homogenized (Krupp & Levine, 2010). Homogenate was then used for stepwise serial dilutions. 5 µL of each dilution in the series was then spot plated in triplicate on LB plates with 100µg/ml ampicillin. Colonies were then counted for each dilution step to determine the concentration of CFUs per fly. For plasmid loss determinations, homogenate and serial dilution was prepared in the manner described above. Next 90µl of solution from each dilution in the series was plated on individual LB plates not supplemented ampicillin. Plates containing colonies were then counted for each dilution step to determine the concentration of CFUs per fly and imaged for luminescence output as described in the "imaging parameters" section. Thus, emissive and non-emissive colonies could be distinguished to determine the proportion of colonies that had lost the *ilux* plasmid.

3.5.6 Imaging parameters

All imaging analyses were performed in black 96-well plates (Grenier Bio One) prepared as described above, or on agar plates for plasmid loss studies. Biological replicates for different experiments were performed on different days. Plates containing flies were imaged immediately

post-injection unless otherwise stated. Injected flies and agar plates containing *ilux-Ecoli* were imaged using an IVIS Lumina II (Xenogen) CCD camera chilled to -90 °C. The stage was kept at room temperature (25 °C) during the imaging session, and the camera was controlled using Living Image Software. The exposure time was 1 s - 10 min depending on the brightness of the sample, and the data binning levels were set to medium. Radiance was integrated over regions of interest and quantified to total flux values using the Living Image software. Raw luminescent images were analyzed using FIJI (Schindelin 2012) . For plasmid loss studies, CFUs on plates were quantified by hand, luminescent colonies were quantified by importing for analysis and counting in FIJI.

3.5.7 Data analysis

All flux data was imported into R 3.6.0 for analysis and visualization (R Core Team. 2019, Wickham H. 2016, Dowle M(2019), Wickham H. 2021, Wickham et al., 2019). Model2 regressions were performed using the package lmodel2 1.7-3 (Legendre 2018). Euclidean dissimilarity and hierarchical clustering were performed using package TSclust 1.3.1 (Montero & Vilar, 2014).

3.6 Data access

All flux data generated during this study as well as the code used to analyze the data and generate figures are available for download at   github.com/WunderlichLab/ilux_infection_tracking. \

3.7 Acknowledgments

3.8 References

Ayres, J. S., Freitag, N., & Schneider, D. S. (2008). Identification of Drosophila Mutants Altering Defense of and Endurance to *Listeria monocytogenes* Infection. *Genetics, 178*(3), 1807-1815. doi:10.1534/genetics.107.083782

Ayres, J. S., & Schneider, D. S. (2008). A Signaling Protease Required for Melanization in *Drosophila* Affects Resistance and Tolerance of Infections. *PLOS Biology, 6*(12), e305. doi:10.1371/journal.pbio.0060305

Brandes, C., Plautz, J. D., Stanewsky, R., Jamison, C. F., Straume, M., Wood, K. V., . . . Hall, J. C. (1996). Novel Features of *Drosophila* period Transcription Revealed by Real-Time Luciferase Reporting. *Neuron, 16*(4), 687-692. doi:https://doi.org/10.1016/S0896-6273(00)80088-4

Buchon, N., Silverman, N., & Cherry, S. (2014). Immunity in *Drosophila melanogaster* — from microbial recognition to whole-organism physiology. *Nature Reviews Immunology, 14*(12), 796 810. doi:10.1038/nri3763

Carruthers, J., Lythe, G., López-García, M., Gillard, J., Laws, T. R., Lukaszewski, R., & Molina-París, C. (2020). Stochastic dynamics of *Francisella tularensis* infection and replication. *PLOS Computational Biology, 16*(6), e1007752. doi:10.1371/journal.pcbi.1007752

Chambers, M. C., Jacobson, E., Khalil, S., & Lazzaro, B. P. (2019). Consequences of chronic bacterial infection in *Drosophila melanogaster*. *PLOS ONE, 14*(10), e0224440. doi:10.1371/journal.pone.0224440

Chambers Moria, C., Jacobson, E., Khalil, S., Lazzaro Brian, P., & Bäumler, A. J. (2014). Thorax Injury Lowers Resistance to Infection in *Drosophila melanogaster*. *Infection and Immunity, 82*(10), 4380-4389. doi:10.1128/IAI.02415-14

Contag, C. H., & Bachmann, M. H. (2002). Advances in *In Vivo* Bioluminescence Imaging of Gene Expression. *Annual Review of Biomedical Engineering, 4*(1), 235-260. doi:10.1146/annurev.bioeng.4.111901.093336

Cronin, M., Akin, A. R., Collins, S. A., Meganck, J., Kim, J.-B., Baban, C. K., . . . Tangney, M. (2012). High Resolution *In Vivo* Bioluminescent Imaging for the Study of Bacterial Tumour Targeting. *PLOS ONE, 7*(1), e30940. doi:10.1371/journal.pone.0030940

Dowle, M., & Srinivasan, A. (2019). data.table: Extension of `data.frame`. R package version 1.12.2. https://CRAN.R-project.org/package=data.table

Duneau, D., Ferdy, J.-B., Revah, J., Kondolf, H., Ortiz, G. A., Lazzaro, B. P., & Buchon, N. (2017). Stochastic variation in the initial phase of bacterial infection predicts the probability of survival in *D. melanogaster*. *eLife, 6*, e28298. doi:10.7554/eLife.28298

Ellner, S. P., Buchon, N., Dörr, T., & Lazzaro, B. P. (2021). Host–pathogen immune feedbacks can explain widely divergent outcomes from similar infections. *Proceedings of the Royal Society B: Biological Sciences, 288*(1951), 20210786. doi:10.1098/rspb.2021.0786

Frochaux, M. V., Bou Sleiman, M., Gardeux, V., Dainese, R., Hollis, B., Litovchenko, M., . . . Deplancke, B. (2020). cis-regulatory variation modulates susceptibility to enteric infection in the *Drosophila* genetic reference panel. *Genome Biology, 21*(1), 6. doi:10.1186/s13059-019-1912-z

Gregor, C., Gwosch, K. C., Sahl, S. J., & Hell, S. W. (2018). Strongly enhanced bacterial bioluminescence with the *ilux* operon for single-cell imaging. *Proceedings of the National Academy of Sciences, 115*(5), 962. doi:10.1073/pnas.1715946115

Hotson, A. G., & Schneider, D. S. (2015). Drosophila melanogaster Natural Variation Affects Growth Dynamics of Infecting *Listeria monocytogenes*. *G3 (Bethesda, Md.), 5*(12), 2593-2600. doi:10.1534/g3.115.022558

Hunter, P. A., Dawson, S., French, G. L., Goossens, H., Hawkey, P. M., Kuijper, E. J., . . Piddock, L. J. V. (2010). Antimicrobial-resistant pathogens in animals and man: prescribing, practices and policies. *Journal of Antimicrobial Chemotherapy, 65*(suppl_1), i3-i17. doi:10.1093/jac/dkp433

Ja, W. W., Carvalho, G. B., Mak, E. M., de la Rosa, N. N., Fang, A. Y., Liong, J. C., . . . Benzer, S. (2007). Prandiology of *Drosophila* and the CAFE assay. *Proceedings of the National Academy of Sciences, 104*(20), 8253. doi:10.1073/pnas.0702726104

James, M. L., & Gambhir, S. S. (2012). A Molecular Imaging Primer: Modalities, Imaging Agents, and Applications. *Physiological Reviews, 92*(2), 897-965. doi:10.1152/physrev.00049.2010

Johnson, M. B., & Butterworth, F. M. (1985). Maturation and aging of adult fat body and oenocytesin *Drosophila* as revealed by light microscopic morphometry. *Journal of Morphology, 184*(1), 51-59. doi:https://doi.org/10.1002/jmor.1051840106

Kaskova, Z. M., Tsarkova, A. S., & Yampolsky, I. V. (2016). 1001 lights: luciferins, luciferases, their mechanisms of action and applications in chemical analysis, biology and medicine. *Chemical Society Reviews, 45*(21), 6048-6077. doi:10.1039/C6CS00296J

Krupp, J. J., & Levine, J. D. (2010). Dissection of oenocytes from adult *Drosophila melanogaster*. *Journal of visualized experiments : JoVE*(41), 2242. doi:10.3791/2242

Kutzer, M. A. M., & Armitage, S. A. O. (2016). The effect of diet and time after bacterial infection on fecundity, resistance, and tolerance in *Drosophila melanogaster*. *Ecology and Evolution,6*(13), 4229-4242. doi:https://doi.org/10.1002/ece3.2185

Lack, J. B., Lange, J. D., Tang, A. D., Corbett-Detig, R. B., & Pool, J. E. (2016). A Thousand FlyGenomes: An Expanded *Drosophila* Genome Nexus. *Molecular Biology and Evolution, 33*(12), 3308-3313. doi:10.1093/molbev/msw195

Lazzaro, B. P., Sackton, T. B., & Clark, A. G. (2006). Genetic Variation in *Drosophila melanogaster* Resistance to Infection: A Comparison Across Bacteria. *Genetics, 174*(3), 1539-1554. doi:10.1534/genetics.105.054593

Legendre, P. (2018). lmodel2: Model II Regression. R package version 1.7-3. https://CRAN.R-project.org/package=lmodel2

Lemaitre, B., & Hoffmann, J. (2007). The Host Defense of *Drosophila melanogaster*. *Annual Review of Immunology, 25*(1), 697-743. doi:10.1146/annurev.immunol.25.022106.141615

Love, A. C., & Prescher, J. A. (2020). Seeing (and Using) the Light: Recent Developments in Bioluminescence Technology. *Cell Chemical Biology, 27*(8), 904-920. doi:https://doi.org/10.1016/j.chembiol.2020.07.022

Massey, S., Johnston, K., Mott, T., Judy, B., Kvitko, B., Schweizer, H., . . . Torres, A. (2011). *In vivo* Bioluminescence Imaging of *Burkholderia mallei* Respiratory Infection and Treatment in the
Mouse Model. *Frontiers in microbiology, 2*, 174.

Montero, P., & Vilar, J. A. (2014). TSclust: An R Package for Time Series Clustering. *Journal of Statistical Software; Vol 1, Issue 1 (2015)*. doi:10.18637/jss.v062.i01

Morrissey, R., Hill, C., & Begley, M. (2013). Shining light on food microbiology; applications of *Lux*-tagged microorganisms in the food industry. *Trends in Food Science & Technology, 32*(1), 415. doi:https://doi.org/10.1016/j.tifs.2013.05.001

Moy, R. H., & Cherry, S. (2013). Antimicrobial Autophagy: A Conserved Innate Immune Response in *Drosophila*. *Journal of Innate Immunity, 5*(5), 444-455. doi:10.1159/000350326

Pham, L. N., Dionne, M. S., Shirasu-Hiza, M., & Schneider, D. S. (2007). A Specific Primed Immune Response in *Drosophila* Is Dependent on Phagocytes. *PLOS Pathogens, 3*(3), e26. doi:10.1371/journal.ppat.0030026

R Core Team. (2016). R: A Language and Environment for Statistical Computing. Vienna, Austria. Retrieved from https://www.R-project.org/

Sackton, T. B., Lazzaro, B. P., & Clark, A. G. (2010). Genotype and Gene Expression Associations with Immune Function in *Drosophila*. *PLOS Genetics, 6*(1), e1000797. doi:10.1371/journal.pgen.1000797

Schindelin, J., Arganda-Carreras, I., Frise, E., Kaynig, V., Longair, M., Pietzsch, T., … Cardona, A. (2012). Fiji: an open-source platform for biological-image analysis. *Nature Methods*, *9*(7), 676–682. doi:10.1038/nmeth.2019

Schneider, D. (2000). Using *Drosophila* as a model insect. *Nature Reviews Genetics, 1*(3), 218 226. doi:10.1038/35042080

Schneider, D. S., & Ayres, J. S. (2008). Two ways to survive infection: what resistance and tolerance can teach us about treating infectious diseases. *Nature Reviews Immunology, 8*(11), 889-895. doi:10.1038/nri2432

Stanewsky, R., Jamison, C. F., Plautz, J. D., Kay, S. A., & Hall, J. C. (1997). Multiple circadian regulated elements contribute to cycling period gene expression in *Drosophila*. *The EMBO Journal, 16*(16), 5006-5018. doi:https://doi.org/10.1093/emboj/16.16.5006

Stempfl, T., Vogel, M., Szabo, G., Wülbeck, C., Liu, J., Hall, J. C., & Stanewsky, R. (2002). Identification of Circadian-Clock-Regulated Enhancers and Genes of *Drosophila melanogaster* by Transposon Mobilization and Luciferase Reporting of Cyclical Gene Expression. *Genetics, 160*(2), 571-593. doi:10.1093/genetics/160.2.571

Troha, K., Im, J. H., Revah, J., Lazzaro, B. P., & Buchon, N. (2018). Comparative transcriptomics reveals CrebA as a novel regulator of infection tolerance in *D. melanogaster*. *PLOS Pathogens, 14*(2), e1006847. doi:10.1371/journal.ppat.1006847

Wickham, H. (2016). *Ggplot2: Elegant graphics for data analysis* (2nd ed.) [PDF]. Springer International Publishing.

Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L. D., François, R., . . . Yutani H (2019). "Welcome to the tidyverse." *Journal of Open Source Software*, **4**(43), 1686. doi: 10.21105/joss.01686.

Wickham, H., François, R., Henry, L., & Müller, K (2021). dplyr A Grammar of Data Manipulation. R package version 1.0.6. https://CRAN.R-project.org/package=dplyr

Zambito, G., Chawda, C., & Mezzanotte, L. (2021). Emerging tools for bioluminescence imaging. *Current Opinion in Chemical Biology, 63*, 86-94. doi:https://doi.org/10.1016/j.cbpa.2021.02.005

