# Widespread shortening of 3' untranslated regions and increased exon inclusion characterize the human macrophage response to infection

Athma A. Pai[a,1,2], Golshid Baharian[b,c,1], Ariane Page Sabourin[c], Yohann Nédélec[b,c], Jean-Christophe Grenier[c], Katherine J. Siddle[d], Anne Dumaine[c], Vania Yotova[c], Christopher B. Burge[a,e], Luis B. Barreiro[c,f,2]

[a]Department of Biology, Massachusetts Institute of Technology, Cambridge, MA, USA
[b]Department of Biochemistry, Faculty of Medicine, University of Montreal, Montreal, QC, Canada
[c]Sainte-Justine Hospital Research Centre, Montreal, QC, Canada
[d]Department of Organismic and Evolutionary Biology, FAS Center for Systems Biology, Harvard University, Cambridge, MA, USA
[e]Department of Biological Engineering, Massachusetts Institute of Technology, Cambridge, MA, USA
[f]Department of Pediatrics, Faculty of Medicine, University of Montreal, Montreal, QC, Canada

[1]These authors contributed equally to this work.
[2]To whom correspondence should be addressed. E-mail: athma@mit.edu and luis.barreiro@umontreal.ca

**ABSTRACT**

Changes in gene regulation have long been known to play important roles in both innate and adaptive immune responses. However, post-transcriptional mechanisms involved in mRNA processing have been poorly studied despite emerging examples of their role as regulators of immune defenses. We sought to investigate the role of mRNA processing in the cellular responses of human macrophages to live bacterial infections. We used mRNA-sequencing to quantify gene expression and isoform abundances in primary macrophages from 60 individuals, before and after infection with *Listeria monocytogenes* and *Salmonella typhimurium*. We show that immune responses to infection are accompanied by pervasive changes in isoform usage that lead to an overall increase in isoform diversity after infection. In response to both bacteria, we see global shifts towards (*i*) the inclusion of cassette exons and (*ii*) shorter 3' UTRs, with near-universal shifts towards usage of more upstream polyadenylation sites. Using individual-specific variation in RNA processing, we identify candidate splicing factors putatively regulating these post-infection patterns in *trans*. Finally, by profiling microRNA levels, we further show that 3' UTR regions with reduced abundance after infection are significantly enriched for target sites for specific miRNAs - including miR-146b, miR-3661, miR-151b, and miR-125a - that are up-regulated following infection. These results suggest that the prominent shift towards shorter 3' UTRs might be a cellular control mechanism to escape repression by immune-activated miRNAs. Overall, our results support concerted regulation of transcription and RNA processing in the control of gene regulatory programs engaged in the response to immune stressors.

## INTRODUCTION

Innate immune responses depend on robust and coordinated gene expression programs involving the transcriptional regulation of thousands of genes (1-3). These regulatory cascades start with the detection of microbial-associated products by pattern recognition receptors, which include Toll Like Receptors (TLRs), NOD-like receptors and specific C-type lectins (4, 5). These initial steps are followed by the activation of key transcription factors (e.g., NF-κB and interferon regulatory factors) that orchestrate the inflammatory and/or antiviral response signals involved in pathogen clearance and the subsequent development of appropriate adaptive immune responses (4).

Although much attention has been devoted to characterizing transcriptional changes in response to infectious agents or other immune stimuli, we still know remarkably little about the contribution of post-transcriptional changes – specifically, changes in alternative pre-mRNA processing – to the regulation of the immune system (6-8). Alternative splicing can significantly impact cellular function by creating distinct mRNA transcripts from the same gene, which have the potential to encode unique proteins that have distinct or opposing functions. For instance, in human and mouse cells, several alternatively spliced forms of genes involved in the TLR pathway have been shown to function as negative regulators of TLR signaling in order to prevent uncontrolled inflammation (9-11). Additionally, several in-depth studies of individual genes suggest that alternative splicing plays an important role in increasing the diversity of transcripts encoding MHC molecules (12) and in modulating intracellular signaling and intercellular interactions through the expression of various isoforms of key cytokines (13, 14), cytokine receptors (9, 15, 16), kinases (17) and adaptor proteins (18-20). Yet, the extent to which changes in isoform usage are a hallmark of immune responses to infection remains largely unexplored at a genome-wide level (8, 21). Moreover, we still do not know which RNA processing mechanisms contribute most to the regulation of immune responses, or how these disparate mechanisms are coordinated.

To address these questions, we investigated genome-wide changes in transcriptome patterns after independent infections with *Listeria monocytogenes* or *Salmonella typhimurium* in human primary macrophages. Because of the distinct molecular composition of these two pathogens and the way they interact with host cells, they activate distinct innate immune pathways after infection (22, 23). Thus, our study design allowed us to evaluate the extent to which changes in isoform usage are pathogen-specific or more generally observed in response to bacterial infection. Furthermore, the large number of individuals in our study allowed us to use natural variation in RNA processing to gain insight into fine-tuned inter-individual regulation of immune responses. Our results provide a comprehensive picture of the role of RNA processing in regulating innate immune responses to infection in human primary immune cells.

## RESULTS

Infection with either *Listeria* or *Salmonella* induces dramatic changes in mRNA expression levels. Following 2 hours of infection with each bacteria, we collected RNA-seq data from 60 matched non-infected and infected samples, with an average of 30 million reads sequenced per sample (see Methods; Table S1). The first principal component of the complete gene expression data set clearly separated infected from non-infected samples, and both PC1 and PC2 clustered infected samples by pathogen (i.e. *Listeria* or *Salmonella*; Figure 1A). Accordingly, we observed a large number of differences in gene expression levels between infected and non-infected cells, with 5,809 (39%) and 7,618 (51%) of genes showing evidence for differential gene expression (DGE) after infection with *Listeria* and *Salmonella*, respectively (using DESeq2, FDR ≤ 0.1% and $|\log_2(\text{fold change})| > 0.5$; Figure S1A, Table S2). As expected, the sets of genes that responded to either infection were strongly enriched (FDR ≤ $1.8\times10^{-6}$) for genes involved in immune-related biological processes such as the regulation of cytokine production, inflammatory responses, or T-cell activation (Table S3).

In order to uncover changes in isoform usage in response to infection using our RNA-sequencing data, we applied two complementary approaches for analyzing RNA processing. We first measured abundances of full isoforms to gain a holistic perspective of the transcriptome landscape. Second, we used the relative abundances of individual exons within a gene to gain a finer understanding of the specific RNA processing mechanisms likely to be involved in the regulation of immune responses.

