Published in final edited form as:

Nature. 2018 December ; 564(7736): 434–438. doi:10.1038/s41586-018-0794-7.

The Translation of Non-Canonical Open Reading Frames Controls Mucosal Immunity

Ruaidhrí Jackson[1], **Lina Kroehling**[1], **Alexandra Khitun**[2], **Will Bailis**[1], **Abigail Jarret**[1], **Autumn G. York**[1], **Omair M. Khan**[1], **J. Richard Brewer**[1], **Mathias H. Skadow**[1], **Coco Duizer**[1], **Christian C. D. Harman**[1], **Lelina Chang**[1], **Piotr Bielecki**[1], **Angel G. Solis**[1], **Holly R. Steach**[1], **Sarah Slavoff**[2,3,4], **Richard A. Flavell**[1,5,6]

[1]Department of Immunobiology, Yale University School of Medicine, New Haven, CT 06520, USA

[2]Department of Chemistry, Yale University, New Haven, Connecticut 06520, United States

[3]Chemical Biology Institute, Yale University, West Haven, Connecticut 06516, United States

[4]Department of Molecular Biophysics and Biochemistry, Yale University, New Haven, Connecticut 06529, United States

[5]Howard Hughes Medical Institute, Yale University, New Haven, CT 06520, USA

Abstract

The annotation of the mammalian protein coding genome is incomplete. Arbitrary open reading frame (ORF) size restriction and the absolute requirement for a methionine codon as the sole initiator of translation, have constrained identification of potentially important transcripts with non-canonical protein coding potential[1,2]. Using unbiased transcriptomic approaches in macrophages responding to bacterial infection, we show widespread ribosome association with a large number of RNAs that were previously annotated as “non-protein coding”. Although the ability of such non-canonical ORFs to encode functional protein is controversial[3,4], we identify a plethora of novel short and non-ATG initiated ORFs with the ability to generate stable and spatially distinct proteins. Importantly, we show that the translation of a novel ORF ‘hidden’ within the long non-coding RNA *Aw112010* is essential for the orchestration of mucosal immunity during both bacterial infection and colitis. Together this work expands our interpretation of the protein coding genome and demonstrates the critical nature of proteinaceous products generated

[6]Correspondence to: R.A.F. richard.flavell@yale.edu.

Contributions

R.J. conceived the project, performed experiments, analyzed the data and wrote the manuscript. L.K. performed all bioinformatics analysis and aided in writing of the manuscript. A.K. preformed all the mass spectrometry experiments, analyzed data and aided in writing of the manuscript. W.B. performed experiments and contributed major conceptual insight into the work. A.J. and A.G.Y. participated in experimental design, conducted experiments, analyzed data and offered vital conceptual insight. O.M.K., J.R.B., M.H.S. and C.D. performed experiments and analyzed data. C.C.D.H. helped with bioinformatics analysis and provided helpful conceptual discussion. L.C., P.B., A.G.S. and H.R.S. helped with experiments. S.S. supervised all mass spectrometry work and contributed to the overall interpretation of this work. R.A.F. supervised the project, helped interpret the work and supervised writing of the manuscript. No author has a competing financial interest.

Data Availability Statement

RNA-seq, RiboTag RNA-seq, and Ribosome Profiling data that support the findings of this study have been deposited in the GEO repository with the accession code GSE120762. All RNAseq, RiboTagSeq, Ribosome Profiling Seq and analysis can also be found in Supplementary Table 1.

from non-canonical ORFs to the immune response *in vivo*. We therefore propose that the misannotation of non-canonical ORF-containing genes as non-coding RNAs may obscure the essential role of a multitude of previously undiscovered protein coding genes in immunity and disease.