3.9 Figures
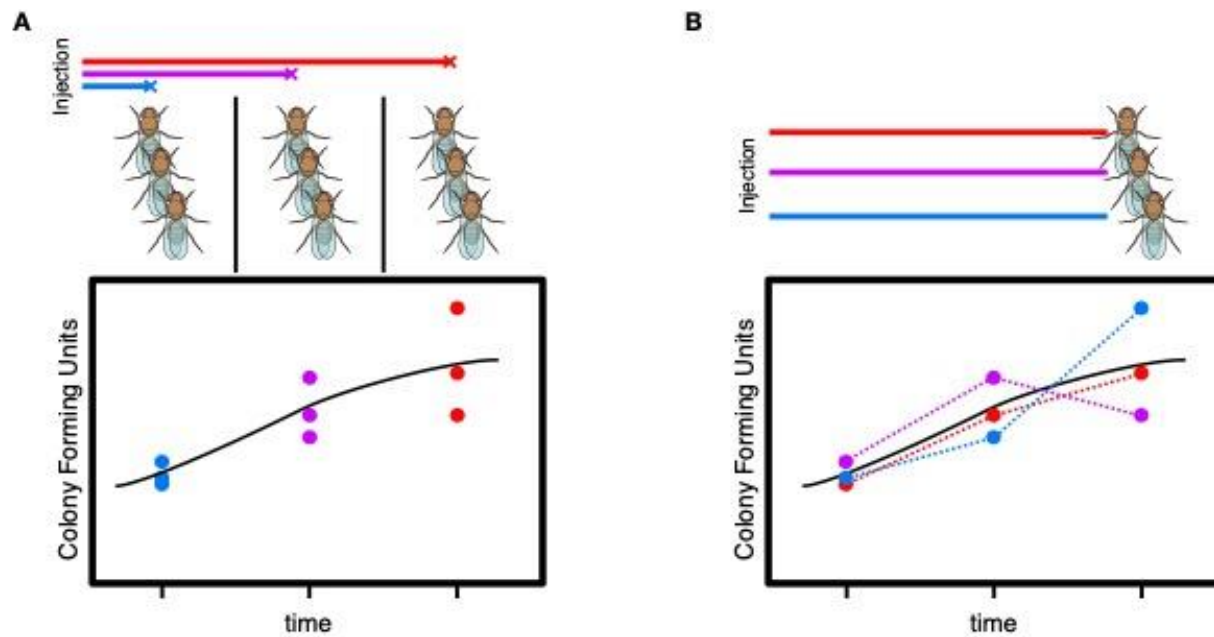
Figure 3.1. A novel method for non-invasive tracking of pathogen load over time. A) Previous methods for determining pathogen load at static time points require animal sacrifice. Larger cohorts are required for experiments as several flies must be sacrificed at desired intervals to check infection progression and clearance. B) This work presents a novel, non-invasive method to track pathogen load over time using bioluminescence. Thus, all flies can be individually monitored over time, allowing for a more comprehensive view of immune response.
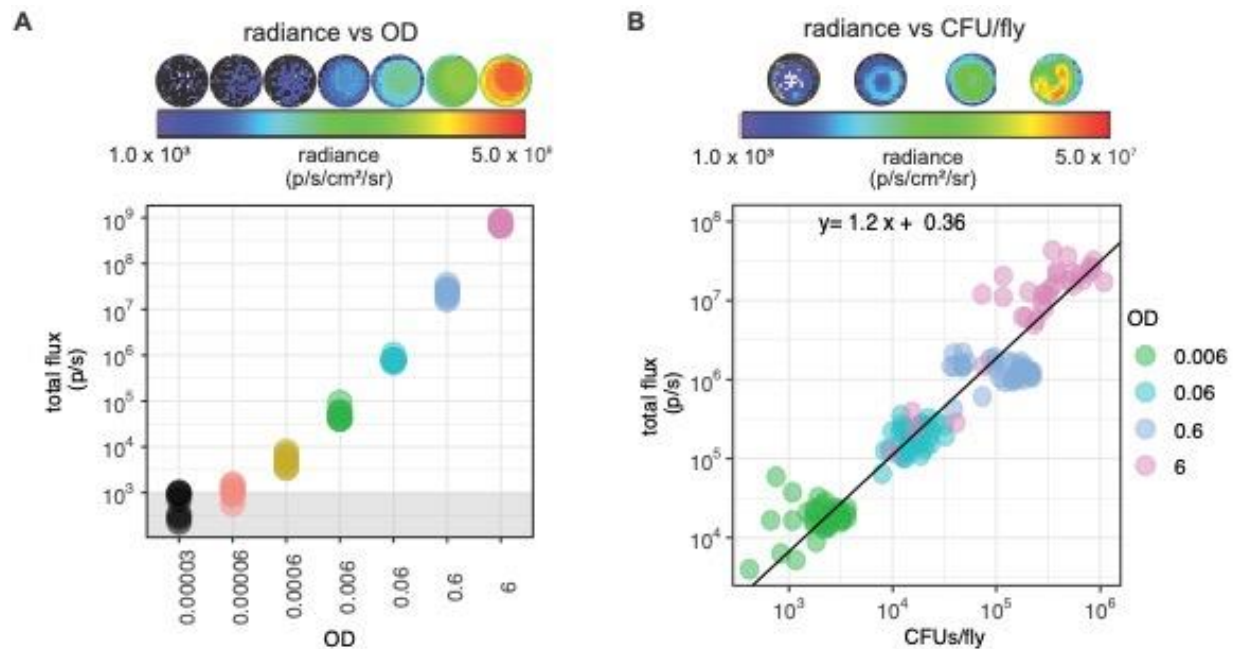
Figure 3.2. Photon flux is positively correlated to bacterial concentration. A) *ilux-Ecoli* were serially diluted in liquid cultures and assayed for photon output. Higher concentrations of bacteria correlate with higher photon fluxes. Plot shows twelve measurements per OD (four technical replicates across three biological replicates). Well images are representative of three biological replicates. B) Wild-type flies were injected with different concentrations of *ilux-Ecoli* and assayed for light emission. A linear correlation was observed between radiance and CFUs injected. Standard major axis regression of the data showed the slope of 1.2 (confidence interval (CI):[1.16,1.29]). Graph shows data of 36 injected flies . Well images above each graph are representative of images of the injected flies. In both experiments, radiance was summed over the entire well to yield flux using the Living Image software.
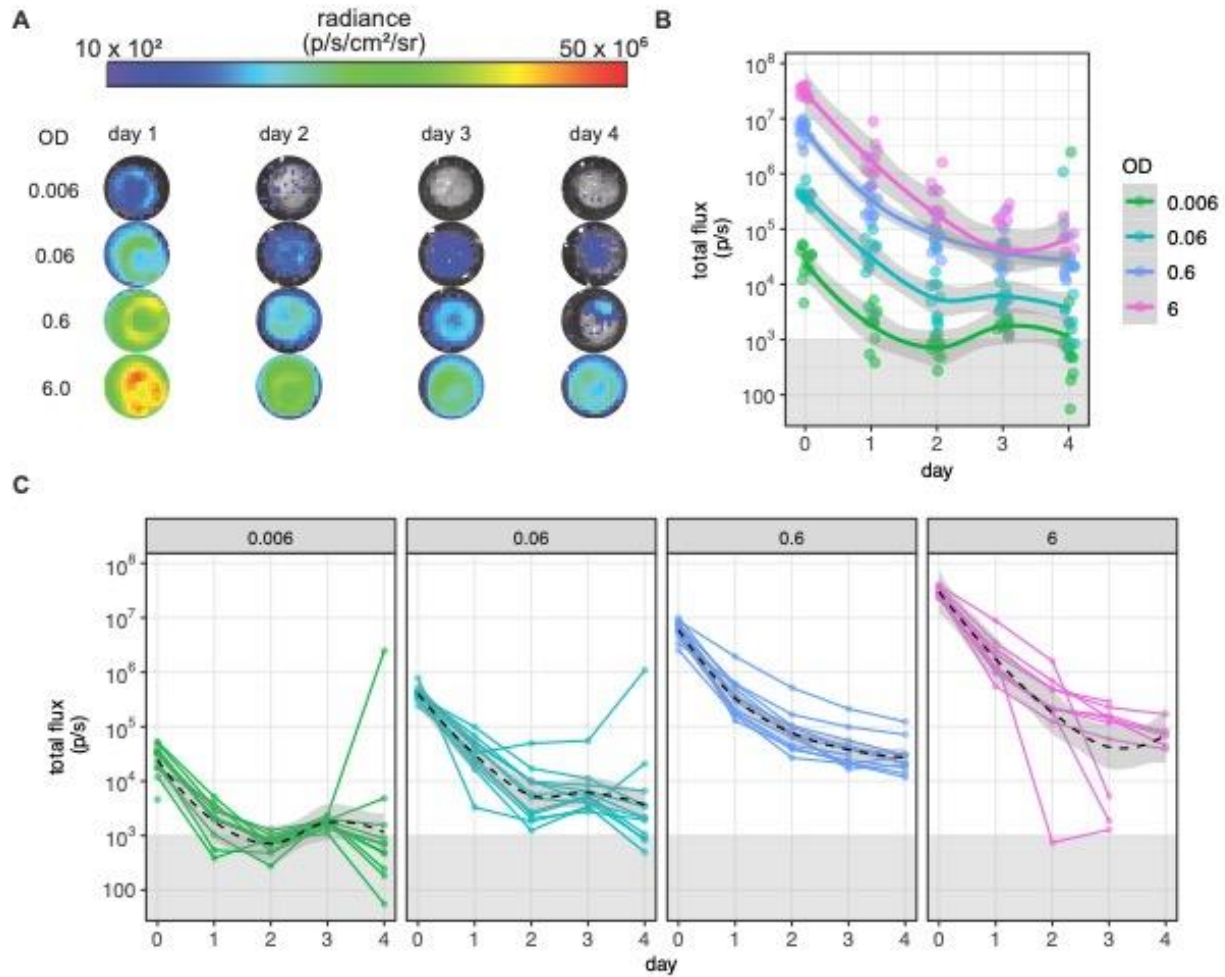
Figure 3.3. Bioluminescence can be used to track changes in pathogen load over time. A) Representative well images for different concentrations of bacteria injected in wild-type flies over four days. Higher initial concentrations were cleared to low, but detectable concentrations. Low initial concentrations were cleared below the limit of detection for the imaging instrument. Well images are representative of 48 injected individuals. B) Average clearance patterns for different concentrations of *ilux-Ecoli* injected in individual flies over time. Photon counts were summed over the entire well where flies resided. Grey box shows the limit of detection of the imaging instrument. Solid colored line represents the average of the cohort. Grey bar represents the standard deviation over replicates, dots represent one individual. Graph shows data of 48 injected flies. C) Individuals display varied routes toward infection clearance, suggesting stochasticity plays a role in infection dynamics. Black dotted line represents the average, and the grey band represents the standard deviation. Grey box shows the limit of detection of the imaging instrument. Solid lines represent routes individuals took toward infection clearance. Data shown were taken from 48 infected individuals.
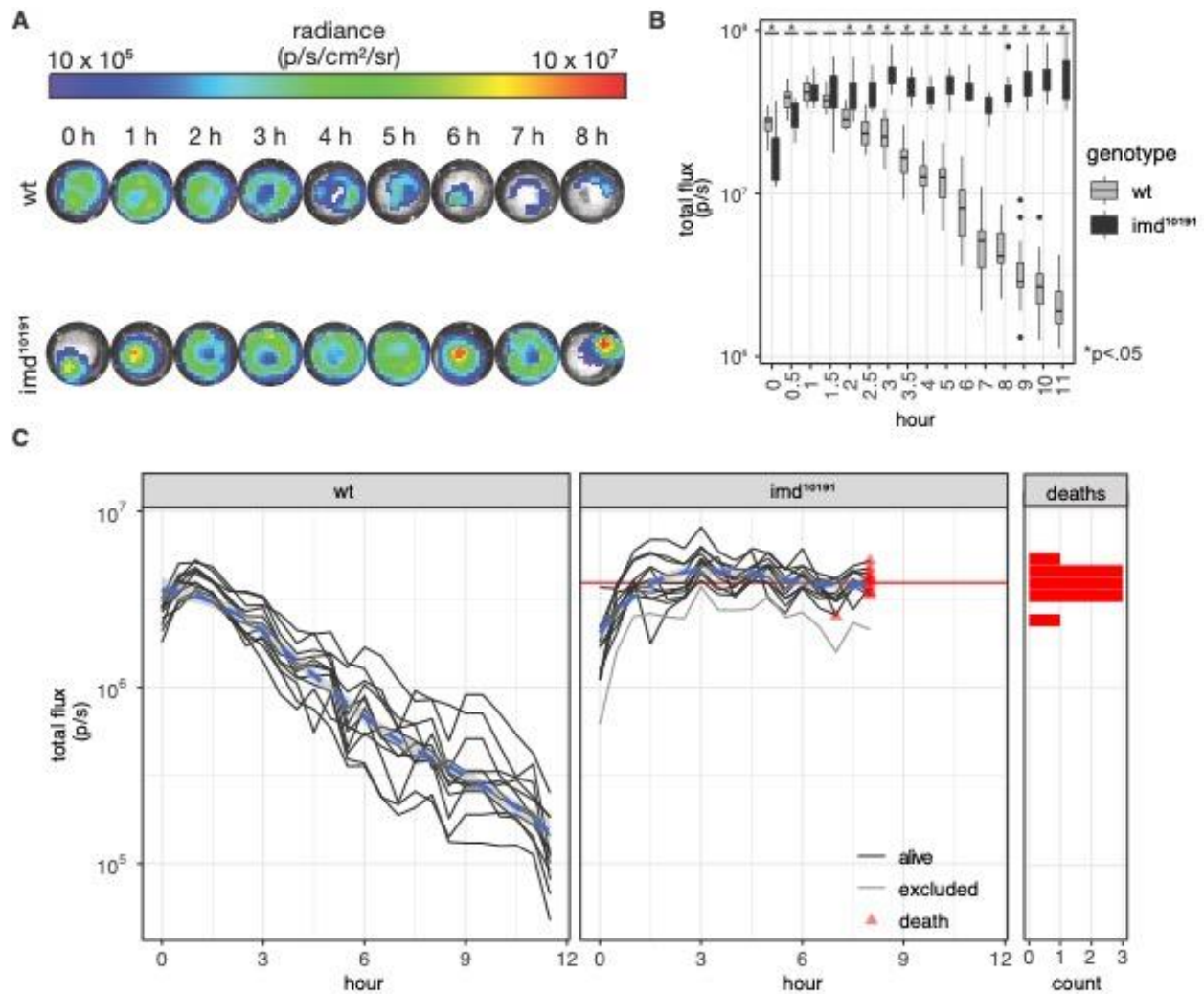
Figure 3.4: Longitudinal tracking of individual flies allows for deconvolution of community dynamics A) Representative images of radiance measurements for wt and imd[10191] injected with .034µl of OD=6 (1,000,000 CFU) *ilux-Ecoli*. Images show the first 8 hours, after which most imd[10191] flies perished. Images are representative of 24 injected individuals. B) Comparison of population level integrated total flux. Timepoints showing difference in mean between the two lines are demarked with an asterisk. Immunodeficient lines received a lower dose of infection than wt flies, but within an order of magnitude difference. This significance was