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

Bryan A. Ramirez-Corona, Stephanie Fruth, Oluchi Ofoegbu, Zeba Wunderlich*

Department of Developmental and Cell Biology, University of California, Irvine, CA

*Corresponding Author:
4107 Natural Sciences 2
University of California, Irvine
Irvine, CA 92697
949-824-5959
zeba@uci.edu

Running title: Expression divergence in the *Drosophila* fat body

Keywords: innate immune response; expression divergence; *Drosophila melanogaster*; fat body

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.

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 adaptative, 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 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 allele-specific expression analysis 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 on 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.

Results

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 25.3 variants per gene. The extensive variation in the coding regions allowed us to map, on average, 11.2% (±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 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 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 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 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 S1A). 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 S1B, S1C). 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 1C). Troha and colleagues identified a subset of immune-responsive genes

as core, i.e., the genes 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 in immune-responsive expression, making them suitable for subsequent F1 hybrid experiments.

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 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 S1). 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 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 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; Supplemental Table S2). 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.

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 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.

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 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.

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 (Figure 2E). The control and *Efae*-infected samples

had a greater proportion of *cis-only* genes than the *Smar* samples (control vs. *Smar* $p = 4.0 \times 10^{-6}$, *Efae* vs. *Smar* $p = 6.8 \times 10^{-14}$, 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* $p = 3.5 \times 10^{-4}$, control vs. *Smar* $p < 1.5 \times 10^{-16}$, *Efae* vs. *Smar* $p = 3.1 \times 10^{-11}$, 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* $p < 1.5 \times 10^{-16}$, 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 S1). 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.

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* encodes 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 protein 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 nonsynonymous changes, some of which fall into functional domains (Supplemental Figure S3, Supplemental Table S3 and S4). 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 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 3A-B). By comparison, the non-DE immune genes had a lower proportion of nonsynonymous changes (19 %, p=1.6 $\times 10^{-4}$, 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 nonsynonymous changes between our genomes of interest than fat body detected or non-differentially expressed immune genes. Some of these nonsynonymous changes may be capable of altering the function of these proteins and therefore drive expression divergence of downstream genes in a *trans*-acting fashion.

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 *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 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 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.

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 key 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 S1). 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.

Methods

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).

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℃ 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.

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 S5). Code and accompanying files related to this section are in Supplemental Code as both R-notebooks and HTML documents (Script1_fig1).

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)

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 S1. Code and accompanying files related to this section are available Supplemental Code as both R-notebook and HTML document form (Script2_fig2).

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.

Analysis of SNPs in coding sequences

To better understand the effects of sequence changes on coding regions between our lines we used the Ensembl 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).

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 S4). Code and accompanying files related to this section can be found in Supplemental Code as both R-notebooks and HTML document form (Script4_fig4).

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.

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.

Acknowledgements

We would like to thank J.J. Emerson and Xinwin Zhang for their thoughtful comments and suggestions on this work, Ali Mortazavi and Lorrayne Serra for their insight and access to sequencing equipment, Tom Schilling for access to microinjection equipment, Joseph Coolon, Carl de Boer and Rabi Murad for sharing their scripts as well as general guidance with computational protocols, and Anthony Long and Mahul Chakraborty for their insight helpful comments on the genomes utilized. This work was funded in part by the National Science Foundation, Award 1953324 to Z.W. B.R. was supported by the California LSAMP Bridge to the Doctorate Program, NSF Award 1500284. S.F. was supported by a UCI UROP award. O.O. was supported by NIH Grant R25 GM055246, T34 GM136498, and an UCI UROP award.

Disclosure Declarations

The authors have no conflicts of interest to declare.

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: 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

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

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

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

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

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

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

Table 1

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	Hematopoiesis
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
FBgn0003495	*spz*	Signaling	0.68	B6	72.8	45.5	Positive regulator of Toll pathway

Table 2

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

Figure 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 2), after (Group 3), and both before and after infection (Group 4). 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.

Figure 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.

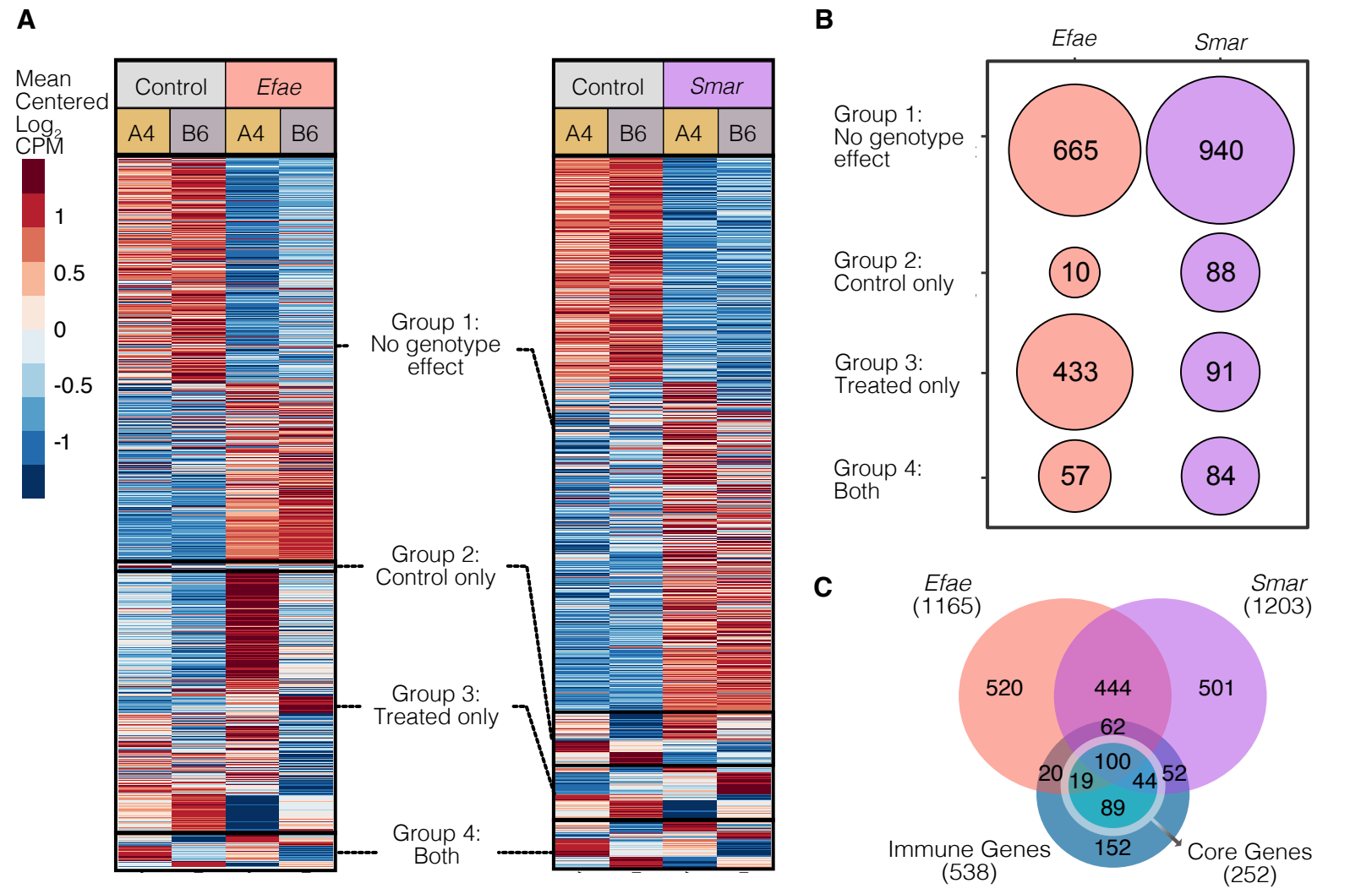
Figure 3: There is a greater proportion of nonsynonymous SNPs in previously identified immune-responsive genes.
A) To look for the prevalence of nonsynonymous SNPs in our genotypes and genes of interest, we defined four gene sets. Among genes detected in the fat body samples, we separated genes into 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 nonsynonymous SNPs. B) DE immune genes have a higher proportion of nonsynonymous SNPs than the fat body expressed genes, which suggests they may carry function-altering SNPs at a higher rate