Ribosome association with mRNA is essential for protein synthesis[5]. Here we investigated the genome wide association of mRNA with ribosomes in macrophages upon bacterial infection by generating RiboTag-LysM-Cre mice[6] (RiboTag^{LysM})[7]. BMDMs from WT and RiboTag^{LysM} were generated and stimulated with LPS for 6 or 24 hr. Immunoprecipitation using HA-conjugated magnetic beads and subsequent RNA isolation yielded high-quality RNA from BMDMs from RiboTag^{LysM} mice and no detectable ribosome associated RNA in WT cells (Extended Data Fig. 1a, b). RiboTag-RNAseq revealed widespread differential ribosome assembly on both protein-coding RNAs and, unexpectedly, on transcripts mapping to “non-coding” RNAs (Fig. 1a)[8–10]. In fact, almost 10% of all ribosome associated RNAs were annotated as “non-coding”, with pseudogenes and long non-coding RNAs (lncRNAs) representing the most abundant classes (Fig. 1b, c). lncRNAs have been described to have major functions in diverse biological systems, including immune cell development and function[11,12]. According to canonical open reading frame designations, lncRNA lack the ability to code for protein[11,12], however we found that over 35% of highly expressed macrophage lncRNAs interact with ribosomes during bacterial infection (Fig. 1d). To identify potentially important protein coding lncRNAs in the innate immune response, we identified genes significantly altered following LPS stimulation (Fig. 1e, f). We confirmed ribosome association with candidate lncRNAs in RiboTag^{LysM} BMDMs stimulated with LPS or infected with *Salmonella enterica* serovar Typhimurium (*S.* Typhimurium) using qPCR. Both LPS stimulation and *S.* Typhimurium infection induced differential ribosome association and transcription of these genes (Fig. 1g and Extended Data Fig. 1c). Finally, we infected RiboTag^{LysM} mice with *S.* Typhimurium and isolated ribosome-associated RNA from colonic macrophages 24hr post bacterial gavage. qPCR analysis revealed significant increases in ribosome-association for the lncRNAs Aw112010 and GM9895 after infection (Fig. 1h).

Although RiboTag-RNAseq reveals widespread association of lncRNAs with ribosomes during bacterial infection, ribosome-association per se does not necessarily indicate if a given RNA is being actively translated[13]. Ribosome profiling techniques have emerged as powerful tools to address such caveats[14]. Using steady state and LPS stimulated WT BMDMs, we generated genome wide ribosome profiles in tandem with conventional Poly-A $^{+}$ RNA sequencing. This allowed the successful discrimination of the known protein coding genes and non-coding RNAs (Extended Data Fig. 2a, b). Interestingly however, we also identified a plethora of lncRNA with distinct ribosome profiles similar to that of known protein coding genes (Extended Data Fig. 2c). We next sought to identify actively translated ORFs in an unbiased manner and conducted RibORF analysis[15]. RibORF correctly identified transcripts undergoing active translation (Extended Data Fig. 2d). During the classical annotation of the genome, protein-coding genes were accurately identified by the presence of an ATG methionine start codon and an ORF greater than 300 nucleotides[16,17]. Interestingly however, both proteinaceous products as little as 11 amino acids[18] and single

nucleotide promiscuity in near cognate ATG codons that facilitate translation initiation[19] have been reported. Using a custom ORFfinder search to identify all ORFs more than 30 nucleotides long using start codons ATG, CTG, TTG, or GTG[20], we generated a library of all potential non-canonical ORFs within BMDM expressed lncRNAs. RibORF analysis identified 224 non-canonical ORFs with the same translation hallmarks (percentage of maximum entropy (PME) ≥ 0.6[21]) as protein coding genes and predicted they undergo active translation (Fig. 2a). Additionally, we wished to identify ORFs using a method which does not require the ORF to be initially defined with ORFfinder. RiboCode analysis[22] identified 85 non-canonical ORFs within lncRNAs de novo (Extended Data Fig. 3a). As previously reported[22], RibORF and RiboCode analyses identify both similar and unique ORFs (Extended Data Fig. 3b). In order to further investigate the translational veracity of these ORFs, we conducted another analytical strategy previously used to challenge the assertion that lncRNAs can encode translated proteins. Employing Translational Efficiency (TE)[4] and Ribosome Release Score (RRS) analysis to both the 3′ untranslated regions (UTRs) and the known protein coding ORF within these genes, we generated a frame of reference for macrophage translated and non-translated transcripts. Using the 95th percentiles of the 3′UTRs TE and RRS, we can predict the ability of a given open reading frame to encode a translated protein (Fig. 2b). By selecting non-canonical ORFs identified by RibORF and RiboCode analysis, we identified 96 lncRNAs which share RRS and TE values with those of the known protein coding genome. Nearly half of these ORFs utilize the ATG start codon while the remaining ORFs shared a relatively even distribution of start codon usage between, TTG, GTG and CTG (Fig. 2c). The vast majority are small ORFs under 300 nt in length (Fig. 2d). To identify putative ORFs with the ability to encode functionally important proteins in the immune response, we performed differential gene expression analysis on these non-canonical ORFs in LPS stimulated macrophages from the ribosome profiling sequencing dataset (Fig. 2e and Extended Fig. 3c). Aw112010 showed a dramatic and significant upregulation by LPS and as it was identified as protein coding by RRS+TE+, RibORF and RiboCode analysis, we choose to investigate this non-canonical ORF further. Overexpression of Aw112010-ORF in HEK293 cells demonstrated robust protein expression and distinct subcellular localization (Fig. 2f), while overexpression of other identified ORFs, Gm9895 and Gm7160, showed strong vesicular localization and predominantly cytoplasmic staining (Extended Data Fig. 4). As such overexpression systems are artificial and cannot definitively establish real translation of an ORF in its natural context, we wished to determine whether translation of Aw112010 occurs endogenously in response to LPS. As no antibodies exist for this uncharacterized ORF, we generated an epitope tagged mouse for Aw112010. RRS analysis for Aw112010 identified the native translation termination codon and using a guide RNA targeting this locus and a donor ssDNA template, we successfully introduced a C-terminal HA-epitope tag into the Aw112010 gene in mice using CRISPR/Cas9 (Extended Data Fig. 5a–c). Generation of BMDMs from WT and Aw112010^{HA} mice revealed that the Aw112010 “lncRNA” does in fact generate a stable protein in response to LPS even 24 hr post stimulation (Fig. 2g). We next wished to characterize in more detail the HA-tagged protein observed in the Aw112010^{HA} macrophages. In order to enrich for the endogenous protein, we stimulated Aw112010^{HA} BMDMs for 6hr with LPS. Cell lysates were generated and subjected to anti-HA immunoprecipitation (Extended Data Fig. 5d). Aw112010^{HA}-purified fractions were subjected to mass spectrometry[23]. Endogenous Lys-C

