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.

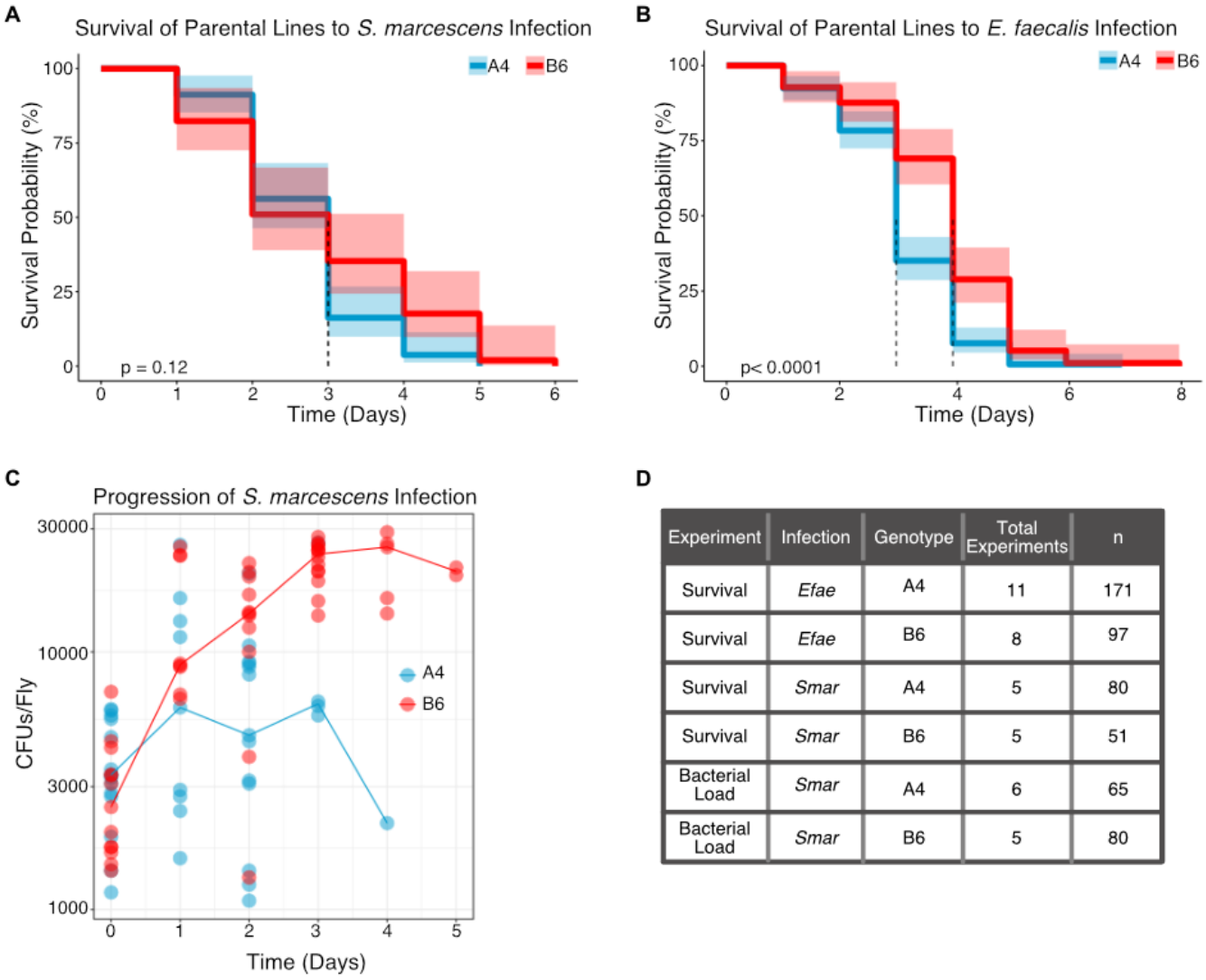
To assess the quality of our annotations and remove genes with poor annotations, genomic sequencing reads for A4 and B6 from the DSPR website were downloaded and aligned to our transcriptome files using Salmon 0.12.0 aligner (Thurmond et al., 2019). We hypothesized that well-annotated genes would show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