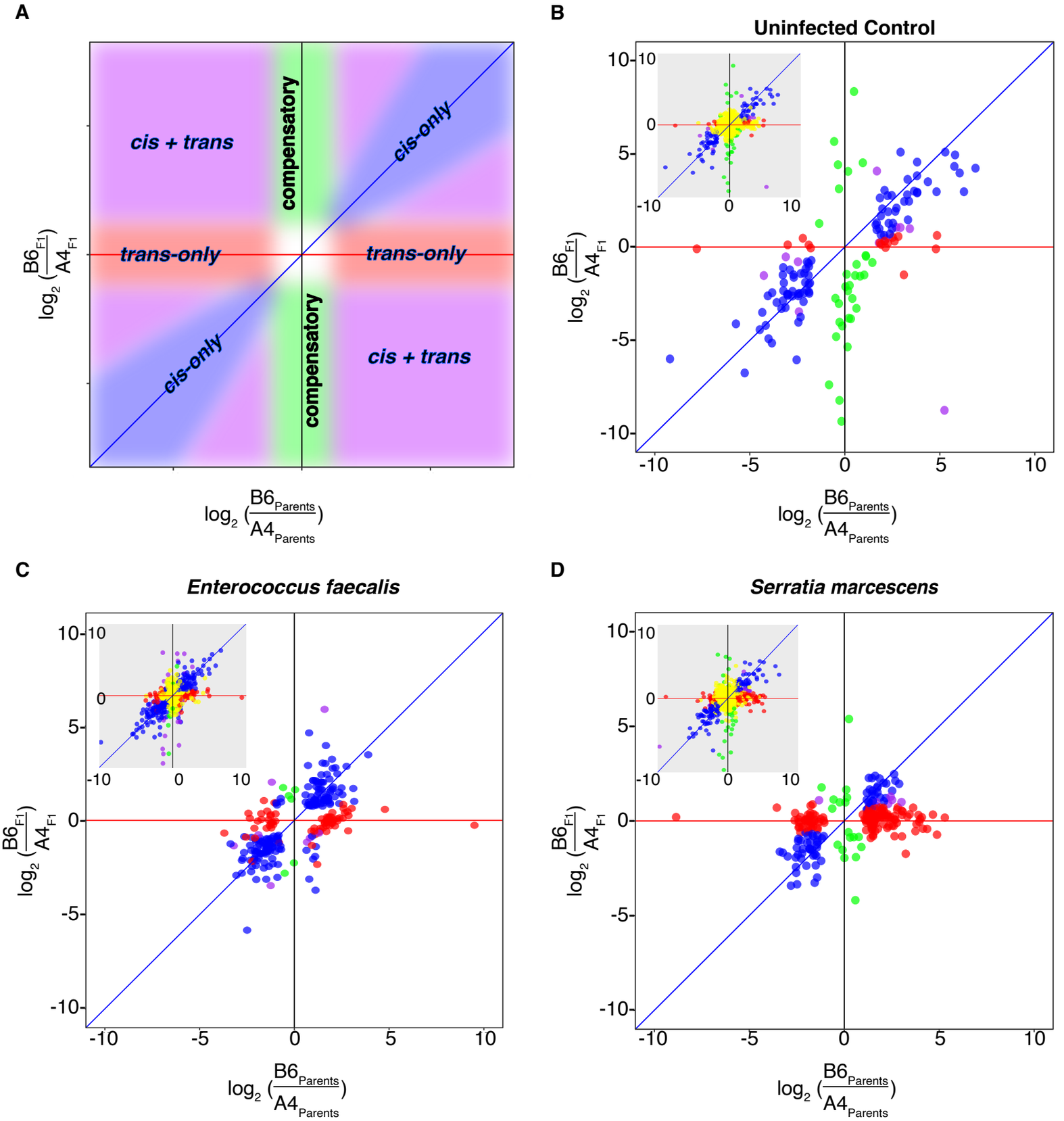
than the fat body expressed genes. *p*-values are Bonferroni-corrected from Chi-square tests with the proportion of nonsynonymous SNPS relative to the fat body expressed gene set.

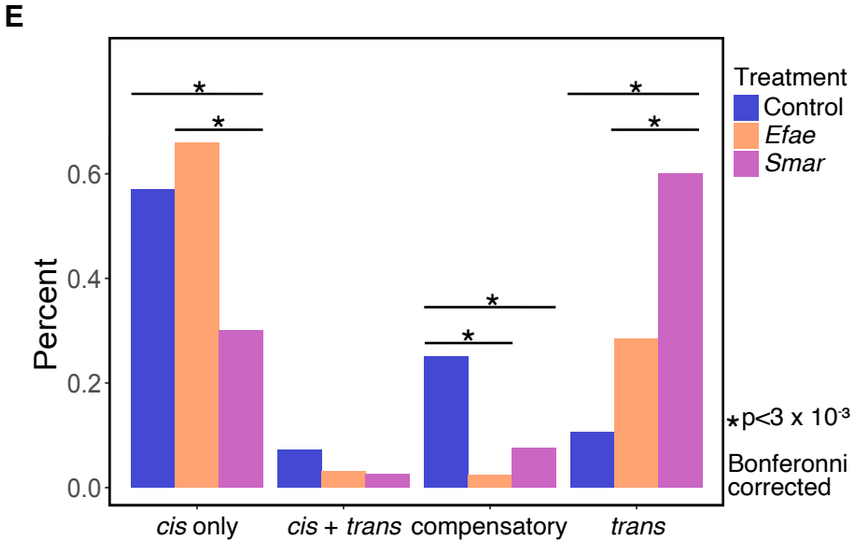
Figure 4: There are greater differences in TFBS in *cis* affected genes than *trans* affected genes.
A) We identified TFBS for 4 immune responsive transcription factors Dl, Rel, Srp and CrebA in 1kb region upstream of 219 *cis*-affected genes and 199 *trans*-affected genes. Differences in total TFBS 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, Srp and CrebA TFBS, there was a broader 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 differences in total TFBS than genes showing *trans* effects, though the differences in these proportions were not significant.