protease digested peptides mapped uniquely to the predicted Aw112010 ORF encoded protein with over 50% total protein coverage (Extended Data Fig. 6a, b). Furthermore, we validated one of the peptide assignments with an isotopically labeled standard (labeled with $^{15}N(2)^{13}C(6)$-lysine at the C-terminal residue) which showed co-elution and co-fragmentation with the unlabeled endogenous peptide from the stimulated Aw112010^{HA} macrophage proteome (Fig. 2h and Extended Fig. 6c). Together, these data indicate that non-canonical ORFs can generate abundant and stable proteins that exhibit discrete subcellular localizations. Furthermore, we demonstrate that Aw112010 is a *bona fide* non-canonical ORF protein-coding gene that is translated during the innate immune response to bacterial infection.

To investigate whether the translation of the non-canonical ORF within Aw112010 was physiologically important in the immune response, we generated Aw112010^{Stop} knock-in mice. We used CRISPR/Cas9 to insert a small frameshifting stop cassette sequence into an area of high ribosomal occupancy to abrogate its protein coding potential (Extended Data Fig. 7a–d). Since bacterial infection can induce robust translation of Aw112010, we infected WT and littermate Aw112010^{Stop} mice with 1×10^3 colony forming units (CFU) of *S.* Typhimurium via oral gavage. Aw112010^{Stop} mice displayed accelerated weight loss compared to WT littermates (Fig. 3a). Indeed, disruption of the Aw112010 ORF resulted in increased fecal CFUs 24hr post infection (Fig. 3b). Likewise, mice sacrificed 4 days post infection showed increased bacterial load in the cecum of Aw112010^{stop} mice (Fig. 3c). Furthermore, Aw112010^{Stop} mice presented with higher bacterial burden and dissemination to the liver and spleen than WT littermates (Fig. 3d, e). Additionally, when Aw112010^{Stop} mice were infected with 1×10^2 CFUs of *S.* Typhimurium they displayed accelerated bacterial dissemination and became moribund with bacterial infection significantly quicker than WT animals (Fig. 3f, g). Finally, in order to investigate if Aw112010 also contributed to mucosal auto-inflammatory disorders such as models of inflammatory bowel disease (IBD), we administered 2.5% dextran sulfate sodium (DSS) to the drinking water of WT and Aw112010^{Stop} littermate mice for 5 days. Aw112010^{Stop} mice were significantly protected from colitis as measured by weight loss and colonic shortening compared to WT counterparts (Fig. 3h, i). Taken together these data demonstrate that the translation of the non-canonical ORF in Aw112010 is required for the mucosal inflammatory response.