**Pervasive differential isoform usage in response to infection.**

We initially sought to assess changes at in isoform usage after infection using proportional abundances of the different isoforms encoded by the same gene. To do so, we used the transcripts per million (TPM) values reported by the RSEM software (24) for each gene to calculate the relative proportions by dividing isoform-level expression by the overall expression level of the gene (i.e., TPM summed across all possible isoforms). We then quantified differential isoform usage (DIU) between conditions for genes with at least two annotated isoforms (N = 11,353) using a multivariate generalization of the Welch's t-test that allowed us to test whether, as a group, the proportional abundances of the different isoforms in each gene were significantly different between infected and non-infected cells (Methods). We found that 1,456 (13%; *Listeria*) and 2,862 (25%; *Salmonella*) genes show evidence for significant DIU after (FDR ≤ 1%; Figure 1B for an examples of an DIU gene, Figure S1B, Table S4). DIU genes were significantly enriched for genes involved in immune responses (FDR ≤ 0.01; Table S3), including several cytokines (e.g. *IL1B*, *IL7*, *IL20*), chemokines (*CCL15*), regulators of inflammatory signals (e.g. *ADORA3* in Figure 1B, *MAP3K14*) and genes encoding co-stimulatory molecules required for T cell activation and survival (e.g. *CD28*, *CD80*) (Figure S2). When we assessed DIU using isoform-specific expression levels calculated with Kallisto (an alignment-free quantification method (25)) and a relaxed FDR threshold ≤ 5%, we observed a 74% (*Listeria*) and 78% (*Salmonella*) agreement with DIU genes identified using RSEM.

This result confirms that identification of DIU genes is largely robust to the method used for isoform quantification. Notably, despite the larger number of DIU genes identified after infection with *Salmonella*, 86% of genes with DIU upon infection with *Listeria* were also classified as having DIU after infection with *Salmonella*, suggesting that changes in isoform usage are likely to be common across a wide variety of immune triggers.

Previously, it had been reported that DGE and DIU generally act independently to shape the transcriptomes of different tissues (26). In contrast, we found that DIU genes were significantly enriched for genes with DGE (Fisher's exact test}, $P < 10^{-15}$ for both *Listeria* and *Salmonella*) (Figure S1C). Despite such enrichment, a substantial proportion of genes (47% (690 genes) in *Listeria* and 39% (1,105 genes) in *Salmonella*) with significant changes in isoform usage after infection do not exhibit DGE following infection (Figure 1C). Even after relaxing the FDR threshold for DGE by 100-fold, 685 (*Listeria*) and 1096 (*Salmonella*) DIU genes still exhibited no significant DGE. Thus, a considerable fraction of transcriptome changes in response to infection does occur solely at the level of RNA processing, and independently of changes in mean expression levels.

Given the large proportion of genes with significant changes in isoform usage after infection, we sought to understand whether these changes arose from a shift in the usage of a dominant isoform or increased isoform diversity within a given gene. To do so, we calculated the Shannon diversity index, which measures the evenness of the isoform usage distribution for each gene before and after infection (low values reflect usage of one or few isoforms; high values reflect either the usage of a more diverse set of isoforms or more equal representation of the same set of isoforms). $\Delta_{Shannon}$ thus quantifies the change in isoform diversity from before to after infection (Table S5). The majority of genes (63% in *Listeria* and 68% in *Salmonella*) show an increase in isoform diversity ($\Delta_{Shannon} > 0$) after infection (Figure 1D), indicating a shift away from usage of the primary (pre-infection) isoform after infection. Notably, extreme $\Delta_{Shannon}$ values also show a relationship with differential gene

expression, with an enrichment of DGE among genes at both tails of the distribution of $\Delta_{Shannon}$ values (Figure 1E). More specifically, increased Shannon diversity tends to be associated with down-regulated genes whereas decreased Shannon diversity occurs in genes that are up-regulated in response to infection (Figure S1D). This observation suggests that the diversification of isoform usage could play a large role in the regulation of gene expression levels after infection.

**Directed shifts in 3' UTR length and alternative splicing following infection.**

Though using proportional isoform-specific abundances illustrated the substantial transcriptome changes upon infection, aggregate isoform-specific levels calculated by RSEM do not distinguish between different types of RNA processing changes. Thus, we next focused on examining changes in five distinct categories of alternative RNA processing that allowed us to zoom in on particular molecular mechanisms that underlie differential isoform usage in response to infection. Specifically, we used human transcript annotations from Ensembl to quantify usage of (1) alternative first exons (AFEs), (2) alternative last exons (ALEs), (3) alternative polyadenylation sites, leading to tandem 3' untranslated regions (TandemUTRs), (4) retained introns (RIs), and (5) skipped exons (SEs). AFEs, ALEs, and TandemUTRs correspond to the alternative usage of terminal exons primarily affecting UTR composition, whereas RIs and SEs correspond to internal splicing events that usually affect the open reading frame. For each gene or exon within each RNA processing category, we used the MISO software (27) to calculate a "percent spliced in" (PSI or $\Psi$) value (Table S6), defined as the proportion of transcripts from a gene that contain the "inclusion" isoform (defined as the longer isoform for RIs, SEs, and TandemUTRs or use of the exon most distal to the gene for AFEs and ALEs). $\Delta\Psi$ values thus represent the difference between PSI values calculated for the infected versus non-infected samples. Overall, we observed many significant changes in RNA processing ($N \geq 1{,}098$) across all categories in both bacteria; with significance defined as Bayes Factor > 5 in at least 10% of individuals and $|\text{mean } \Delta\Psi| > 0.05$), compared to null

expectations derived from measuring changes among pairs of non-infected samples (N=29) (Figure 2A, Table S7). The greatest relative number of changes after infection occurred among retained intron and TandemUTR events. PSI values for 7.3% and 14% of retained introns and 7.6% and 16.7% of TandemUTRs significantly changed after infection with *Listeria* or *Salmonella*, respectively. When we considered the set of genes associated with at least one significant change in RNA processing, we found an over-representation of Gene Ontology categories involved in immune cell processes and the cellular response to a stimulus (Figure 2B, Table S8).

Genome-wide shifts towards more prevalent usage of inclusion or exclusion isoforms have previously been observed as signatures of cellular stress responses or developmental processes (28-32). To examine directional shifts in particular RNA processing mechanisms, we used mean ΔΨ values (defined as the mean ΔΨ across all individuals for each gene or exon) and observed a striking global shift towards the usage of upstream polyadenylation sites in Tandem 3' UTR regions, indicating pervasive 3' UTR shortening following infection with either *Listeria* or *Salmonella* (Figure 2C; both $P < 2.2\times10^{-16}$ with Mann-Whitney U test; 94.4% and 98% of significant TandemUTRs changes in *Listeria* and *Salmonella*). Additionally, we found a substantial increase in the inclusion of skipped exons after infection (both $P < 2.2\times10^{-16}$ with Mann-Whitney U test; 79.8% and 77.3 % of significant SE changes in *Listeria* and *Salmonella*). The consistency of these patterns suggests that a dedicated post-transcriptional regulatory program might underlie genome-wide shifts in RNA processing after bacterial infection.