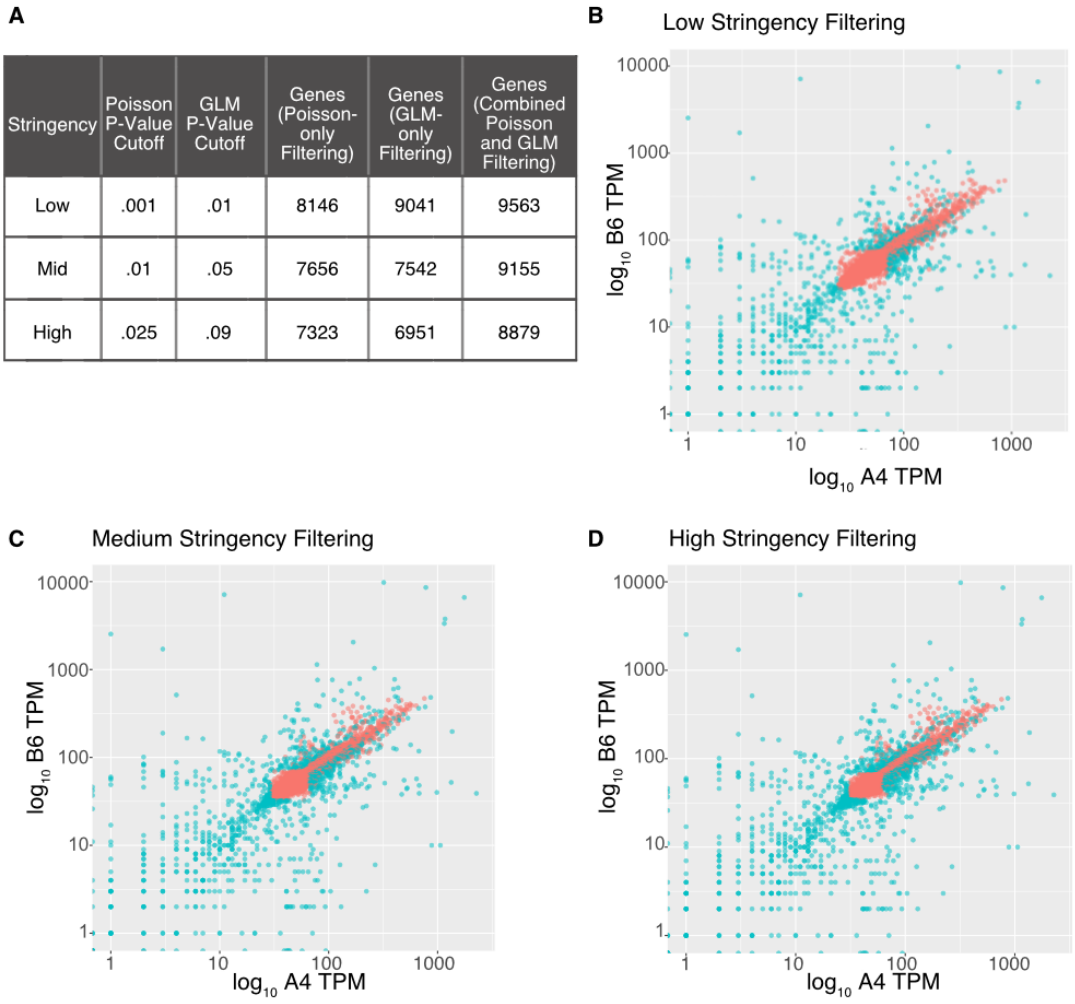


Experiment	Infection	Genotype	Total Experiments	n
Survival	*Efae*	A4	11	171
Survival	*Efae*	B6	8	97
Survival	*Smar*	A4	5	80
Survival	*Smar*	B6	5	51
Bacterial Load	*Smar*	A4	6	65
Bacterial Load	*Smar*	B6	5	80