Table 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.

Table 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 < 0.05 and 'False' indicates an insignificant test result.


A
Mean Centered Log2 CPM
1
0.5
0
-0.5
-1
Control
Efae
A4
B6
A4
B6
Control
Smar
A4
B6
A4
B6
Group 1: No genotype effect
Group 2: Control only
Group 3: Treated only
Group 4: Both
B
Efae
Smar
Group 1: No genotype effect
665
940
Group 2: Control only
10
88
Group 3: Treated only
433
91
Group 4: Both
57
84
C
Efae (1165)
Smar (1203)
520
444
501
62
100
20
19
44
52
89
152
Immune Genes (538)
Core Genes (252)

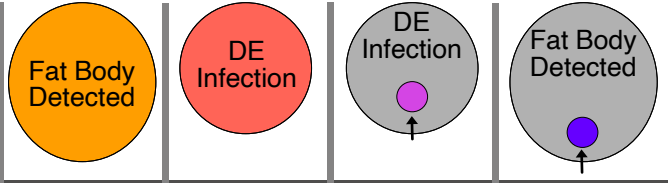




F

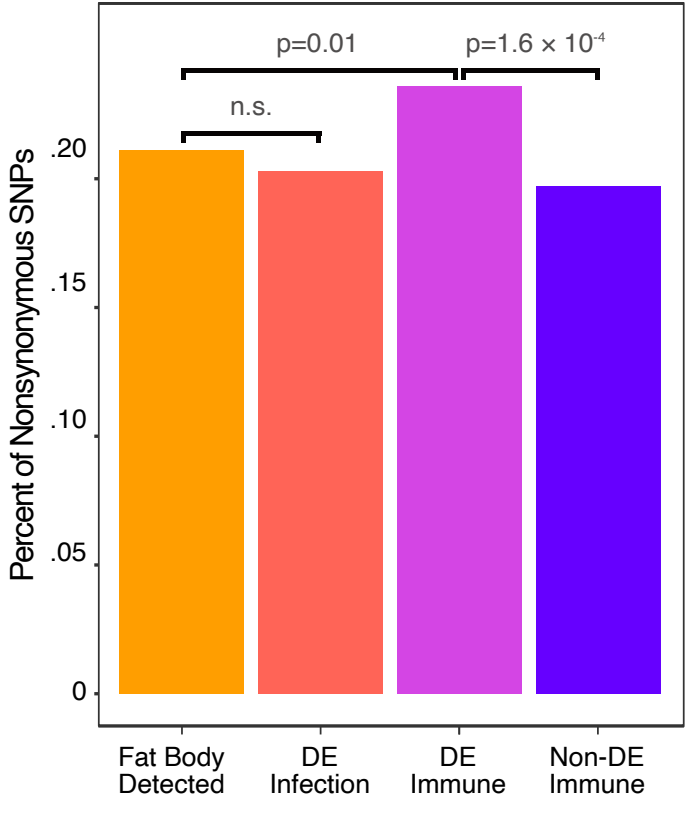
Category	Control	*Efae*	*Smar*
Conserved	3802	2580	4106
cis-only	87	174	79
cis + trans	11	8	7
compensatory	37	6	16
trans-only	16	75	149