We next wished to elucidate the mechanism by which Aw112010 mediates its anti-bacterial and pro-inflammatory function in BMDMs. WT and Aw112010^{stop} BMDMs showed no difference in their ability to phagocytose or initiate phagosome acidification of pH sensitive bacterial BioParticles (Fig. 4a). Similarly, intracellular killing/survival of *S.* Typhimurium was also comparable between WT and Aw112010^{Stop} BMDMs (Fig. 4b). Furthermore, the ability of WT and Aw112010^{Stop} to undergo the inflammatory cell death pathway, pyroptosis, was unaltered in Aw112010^{Stop} macrophages compared to WT counterparts (Fig. 4c). We next investigated if Aw112010 was essential for the production of known cytokines responsible for anti-Salmonella defense and that contributed physiologically to the intestinal inflammation and IBD. Strikingly, although WT BMDMs were able to generate a robust LPS induction of IL-12p40 and IL-6, Aw112010^{Stop} macrophages showed a significant deficiency in their production, while release of anti-inflammatory IL-10 was unaltered in Aw112010^{Stop} macrophages (Fig. 4d and Extended Data Fig. 8a–c). In order to confirm this

in vivo, WT and Aw112010^{Stop} mice were administered LPS by intraperitoneal injection and serum collected after 6 hr. Again, WT animals were able to induce a robust IL-12p40 and IL-6 response that was significantly curtailed in Aw112010^{stop} animals (Fig. 4e and Extended Data Fig. 8d). Interestingly, patients with deficiencies in the IL-12R are characterized by severe and recurrent Salmonella infections and mice deficient in IL-12p40 cytokine subunit are also susceptible to *S*. Typhimurium challenge[24,25]. Furthermore, IL-12p40 plays a critical role in IBD and a neutralizing monoclonal anti-IL12p40 has been shown to be an efficacious treatment in Crohn's patients and in experimental models of colitis[26,27]. Taken together, the introduction of a stop codon into Aw112010 causes a major defect in IL-12p40 production and the ability of mice to combat Salmonella infection and undergo mucosal inflammation. However, as premature stop codon introduction into a protein coding gene can lead to non-sense-mediated decay (NMD)[28], we investigated this phenomenon in Aw112010^{Stop} macrophages. Interestingly, premature stop codon insertion into Aw112010 does indeed trigger NMD that can be rescued with the administration of cycloheximide or a specific inhibitor of NMD (iNMD)[29] (Fig. 4f, g). Although this demonstrates the vital importance of translation for Aw112010 gene expression, it presents a problem in that we cannot distinguish a potential long non-coding RNA function from a potential protein function. To this end, we generated an expression plasmid containing the natural Aw112010 ORF transcript (WT) and an extensively mutated transcript with synonymous nucleotide substitutions in all codons except the CTG start codon (Mut). As expected these RNAs display a very different folding behavior but as predicted they generate the same protein product (Fig. 4h and Extended Data Fig. 9a–c). Importantly, we could successfully reintroduce Aw112010 protein expression in Aw112010^{Stop} BMDMs with nuclear electroporation (Fig. 4i). Strikingly, the loss of IL-12p40 expression and release of this cytokine were almost completely restored with both the WT and Mut Aw112010 expressing rescue plasmids (Fig. 4j and Extended Fig. 9d). Taken together, our data provides clear evidence that the ability of Aw112010 to drive IL-12p40 production is wholly dependent on the non-canonical ORF encoded protein and does not act as its long non-coding RNA annotation dictates.

We demonstrate that the translation of functional non-canonical ORFs is a crucial event during the innate immune response to infection and inflammation. Interestingly, a significant fraction of non-coding RNA genes not only associate with the ribosome, but also undergoes active protein translation. Although we have only shown this in murine macrophages, it is highly likely that the same will hold true for most cells and tissues in higher eukaryotes. Importantly, we have shown the translation of one of these ORFs, and the protein it encodes is functional and plays a critical role in host defense and inflammatory disease. Further work is ongoing to uncover the roles of the other identified non-canonical ORFs. Together, we propose that a re-evaluation of the human protein coding genome is required to identify cryptic non-canonical ORF protein products which may have major implications for human health and disease.