lost by hour 1, with imd[10191] bacterial load surpassing that of *wt* by hour 2. Data comprise 12 injected individuals. C) Comparison of wt and imd[10191] individual tracks. Individual variation of immune response and pathogen clearance was observed in living flies (solid lines). Deaths are marked by red triangles, and the lines end. The blue dotted line shows the average of the cohort, while dark grey lines show the individual paths toward clearance or death. Death histogram shows the effective pathogen load upon death as a function of total integrated flux. The red line on the imd[10191] graph also demarcates the average of these values. The light grey line indicates a fly that received a lower-than-average initial dose. Thus, these data were filtered out in the subsequent analysis. Data shown comprise 12 injected individuals.
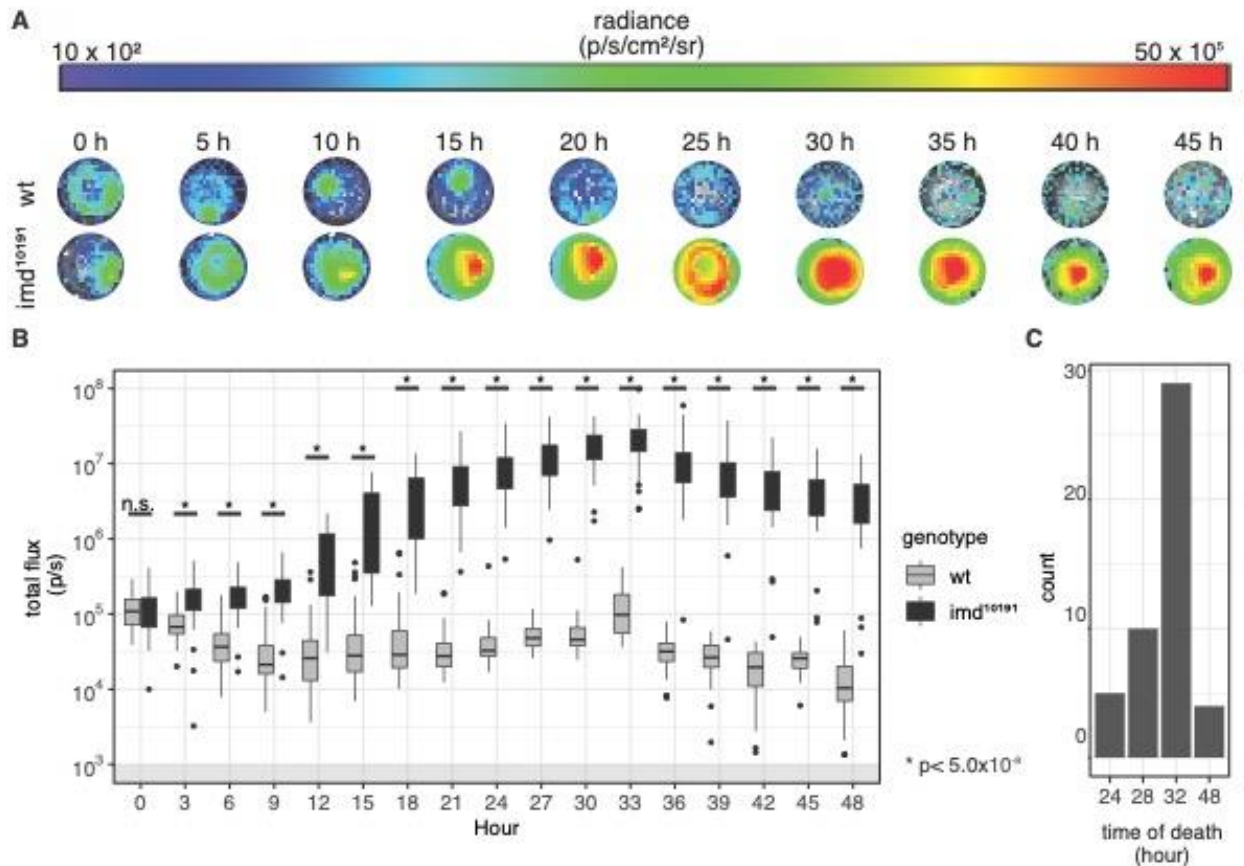
Figure 3.5: Individual infection tracking of immune-deficient flies allows for deconvolution of infection variation. Both wt and imd[10191] flies were injected with 0.034µL of OD=0.06 (10,000 CFU) *ilux-Ecoli* (n=48 for each genotype) A) Representative images of radiance measurements for wt and imd[10191] flies injected with *ilux-Ecoli*. Images are representative for 48 individuals and are shown in 5-hour intervals for the first 45 hours of infection. B) Summary of integrated total flux values for the wt and imd[10191] in 5-hour intervals for the first 45 hours. Timepoints showing difference in mean between the two lines are demarked with an asterisk. By hour 5, both lines show differences in the ability to fight off infection with wild-type flies observed clearing the infection and imd[10191] flies much higher bacterial loads. C) Histogram displaying time of death statistics for imd[10191] flies. The majority of imd[10191] flies died at hour 32, when integrated total flux reached its highest peak. Death data were compiled from 48 imd[10191] individuals.
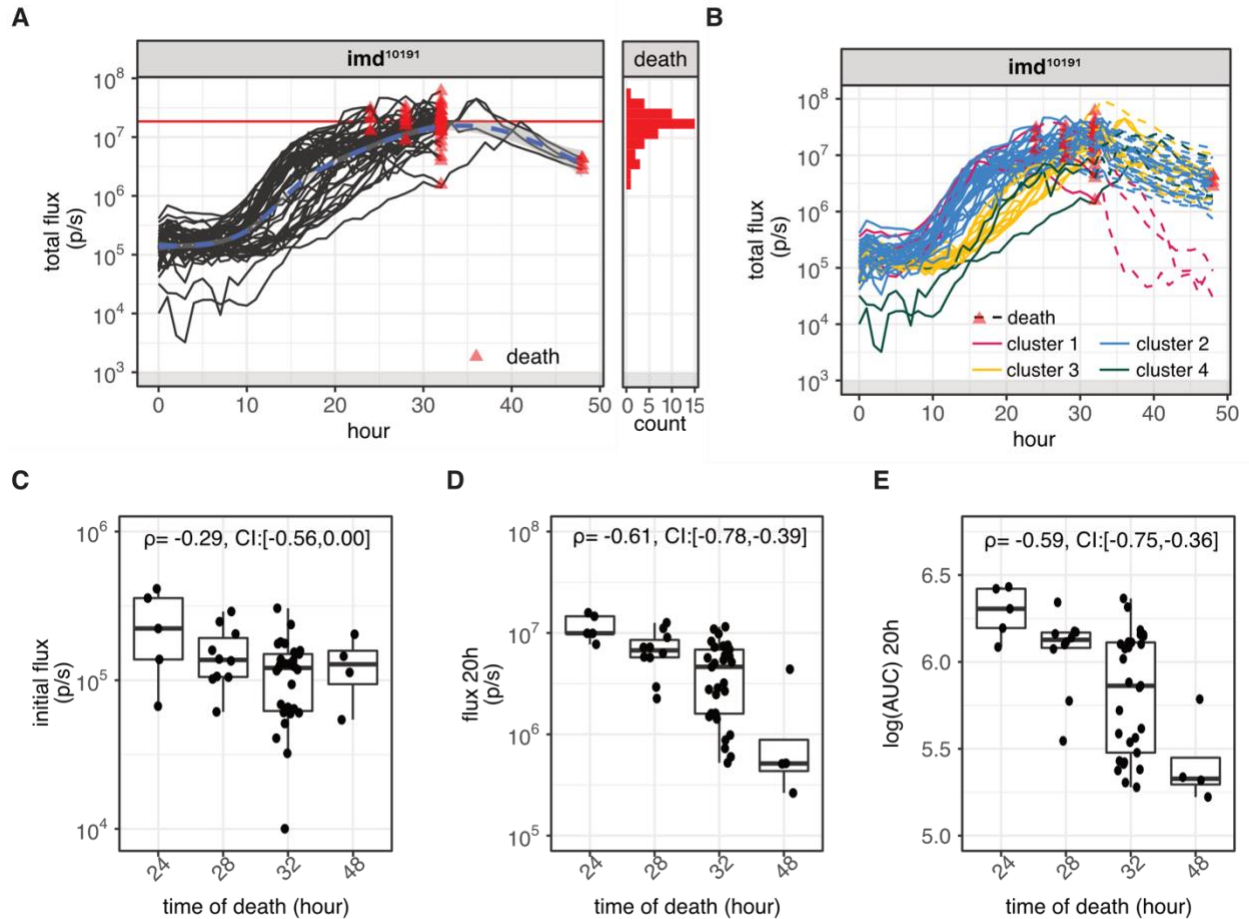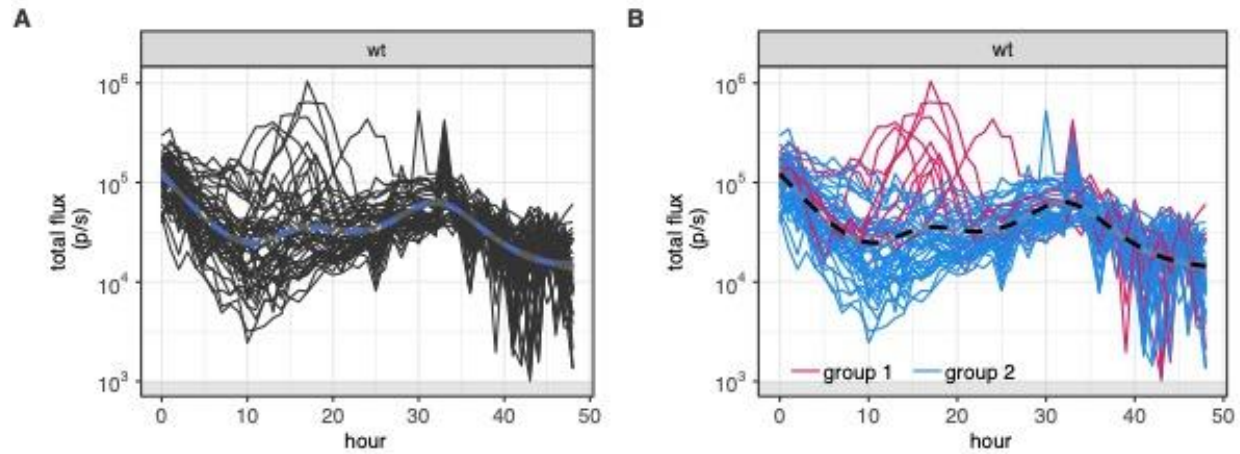
Figure 3.6: Individual infection tracking of oregon-R flies shows two distinct pathways towards bacterial clearance. A) Individual tracks of infection in live imd[10191] flies (black lines, n=48); deaths are marked as a red triangle and the end of the line. All flies died by hour 48 and the mean radiance upon death was 18.6×10^6 (solid red line). Threshold of accurate detection demarcated with a grey box. Histogram shows the distribution of integrated photon flux (serving as a proxy for pathogen count) of imd[10191] flies upon death. B) Individual tracks of infection colored by clusters. Clusters were assigned via hierarchical clustering using Euclidean dissimilarity. Four distinct groups were assigned with cluster 2 and 3 containing the majority of samples (cluster 2= 29, cluster 3=13) and clusters 1 and 4 containing 3 samples each. Data are for 48 imd[10191] individuals. C) Spearman rank correlation (ρ) and the 95% bootstrapped confidence interval (CI) between initial flux and time of death for imd[10191] flies. D) Spearman rank correlation and CI between total flux at hour 20 and time of death. E) Spearman rank correlation and CI between the log transformed area under the curve (AUC) up until hour 20 and the time of death. The curve here refers to the bacterial load (CFU) vs time curve. In all cases CI of Spearman correlation coefficients were computed by bootstrapping 10,000 synthetic datasets and computing the correlations on these datasets. 2.5 and 97.5 percentile values of the sampled correlations are reported.