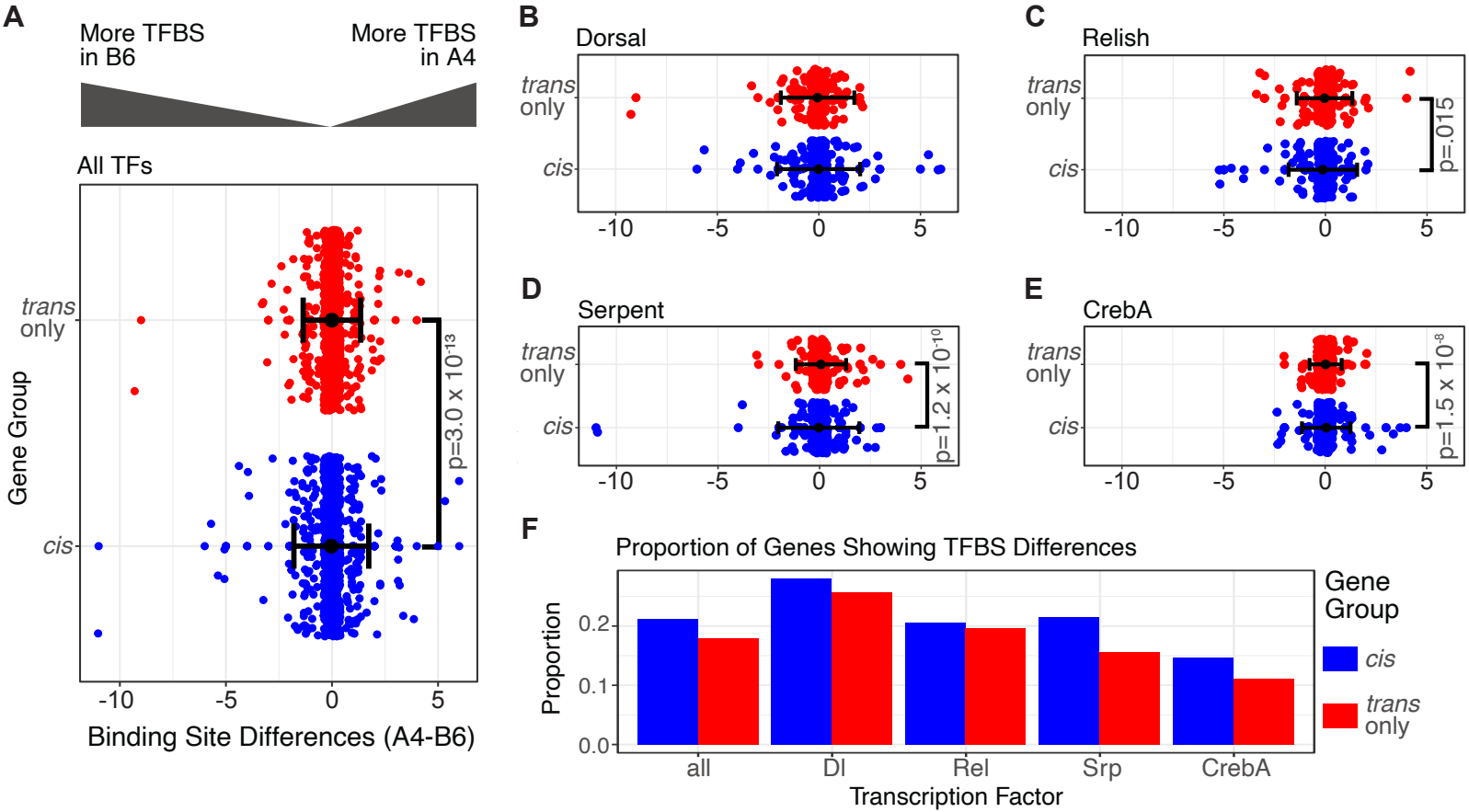
A



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




A
More TFBS in B6
More TFBS in A4
All TFs
trans only
cis
Gene Group
p=3.0 x 10-13
-10
-5
0
5
Binding Site Differences (A4-B6)
B
Dorsal
C
Relish
p=.015
D
Serpent
p=1.2 x 10-10
E
CrebA
p=1.5 x 10-8
F
Proportion of Genes Showing TFBS Differences
Proportion
0.0
0.1
0.2
all
Dl
Rel
Srp
CrebA
Transcription Factor
Gene Group
cis
trans only

Supplemental Materials

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

Bryan A. Ramirez-Corona, Stephanie Fruth, Oluchi Ofoegbu, Zeba Wunderlich*

Department of Developmental and Cell Biology, University of California, Irvine, CA

*Corresponding Author:
zeba@uci.edu

Table of Contents:

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 S1). 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.

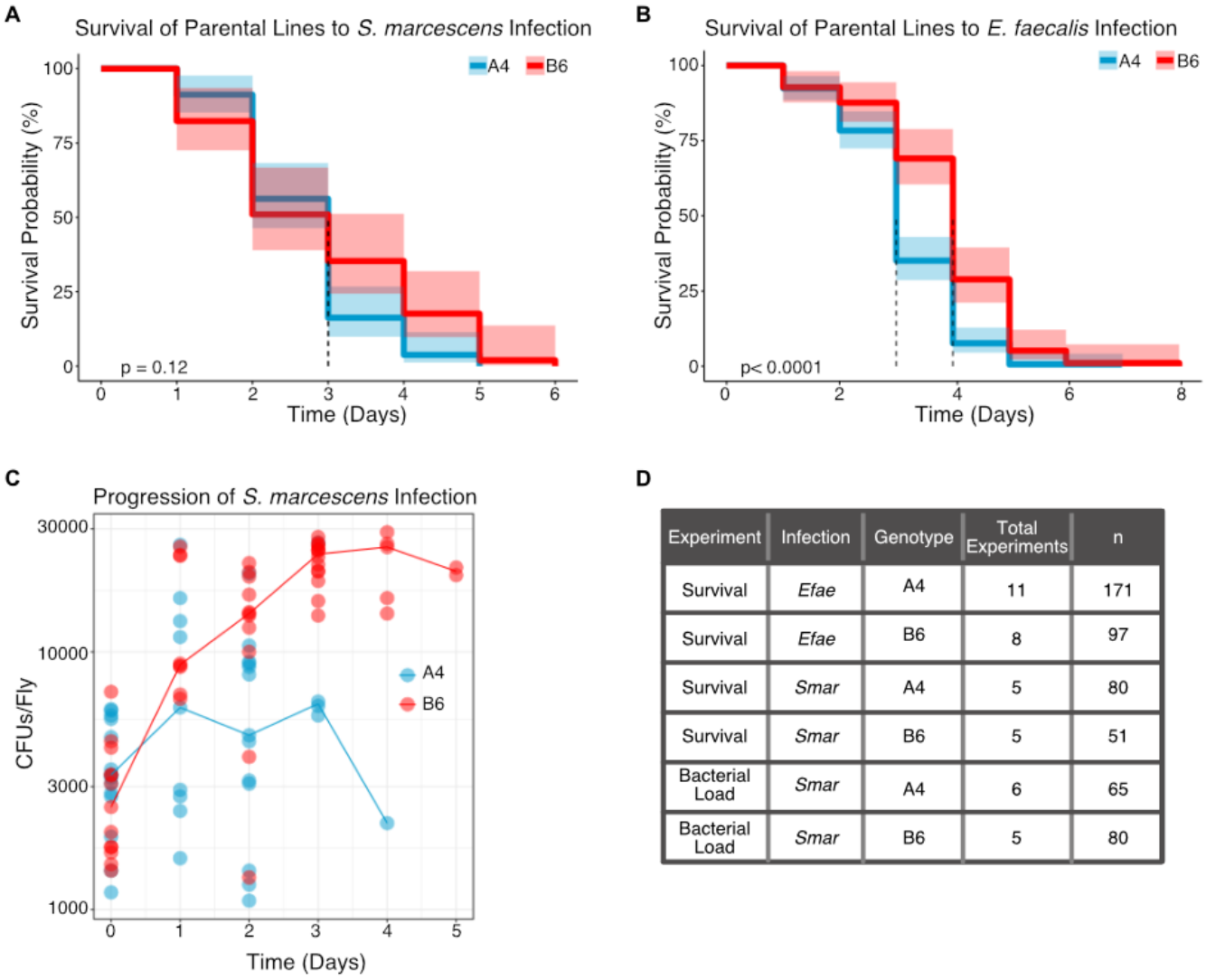
To assess the quality of our annotations and remove genes with poor annotations, genomic sequencing reads for A4 and B6 from the DSPR website were downloaded and aligned to our transcriptome files using Salmon 0.12.0 aligner (Thurmond et al., 2019). We hypothesized that well-annotated genes would show 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). 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 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.