**RNA processing patterns persist 24 hours after infection.** To evaluate whether the strong global shifts we observed were the result of a sustained immune response to the bacteria rather than a short-lived cellular stress response, we sequenced RNA from a subset of our individuals (n=6) after 24 hours of infection with *Listeria* or *Salmonella*. Samples sequenced after 24 hours were paired with matched non-infected controls cultured for 24 hours to calculate ΔΨ values for

alternative RNA processing categories using MISO (27). Across all categories, we found many more significant changes in RNA processing after infection for 24 hours than after 2 hours (Figure S2A), and stronger shifts in the distributions of ΔΨ values (Figure S2B). These results further support a key role for RNA processing in controlling innate immune responses to infection. For genes or exons that changed significantly after 2 hours of infection in this set of 6 individuals (see Methods), the majority of changes (67%) remained detectable at 24 hours, with a similar magnitude and the same direction of effect (Figure 2D). The one notable exception is that introns that are retained at 2 hours in infected versus non-infected cells tend to be spliced out at 24 hours. Our results thus suggest that the splicing of introns during the course of the immune response to infection is temporally dynamic and patterns at 2 hours after infection might be reflective of cellular responses to both bacterial infection and stress (32).

**Increased exon inclusion is associated with increased in gene expression levels**

Since we observe strong connections between overall isoform usage and differential gene expression, we asked whether there was any relationship between the occurrence of particular RNA processing changes and the direction of gene expression changes in response to infection. We found that genes with significant differences in skipped exon usage pre- and post-infection (up to 80% of which have increased skipped exon inclusion after infection) are more likely to be up-regulated after infection (T-test, both $P \leq 1\times10^{-4}$} for *Listeria* and *Salmonella*; Figure 3A; Figure S4A), while genes with other types of significant changes show no consistent patterns of differential gene expression. This is consistent with a previous study of the evolution of skipped exon usage, which reported increased gene expression in tissues with increased inclusion of novel SEs (33). Since gene expression levels are calculated using a composite of reads from all isoforms, we were concerned that apparent increases in gene expression levels could arise as an artifact of a larger number of reads mapping to the skipped exon. However, even when gene expression levels were calculated using reads mapping to constitutive exons only,

we still observed an association between significant skipped exon changes and gene expression up-regulation in response to infection (Figure S4A).

Genes with significant skipped exon usage are enriched for Gene Ontology categories such as "response to stimulus" and, intriguingly, "RNA processing" and "RNA stability" (Figure S5A). Included within these latter categories are several members of 2 splicing regulatory factor families, SR proteins and hnRNPs, which show increased inclusion of skipped exons after infection, as well as up-regulation of their own gene expression levels (Figure S5B). The SR protein and hnRNP families are known to function in complex auto-regulatory pathways (34, 35), competing for use of splicing regulatory elements (SREs) within regulated exons and surrounding intronic regions (36-38). Consistent with the slightly higher up-regulation of members of the hnRNP family, we observe a consistent enrichment for intronic splicing silencers (canonically bound by hnRNPs to promote exclusion of a cassette exon (37)) in upstream intronic regions around significantly excluded skipped exons (Figure S5C).

**Variation across individuals provides insight into putative trans-regulatory factors**

A unique feature of our study design is our sampling of 60 individuals after infection with both *Listeria* and *Salmonella*. By taking advantage of natural variation in the regulation of RNA processing to infection, we aimed to gain further insight into the connections between RNA processing and changes in gene expression levels in response to infection. For each gene, we calculated the correlation between the inter-individual variation in RNA processing changes ($\Delta\Psi$) and the fold changes in gene expression levels upon infection. For most categories, we observe shifts in the distribution of correlations between RNA processing and gene expression levels relative to permuted controls (Figure S4B), with skipped exons and TandemUTRs showing the most consistent patterns (Kolmogorov-Smirnov test, $P \leq 0.005$; Figure 3C). Increased skipped exon usage correlates with increased gene expression of the

gene across individuals in 69% of the genes with significant skipped exons. In addition, 3' UTR shortening tends to be correlated with increased up-regulation of the associated genes in response to infection. Our findings thus suggest that transcriptional responses to infection are coordinately regulated with changes in RNA processing, either as a direct consequence of the RNA processing change or co-regulation by similar regulatory cascades.

Although the majority of individuals have increased skipped exon inclusion or 3' UTRs shortening overall, we also found substantial inter-individual variation in the genome-wide average amount of inclusion or shortening of skipped exons or 3' UTRs, respectively (Figure S6A). Strikingly, and further supporting a strong link between transcriptional responses and changes in RNA processing, variation in mean RNA processing correlates strongly with average gene expression changes upon infection, particularly for skipped exons and TandemUTRs (Figure S6B).

Next, we sought to exploit our relatively large sample size to identify candidate *trans*-factors controlling the amount of skipped exon and TandemUTR changes observed across individuals. Specifically, we examined the relationship between variation in mean RNA processing changes with fold changes in gene expression for all expressed genes. When we focus on proteins with known RNA binding functions or domains – hypothesizing that these factors are more likely to regulate post-transcriptional mechanisms – we find many significant correlations (FDR ≤ 1%) for known regulators of the corresponding RNA processing category (Figure S7). For instance, genes whose changes in expression correlated with individual-specific mean 3' UTR shortening included several factors previously implicated in regulating polyadenylation site usage (such as cleavage and polyadenylation factors (39), hnRNP H proteins (27), and hnRNP F (40); Figure S7A). Additionally, many known regulators of alternative splicing – including a number of SR proteins and hnRNPs – have increased expression levels in individuals with larger extents of skipped exon inclusion (Figure S7B). Intriguingly, fold changes in hnRNP H1 gene expression also have the highest correlation with skipped exon inclusion, perhaps indicating a dual

regulatory role for the hnRNP H family of factors in alternative splicing and 3' end processing.

**Global shortening of Tandem 3' UTRs to evade regulation by immune-related miRNAs.**

The largest global shift observed in RNA processing was for 3' UTR shortening, where as many as 98% of significantly changing TandemUTRs involved a mean shift towards usage of an upstream polyA site post-infection. Previously, global 3' UTR shortening has been associated with proliferating cells, particularly in the context of cellular transitions in development (28), cell differentiation (41), or cancerous (29) states. However, macrophages do not proliferate, which we confirmed with a BrdU labeling assay in both resting and infected conditions (Figure S8). This result indicates that the pervasive 3' UTR shortening observed in response to infection is due to an active cellular process that is independent of cell division and might be mechanistically distinct from that leading to 3' UTR shortening in proliferating cells.