Figure 3.7: Individual infection tracking of wt flies shows two distinct pathways towards bacterial clearance. A) Individual tracks of infection in live flies (black lines). No flies died during this time course. Threshold of accurate detection demarcated with grey box. B) Individual tracks of infection grouped by the presence of a secondary peak during the infection process. While all flies showed a decrease in bacterial load by the 48 hour mark, a subset of flies (n = 10) showed an increase in bacterial load between 10 - 25 hours (magenta lines). Dashed lines represent the mean trajectory for the genotype.

3.10 Supplementary Materials
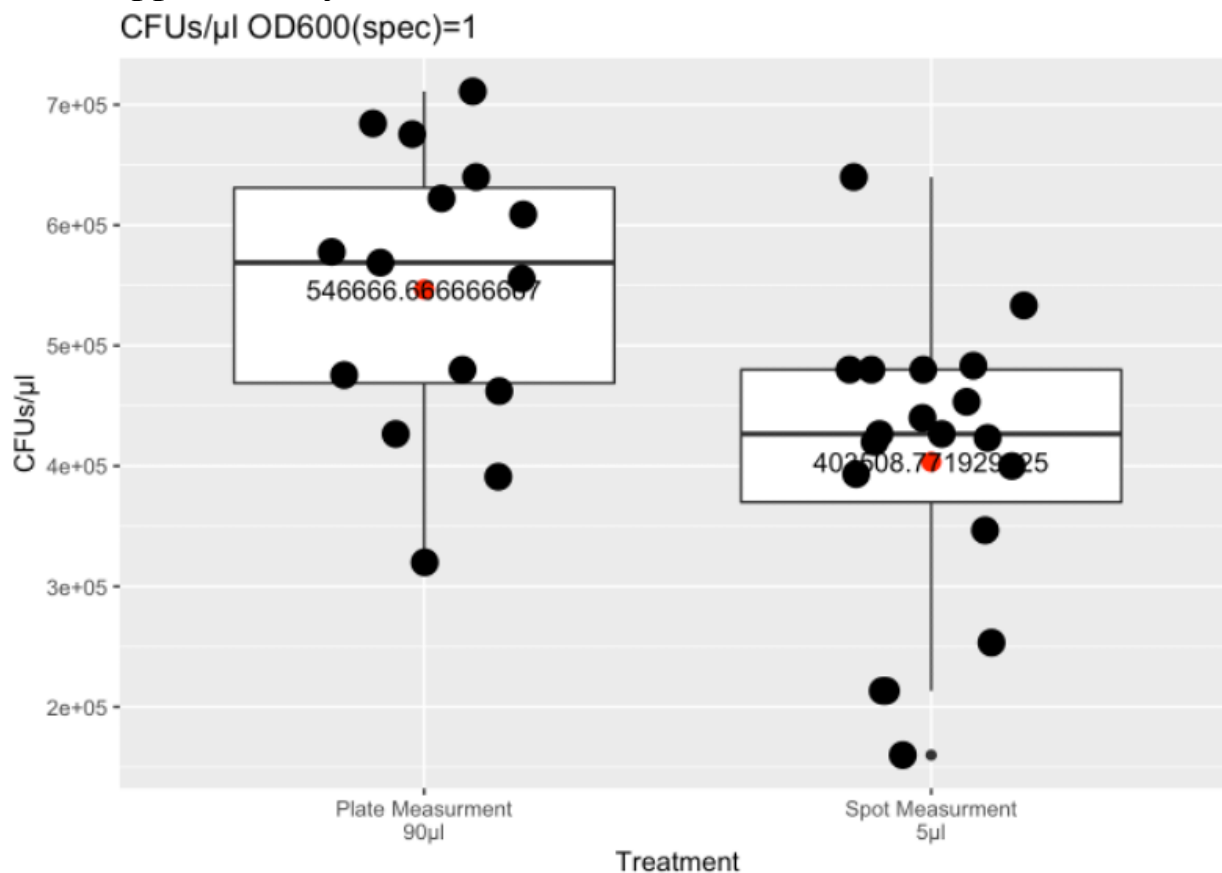
Figure S3.1. *ilux-Ecoli* Calculated to be 546666.67 CFUs/µl. Calculated CFUs/µl for two dilution plating experiments. *ilux-Ecoli* solution were prepared, serially diluted and plated using one of two methods (see Methods). From each dilution, either a single 90µl aliquot was plated on a single plate (left) or six 5µl aliquots were plated on a single plate (right). Calculated CFUs per µl for both measurements were within an order of magnitude of each other. We chose the 90µl plate measurements for our final calculations given that the larger aliquots make them less likely to be skewed by stochastic sampling.
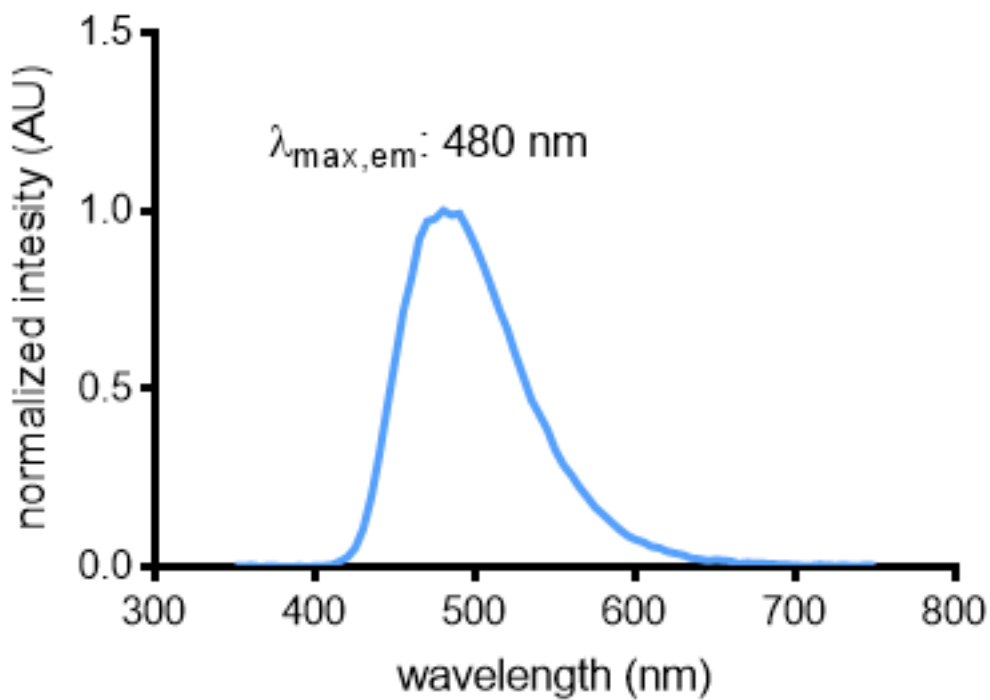
Figure S3.2. Spectroscopic properties of *ilux*. *ilux-Ecoli* were cultured in LB-AMP for 24 h at 37 °C. An aliquot of cells (700 µL) was transferred to a quartz cuvette, and the bioluminescence spectra was recorded at 25 °C. Luminescence values were normalized such that the maximum value equaled one and plotted in GraphPad Prism 5.
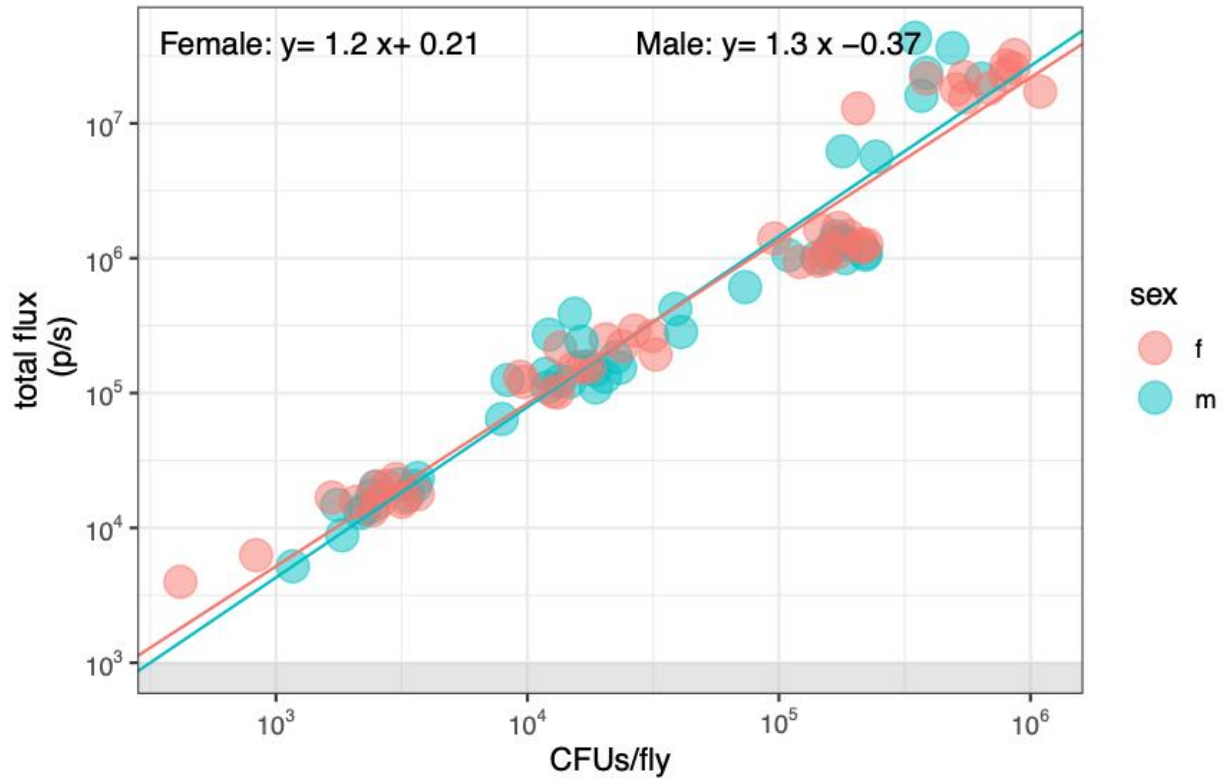
Figure S3.3: Male and female flies show no differences in relationship of CFUs injected to radiance detected. Male and female flies were injected with one of four doses (OD = 6, 0.6, 0.06 or 0.006) of *ilux-Ecoli*. Standard major axis regression of log-transformed data for each sex showed no significant differences between slope of the lines male m= 1.3 (CI: [1.37,1.16]) female m=1.2 (CI:[1.28,1.14]). While there are differences in intercept, for our range of detection, we conclude that there are no differences in radiance detection between female flies and the more pigmented male flies. See Methods for more details.
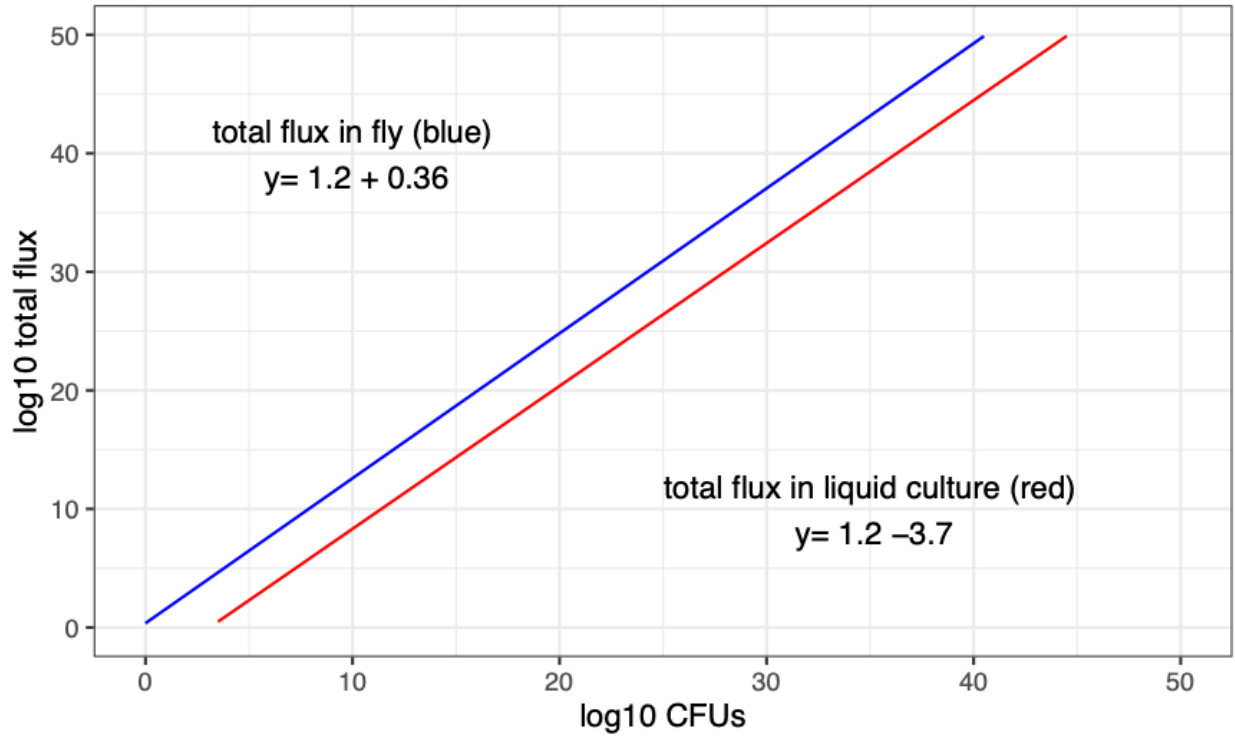
Figure S3.4: Bacteria injected into flies and in liquid culture show a similar relationship between total CFUs and total flux. Standard major axis regression of the relationship between detected radiance and total CFUs injected into flies (blue) and CFUs in liquid culture (red). We observe no difference in slope of the two lines (flies m=1.2, CI:[1.29,1.16] , liquid m=1.2, CI:[1.26,1.15]), demonstrating the positive correlation between flux and concentration is upheld *in vivo*. We suspect that the lower intercept of the bacteria in liquid culture is likely due to the greater area these bacteria occupy, thus making the signal harder to detect (ie more diffuse).
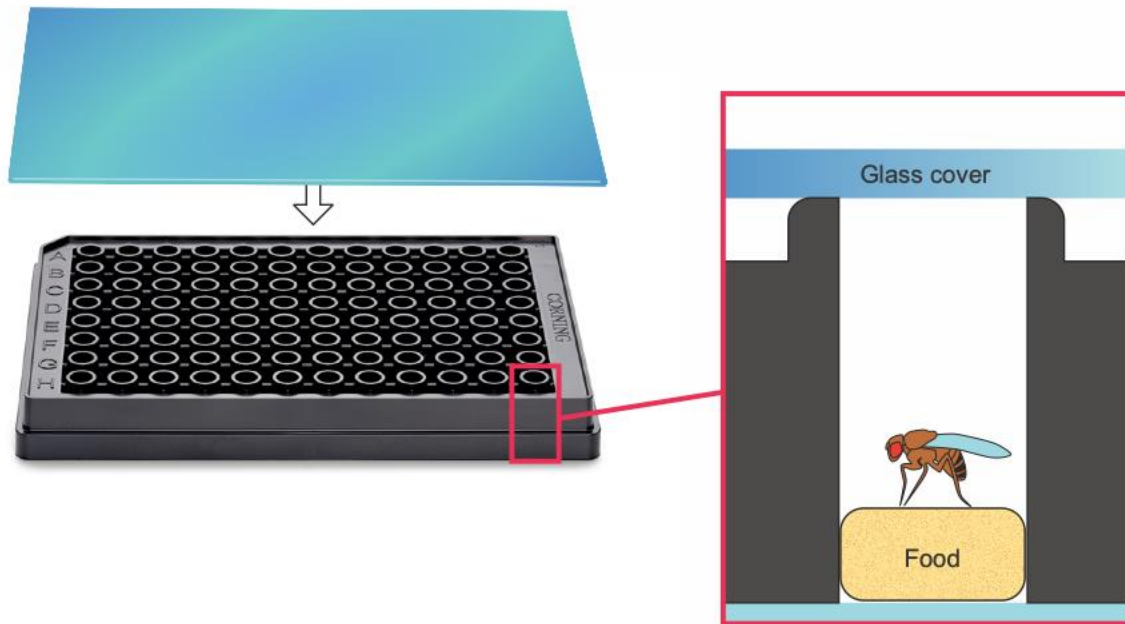
Figure S3.5: Schematic of the plate set up used for housing and imaging flies. Black-walled, clear-bottomed 96-well plates with a 2 mm thick glass cover were used to house flies for the duration of the imaging time course. For food, 2-5 mm thick disks of standard cornmeal media were placed at the bottom of each well before placing individual flies inside. Disks were prepared by using the end of a 5 mL serological pipette to "punch out" disks from solid media.
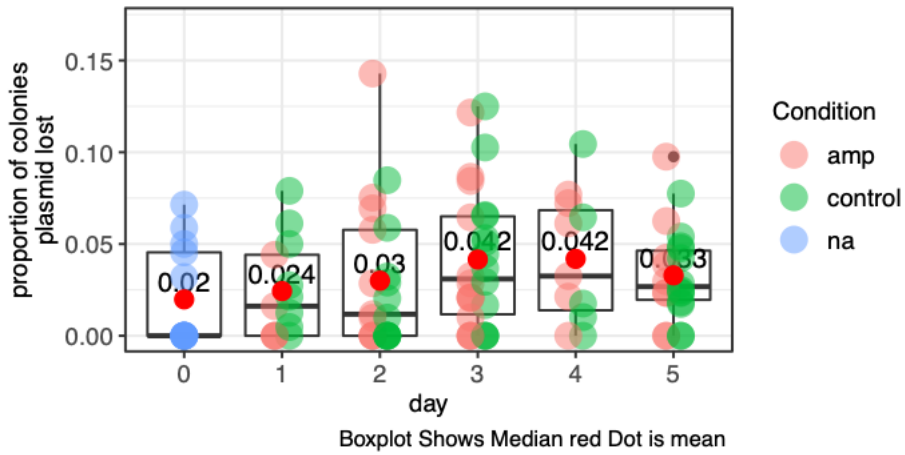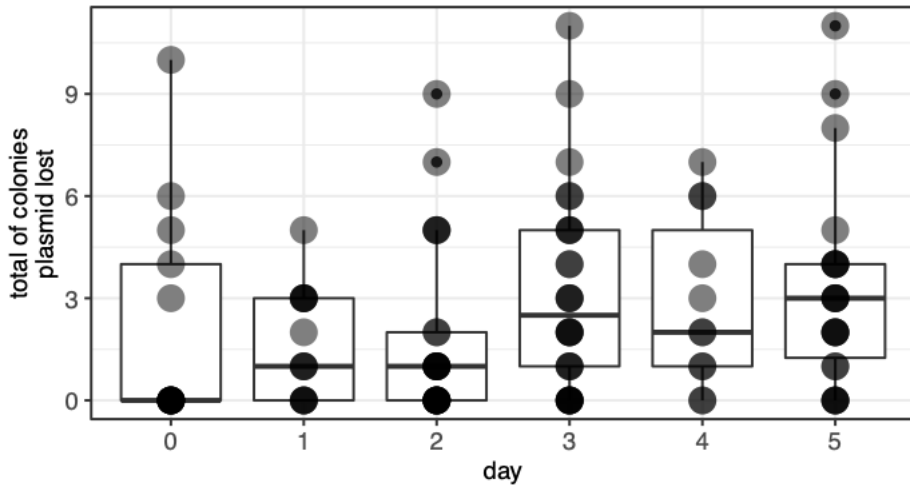
A