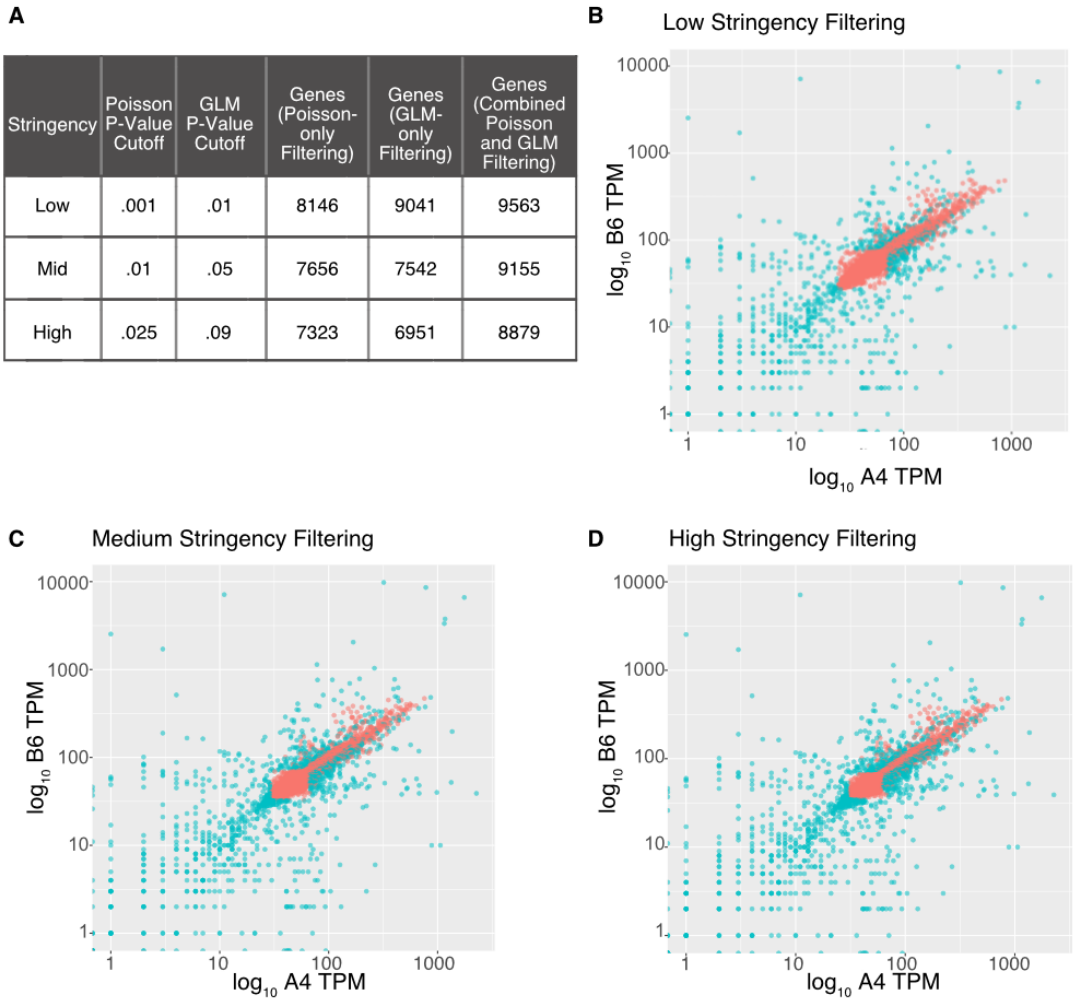
D

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

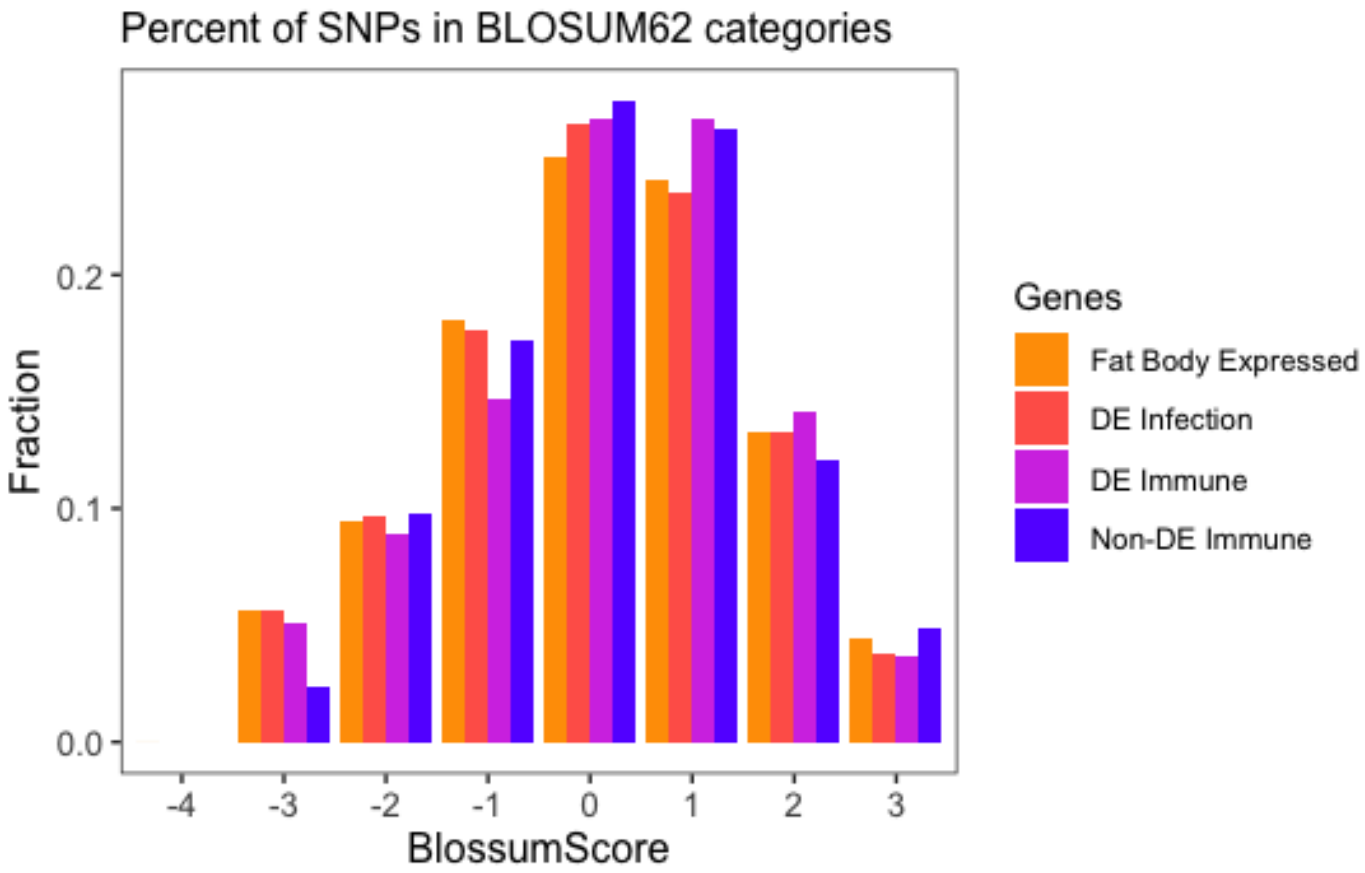
Supplemental Figure S1: A4 and B6 lines show differences in survival in response to Gram-positive but not Gram-negative infection.
A) Survival curves and confidence intervals for flies infected with an average 1000 CFUs of *S. marcescens*, observed once per day. P-value was calculated using a log-rank 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