Previous studies in proliferating cells postulated that 3' UTR shortening can act as a way to evade regulation by micro RNAs (miRNAs), since crucial miRNA target sites are most often found in 3' UTR regions (28). To evaluate this hypothesis we performed small-RNA sequencing in 6 individuals after 2 hours and 24 hours of infection with *Listeria* and *Salmonella*. When focusing specifically on miRNAs expressed in macrophages (Table S9), we found that the extended UTR regions of infection-sensitive 3' UTRs have a significantly higher density of miRNA target sites compared to TandemUTRs that do not change in response to infection (Figure 4A). This increased density of miRNA target sites is restricted to the region that gets shortened after infection: no increase in the density of target sites was observed in the common core 3' UTR regions of the same genes (Figure S9).

Next, we tested if the increased density of target sites was driven by the enrichment of target sites for particular miRNAs expressed in macrophages. For each miRNA

that was expressed in either non-infected or infected macrophages, we calculated an enrichment score assessing whether the target sites of that miRNA were significantly enriched in the extended region of significantly shortened 3' UTRs (see Supplemental Methods). We found 10 miRNAs with significantly enriched target sites (FDR ≤ 10% and enrichment ≥ 1.5 fold) in either the *Listeria* or *Salmonella* conditions, with 5 miRNAs over-represented in both bacterial conditions (Figure 4B). Interestingly, of the the miRNAs with the highest enrichment after infection with either *Listeria* or *Salmonella*, two (miR-146b and miR-125a) have previously been shown to be important regulators of the innate immune response (42-46). Accordingly, we found that these three miRNAs are all up-regulated following infection, and strongly so after 24 hours of infection (Figure 4B). Along with the observation that, in the majority of cases, more 3' UTR shortening is associated with an increase in gene expression across individuals, this suggests that shifts towards shorter 3' UTRs after infection are likely to be a cellular control mechanism that allows these transcripts to escape from regulation by specific immune-induced miRNAs.

## DISCUSSION

Taken together, our results provide strong evidence that RNA processing plays a key role in the regulation of innate immune responses to infection. To our knowledge, our study is the first to comprehensively characterize the prevalence of differential isoform usage following an immune stress in mammals. Despite known differences in the regulatory pathways elicited by macrophages in response to *Listeria* and *Salmonella*, our data revealed striking similarities in the overall patterns of RNA processing induced in response to both pathogens. Given the strong overlaps, we chose throughout this study to focus on consistent patterns across the two bacteria.

Though genes that have differential isoform usage in response to infection are more likely to show changes in gene expression, we found that as many as 47% of genes showing differential isoform usage have no evidence for differences in gene

expression upon infection. This observation suggests that a considerable proportion of genes are regulated only by post-transcriptional mechanisms, without noticeable changes in transcriptional regulation, to create a greater pool of isoform diversity in the cell. Overall, genes with increased diversity upon infection tend to be down-regulated (Figure S1D), raising the possibility that increased isoform diversification functions as a post-transcriptional regulator of steady-state mRNA abundance (47, 48).

In conjunction with our initial observations of overall increases in isoform diversity, it is plausible that increased inclusion of skipped exons is explained by a general up-regulation of splicing machinery following infection (49), which would allow the cell to recognize and use additional splice sites. Yet, we note that 20% of the significant skipped exon changes result in the exclusion of the exon, which indicates that other regulatory mechanisms must also be at play. In support of this argument, members of both the hnRNP and SR protein splicing factor families show up-regulation of their overall gene expression levels (Figure S5B) and are known to bind splicing regulatory silencing elements that are enriched around excluded exons, perhaps explaining how some exons are skipped even in a cellular milieu that favors exon inclusion.

The most prominent hallmark of RNA processing changes after infection is a global shortening of 3' UTRs. Interestingly, this effect resembles the global 3' UTR shortening seen in proliferating (28), cancerous (29), or differentiating (41) cells, suggesting that immune pathways might co-opt innate developmental mechanisms as a way to escape regulation by dynamically changing miRNAs (28, 29). Alternatively, the shift towards shorter 3' UTRs could be due to increased miRNA-dependent degradation of transcripts with longer 3' UTRs post-infection. However, genes that shift towards shorter 3' UTRs do not tend to be more lowly expressed post-infection, making this explanation unlikely. More specifically, our data suggest that infected macrophages specifically seek to escape regulation by a subset of miRNAs that are up-regulated after infection, including miR-146b, miR-3661, and

miR-151b (Figure 4B). These miRNAs target the extended 3' UTR regions of several critical activators of the innate immune response – including the IRF5 transcription factor (50) and the MAP kinases MAPKAP1 and MAP2K4 (51) – all of which shift to usage of shorter 3' UTRs after infection. It is tempting to speculate that without such 3'UTR shortening, the binding of one or more of these miRNAs to the 3' UTR region of these genes would compromise their up-regulation and the subsequent activation of downstream immune defense pathways.

The convergence towards similar isoform outcomes across many disparate genes suggests the activation of factors acting in *trans* to drive global shifts towards inclusion of cassette exons or usage of upstream proximal polyadenyalation sites. Taking advantage of our relatively large sample size, we were able to identify candidate *trans*-factors whose transcriptional up-regulation upon infection might have larger downstream consequences for widespread RNA processing patterns. Intriguingly, many candidate RNA-binding proteins are promising due to their known roles in regulating either alternative splicing or 3'end processing (Figure S7). Furthermore, individuals who exhibited larger global shifts in RNA processing also tended to exhibit larger global shifts in gene expression levels upon infection (Figure S6B). These relationships suggest that one or several *trans*-factor might have functional consequences on the canonically tight regulation of changes in splicing and gene expression of many genes. These patterns might be due to stochastic variation in levels of the *trans*-factor(s) or to genetic differences among individuals. The latter case would be particularly interesting because it might help to explain variation in individual susceptibility to infection. Thus, using genetic mapping approaches to identify loci that are associated with both *cis* and/or *trans* regulated variation in alternative RNA processing might provide further insights into the molecular mechanisms underlying inter-individual variation in immune responses to infection.