Boxplot Shows Median red Dot is mean
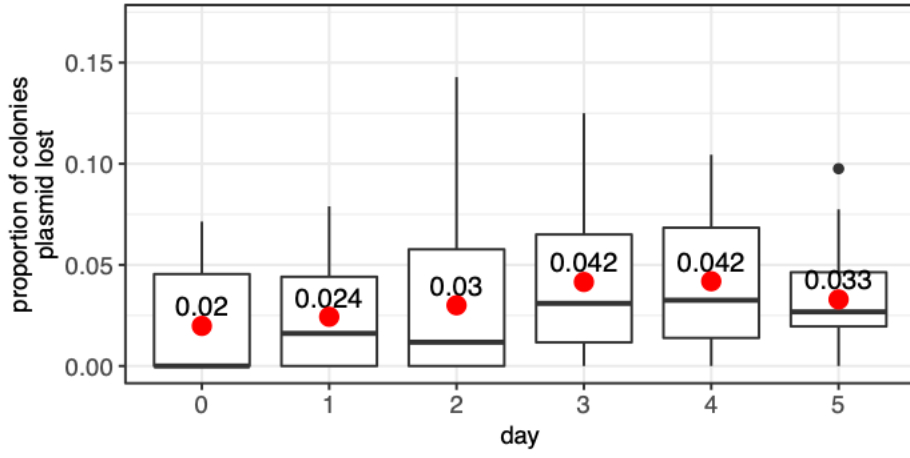
B

Figure S3.6: *ilux-Ecoli* **injected into flies show minimal plasmid loss over time.** To ensure *ilux-Ecoli* maintain plasmid expression during time course experiments, we measured the ratio of luminescent vs non-luminescent colonies from injected flies over 5 days. To this end, we injected flies with 34 nL of *ilux-Ecoli* at an optical density of 6 (approximately 680,000 CFUS) and measured the number of *ilux⁺* and *ilux⁻* colonies over the course of 5 days. A) We found over the course of 5 days mean plasmid loss remained below 5%. B) Total counts of colonies with plamid loss. C) We also measured the amount of plasmid loss experienced by *ilux-Ecoli* injected into ampicillin fed flies and we found no differences. Red dots listed on plots represent the mean of each condition. The first time point (day 0) was marked as NA since flies were homogenized within 30 minutes of injection and not expected to show differences due to presence or absence of antibiotic in the food.
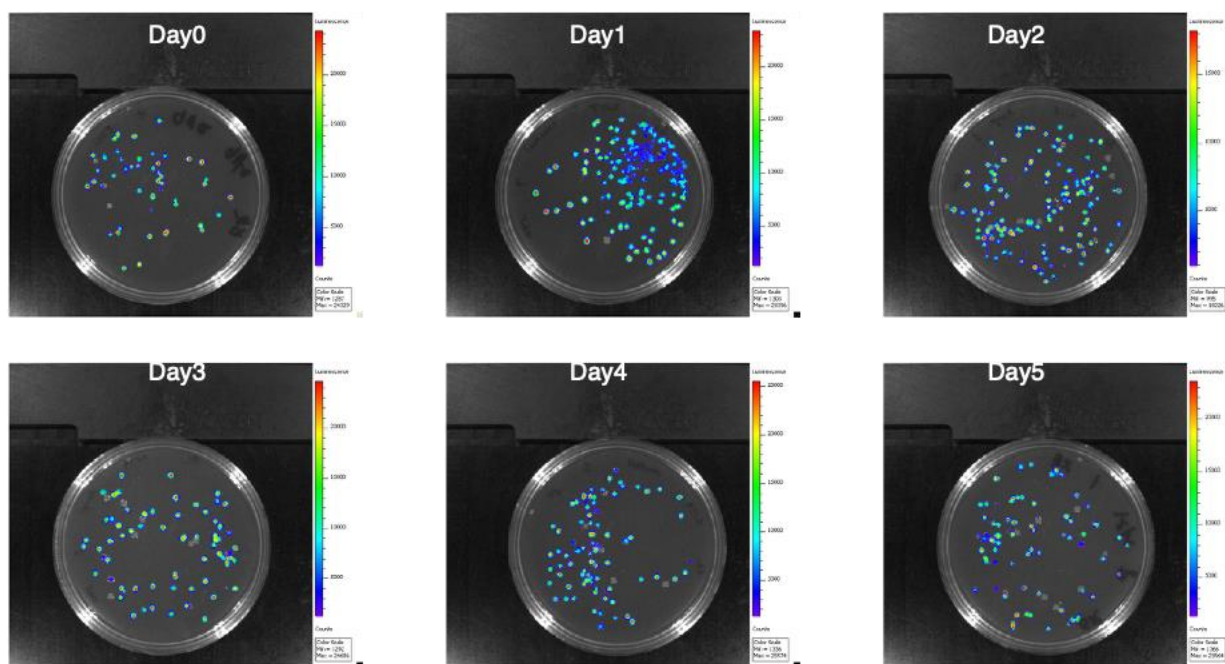
Figure S3.7 : Representative images of colonies used to measure plasmid loss. To assess ilux+ and ilux- colonies, plates were imaged and colonies assessed for bioluminescence emission. Positive colonies were determined as any colonies showing signal above the background level of radiance.
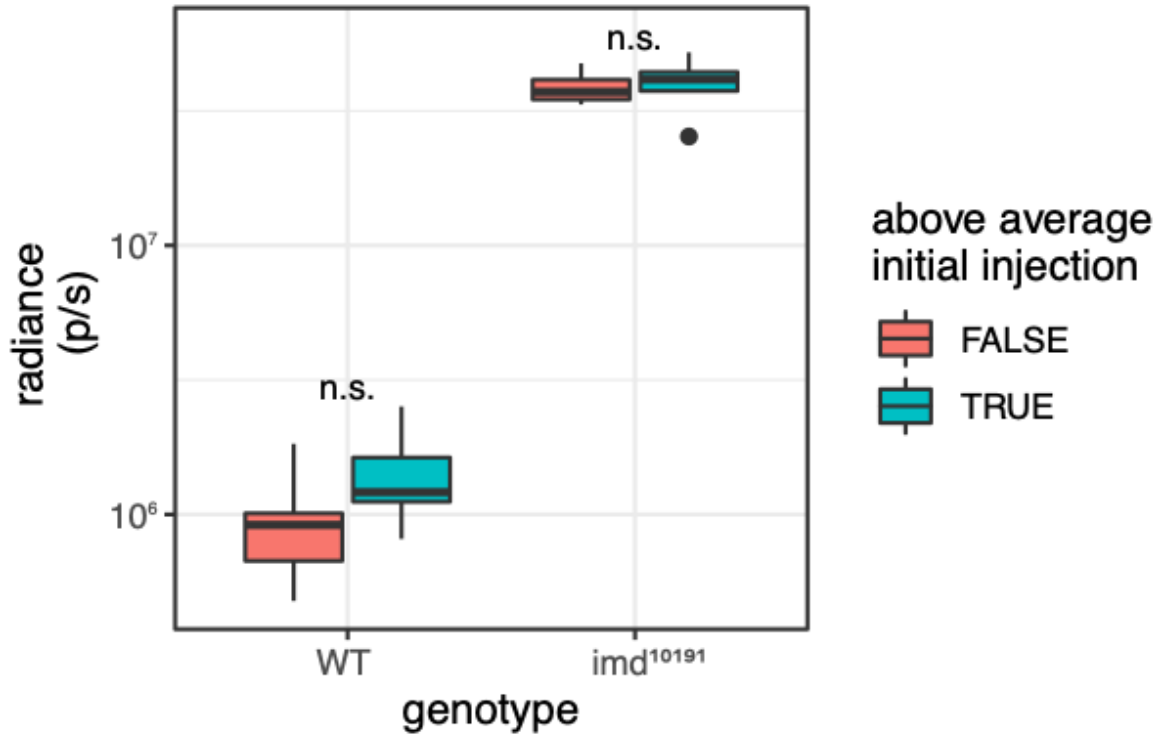
Figure S3.8: Initial dose of infection does not result in differences in final load in Oregon-R or imd[10191]. In order to determine whether the initial load of infection for each genotype infected in Figure 4 resulted in differences in the final radiance (upon death or the end of the time course), we took the mean injected radiance for each genotype and categorized flies as having received below average load (pink) or above average load (cyan). We then tested for differences in the mean using a Welch unpaired two sample t-test. We found that for both Oregon-R and imd[10191] initial bacterial load did not result in differences in the final load. Oregon-R high vs low: $p=0.18$, imd[10191] high vs low: $p=0.78$.
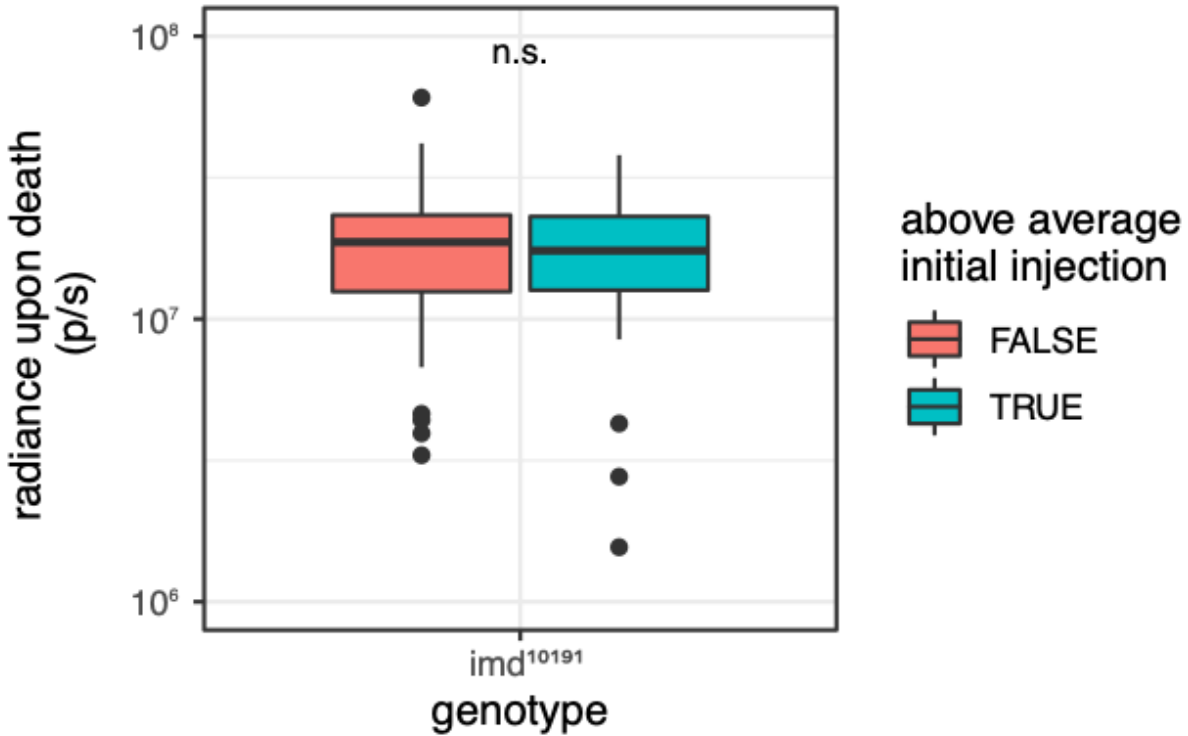
Figure S3.9: Initial dose of infection does not result in differences in final load in imd[10191].
Given the higher number of biological replicates(n=48) of this experiment (Figure 5) over the previous (n=12) (Figure 4), we asked if the initial load of infection for imd[10191] resulted in differences in the radiance upon death. We took the mean injected radiance and categorized flies as having received below average load (pink) or above average load (cyan). We found that the initial bacterial load did not result in differences in the final load. imd[10191] high vs low: p=0.78, Welch unpaired two sample t-test.
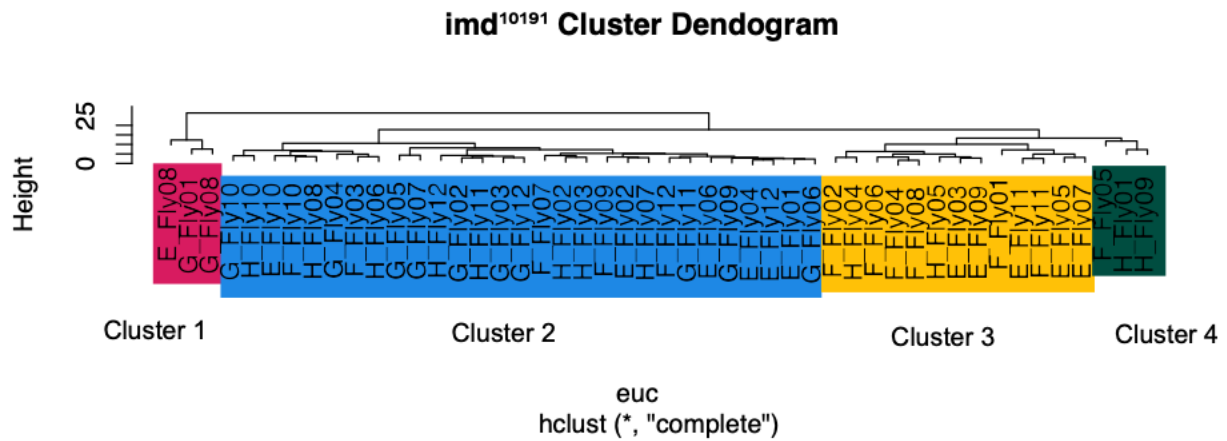
Figure S3.10: Unsupervised clustering of imd^{10191} results in 4 distinct clusters.
In order to group infection profiles we performed hierarchical clustering using Euclidean dissimilarity on log transformed radiance data. Clusters were assigned based on the groups resulting from the first 3 branching points of the dendrogram. For this analysis individual flies were named after their position in the plate with the letter corresponding to a row in the plate and the fly number corresponding with the plate column.
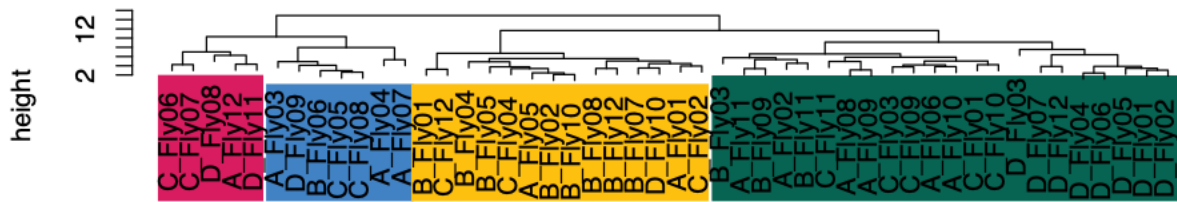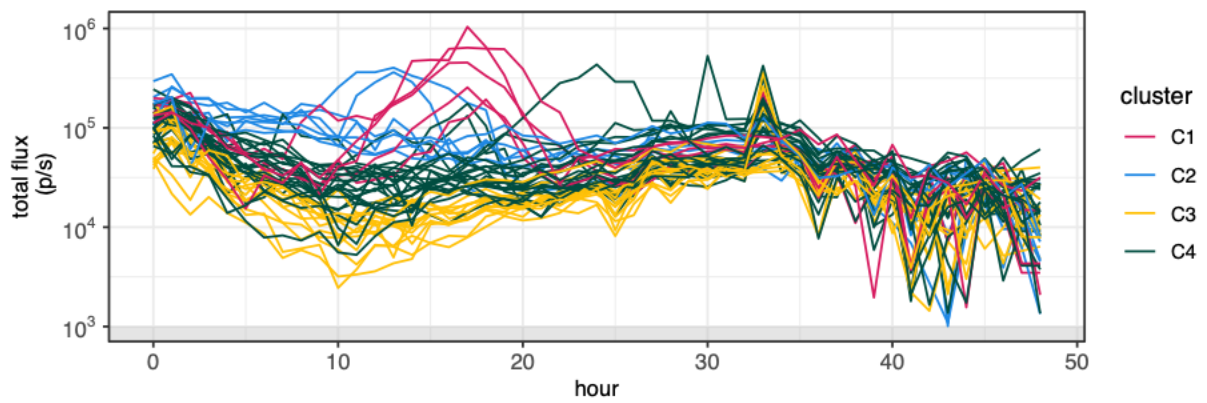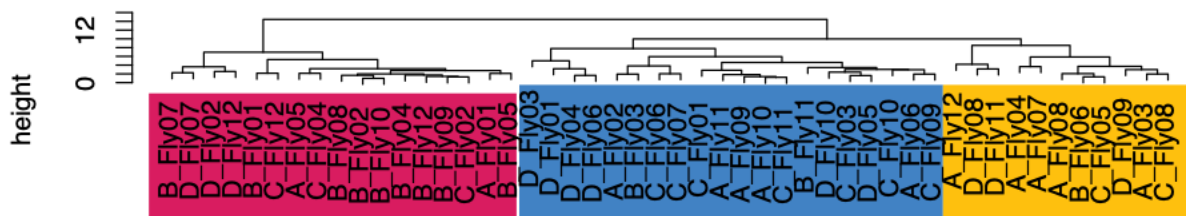
Figure S3.11: Time of death does not correlate with assigned cluster. We asked if there was any cluster that was enriched for a particular time of death. We performed Welch unpaired two sample t-test on every cluster combination and found that no cluster mean appeared to be significantly different from the rest. C1 vs. C2 p=0.7796, C1 vs. C3 p=0.2054, C1 vs. C4 p=0.2739, C2 vs. C3 p=0.05494, C2 vs. C4 p=0.3123, C3 vs. C4 p=0.6186.