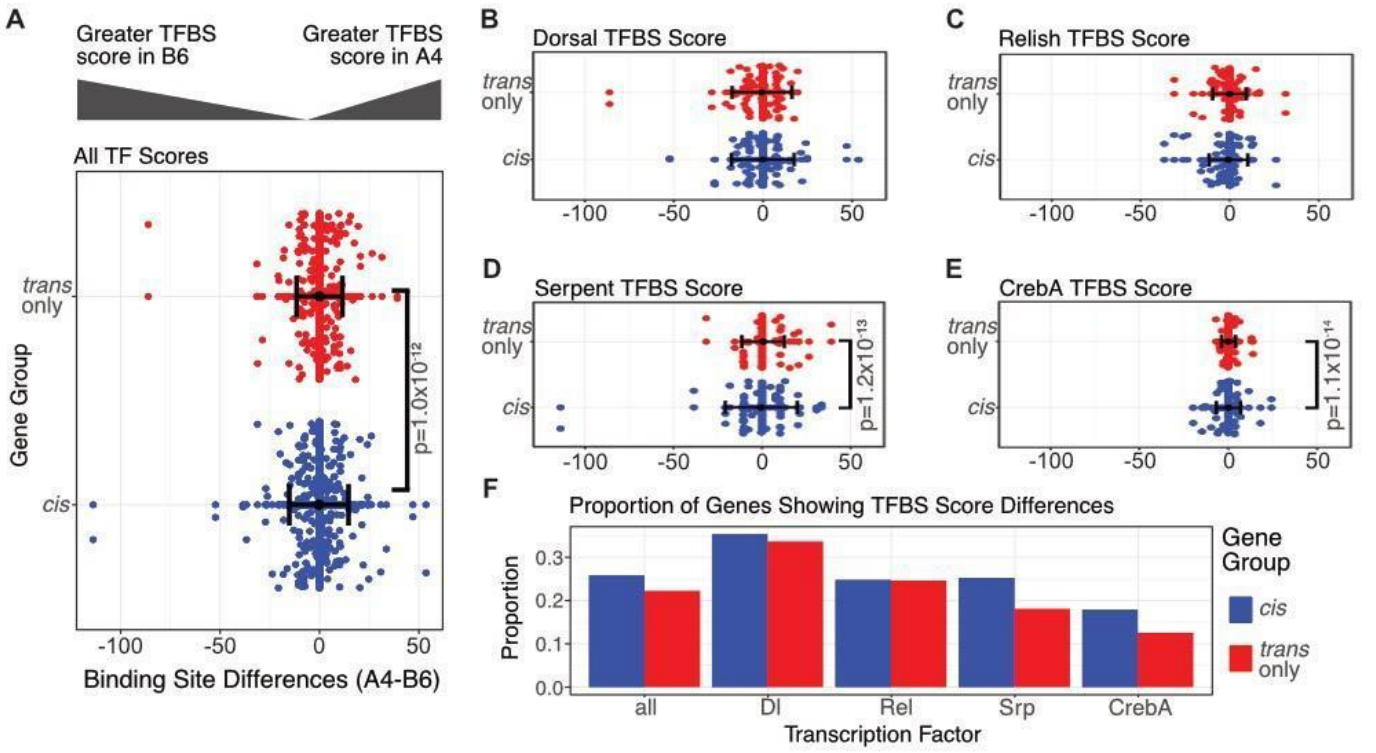
Stringency	Poisson P-Value Cutoff	GLM P-Value Cutoff	Genes (Poisson-only Filtering)	Genes (GLM-only Filtering)	Genes (Combined Poisson and GLM Filtering)
Low	.001	.01	8146	9041	9563
Mid	.01	.05	7656	7542	9155
High	.025	.09	7323	6951	8879



Supplemental Figure S2: 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.

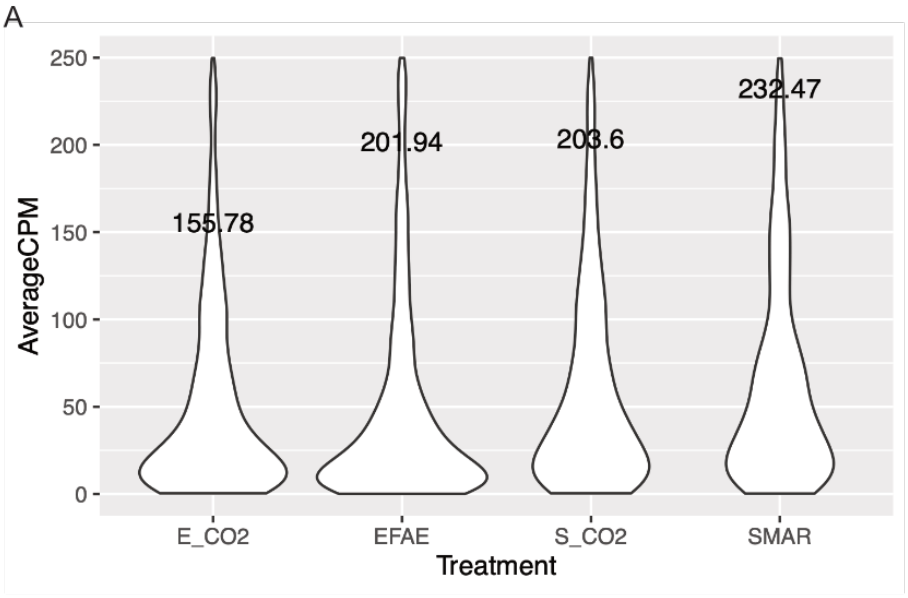


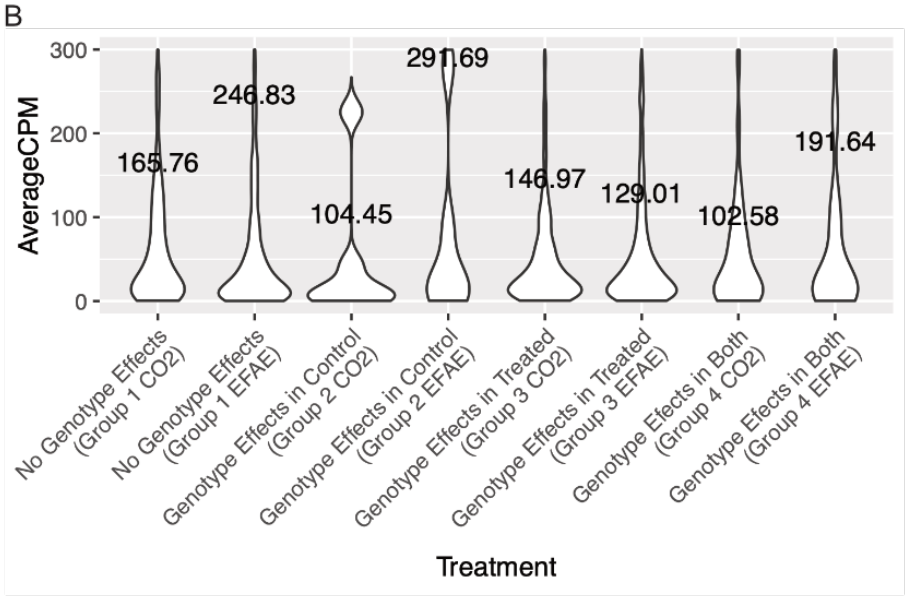
Supplemental Figure S3: Most nonsynonymous mutations have non-negative BLOSUM62 scores. As a coarse-grained approximation of the effects of non-synonymous changes on protein function, we analyzed the distribution of BLOSUM62 scores for the four gene sets described. The BLOSUM62 score is a homology-based 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.