## Materials and Methods

Complete details of the experimental and statistical procedures can be found in SI Materials and Methods. Briefly, blood samples from 60 healthy donors were obtained from Indiana Blood Center with informed consent and ethics approval from the ethics committee at the CHU Sainte-Justine (protocol #4022). All individuals recruited in this study were healthy Caucasian males between the ages of 21 and 55 y old. Blood mononuclear cells from each donor were isolated by Ficoll-Paque centrifugation and blood monocytes were purified from peripheral blood mononuclear cells (PBMCs) by positive selection with magnetic CD14 MicroBeads (Miltenyi Biotec). In order to derive macrophages, monocytes were then cultured for 7 days in RPMI-1640 (Fisher) supplemented with 10% heat-inactivated FBS (FBS premium, US origin, Wisent), L-glutamine (Fisher) and M-CSF (20ng/mL; R&D systems). After validating the differentiation/activation status of the monocyte-derived macrophages we infected them at a multiplicity of infection (MOI) of 10:1 for *Salmonella typhimurium* and an MOI of 5:1 for *Listeria monocytogenes* for 2- and 24-hours. Genome-wide gene expression profiles of untreated and infected macrophages were obtained by RNA-sequencing for both mRNA transcripts and small RNAs. After a series of quality checks (SI Materials and Methods), mRNA transcript abundances were estimated using RSEM (52) and Kallisto (25) and microRNA expression estimates were obtained as previously described(53). To detect genes with differential isoform usage between two groups of non-infected and infected samples we used a multivariate generalization of the Welch's t-test. Shannon entropy $H_{sh}$ (also known as Shannon index) was applied to measure the diversity of isoforms for each target gene before and after infection. Changes across individual RNA processing events were quantified using the MISO software package (v0.4.9) (27) using default settings and hg19 version 1 annotations. Events were considered to be significantly altered post-infection if at least 10% (n >= 6) of individuals had a BF >= 5 and the |mean $\Delta\Psi$| ≥ 0.05 (Table S6). We used a custom gene ontology script to test for enrichment of functional annotations among genes that significantly changed isoform usage in response to infection (SI Materials and Methods). All predicted miRNA target sites within annotated TandemUTR regions were obtained using TargetScan (v6.2) (54).

**Acknowledgements**

We thank P Freese for a categorized list of RNA binding proteins and G McVicker for an initial script to perform iterative gene ontology analyses. We thank Y Katz, JM Taliaferro, P Sudmant, J Tung, and members of the Barreiro and Burge labs for helpful discussions and comments on the manuscript. This work was supported by grants from the Canadian Institutes of Health Research (232519), the Human Frontiers Science Program (CDA-00025/2012) and the Canada Research Chairs Program (950-228993) (to L.B.B). We thank Calcul Quebéc and Compute Canada for managing and providing access to the supercomputer Briaree from the University of Montreal, used to do many computations reported in the manuscript. Additional computational resources were performed on an MIT computational cluster supported by the National Science Foundation (Grant No. 0821391). A.A.P. was supported by a Jane Coffin Childs postdoctoral fellowship. G. B. was supported by a postdoctoral fellowship from the Fonds de la Recherche en Santè du Québec.

## References


1. Huang Q, et al. (2001) The Plasticity of Dendritic Cell Responses to Pathogens and Their Components. *Science* 294(5543):870–875.

2. Smale ST (2010) Selective Transcription in Response to an Inflammatory Stimulus. *Cell* 140(6):833–844.

3. Medzhitov R, Horng T (2009) Transcriptional control of the inflammatory response. *Nature Reviews Immunology* 9(10):692–703.

4. Medzhitov R, Janeway CA Jr (1998) Innate immune recognition and control of adaptive immune responses. *Seminars in Immunology* 10(5):351–353.

5. Kawai T, Akira S (2010) The role of pattern-recognition receptors in innate immunity: update on Toll-like receptors. *Nat Immunol* 11(5):373–384.

6. Lynch KW (2004) Consequences of regulated pre-mRNA splicing in the immune system. *Nature Reviews Immunology* 4(12):931–940.

7. Martinez NM, Lynch KW (2013) Control of alternative splicing in immune responses: many regulators, many predictions, much still to learn. *Immunological Reviews* 253(1):216–236.

8. Alasoo K, Estrada FM, Hale C, Gordon S, Powrie F (2015) Transcriptional profiling of macrophages derived from monocytes and iPS cells identifies a conserved response to LPS and novel alternative transcription. *bioRxiv*.

9. Rao N, Nguyen S, Ngo K, Fung-Leung W-P (2005) A novel splice variant of interleukin-1 receptor (IL-1R)-associated kinase 1 plays a negative regulatory role in Toll/IL-1R-induced inflammatory signaling. *Molecular and Cellular Biology* 25(15):6521–6532.

10. Wells CA, et al. (2006) Alternate transcription of the Toll-like receptor signaling cascade. *Genome Biology* 7(2):R10.

11. O'Connor BP, et al. (2015) Regulation of Toll-like Receptor Signaling by the SF3a mRNA Splicing Complex. *PLoS Genet* 11(2):e1004932.

12. Ishitani A, Geraghty DE (1992) Alternative splicing of HLA-G transcripts yields proteins with primary structures resembling both class I and class II antigens. *Proceedings of the National Academy of Sciences* 89(9):3947–3951.

13. Bihl MP, et al. (2002) Identification of a novel IL-6 isoform binding to the endogenous IL-6 receptor. *American journal of respiratory cell and molecular biology* 27(1):48–56.

14. Nishimura H, et al. (2000) Differential roles of interleukin 15 mRNA isoforms generated by alternative splicing in immune responses in vivo. *The Journal of experimental medicine* 191(1):157–170.

15. Goodwin RG, et al. (1990) Cloning of the human and murine interleukin-7 receptors: demonstration of a soluble form and homology to a new receptor superfamily. *Cell* 60(6):941–951.

16. Koskinen LL, et al. (2009) Association study of the IL18RAP locus in three European populations with coeliac disease. *Hum Mol Genet* 18(6):1148–1155.

17. Jensen LE, Whitehead AS (2001) IRAK1b, a novel alternative splice variant of interleukin-1 receptor-associated kinase (IRAK), mediates interleukin-1 signaling and has prolonged stability. *Journal of Biological Chemistry* 276(31):29037–29044.

18. Gray P, et al. (2010) Identification of a novel human MD-2 splice variant that negatively regulates Lipopolysaccharide-induced TLR4 signaling. *The Journal of Immunology* 184(11):6359–6366.

19. Ohta S, Bahrun U, Tanaka M, Kimoto M (2004) Identification of a novel isoform of MD-2 that downregulates lipopolysaccharide signaling. *Biochem Biophys Res Commun* 323(3):1103–1108.