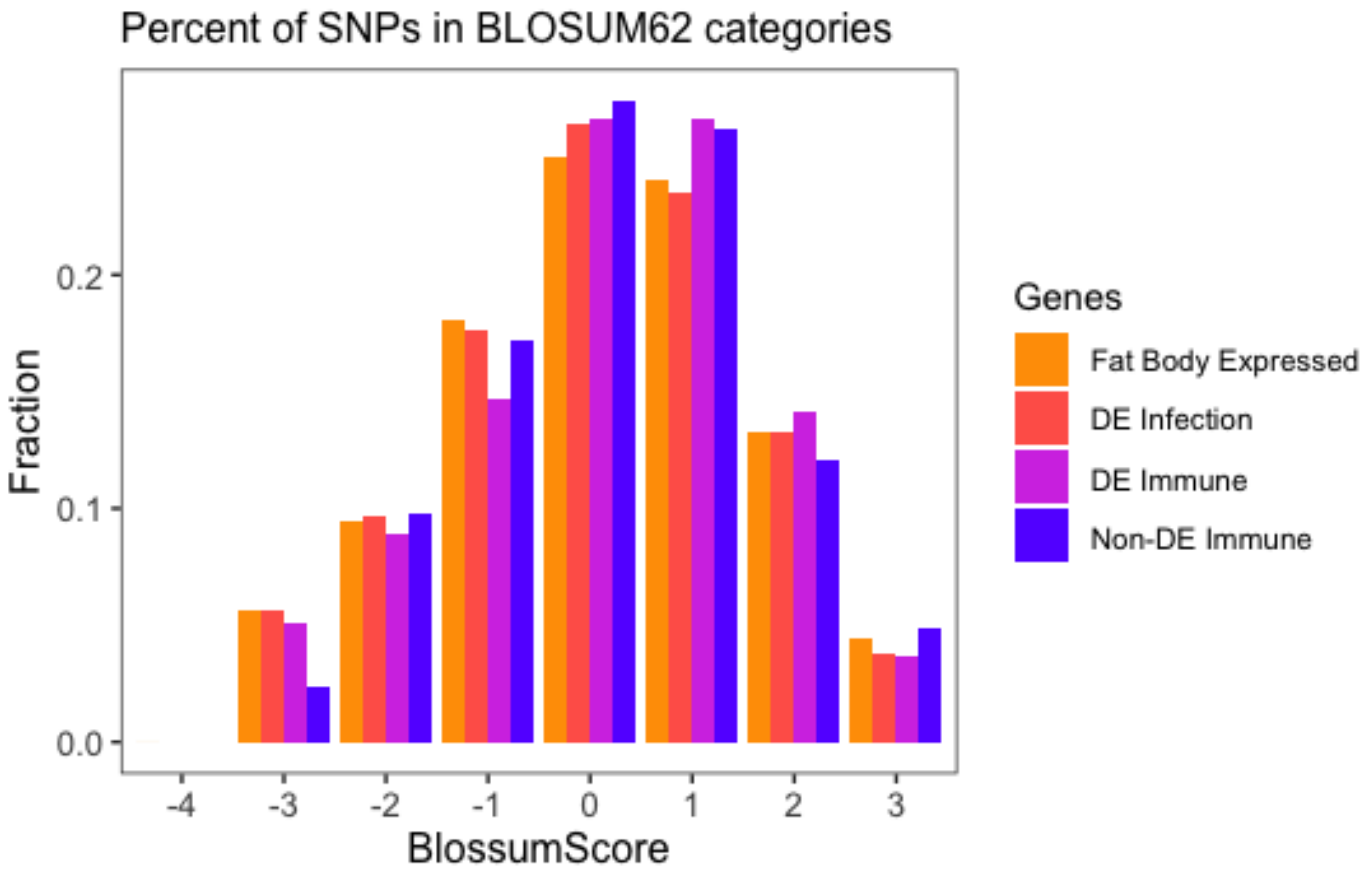
Supplemental Figure S1: A4 and B6 lines show differences in survival in response to Gram-positive but not Gram-negative infection.
A) Survival curves and confidence intervals for flies infected with an average 1000 CFUs of *S. marcescens*, observed once per day. P-value was calculated using a log-rank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