Figure S3.12: Unsupervised clustering of Oregon-R infection profiles does not separate trajectories by infection resurgence. In order to group infection profiles, we performed hierarchical clustering using Euclidean dissimilarity on log transformed radiance data. We noticed that several flies showed an increase in bacterial load between 10-25 h. A) Clustering using the full data, clusters were assigned based on the first 3 branching points of the dendrogram. For this analysis individual flies were named after their position in the plate with the letter corresponding to a row in the plate a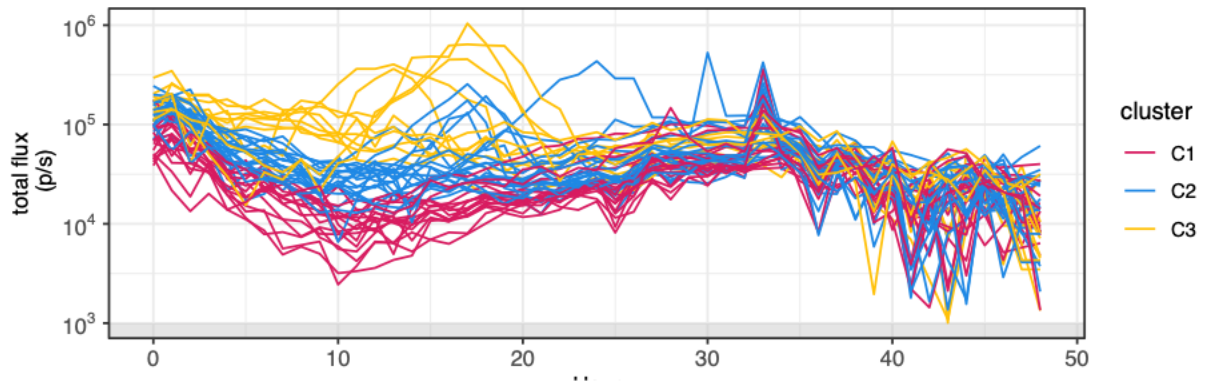nd the fly number corresponding with the plate column. B) Coloring of individual infection profiles shows that initial clustering fails to group profiles that show a resurgence in bacterial load. C) We suspected that the general convergence of the data at around hour 35 and the noise after the 40 h mark may be affecting the clustering, so we performed a second round of clustering using only data from before hour 30. D) Coloring of individual infection profiles based on clustering shows that the clustering of the censored data similarly does not group profiles based on our feature of interest.
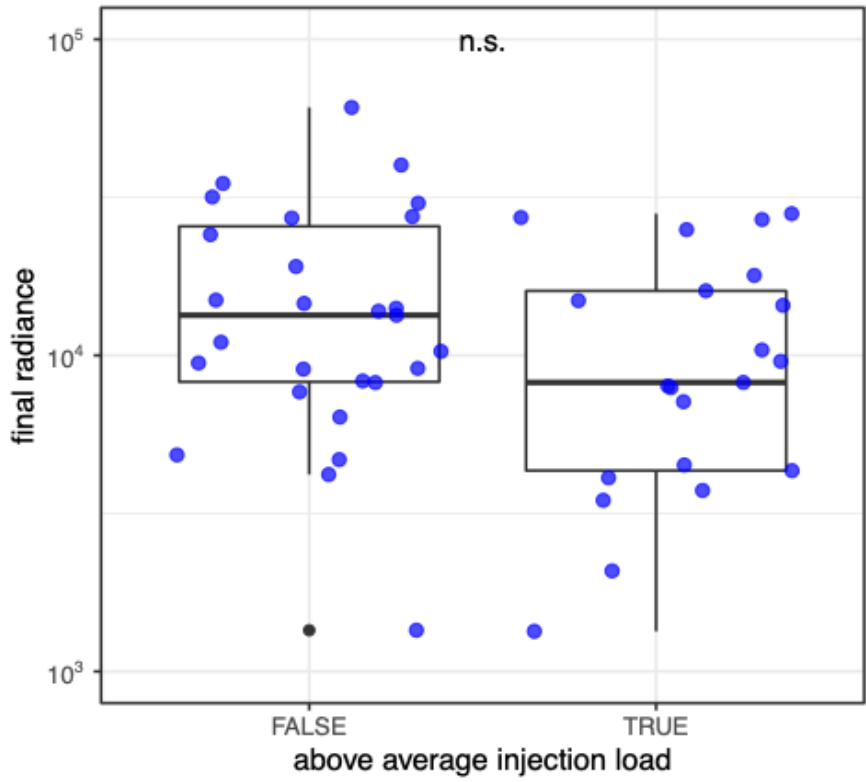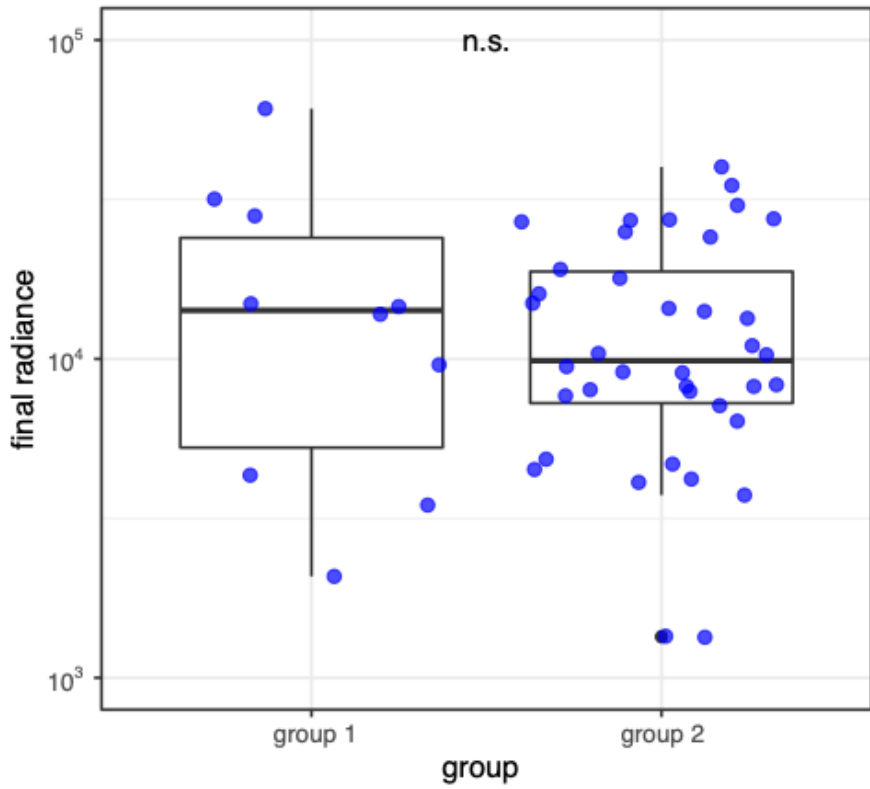
Figure S3.13: Initial load of infection and cluster do not result in differences in final radiance.
A) In order to determine if the initial load of infection for Oregon-R resulted in differences in the final radiance, we took the mean injected radiance for the line and categorized flies as having received below average load or above average load. We found that the difference in initial bacterial load did not result in differences in the final load. Oregon-R high vs low: p=0.51, Welch two sample t-test. B) To determine whether the unusual increase in bacterial load between hours 10-25 resulted in differences in the final radiance, we took the final loads of flies falling into group 1 (resurgence in infection) and cluster 2 (monotonically decreasing infection). We performed a Welch two sample t-test for group1 vs group2 and found that they were not significantly different (p=0.41).
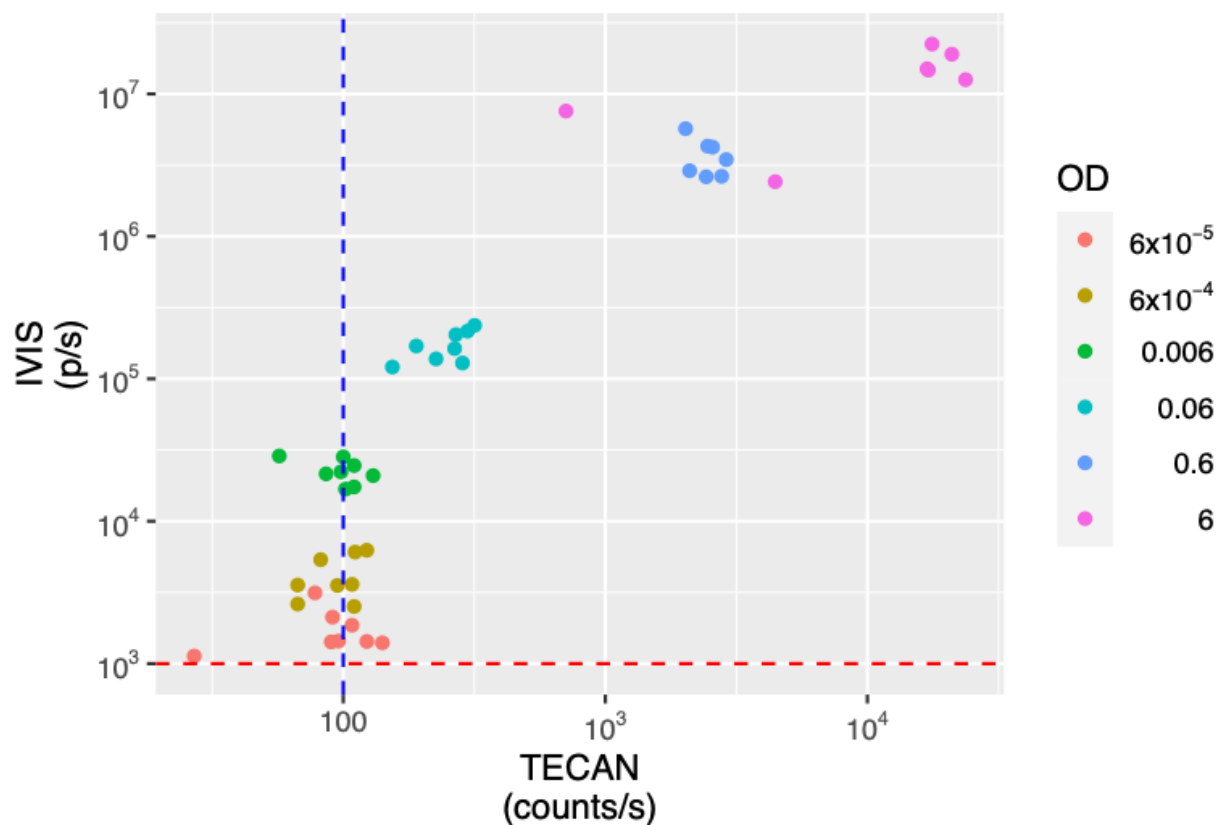
Figure S3.14: Autobioluminecent bacteria is detectable by both IVS and plate readers. To determine how amenable this method is to more cost effective equipment set-ups we compared readings from flies between our set up (IVIS) and a TECAN plate reader. Flies were injected with 6 doses of bacteria ranging from approximately 680,000 CFUs/fly (OD=6) to less than 10 CFUs (OD=6x10^{-5}). Thresholds of detection for both equipment are marked as a red dashed line for the IVIS and blue dashed line for the TECAN plate reader. We observe that the IVIS is able to distinguish injection doses above an OD of 6x10^{-4} while the plate reader set up is able to distinguish doses above an OD of 0.006. Thus while the IVIS is more sensitive at the lower range of detection both equipment are able to distinguish doses at above an approximately 600CFUs/fly.

CHAPTER 4
Tissue-specific labeling of nascent RNA in *Drosophila melanogaster*

4.1 Abstract

Expression profiling of fragile tissues and rare cell types remains challenging, even with single-cell methodologies. One potential solution to this is the precise metabolic labeling and isolation of RNA from tissue or cell types of interest. In *Drosophila,* the chemical tools available for such labeling is limited. Here we adapted a previously described system for the specific labeling and isolation of nascent RNA in the immune responsive fat body. This method uses a protected adenosine analog carrying a 2' azido moiety (2'AZ), in combination with exogenous penicillin G amidase (PGA) under the control of a Gal4-UAS system. We find that orally administered nucleoside analog can be effectively taken up by adult *Drosophila* to achieve RNA integration. Preliminary data suggests that this can be used to specifically label tissue specific RNA though additional verification and optimization is necessary.

4.2 Introduction

Since the emergence of ultra-high-throughput mRNA sequencing, transcriptome profiling has become an essential tool for the understanding of cellular function and identity (Mortazavi 2008, Nguyen 2018). With the development of single-cell sequencing technologies we can phenotype highly heterogenous populations of cells. From single cell studies we have gained invaluable insight into many biological processes, but in a more general sense it has highlighted the large amounts of cellular heterogeneity present even within seemingly homogenous tissues (Nguyen 2018). This has further emphasized need for targeted expression profiling of specific cell groups to disentangle biological processes without confounding signal from extraneous cell populations.

Current methods for targeted expression profiling generally rely on the physical isolation of cell populations of interest (Wang 2015, Nguyen 2018). These methods will dissociate a tissue or organ into individual cells, individual cells are then lysed, and their RNA is barcoded for sequencing. Populations that are rare or difficult to isolate can be enriched, using cell sorting methods or by simply sequencing a large enough population to detect these (Nguyen 2018). In the cases of cell types that are not amenable to dissociation methods, single nucleus sequencing is also available (Nguyen 2018). However, while this method is useful for accurately quantifying relative expression levels it has been observed that single nucleus samples are enriched for nuclear RNAs, lncRNAs and intronic sequences (Nguyen 2018). Therefore, there is a need for a method to specifically profile expression in cell types of interest that does not rely on physical isolation of cells.

One such system that exists to address this need is the metabolic labeling of RNA using protected nucleoside analogs (Singha 2020). This is a two-part system that relies on an inert nucleoside analog and an exogenous enzyme that can activate the nucleoside for integration into RNA (Nguyen 2016, Beasley 2019). Two existing analogs are the 5-ethynyl-uridine (5EU) and 2'azido-adenosine (2'AZ). Both of these can be readily biotinylated via Cu(I)-catalyzed azide-alkyne cycloaddition (CuAAC). However, the addition of a protecting phenylacetyl group (PAC) to these molecules prevents them from being integrated into RNA, rendering the molecule inert (Nguyen 2016, Beasley 2019). The bacterial enzyme Penicillin G amidase (PGA) can cleave off the PAC moiety and endogenous kinases will then phosphorylate the analog, thus activating the analog for RNA integration (Nguyen 2016, Beasley 2019). To confer specific labeling, we can place the PGA gene under control a tissue or cell specific *cis*-regulatory element. Though the protected analog will diffuse indiscriminately throughout the system, PGA will only be expressed

129

in a subset of cells. In these, the analog will be uncaged and thus we confer specificity of RNA labeling (Nguyen 2016, Beasley 2019). Labeled RNA can subsequently be biotinylated, enriched and identified using RNA sequencing (Singha 2020).

The aforementioned system, while powerful, has only been established in cell culture and has yet to be adapted for whole organism use (Nguyen 2016, Beasley 2019, Singha 2020). In *Drosophila,* genetic tools already exist for the specific expression of genes in select cells and tissues. We posited that the GAL4-UAS would be well suited for this RNA labeling method. This is because of the large number of publicly available GAL4 driver lines makes it easy to express a gene under the control of a UAS-sequence in a cell population of interest.

Here we describe the adaptation of the PGA-PAC system for use in adult *Drosophila melanogaster.* We find that orally administered nucleoside analog 2'AZ is able to be detected in as little as 4 hours after exposure. We then create a UAS-PGA fly line and show preliminary results that we can specifically uncage protected analog PAC-2'AZ in a tissue specific manner using existing GAL4 driver lines. The specificity of this method remains to be verified via RNA sequencing; however, this work provides promising preliminary results to the utility of this method.

4.3 Results

To determine if we can detect RNA incorporation of nucleoside analogs via feeding alone, we supplemented standard corn meal media with the unprotected 5EU or 2'AZ. Previous research demonstrated that *Drosophila melanogaster* larvae could uptake and integrate modified nucleobases simply by feeding (Miller 2009, Ghosh 2015). Though it seemed likely that nucleoside analogs delivered by feeding should be similarly incorporated in RNA, this remained to tested. To

assess this, we placed 2nd instar wild type (WT) larvae on 1g of standard fly food supplemented with 15mg of either 5EU (63mM) or 2'AZ (54mM) for 16 hours. We then extracted RNA from batches of 10 larvae and for comparison we extracted RNA from larvae that were not fed any analog. RNA was then biotinylated, and integration of the analog into RNA was visualized using streptavidin-HRP with a chemiluminescent substrate (see Methods, Figure 4.1 A, B). we found that both 5EU and 2'AZ should strong signal of incorporation above background.

Next, we wanted to test if we could achieve uptake and RNA integration of the modified nucleoside in adult flies. We suspected that adult flies would likely eat less than larvae and as such may not be consume enough food, and therefore analog, to achieve detectable signal. To test this, we placed 4-day old WT on 1g of standard fly food supplemented with 15mg of 5EU (63mM) for 16 hours. We found very strong signal for RNA labeling with nucleotide analog in comparison to unfed flies (Figure 4.1 C). This showed that the nucleoside analog can be orally administered via fly food to achieve detectable levels of RNA labeling. Given that azido functional groups can be biotinylated via copper-free strain-promoted azide-alkyne cycloaddition (SpAAC) which is less damaging to RNA and shows lower levels of background labeling we decided to continue with 2'AZ for future experiments rather that 5EU (Nguyen 2016, Nainar 2016).

Next, we wanted to determine how low a concentration of nucleotide analog could be used to achieve detectable RNA labeling. We tested out three concentrations of 2'AZ by adding either 15mg, 10mg or 5mg to 1 g of fly food to achieve concentrations of 54mM, 36mM or 18mM of nucleoside. We allowed adult flies to feed for 36 hours and found labeling for all three concentrations (Figure 4.1 D). Lastly, we wondered what the lowest concentration and exposure time could be used to achieve detectable labeling. We found we could detect RNA labeling in as little as 4 hours with a concentration of 14mM (Figure 4.1 E). Despite seeing signal at a

concentration of 3.6mM at 4-8 hours this signal is still low and is entirely lost by 12 hours. Additionally, all tested concentrations showed a decrease in signal at 24 hours of exposure. It is possible that this is an artifact of this particular dot blot experiment or that the nucleoside analog may display a degree of photosensitivity. Additional experiments will be necessary to ascertain this.

Having shown that we could achieve detectable levels of RNA labeling by simply feeding *Drosophila* adults 2'AZ, we wanted to assess the level of background uncaging of our protected analog. Given that this systems utility hinges on the fact that the protected modified nucleoside cannot be incorporated into RNAs, we wanted to determine if there was any degree of background incorporation. We posited that if the protecting group of the analog were to be spontaneously lost that this would allow for the integration of the analog into the nucleotide and result in nonspecific labeling of RNA. To test the level of background incorporation, we fed adult WT flies with the protected PAC-2'AZ for either 12 or 24 hours. For comparison we also included flies fed with unprotected 2'AZ as well as a sample of flies that were not fed any nucleoside (Figure 4.2 A). We found that flies fed with PAC-2'AZ showed a minimal amount of signal just above that of background. This experiment shows that nonspecific RNA labeling is present at relatively low level when using PAC-2'AZ. However, more quantitative measurements of the extent of this labeling still need to be done to determine the extent of this labeling.

Having established oral uptake of analog nucleosides as a feasible way to achieve RNA labeling, we then needed to generate PGA lines under the control of a UAS promoter. Two lines were generated, we referred to these fly lines as PGA1 and PGA2 and used PGA2 for our fly crosses with GAL4 drivers (see Methods). Once these lines where generated, we decided to test the potential side effects of labeling RNA on the PGA2 line. Previous work in cells has shown that

132

the addition of nucleobase analogs demonstrates some degree of toxicity, partially due to the DMSO used to suspend the analog (Singha 2020). To test for detrimental effects, we fed PAC-2'AZ to adult PGA2 flies controlled by a ubiquitous GAL4 driver (UB-PGA2) and to WT flies that are incapable of uncaging the PAC-2'AZ. To assess the negative effects of feeding flies solvent, which is used to suspend PAC-2'AZ, we included a condition of flies fed with 5% DMSO by volume of fly food (Figure 4.2 B). We monitored survival over the course of 15 days and found that WT flies fed with only the DMSO died faster than flies fed with 2'AZ in DMSO. The PGA expressing flies supplied with the PAC-2'AZ in DMSO, showed the highest survival. We suspect that the differences in the survival of flies fed only DMSO are because these flies received a higher concentration of the solvent than flies that where fed with the analog suspended in DMSO. Nonetheless among the flies fed either the protected or unprotected nucleotide there does not appear to be a difference in survival for timescales shorter than 9 days. Given that we can detect strong signal in adults within 12 hours of exposure we do not foresee this difference in survival being a problem for the application of this system.