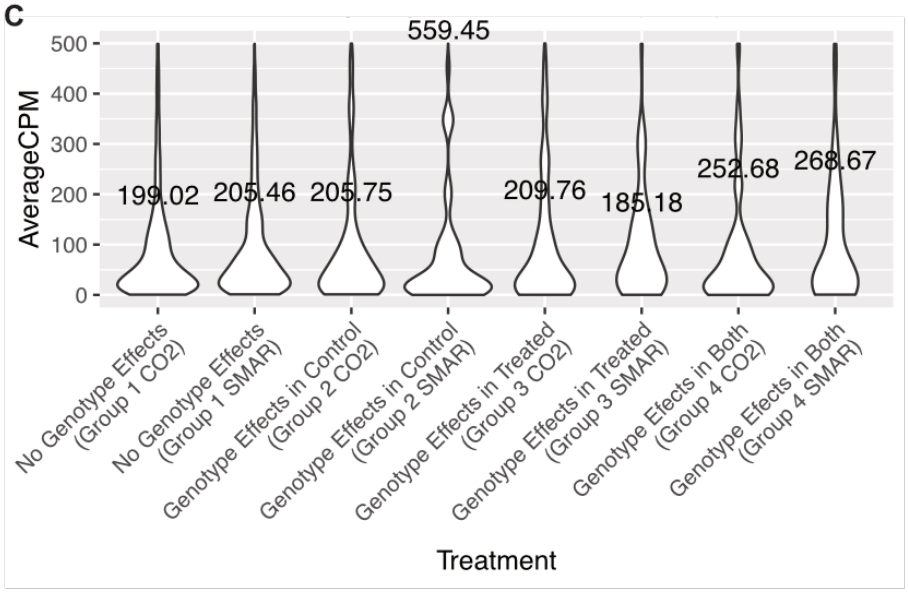


Supplemental Figure S4: There are greater differences in TFBS score in *cis* affected genes than *trans* affected genes.

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.


A
AverageCPM
250
200
150
100
50
0
155.78
201.94
203.6
232.47
E_CO2
EFAE
S_CO2
SMAR
Treatment


B
AverageCPM
300
200
100
0
165.76
246.83
104.45
291.69
146.97
129.01
102.58
191.64
No Genotype Effects (Group 1 CO2)
No Genotype Effects (Group 1 EFAE)
Genotype Effects in Control (Group 2 CO2)
Genotype Effects in Control (Group 2 EFAE)
Genotype Effects in Treated (Group 3 CO2)
Genotype Effects in Treated (Group 3 EFAE)
Genotype Efects in Both (Group 4 CO2)
Genotype Efects in Both (Group 4 EFAE)
Treatment


C
AverageCPM
500
400
300
200
100
0
199.02
205.46
205.75
559.45
209.76
185.18
252.68
268.67
No Genotype Effects (Group 1 CO2)
No Genotype Effects (Group 1 SMAR)
Genotype Effects in Control (Group 2 CO2)
Genotype Effects in Control (Group 2 SMAR)
Genotype Effects in Treated (Group 3 CO2)
Genotype Effects in Treated (Group 3 SMAR)
Genotype Efects in Both (Group 4 CO2)
Genotype Efects in Both (Group 4 SMAR)
Treatment

Supplemental Figure S5: 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 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 1). No group shows significance in average CPM value between the treated and control samples (two sample Wilcoxon rank sum test, Bonferroni corrected).

Supplemental Table S1: 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).

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

Supplemental Table S2: 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.

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

Supplemental Table S3: 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.

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:13975515	G	*NimC1*	FBgn0259896	FBtr0080615	578	193	V/A	gTc/gCc	0
2L:13975735	T	*NimC1*	FBgn0259896	FBtr0080615	358	120	G/S	Ggc/Agc	0
2L:13976157	C	*NimC1*	FBgn0259896	FBtr0343644	40	14	S/A	Tca/Gca	1
2R:8717036	G	*PGRP-SC2*	FBgn0043575	FBtr0088709	70	24	I/V	Atc/Gtc	3
2R:10207902	C	*Hr3*	FBgn0000448	FBtr0330609	1570	524	P/A	Cca/Gca	-1
2R:10232873	T	*Hr3*	FBgn0000448	FBtr0452140	439	147	S/T	Tcg/Acg	1
2R:10237018	G	*Hr3*	FBgn0000448	FBtr0112799	23	8	N/T	aAc/aCc	0
3L:7651752	T	*PGRP-SD*	FBgn0035806	FBtr0076807	548	183	S/F	tCc/tTc	-2
3L:9441876	A	*Nf-YA*	FBgn0035993	FBtr0076504	17	6	S/I	aGc/aTc	-2
3R:7148618	C	*gfzf*	FBgn0250732	FBtr0334671	1480	494	H/D	Cac/Gac	-1
3R:7150621	A	*gfzf*	FBgn0250732	FBtr0091512	10	4	P/S	Ccc/Tcc	-1
3R:23378558	T	*CG4393*	FBgn0039075	FBtr0339617	3322	1108	L/I	Tta/Ata	2
3R:2337	C	*CG4393*	FBgn00	FBtr030	3313	1105	P/A	Cca/Gc	-1

8567			39075	1085				a	
3R:23378571	A	*CG4393*	FBgn0039075	FBtr0339616	3309	1103	E/D	gaG/gaT	2
3R:23379640	G	*CG4393*	FBgn0039075	FBtr0339617	2375	792	Q/P	cAa/cCa	-1
3R:23379641	T	*CG4393*	FBgn0039075	FBtr0301085	2374	792	Q/K	Caa/Aaa	1
3R:23381986	T	*CG4393*	FBgn0039075	FBtr0301085	548	183	T/N	aCc/aAc	0
3R:27066830	G	*spz*	FBgn0003495	FBtr0085137	199	67	T/P	Acc/Ccc	-1
3R:30773707	A	*zfh1*	FBgn0004606	FBtr0331180	232	78	Q/K	Cag/Aag	1
3R:30774111	T	*zfh1*	FBgn0004606	FBtr0085701	386	129	K/M	aAg/aTg	-1
3R:30774123	T	*zfh1*	FBgn0004606	FBtr0331180	398	133	A/V	gCc/gTc	0
3R:30774165	C	*zfh1*	FBgn0004606	FBtr0085701	440	147	S/T	aGc/aCc	1
3R:30785831	T	*zfh1*	FBgn0004606	FBtr0085701	2861	954	A/V	gCg/gTg	0

Supplemental Table S4: 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.