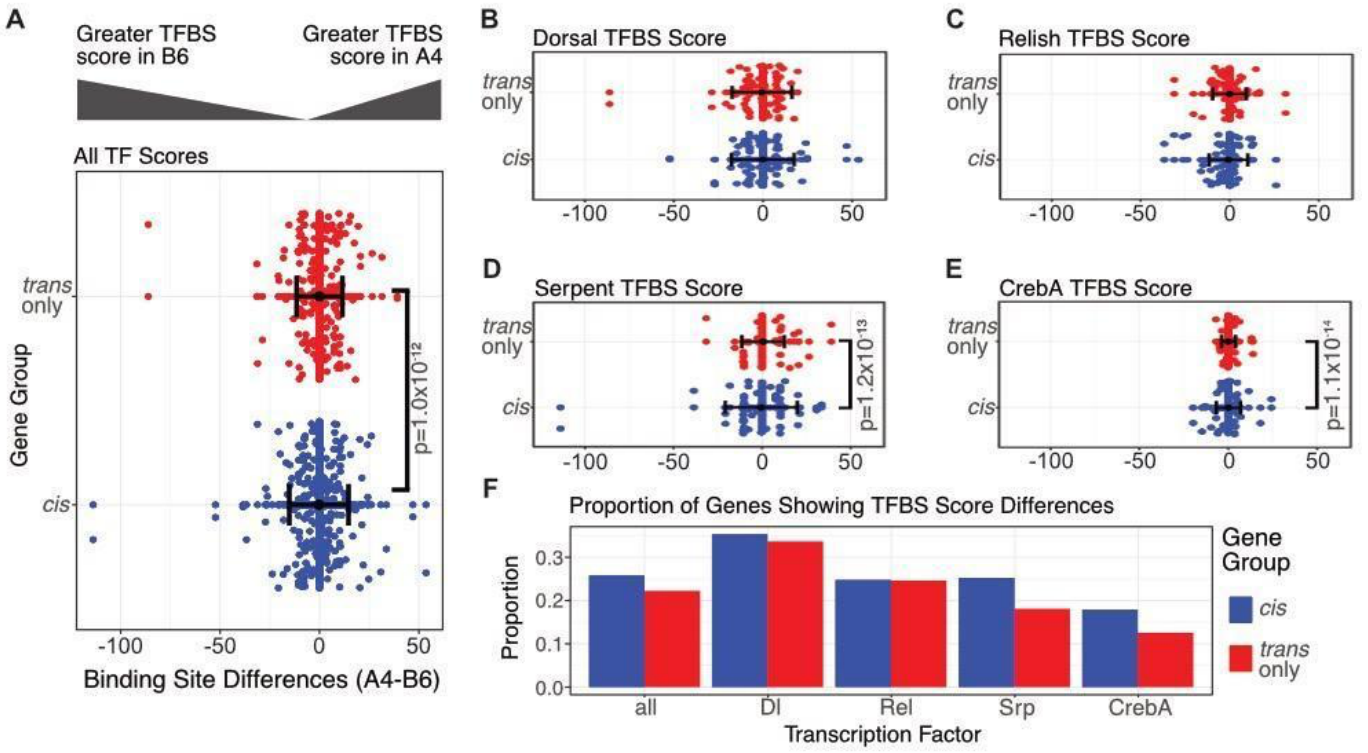
Stringency	Poisson P-Value Cutoff	GLM P-Value Cutoff	Genes (Poisson-only Filtering)	Genes (GLM-only Filtering)	Genes (Combined Poisson and GLM Filtering)
Low	.001	.01	8146	9041	9563
Mid	.01	.05	7656	7542	9155
High	.025	.09	7323	6951	8879



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.