Methods

Animals

RiboTag (B6N.129 strain) and LysM-Cre (B6.129P2 strain) mice have previously been described and were obtained from Jackson Laboratories[6,7]. Crossing these mice facilitates the inclusion of a HA-epitope Tag on the ribosomal protein RPL22 in all LysM expressing cells, such as bone marrow derived macrophages (BMDMs) and colonic macrophages. Aw112010^{HA} and Aw112010^{Stop} mice were generated as previously described with CRISPR/Cas9[30] into C57/B6N embryos. To generate Aw112010^{HA} mice, a ssDNA donor oligo containing a flexible linker sequence followed by an HA epitope Tag (HA-tag) sequence flanked with homology arms was provided to facilitate homology directed repair (HDR) insertion of the sequence into the C Terminus of the Aw112010 ORF. Briefly, a guide RNA (**AGAAGGAAGAGGACTTATTT**GTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCTTTTTT) and ssDNA Ultramer (IDT) donor template (AACCTCAAGTGGAAAAAGCCACCCACTGGGTCGTTCAGGAGAGATCCAGTCTTTAAAGAAGCAAAACggtggttctggtggtggttctggtggtggttctggttacccatacgatgttccagattacgctTAGAGAgCAAAATAAGTCCTCTTCCTTCTAGATGTGCATCATCTGCTTCTTCCTTCCCTAGAAGACT) was designed targeting the exon 3 stop codon locus of Aw112010 to generate Aw112010^{HA} mice. To generate Aw112010^{Stop} knock in mice, CRISPR/Cas9 was used to introduce a dsDNA in the second exon of the transcript, in area of high ribosomal protection. A ssDNA donor oligo containing a small 14-nucleotide sequence (TAA TTA ATT AAT TA) sequence flanked with homology arms was provided to facilitate homology directed repair (HDR) insertion into the ORF. This sequence contains a stop codon in all three frames and due to its even number of nucleotides, it will additionally frameshift any protein coding sequence upon insertion. A guide RNA (**CTGCCTGATGCAACAATACC**GTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCTTTTTT) targeting exon 2 of Aw112010 was designed and co-injected into fertilized C57/B6N eggs with a ssDNA Ultramer (IDT) donor template (TCCTATTCATCTGATCTGCTTCCAGATCCCTCTGATATTTATCTTTGGTGGTGTGCTCATCATCTGCCTGATGCAAtaattaattaattaCAATACCTGGCGTATAAGTCTTCTAAGAACGTCGTTAAAGTCTTCTGCCATCAAGCCAATGATGTGAGTGCTGGG) containing a frameshifting stop insertion and 2 homology arms to allow for HDR mediated genomic integration to generate A112010^{Stop} codon knock in mice. Single heterozygous founder mice were generated and backcrossed to C57/B6N mice. Experimental groups of WT and Aw112010^{Stop} knock in mice were generated by heterozygote by heterozygote breeding. All experiments were performed using littermate control, cohoused mice. All animal experimentation was performed in compliance with Yale Institutional Animal Care and Use Committee protocols. No formal blinding or randomization was conducted, however, control and treated groups were chosen arbitrarily for each experiment. Samples sizes were chosen in line with pervious experimental experience and consistent with the broader literature. Animal weights and CFUs were measured in a blinded manner.