Lastly, to determine the potential effectiveness for this method to specifically label RNA, we measured the signal detected from each of the individual parts of the system (Figure 4.3 A). We chose a ubiquitous GAL4 driver (UB) as well as fat-body specific GAL4 driver (FB) to test this (see Methods). We found that crossing of our GAL4 lines to the PGA2 line resulted in flies that show uncaging and incorporation of the protected analog {Figure reference?}. The level of signal achieved using the 2-part system was not as high as the signal produced using the unprotected nucleoside alone, but it was still above background signal. Additionally, the individual uncombined FB, UB and PGA2 lines fed with PAC-2'AZ do show a low level of RNA labeling, despite the absence of PGA expression. We found that the fat-body specific driver showed greater

signal of RNA labeling (albeit still low) than the ubiquitous driver. Two potential explanations exist for this that are not necessarily mutually exclusive. The first is that the FB driver may have a higher level of nonspecific uncaging of the protected nucleoside, even in the absence of the PGA gene. Evidence for this can be seen by the relatively higher signal observed from the FB line as opposed to the UB line which shows a signal closer to the background. It is unclear why this would be the case since we do not expect PGA homologs to be present in the background of any of these lines. The second explanation could be the FB driver is simply more active than the UB driver. This seems very likely as strong UB GAL4 drivers often negatively impact the health of a line. In either case the GAL4 driven expression of PGA is able uncage and integrate orally administered PAC-2'AZ in adult *Drosophila*. The exact extent of this specificality remains to be determined.

4.4 Discussion

Here we show that we can adapt a system for specific RNA labeling in *Drosophila melanogaster* using PAC-2'AZ and the exogenous PGA gene. We find that we can achieve detectable levels of RNA labeling using orally administered nucleoside analogs in adult flies. We observe minimal levels of PAC-2'AZ background labeling and we show that we can use the PGA gene under the control of the GAL/UAS system to selectively unblock the protected analog. The data gathered here shows promise for utilizing this method in *Drosophila*.

Before this method can be widely used to garner biological insight in *Drosophila*, some amount of optimization and experimentation remains to be done. One simple improvement that can be made is to the UAS-PGA lines. This may help achieve stronger specificity and signal. Currently, both PGA expressing lines control the gene's expression using a 5x UAS cassette. However, it has been shown that inclusion of up to 40x UAS sites can achieve stronger expression

134

of the gene being controlled (Pfeiffer 2010). In addition to this, the specificity of labeling RNA remains to be quantitatively measured. Sequencing of biotinylated RNA from lines fed with PAC-2'AZ that either express or do not express the PGA gene is still necessary. This will allow us to directly quantify the amount of nonspecific RNA labeling that occurs simply by the inclusion of the protected nucleoside analog. Determining the levels and sources of non-specific RNA labeling by this method will be critical for assessing the viability of this method.

4.5 Methods

4.5.1 Media Preparation

All flies were grown on standard corn meal media (Brent and Oster 1974) at 25˚C with 12-hour light and dark cycles. For flies and larvae fed with analog nucleoside, media was prepared by mixing 1g of warmed corn meal media with nucleotide suspended in DMSO. Media was then poured into 15ml plastic tubes and stopped with cotton plugs. Unprotected and protected nucleosides were provided by the Spitale lab at UCI and were stored at -20˚C.

4.5.2 Fly lines

For experiments wild type fly lines A4 were acquired from the Drosophila Synthetic Population Resource center at UCI (King 2012). The ubiquitous expressing GAL4 line *w;Arm>GAL4/TM3, Sb Ser* was obtained from the Arora lab at the UCI. The fat-body specific GAL4 line *w[1118]; P{w[+mC]=Cg-GAL4.A}2* was ordered from the Bloomington stock center (Stock number 7011).

4.5.3 Generating new fly lines

UAS PGA lines where generated by PCRing the 5' Flag-tagged PGA gene from the pWIG-Flag-coPGA-IRES-GFP plasmid. Using Gibson assembly, we inserted this into the pUAST-attB vector (Pfeiffer 2010). This vector contains a 5X UAS cassette, a heat shock promoter (hsp), and we cloned the 5' Flag-tagged PGA gene downstream of the hsp promoter. The vector also has an attB site, making it suitable for phiC31 mediated integration into a specific attP site, and a mini-white gene to allow for screening of transgenic flies. The full construct (pUAS-attB-FLAG-PGA) was shipped to BestGene. Here the fly lines were made by injecting the full construct into attP2 flies (Bloomington Drosophila Center Stock #8622, full genotype: *y1 w67c23; P{CaryP}attP2*) and made homozygous transgenic stocks using the mini-white marker. The attP site in this stock is on Chr 3L.

4.5.4 Feeding experiments

For feeding experiments flies or larvae were placed on media containing nucleoside analog. Tubes where then kept at 25˚C with 12-hour light and dark cycles for the duration of the experiment. At the end of an experiment flies where anesthetized using a CO2 and 3 flies per tube with 250µl of TRIzol reagent (Thermofisher 15596026) were frozen by placing in -80˚C freezer for later extraction

4.5.5 CuAAC chemistry

To detect the relative levels of labeled RNA, total RNA was extracted using Zymo Research Direct-zol RNA Extraction Kits. RNA was quantified using NanoDrop 3000 (Thermo Fisher).

CuAAC protocol was obtained from Kim Nguyen from the Spitale lab. To biotinylate RNA we mixed 25 µl of "click" chemistry cocktail (1:1:1 solution of 12 mM CuSO4, 2mg/mL tris-hydroxypropyltriazolylmethylamine, and 2.1 mg/mL of NaAscorbate), 5 µl of 10 mMEDTA, 5 µl of 10 mM biotin-azide or biotin alkyne, and 2-10 µg of RNA for a final 50 µl reaction. Mixture

was incubated for 30 min at room temperature with gentle shaking. Total RNA was then purified using Zymo RNA Clean & Concentrator kit, with the final elution step performed twice. RNA was then stored at -80˚C.
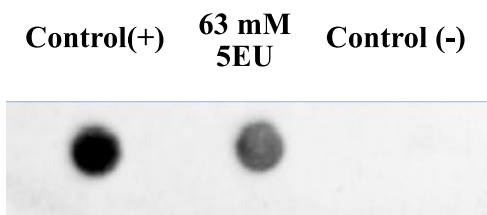
4.5.6 Dot blots

To detect the relative level of biotinylated to non-biotinylated RNA we performed dot blots using Hybond-N+ (GE Healthcare). Membrane was equilibrated in 2x SSC buffer for 5 minutes (20x SSC buffer 87.65 g NaCl, 44.1g Na_3Citrate, final 500 mL volume). Membrane was air dried for 10 min, then 2µg of RNA per sample was loaded onto membrane in 1µl increments. RNA was then UV crosslinked to membrane using Stratalinker UV crosslinker (254 nm UV, 2x 1200 µjoules). Following this membrane was incubated in blocking buffer for 30 minutes with rotation (Blocking buffer : 3.65 g NaCl, 1.2 g Na_2HPO_4, .5 g NaH_2PO_4, 50 g SDS, final volume 500mL water). Blocking buffer was discarded and membrane was then incubated with 1 mL of 1:5000 solution of blocking buffer and Pierce™ High Sensitivity Streptavidin-HRP (thermo fisher) for 5 minutes. Membrane was then washed twice with a 1:10 dilution of blocking buffer, for 20 minutes each with rotation. Lastly, membrane was washed twice with a 1x washing solution for 5 minutes with rotation (washing solution : 6 g Tris base, 2.9 g NaCl, 1 g $MgCl_2$, final volume 500mL, pH 9.5). Membrane was then incubated with 1:1 Pierce™ ECL Western Blotting Substrate (Thermo scientific) and imaged using the ChemiDoc XRS+ imager (Bio-Rad)
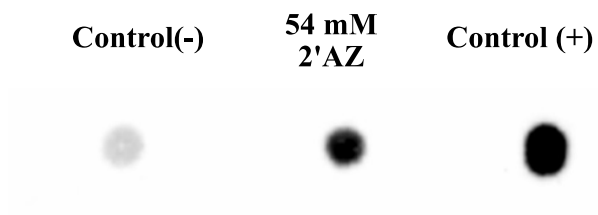
4.6 References

Beasley, S., Nguyen, K., Fazio, M., & Spitale, R. C. (2019). Protected pyrimidine nucleosides for cell-specific metabolic labeling of RNA. *Tetrahedron Lett.*, *59*(44), 3912–3915. https://doi.org/10.1016/j.tetlet.2018.09.040.Protected

Brent, M. M., & Oster, I. I. (1974). Nutritional substitution -- a new approach to microbial control for Drosophila cultures. *Drosophila Information Service*, (52), 155–157.

Ghosh, A. C., Shimell, M., Leof, E. R., Haley, M. J., & O'Connor, M. B. (2015). UPRT, a suicide-gene therapy candidate in higher eukaryotes, is required for Drosophila larval growth and normal adult lifespan. *Scientific Reports*, *5*(July), 1–11. https://doi.org/10.1038/srep13176

King, E. G., Merkes, C. M., McNeil, C. L., Hoofer, S. R., Sen, S., Broman, K. W., … Macdonald, S. J. (2012). Genetic dissection of a model complex trait using the Drosophila Synthetic Population Resource. *Genome Research*, *22*(8), 1558–1566. https://doi.org/10.1101/gr.134031.111

Miller, M. R., Robinson, K. J., Cleary, M. D., & Doe, C. Q. (2009). TU-tagging: Cell type-specific RNA isolation from intact complex tissues. *Nature Methods*, *6*(6), 439–441. https://doi.org/10.1038/nmeth.1329

Mortazavi, A., Williams, B. A., McCue, K., Schaeffer, L., & Wold, B. (2008). Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nature Methods*, *5*(7), 621–628. https://doi.org/10.1038/nmeth.1226

Nainar, S., Beasley, S., Fazio, M., Kubota, M., Dai, N., Corrêa, I. R., & Spitale, R. C. (2016). Metabolic Incorporation of Azide Functionality into Cellular RNA. *ChemBioChem*, *17*(22), 2149–2152. https://doi.org/10.1002/cbic.201600300

Nguyen, Q. H., Pervolarakis, N., Nee, K., & Kessenbrock, K. (2018). Experimental considerations for single-cell RNA sequencing approaches. *Frontiers in Cell and Developmental Biology*, *6*(SEP), 1–7. https://doi.org/10.3389/fcell.2018.00108

Pfeiffer, B. D., Ngo, T. T. B., 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*(2), 735–755. https://doi.org/10.1534/genetics.110.119917

Singha, M., Spitalny, L., Nguyen, K., Vandewalle, A., & Spitale, R. C. (2021). Chemical methods for measuring RNA expression with metabolic labeling. *Wiley Interdisciplinary Reviews: RNA*, *12*(5), 1–13. https://doi.org/10.1002/wrna.1650
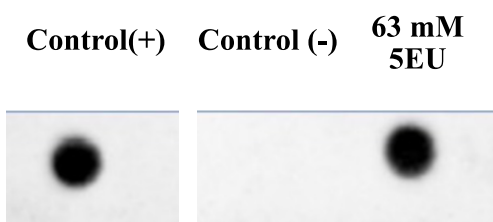
4.7 Figures

A)

Control(+)　　63 mM 5EU　　Control (-)

B)

Control(-)　　54 mM 2'AZ　　Control (+)

C)

Control(+)　　Control (-)　　63 mM 5EU

D)

54 mM　　54 mM　　36 mM

36 mM　　18 mM　　18 mM

Control(+)

E)

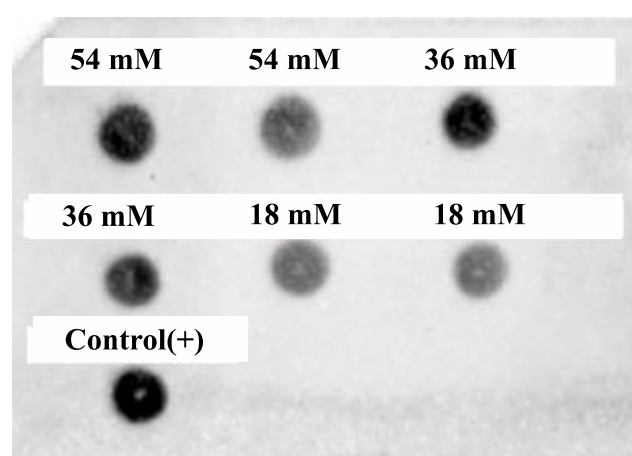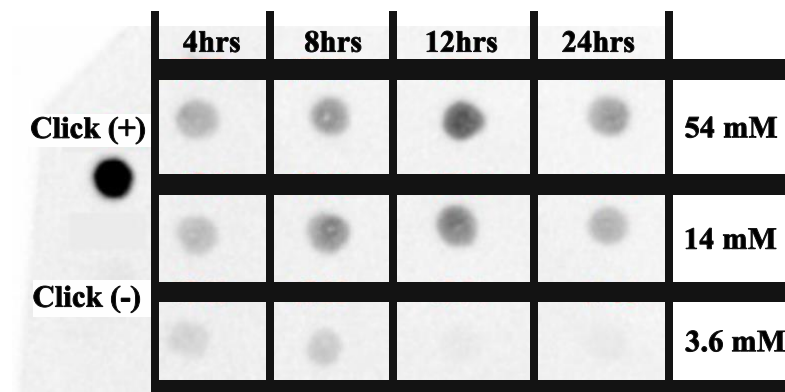|	4hrs	8hrs	12hrs	24hrs	
Click (+)					54 mM
					14 mM
Click (-)					3.6 mM