20. Janssens S, Burns K, Tschopp J, Beyaert R (2002) Regulation of interleukin-1- and lipopolysaccharide-induced NF-κB activation by alternative splicing of MyD88. *Current Biology* 12(6):467–471.

21. Rodrigues R, Grosso AR, Moita L (2013) Genome-wide analysis of alternative splicing during dendritic cell response to a bacterial challenge. *PLoS One*.

22. Haraga A, Ohlson MB, Miller SI (2008) Salmonellae interplay with host cells. *Nature Reviews Microbiology* 6(1):53–66.

23. Pamer EG (2004) Immune responses to Listeria monocytogenes. *Nature Reviews Immunology* 4(10):812–823.

24. Li B, Ruotti V, Stewart RM, Thomson JA, Dewey CN (2010) RNA-Seq gene expression estimation with read mapping uncertainty. *Bioinformatics* 26(4):493–500.

25. Bray N, Pimentel H, Melsted P, Pachter L (2015) Near-optimal RNA-Seq quantification. *arXiv* q-bio.QM.

26. Barbosa-Morais NL, et al. (2012) The evolutionary landscape of alternative splicing in vertebrate species. *Science* 338(6114):1587–1593.

27. Katz Y, Wang ET, Airoldi EM, Burge CB (2010) Analysis and design of RNA sequencing experiments for identifying isoform regulation. *Nat Meth* 7(12):1009–1015.

28. Sandberg R, Neilson JR, Sarma A, Sharp PA, Burge CB (2008) Proliferating Cells Express mRNAs with Shortened 3' Untranslated Regions and Fewer MicroRNA Target Sites. *Science* 320(5883):1643–1647.

29. Mayr C, Bartel DP (2009) Widespread Shortening of 3´UTRs by Alternative Cleavage and Polyadenylation Activates Oncogenes in Cancer Cells. *Cell* 138(4):673–684.

30. Wong JJL, et al. (2013) Orchestrated Intron Retention Regulates Normal Granulocyte Differentiation. *Cell* 154(3):583–595.

31. Lackford B, et al. (2014) Fip1 regulates mRNA alternative polyadenylation to promote stem cell self-renewal. *The EMBO Journal* 33(8):878–889.

32. Shalgi R, Hurt JA, Lindquist S, Burge CB (2014) Widespread inhibition of posttranscriptional splicing shapes the cellular transcriptome following heat shock. *Cell Reports* 7(5):1362–1370.

33. Merkin JJ, Chen P, Alexis MS, Hautaniemi SK (2015) Origins and impacts of new mammalian exons. *Cell Reports* 10(12):1992–2005.

34. Huelga SC, et al. (2012) Integrative Genome-wide Analysis Reveals Cooperative Regulation of Alternative Splicing by hnRNP Proteins. *Cell Reports* 1(2):167–178.

35. Pandit S, et al. (2013) Genome-wide analysis reveals SR protein cooperation and competition in regulated splicing. *Molecular Cell* 50(2):223–235.

36. House AE, Lynch KW (2008) Regulation of Alternative Splicing: More than Just the ABCs. *Journal of Biological Chemistry* 283(3):1217–1221.

37. Wang Z, Burge CB (2008) Splicing regulation: From a parts list of regulatory elements to an integrated splicing code. *RNA* 14(5):802–813.

38. Busch A, Hertel KJ (2012) Evolution of SR protein and hnRNP splicing regulatory factors. *WIREs RNA* 3(1):1–12.

39. Zheng D, Tian B (2014) RNA-binding proteins in regulation of alternative cleavage and polyadenylation. *Systems Biology of RNA Binding Proteins* (Springer), pp 97–127.

40. Veraldi KL, et al. (2001) hnRNP F Influences Binding of a 64-Kilodalton Subunit of Cleavage Stimulation Factor to mRNA Precursors in Mouse B Cells.

*Molecular and Cellular Biology* 21(4):1228–1238.

41. Lackford B, et al. (2014) Fip1 regulates mRNA alternative polyadenylation to promote stem cell self-renewal. *The EMBO Journal* 33(8):878–889.

42. Taganov KD, Boldin MP, Chang KJ (2006) NF-κB-dependent induction of microRNA miR-146, an inhibitor targeted to signaling proteins of innate immune responses.

43. Perry MM, Williams AE, Tsitsiou E, Larner-Svensson HM, Lindsay MA (2009) Divergent intracellular pathways regulate interleukin-1β-induced miR-146a and miR-146b expression and chemokine release in human alveolar epithelial cells. *FEBS Letters* 583(20):3349–3355.

44. Boldin MP, Baltimore D (2012) MicroRNAs, new effectors and regulators of NF-κB. *Immunological Reviews* 246(1):205–220.

45. Kim SW, Ramasamy K, Bouamar H (2012) MicroRNAs miR-125a and miR-125b constitutively activate the NF-κB pathway by targeting the tumor necrosis factor alpha-induced protein 3 (TNFAIP3, A20) doi:10.1073/pnas.1200081109/-/DCSupplemental/pnas.201200081SI.pdf.

46. Banerjee S, et al. (2013) miR-125a-5p Regulates Differential Activation of Macrophages and Inflammation. *Journal of Biological Chemistry* 288(49):35428–35436.

47. Lewis BP, Green RE, Brenner SE (2003) Evidence for the widespread coupling of alternative splicing and nonsense-mediated mRNA decay in humans. *PNAS* 100(1):189–192.

48. Green RE, et al. (2003) Widespread predicted nonsense-mediated mRNA decay of alternatively-spliced transcripts of human normal and disease genes. *Bioinformatics* 19(suppl 1):i118–i121.

49. Munding EM, Shiue L, Katzman S, Donohue JP, Ares M Jr (2013) Competition between Pre-mRNAs for the Splicing MachineryDrives Global Regulation of Splicing. *Molecular Cell* 51(3):338–348.

50. Lawrence T, Natoli G (2011) Transcriptional regulation of macrophage polarization: enabling diversity with identity. *Nature Reviews Immunology* 11(11):750–761.

51. Arthur JSC, Ley SC (2013) Mitogen-activated protein kinases in innate immunity. *Nature Reviews Immunology* 13(9):679–692.

52. Li B, Dewey CN (2011) RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics* 12(1):323.

53. Siddle KJ, Deschamps M, Tailleux L (2014) A genomic portrait of the genetic architecture and regulatory impact of microRNA expression in response to infection. *Genome Research*. doi:10.1101/gr.161471.113.

54. Friedman RC, Farh KK-H, Burge CB, Bartel DP (2009) Most mammalian mRNAs are conserved targets of microRNAs. *Genome Research* 19(1):92–105.