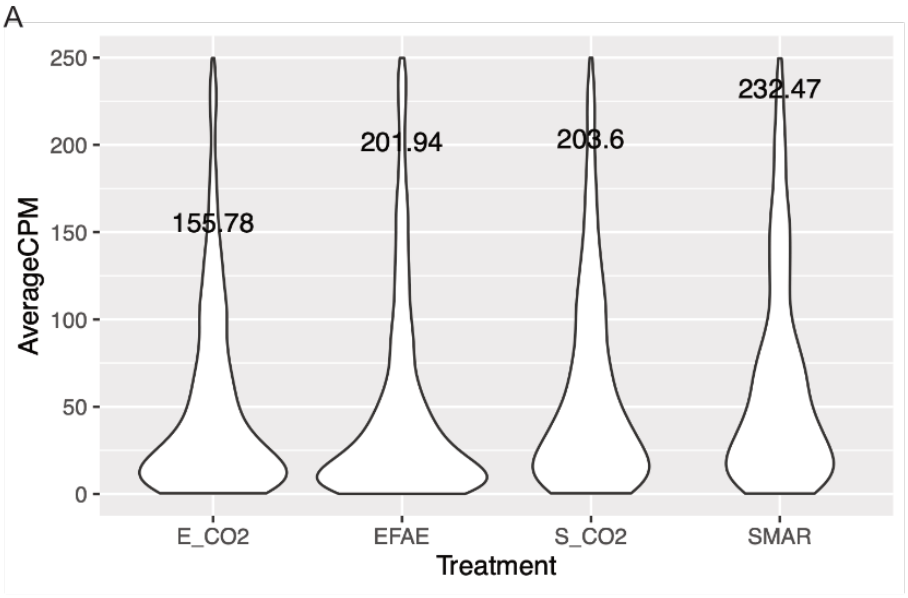


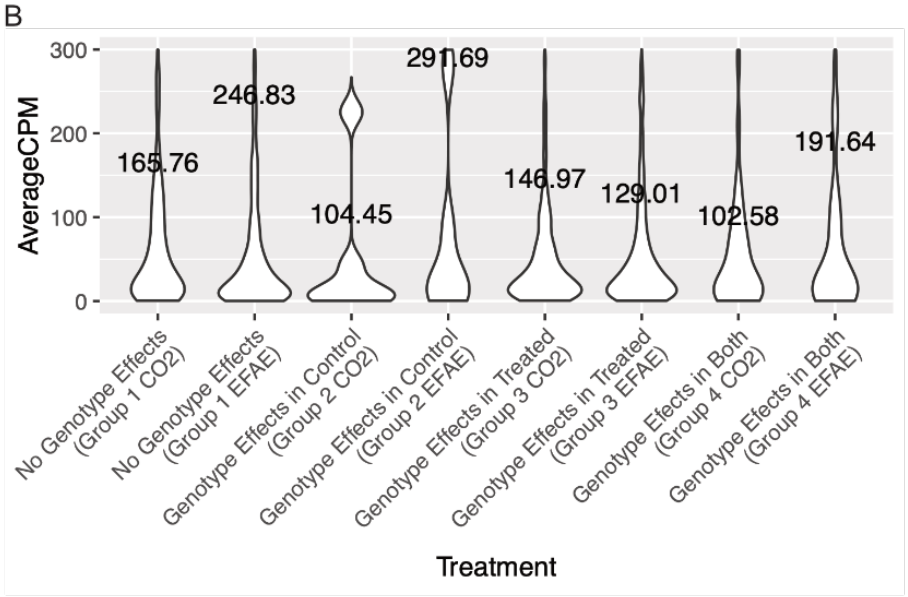
Supplemental Figure S3: Most nonsynonymous mutations have non-negative BLOSUM62 scores. As a coarse-grained approximation of the effects of non-synonymous changes on protein function, we analyzed the distribution of BLOSUM62 scores for the four gene sets described. The BLOSUM62 score is a homology-based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

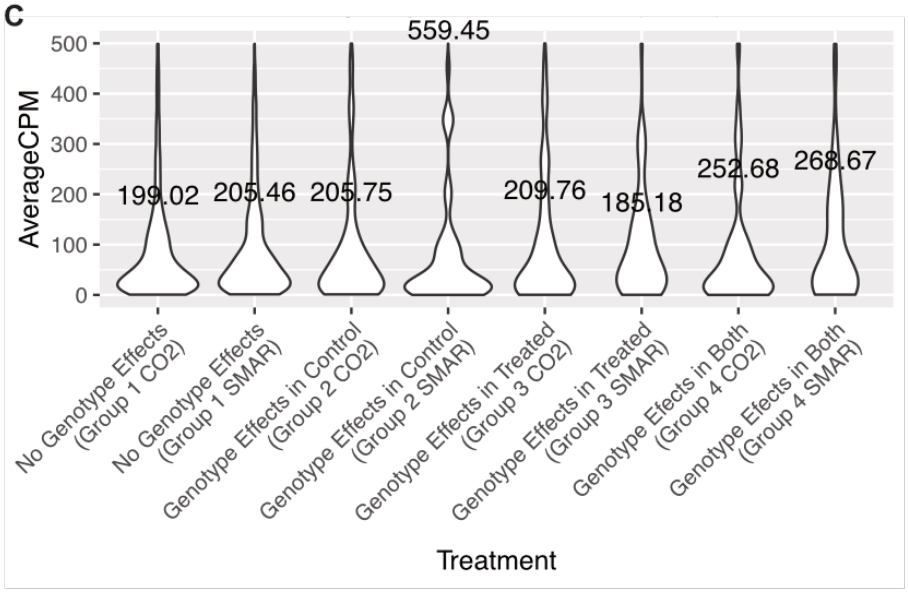


Supplemental Figure S4: There are greater differences in TFBS score in *cis* affected genes than *trans* affected genes.

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