Figure 4.1 Orally administered nucleoside analog can be detected in RNA of larval and adult *Drosophila*. Unprotected nucleoside analogs were mixed into standard cornmeal media for oral uptake. RNA from animals fed for 16 hours was then extracted, biotinylated and loaded onto blot paper. The relative amount of biotinylated RNA was then detected using streptavidin horseradish peroxidase and a light emitting substrate (see methods) A) 5EU fed- and B) 2'AZ fed 2nd instar larvae both showed signal above background (RNA extracted from non-fed animals). RNA for positive controls were from cells in culture after being supplied with the nucleoside analog. C) Adult flies show uptake of 5EU and incorporation above the background. D) Adult flies fed 2'AZ for 36 hours show RNA labeling in concentrations as low as 18mM of modified nucleoside. As positive control RNA from 2nd instar larvae fed with 54 mM of 2'AZ was used. E) We quantified the lowest exposure time and concentration of nucleoside needed to detect RNA labeling via dot blot. We detect labeling in as little as 4 hours with concentrations of 14mM.
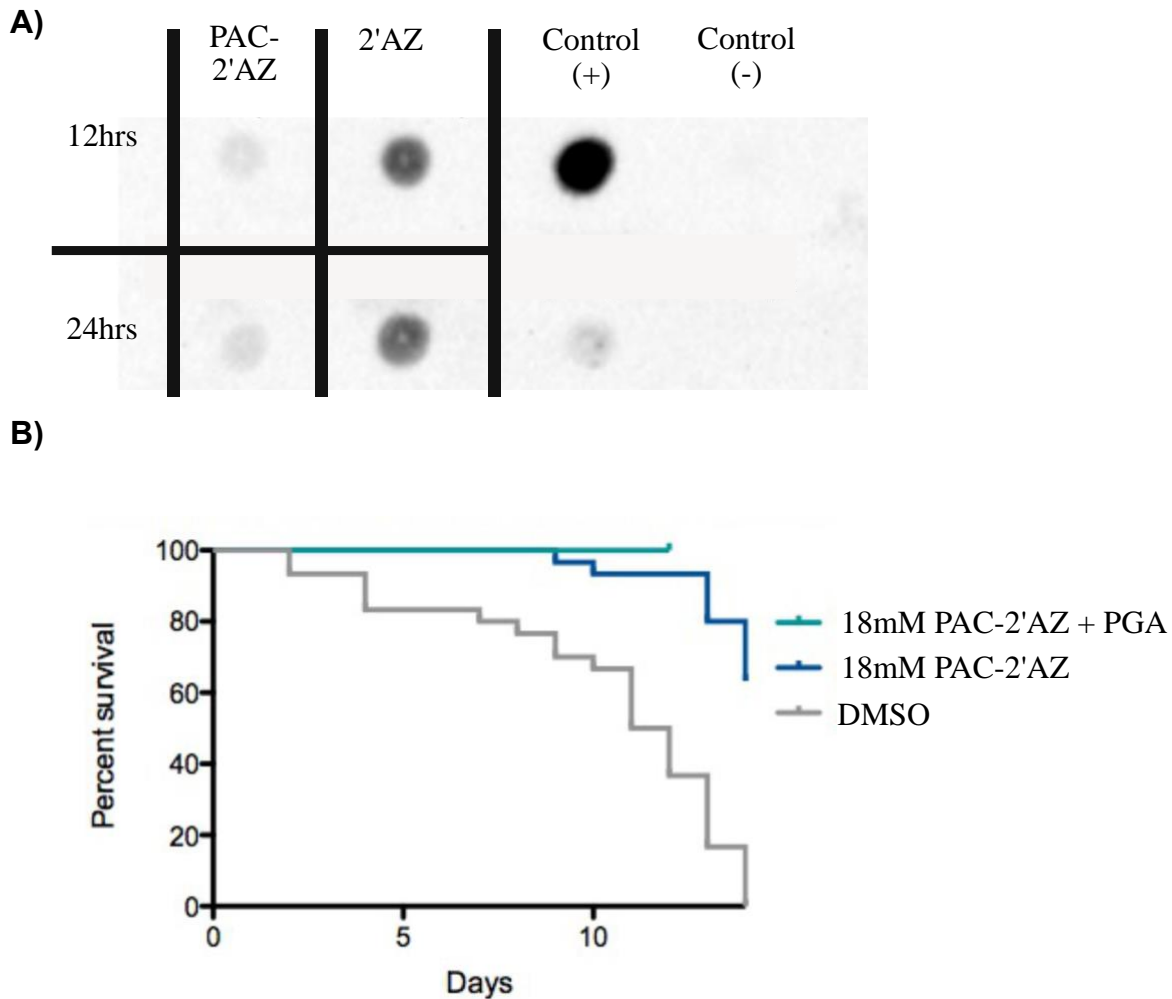
Figure 4.2 PAC-2'AZ shows minimal labeling and survival effects. A) To assess the level of background uncaging of PAC-2'AZ we fed adult flies with 12mM concentration of the nucleoside for either 12 or 24 hours. For comparison we included flies fed with 18mM of 2'AZ as well as a positive control, RNA from 2nd instar larvae fed with 54 mM 2'AZ, and a negative control, adults fed no nucleoside. B) To determine any potential detrimental effects that the supplied protected nucleoside may be having on adult flies we fed flies expressing the PGA gene throughout the body with PAC-2'AZ. For comparison we included non-PGA expressing WT flies and flies fed DMSO. Uncaging and integration of the 2'AZ does not appear to negatively impact the survival of flies.
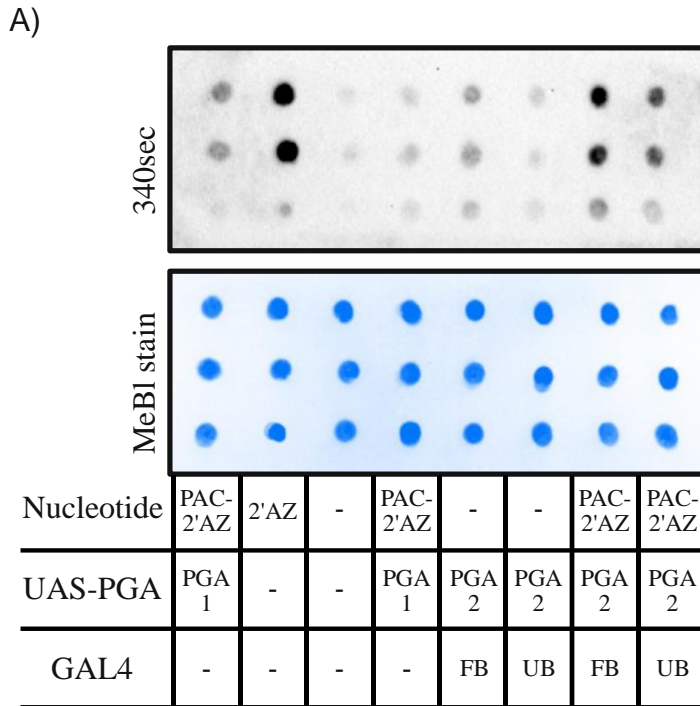
A)

Nucleotide	PAC-2'AZ	2'AZ	-	PAC-2'AZ	-	-	PAC-2'AZ	PAC-2'AZ
UAS-PGA	PGA 1	-	-	PGA 1	PGA 2	PGA 2	PGA 2	PGA 2
GAL4	-	-	-	-	FB	UB	FB	UB

Figure 4.3 PGA-2 line can be used to specifically label RNA. A) To determine if the ability of the UAS-PGA system to specifically label RNA we tested 2 GAL4 drivers with the PGA lines. We supplied flies with either PAC-2'AZ or 2'AZ. We find that the GAL4-UAS+PGA system is able to uncage PAC-2'AZ and label RNA. Two UAS+PGA lines where generated a "weak" PGA line (PGA1) and a "strong" PGA line (PGA2). Additionally, 2GAL4 drivers were tested a ubiquitous driver (UB) and a fat body specific driver (FB). Three biological replicates (rows) were tested per condition (columns). Methylene blue staining was also performed to ensure equal amounts of loaded RNA.

CHAPTER 5
Future directions

In the previous chapters, I detailed my works on quantifying variation in the *Drosophila* immune response. In Chapter 2, I quantified genetic sources of variation between two genetic lines of *D. melanogaster*. I showed that the relative contribution of *cis* and t*rans* effects to expression divergence in the *Drosophila* fat body is condition specific. In Chapter 3, I demonstrated a new method for non-invasively tracking bacterial infection over time using autoluminescent bacteria. I showed the utility of this method to elucidate sources of variation for individual infection outcomes. In Chapter 4, I presented a novel method for the precise chemical labeling and isolation of cell-specific RNA in *D. melanogaster*. I adaptive this chemical-genetic system for use *in vivo* and showed preliminary data that suggests that can be used to precisely label nascent RNA in a tissue specific manner. Taken together these works identify genetic sources of variation in immune response and provide methods for exploration of non-genetic sources. Below I outline two logical extensions of the work discussed in chapters 2 and 3.

5.1 Pathway specific *cis* and *trans* variation

A great deal of work has been invested into understanding the mode and tempo of expression divergence both within and between species (Wittkopp 2012). While previous work has been instrumental for establishing a foundational understanding of this subject, we are still missing a broader understanding of the rules and mechanisms that dictate the mode in which expression divergence arises (Signor 2018). One major limitation in the field is that a great deal of the work to date has averaged signal across multiple tissues and therefore biological pathways (Wittkopp 2004, Landry 2005, Wittkopp 2008, McManus 2010, Gonclaves 2012, Coolon 2014, Osada 2017, Benowitz 2020, Frochaux 2020). Different biological processes may be under distinct evolutionary pressures which in turn may affect the relative contributions of *cis* and *trans* changes, but these

differences cannot be examined in these data. To achieve this broader level of understanding, we require a more precise view of the mode and tempo of expression divergence that is at the level of individual biological pathways. This view must include diverse biological pathways across multiple conditions and evolutionary distances. By elucidating the relative contributions of *cis* and *trans* effects in these, we can build a framework to better understand the broader mechanisms that dictate how expression variation arises.

In Chapter 2 I aimed to examine the relative contributions of *cis* and *trans* changes specifically in the immune response pathway. In my work I homed in on this pathway by examining expression only in the abdominal fat body, which is the primary immune responsive organ in *Drosophila.* Additionally, by examined expression divergence in this tissue under infected and non-infected settings I was able to show that the relative proportion of *cis* and *trans* effects are condition specific. While this experimental design is an improvement on previous studies that used expression data from whole bodies, it is still not entirely pathway specific. In the same way that previous studies averaged signal across multiple tissue this study is likely averaging across multiple cell types. In larvae the fat body has been observed to have functionally distinct regions, so it is not unreasonable to believe that this may be the case in adults as well (Haunerland 1995). To account for populations of cells that may be involved in distinct biological pathways more precise methods of expression profiling must be used when quantifying the sources of expression divergence.

The use of single cell (SC) and single nucleus (SN) methods in combination with an F1 hybrid experimental design from Chapter 2, will be necessary to precisely quantify the contributions of *cis* and *trans* expression variation at the level of pathway. Pathway level resolution of mode of gene expression divergence will be crucial to advance our understanding in *Drosophila* and

145

multicellular organisms in general. Coupling of these methods will allow for the exploration of pathway specific expression variation the field is currently lacking.

Specific to the fat-body, SN sequencing will be necessary, since the tissue itself is not amenable to SC methods. In fact, work is already being done to optimize SN methodologies for the *Drosophila* fat body (Gupta 2021). SN data can be used to cluster populations of cells across samples prior to performing allele specific expression analysis, ensuring that cells involved in similar pathways are compared to each other. In this manner we can describe the mode of expression divergence in a pathway specific manner. Expanding this experimental design to examine multiple pathways, conditions and evolutionary distances is the next logical step for the field.

From this type of experiment, we would gain invaluable insight into the general biology of the fat body and a broader view of the mode of expression divergence. We would expect that in performing SN in the fat body we should find distinct populations of cells specialized for specific biological roles such as immune response and general metabolism. Moreover, we would be able to compare how conserved these populations are across evolutionary distances. This information would greatly improve our understanding of the general biology of the fat body.

Secondly, we would observe how the mode of expression divergence differs between biological pathways that are under different selections pressures. We would expect that pathways under more intense selection pressures should generally show greater changes in *trans* whereas those under less intense selection should show greater changes in *cis*. Should this expectation be met, this would provide strong evidence to support that the mode of expression divergence may be dependent on the intensity of selection that is occurring on a biological process. Whereas if this expectation is not met then this would suggest that something else must also be involved.

5.2 Using autobioluminescence to examine sources of variation

In *Drosophila* the understanding of sources of inter individual variation in immune response remains underexplored (Duneau 2017, Ellner 2021). This is due in no small measure, to the limitations of the methods used to assess the immune response which generally rely on animal sacrifice. In Chapter 3 of my thesis, I provide a new method for longitudinal tracking of bacterial infection using an autobioluminescent reporter that allows for the observation of individual infection trajectories. This method also allows for the direct linking of individual infection histories to infection outcomes. Additionally, the autobioluminescent construct used in this method is amenable for use in other bacterial strains making this method extremely flexible to the pathogen of choice. This method opens a door to explore genetic and non-genetic sources of infection variation in a high throughput manner.

Given that this area is underexplored it is relatively rich for discovery, but of particular interest is the identification of predictors of infection outcome. Work by Duneau et al 2017 identified the bacteria of *Enterococcus faecalis*, *Providencia burhodogranariea*, and *Providencia rettgeri* as showing a bimodal infection outcome in *Drosophila* lines. A logical next step would be to use this system to identify potential predictors for infection outcome. By transforming these bacteria with the autobioluminescent reporter, measuring infection progression and outcome in the previous lines, we can leverage this bimodal outcome to pinpoint the sources that dictate the observed outcome. We would expect that variation in the inoculation load, initial injury and immune activation will all contribute to the infection outcome. However, it will be interesting to see the effect size of each of these factors. This would provide an unprecedented level of resolution

into infection progression that can be used to build more refined predictive models of infection outcome.

5.3 Concluding remarks

Above I discuss logical extensions to the research presented in previous chapters of my thesis. Specifically, I suggest the utilization of single cell and single nucleus methods in combination with allele specific expression analysis as a means to quantify pathway specific expression divergence. This will hopefully lead to a better understanding of the broader rules that dictate the way in which expression and thus phenotypic divergence. I also suggest an experimental design set up for which to use the method described in Chapter 3, that may aid in the designation of potential predictors of infection out come between genetically identical individuals. The research presented in this thesis only begins to address the sources of variation in in immune response. It expands on decades of previous work into the contribution of *cis* and *trans* effects in immune response, but also points out areas that remain to be expanded on. Similarly, it provides a mere glimpse into the relatively unexplored field of interindividual variation and raises a great many questions as to the sources of this variation that can now be addressed. As the colloquial expression goes, "good research will provide some answers but also open more questions."

5.4 References

Benowitz, K. M., Coleman, J. M., Allan, C. W., & Matzkin, L. M. (2020). Contributions of cis- And trans-Regulatory Evolution to Transcriptomic Divergence across Populations in the Drosophila mojavensis Larval Brain. *Genome Biology and Evolution*, *12*(8), 1407–1418. https://doi.org/10.1093/GBE/EVAA145

Coolon, J. D., McManus, C. J., Stevenson, K. R., Graveley, B. R., & Wittkopp, P. J. (2014). Tempo and mode of regulatory evolution in Drosophila. *Genome Research*, *24*(5), 797–808. https://doi.org/10.1101/gr.163014.113

Duneau, D., Ferdy, J.-B., Revah1, J., Kondolf, H., Ortiz1, G. A., Lazzaro, B. P., & Buchon, N. (2017). Stochastic variation in the initial phase of bacterial infection predicts the probability of survival in D. melanogaster. *ELIFE*, *129*(1), 76–82. https://doi.org/https://doi.org/10.7554/eLife.28298.001

Ellner, S. P., Buchon, N., Dörr, T., & Lazzaro, B. P. (2021). Host-pathogen immune feedbacks can explain widely divergent outcomes from similar infections. *Proceedings of the Royal Society B: Biological Sciences*, *288*(1951). https://doi.org/10.1098/rspb.2021.0786

Frochaux, M. V., Bou Sleiman, M., Gardeux, V., Dainese, R., Hollis, B., Litovchenko, M., … Deplancke, B. (2020). cis-regulatory variation modulates susceptibility to enteric infection in the Drosophila genetic reference panel. *Genome Biology*, *21*(1), 6. https://doi.org/10.1186/s13059-019-1912-z

Goncalves, A., Leigh-Brown, S., Thybert, D., Stefflova, K., Turro, E., Flicek, P., … Marioni, J. C. (2012). Extensive compensatory cis-trans regulation in the evolution of mouse gene expression. *Genome Research*, *22*(12), 2376–2384. https://doi.org/10.1101/gr.142281.112

Gupta, V., & Lazzaro, B. P. (2021). A robust method to isolate Drosophila fat body nuclei for transcriptomic analysis. *BioRxiv*, 1–10.

Haunerland, N. H., & Shirk, P. D. (1995). Regional and Functional Differentiation in the Insect Fat Body. *Annual Review of Entomology*, *40*(1), 121–145. https://doi.org/10.1146/annurev.ento.40.1.121

Landry, C. R., Wittkopp, P. J., Taubes, C. H., Ranz, J. M., Clark, A. G., & Hartl, D. L. (2005). Compensatory cis-trans evolution and the dysregulation of gene expression in interspecific hybrids of drosophila. *Genetics*, *171*(4), 1813–1822. https://doi.org/10.1534/genetics.105.047449

Mcmanus, C. J., Coolon, J. D., Duff, M. O., Eipper-mains, J., Graveley, B. R., & Wittkopp, P. J. (2010). Regulatory divergence in Drosophila revealed by mRNA-seq, 816–825. https://doi.org/10.1101/gr.102491.109.

Osada, N., Miyagi, R., & Takahashi, A. (2017). Cis- and trans- regulatory effects on gene expression. *Genetics*, *206*(August), 2139–2148. https://doi.org/10.1534/genetics.117.201459/-/DC1.1

Signor, S. A., & Nuzhdin, S. V. (2018). The Evolution of Gene Expression in cis and trans. *Trends in Genetics*, *34*(7), 532–544. https://doi.org/10.1016/j.tig.2018.03.007

Wittkopp, P. J., Haerum, B. K., & Clark, A. G. (2008). Regulatory changes underlying expression differences within and between Drosophila species. *Nature Genetics*, *40*(3), 346–350. https://doi.org/10.1038/ng.77

Wittkopp, P. J., Haerum, B. K., & Clark, A. G. (2004). Evolutionary changes in cis and trans gene regulation. *Nature*, *430*(6995), 85–88. https://doi.org/10.1038/nature02698

Wittkopp, P. J., & Kalay, G. (2012). Cis-regulatory elements: Molecular mechanisms and evolutionary processes underlying divergence. *Nature Reviews Genetics*, *13*(1), 59–69. https://doi.org/10.1038/nrg3095