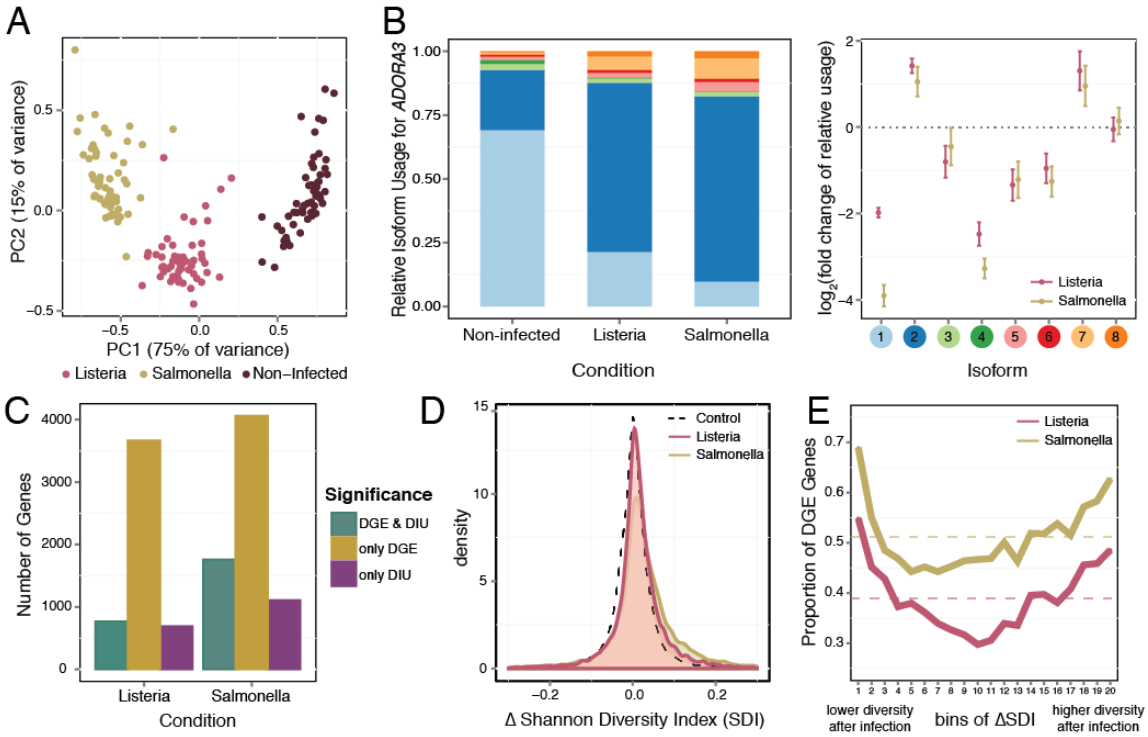


**Figure 1. Gene expression and isoform proportion differences in response to bacterial infection.** (A) Principal component analysis of the gene expression data from all the samples (PC1 and PC2 are plotted on the *x*- and *y-axis*, respectively), illustrating three separate clusters based on sample conditions. (B) Example of a gene, *ADORA3*, that shows significant changes in isoform usage upon infection with *Listeria* and *Salmonella*. The left panel shows the average relative isoform usage across samples and the right panel shows the corresponding fold changes in $\log_2$ scale (with standard error bars) for each of the isoforms of *ADORA3* in response to infection. Isoforms are ordered by their relative abundance in the non-infected condition, and the colored circles on the right panel correspond to the bar colors in the left panel. (C) Number of genes showing only DIU, only DGE, and both DIU and DGE upon infection with *Listeria* and *Salmonella*, out of the 11,353 genes tested. (D) Distributions of $\Delta_{\text{Shannon}}$ diversity index following infection with *Listeria* and *Salmonella*. Isoform diversity increases in both cases relative to the null distribution (*black dotted line*) generated by permuting samples across conditions. (E) The proportion of DGE genes (*y-axis*) among 5% quantile bins of $\Delta_{\text{Shannon}}$ values (*x-axis*), for both *Listeria* (*pink*) and *Salmonella* (*gold*). Dotted lines indicate the overall proportion of DGE genes in each condition.