Location	Allele	Gene Symbol	Domains
2L:4122351	T	*Sr-CI*	Gene3D:2.60.120.200,Pfam:PF00629,PROSITE_profiles:PS50060,PANTHER:PTHR23282,SMART:SM00137,Superfamily:SSF49899,CDD:cd06263
2L:4122897	C	*Sr-CI*	Gene3D:2.60.120.200,PANTHER:PTHR23282
2L:4123356	T	*Sr-CI*	PANTHER:PTHR23282,MobiDB_lite:mobidb-lite,Low_complexity_(Seg):seg
2L:8005499	A	*Spn28Dc*	Gene3D:3.30.497.10,Pfam:PF00079,PANTHER:PTHR11461,PANTHER:PTHR11461:SF281,SMART:SM00093,Superfamily:SSF56574,CDD:cd00172
2L:8005523	G	*Spn28Dc*	Gene3D:3.30.497.10,Pfam:PF00079,PANTHER:PTHR11461,PANTHER:PTHR11461:SF281,SMART:SM00093,Superfamily:SSF56574,CDD:cd00172
2L:8005549	G	*Spn28Dc*	Gene3D:3.30.497.10,Pfam:PF00079,PANTHER:PTHR11461,PANTHER:PTHR11461:SF281,SMART:SM00093,Superfamily:SSF56574,CDD:cd00172
2L:8006451	A	*Spn28Dc*	Gene3D:3.30.497.10,Pfam:PF00079,PANTHER:PTHR11461,PANTHER:PTHR11461:SF281,SMART:SM00093,Superfamily:SSF56574,CDD:cd00172
2L:8006864	C	*Spn28Dc*	PANTHER:PTHR11461,PANTHER:PTHR11461:SF281,Superfamily:SSF56574
2L:13968919	C	*NimB4*	PANTHER:PTHR24047,Gene3D:2.10.25.10,SMART:SM00181
2L:13974306	G	*NimC1*	PANTHER:PTHR24047,PANTHER:PTHR24047:SF29,Transmembrane_helices:TMhelix
2L:13974690	T	*NimC1*	PANTHER:PTHR24047,PANTHER:PTHR24047:SF29
2L:13974703	G	*NimC1*	PANTHER:PTHR24047,PANTHER:PTHR24047:SF29
2L:13975363	T	*NimC1*	Gene3D:2.10.25.10,PANTHER:PTHR24047,PANTHER:PTHR24047:SF29,SMART:SM00181,Superfamily:SSF57184
2L:13975380	T	*NimC1*	Gene3D:2.10.25.10,PANTHER:PTHR24047,PANTHER:PTHR24047:SF29,SMART:SM00181,Superfamily:SSF57184
2L:13975515	G	*NimC1*	Gene3D:2.10.25.10,PANTHER:PTHR24047,PANTHER:PTHR24047:SF29,SMART:SM00181

2L:13975735	T	*NimC1*	Gene3D:2.10.25.10,PROSITE_patterns:PS00022,PANTHER:PTHR24047,PANTHER:PTHR24047:SF29,SMART:SM00181
2L:13976157	C	*NimC1*	Cleavage_site_(Signalp):SignalP-noTM
2R:8717036	G	*PGRP-SC2*	Gene3D:3.40.80.10,PIRSF:PIRSF037945,PANTHER:PTHR11022,SMART:SM00701,Superfamily:SSF55846,Transmembrane_helices:TMhelix
2R:10207902	C	*Hr3*	Low_complexity_(Seg):seg
2R:10232873	T	*Hr3*	-
2R:10237018	G	*Hr3*	PANTHER:PTHR45805,PANTHER:PTHR45805:SF2
3L:7651752	T	*PGRP-SD*	Gene3D:3.40.80.10,PIRSF:PIRSF037945,PANTHER:PTHR11022,PANTHER:PTHR11022:SF67,Superfamily:SSF55846
3L:9441876	A	*Nf-YA*	-
3R:7148618	C	*gfzf*	PANTHER:PTHR43969,PANTHER:PTHR43969:SF7
3R:7150621	A	*gfzf*	PANTHER:PTHR43969,PANTHER:PTHR43969:SF7
3R:23378558	T	*CG4393*	PANTHER:PTHR24174,PANTHER:PTHR24174:SF1,MobiDB_lite:mobidb-lite
3R:23378567	C	*CG4393*	PANTHER:PTHR24174,PANTHER:PTHR24174:SF1,MobiDB_lite:mobidb-lite
3R:23378571	A	*CG4393*	PANTHER:PTHR24174,PANTHER:PTHR24174:SF1,MobiDB_lite:mobidb-lite
3R:23379640	G	*CG4393*	PANTHER:PTHR24174,PANTHER:PTHR24174:SF1,MobiDB_lite:mobidb-lite
3R:23379641	T	*CG4393*	PANTHER:PTHR24174,PANTHER:PTHR24174:SF1,MobiDB_lite:mobidb-lite
3R:23381986	T	*CG4393*	Gene3D:1.25.40.20,PROSITE_profiles:PS50297,PANTHER:PTHR24174,PANTHER:PTHR24174:SF1,Superfamily:SSF48403
3R:27066830	G	*spz*	PANTHER:PTHR23199,PANTHER:PTHR23199:SF4

3R:30773707	A	*zfh1*	Pfam:PF13912,PROSITE_patterns:PS00028,PROSITE_profiles:PS50157,PANTHER:PTHR24391,PANTHER:PTHR24391:SF27,SMART:SM00355
3R:30774111	T	*zfh1*	PANTHER:PTHR24391,PANTHER:PTHR24391:SF27,MobiDB_lite:mobidb-lite,MobiDB_lite:mobidb-lite
3R:30774123	T	*zfh1*	PANTHER:PTHR24391,PANTHER:PTHR24391:SF27,MobiDB_lite:mobidb-lite,MobiDB_lite:mobidb-lite,Low_complexity_(Seg):seg
3R:30774165	C	*zfh1*	PANTHER:PTHR24391,PANTHER:PTHR24391:SF27,MobiDB_lite:mobidb-lite,MobiDB_lite:mobidb-lite,Low_complexity_(Seg):seg
3R:30785831	T	*zfh1*	PANTHER:PTHR24391,PANTHER:PTHR24391:SF27

Supplemental Table S5: 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.

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

Supplemental References

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

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

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