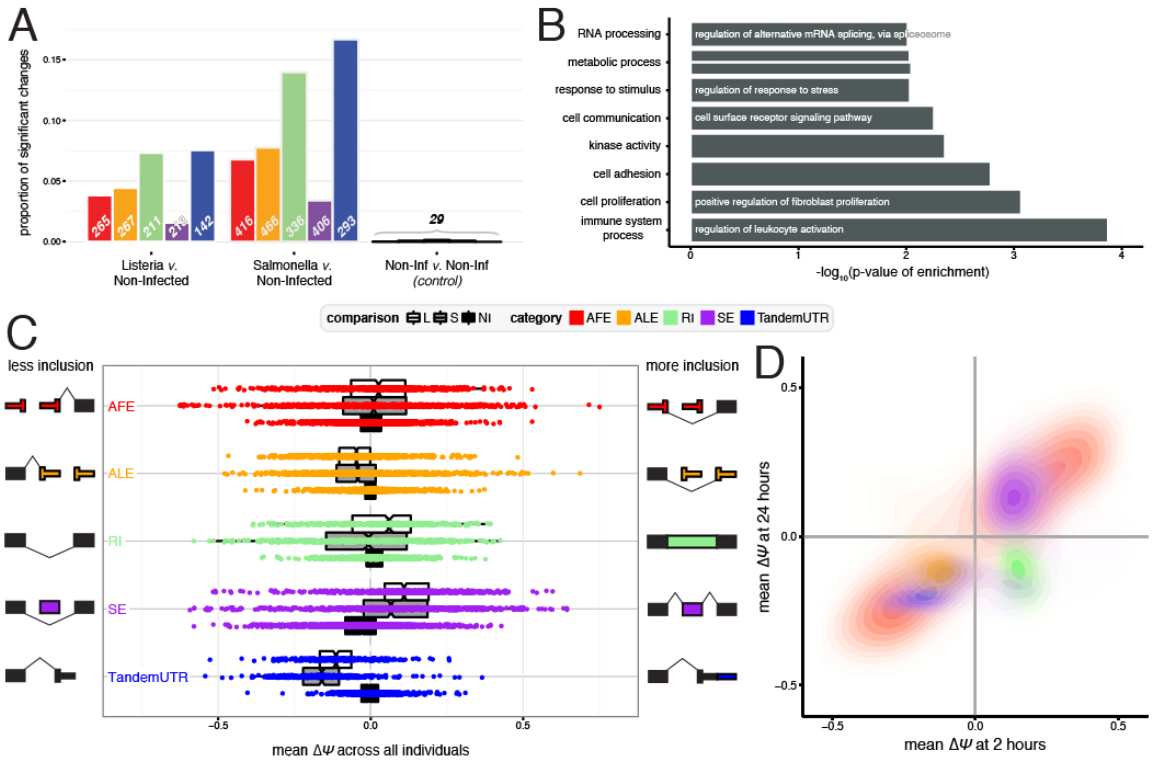


**Figure 2. RNA processing changes in response to bacterial infection.** (A) The proportion of significant events detected for RNA processing category within each comparison: *Listeria*-infected v. non-infected (*left*), *Salmonella*-infected v. non-infected (*middle*), and non-infected v. non-infected as a control (*right*). Numbers indicate the number of significant changes per category. (B) Significantly enriched gene ontology categories for genes with any significant RNA processing change (FDR ≤ 10%). (C) Distribution of ΔΨ values for each RNA processing category per condition. Negative values represent less inclusion, while positive values represent more inclusion, as defined by the schematic representations next to each category. (D) Correlations between ΔΨ values after 2 hours of infection (*x-axis*) vs. ΔΨ values after 24 hours of infection (*y-axis*), presented as density plots per RNA processing category in contrasting colors. Only significant changes after 2 hours of either *Listeria* or *Salmonella* infection are plotted.

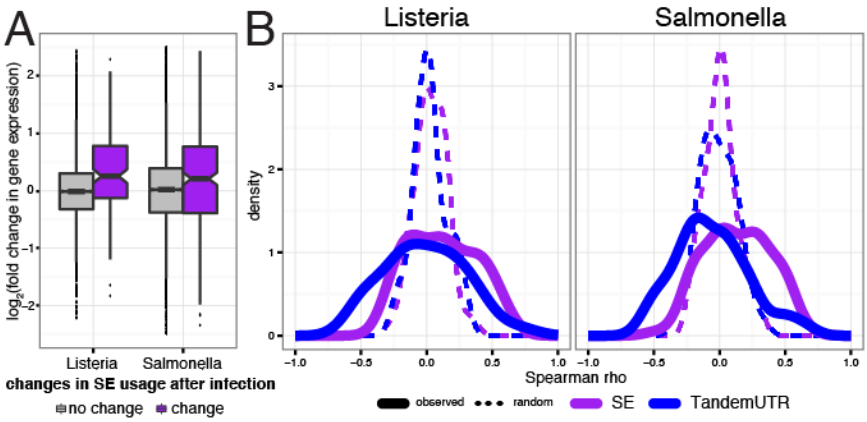


**Figure 3. Relationship between RNA processing changes and gene expression changes.** (A) Distribution of overall fold changes in gene expression (*y-axis*, $\log_2$ scale) for genes that have significant skipped exon changes after infection (purple) and genes with no change in skipped exon usage after infection (*gray*). Genes that show significant changes in skipped exon usage post-infection are more likely to also be upregulated post-infection. (B) Distribution of Spearman correlations between ΔΨ and fold change in gene expression levels measured in n=60 individuals per gene, for genes with only one annotated alternative RNA processing event and significantly changed skipped exon usage (*purple*; $n_L$ = 46 genes and $n_S$ = 97 genes) or tandem 3' UTR usage (*blue*; $n_L$ = 36 genes and $n_S$ = 86 genes). Solid lines show observed values and dotted lines show the distribution of the same correlation coefficients after permuting ΔΨ against gene expression fold-change values.

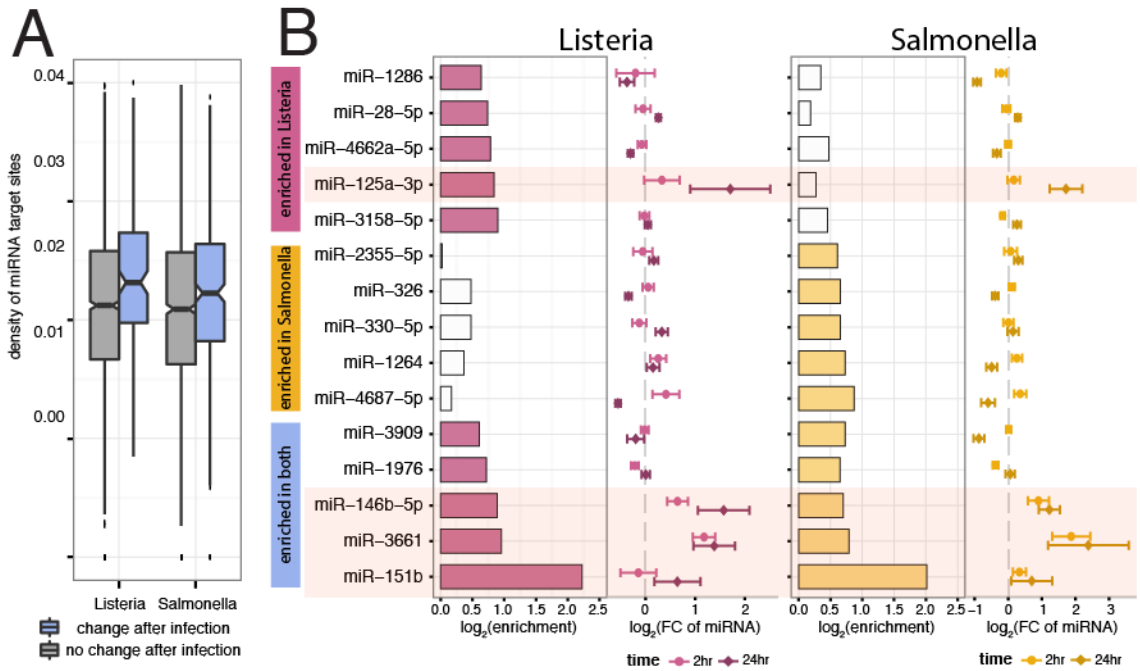


**Figure 4. Tandem 3' UTR shortening allows evasion of regulation by miRNAs.** (A) Distribution of frequency of miRNA target sites per nucleotide in the extended regions of Tandem UTRs that either show no change after infection (*grey*) or significantly change after infection (*blue*). (B) Significantly enriched miRNA target sites (FDR ≤ 10%) in the extended regions of significantly changing Tandem UTRs after infection with *Listeria*-only (*top*), with *Salmonella*-only (*middle*), or both bacteria (*bottom*). For each bacteria, the barplots in the left panels show the fold enrichment (*x-axis*, $\log_2$ scale) of target sites in the extended regions. White bars represent non-significant enrichments. Panels on the right show the fold change in miRNA expression (*x-axis*, $\log_2$ scale with standard error bars) after either 2 hours of infection (*light colors*) or 24 hours of infection (*dark colors*).