RiboTag^{LysM} Macrophage RNA isolation and Processing

Bone marrow derived macrophages (BMDMs) were generated from progenitor cells isolated from the femurs and tibias of RiboTag^{LysM} mice and maintained in macrophage-colony stimulating factor (50 ng/ml) for 7 days. Cell were stimulated with LPS (Serotype O111:B4) at the indicated concentrations and times. BMDMs were infected in antibiotic free media with 1 multiplicity of infection (MOI) of *S.* Typhimurium for 1 hr. Cells were then treated with gentamycin (100 μg/ml) to kill extracellular bacteria and incubated for a further 5 hr prior to harvest. Cells were washed in ice cold PBS twice, and then RiboTag Lysis buffer (containing cycloheximide, heparin and the RNAse inhibitor SuperaseIN) was added directly to the cells on ice as previously described[6,31–33]. Cell lysates were passed through a 26g needle 10 times and incubated for 30 min on ice to ensure complete lysis. For intestinal macrophage ribosome isolation, control and *S.* Typhimurium infected mice were fasted for 4 hours and administered Streptomycin (20 mg per mouse) by oral gavage. 20 hours later mice were again fasted for 4 hours and gavaged with (2×10^8 colony forming units (CFUs) of *S* Typhimurium. 24 hours later, mice were euthanized and the colons removed. After washing and flushing with PBS, the colon was separated in 5 equal sized samples and placed into 1 ml of RiboTag lysis buffer each. Tissue was lysed in a 3 step manner. Firstly, tissue was mechanical disrupted with an electronic tissue homogenizer for 30 sec. Then homogenized tissue was further processed in a Dounce homogenizer with 10 strokes. Finally, colon lysates were passed through a 26g needle 10 times and incubated on ice for 30 min. Ribosome-RNA containing supernatants were clarified by centrifugation at 12,000 g for 10 min at 4˚C. HA-conjugated magnetic beads (Pierce) were added to samples and incubated overnight under gentle inversion at 4˚C. Beads were washed 3 times for 10 min with gentle rotation in high salt buffer containing cycloheximide. RNA was eluted from HA-beads using Qiagen RLT buffer containing 2-Mercaptoethanol and anti-foaming DX reagent (Qiagen) by 30 sec vortex pulsing. RNA was isolated using a RNAeasy micro kit. RNA was sent for sequencing or converted to cDNA using Maxima Reverse transcriptase kit (Thermo). qPCR using Sigma KiCqStart predesigned SYBR green primers was conducted. mRNA for RNAseq analysis was purified using PolyA$^+$ selection and processed by the Yale Centre for Genome analysis using standard methodology. RNA was sequenced on a HIseq2000 with 75bp pair ended reads.

RiboTag RNAseq Analysis

Fastq files from RNA-seq and RiboTag RNAseq were aligned to the Ensemble GRCm38.p5 genome using Tophat2 version 2.1.1, and using a gene annotation file which combined all RefSeq, UCSC, Ensembl, Gencode, mirBase annotations for full genome coverage. Remaining missing annotations were added manually from MGI. Cufflinks version 2.2.1 was used for differential analysis. A cutoff of fpkm ≥ 0.1 was used in the RNA-seq to define a ‘detectable’ lncRNA, and a cutoff of fpkm ≥ 1 in at least one of the three ribotag-RNAseq treatments was defined as an ‘expressed’ lncRNA. featureCounts of the Subread package was used to determine reads per feature in the genome[34], which was then used with DEseq2 and ggplot2 (http://ggplot2.tidyverse.org/index.html) to produce volcano plots[35]. RCircos was used to make a circos plot showing the RiboTag RNAseq data[36], using a differential cutoff of 0.01. Only reads which aligned to a single genome locus were used.

Macrophage Ribosome Profiling and Analysis

BMDMs were generated and plated at 1×10^7 cells per treatment group. Non-treated and LPS stimulated macrophages were treated with cycloheximide (50 μg/ml) for 2 min at 37 °C. Cells were then washed with ice cold PBS containing cycloheximide (50 μg/ml). Ribosome Profiling was conducted using the illumina TruSeq Ribo Profile (mammalian) Kit as per manufacturer's instructions. Prepared libraries were sequencing with a HiSeq 2000 with 75bp single end reads. In tandem, total RNA from paired samples were subjected to Ribosomal RNA removal and convention RNAseq. Reads from ribosome profiling experiments had their adapters trimmed using the FASTX-Toolkit version 0.0.14 (http://hannonlab.cshl.edu/fastx_toolkit/index.html). Reads were aligned to Ensemble GRCm38.p5 genome using Tophat2. The accepted hits were filtered using Samtools version 1.5 (http://samtools.sourceforge.net/) along the criteria of having an alignment score of 0 to −2, and the number of possible alignments in the genome being less than or equal to 2 in order to reduce false alignments to repeat regions. Cufflinks was then used for differential expression. DESeq2 and ggplot2 were used for making volcano plots. Sushi and ggplot2 were used to build RNA-seq and Ribosome Profiling coverage plots for certain features[37]. A pipeline to determine the Translation Efficiency (TE) and Ribosome Release Score (RRS) included finding the number of reads covering each feature using featureCounts. Protein coding gene annotations and their untranslated regions (UTRs) were extracted from our larger GTF annotation file. mRNA sequences from lncRNA exons were extracted using Bedtools version 2.26.0[38]. In order to identify non-canonical ORFs in lncRNAs, we first catalogued all transcripts that harbored high quality and unique ribosome profiling reads mapping to exons in the annotated lncRNA family. We then conducted a custom ORFfinder search, in which we relaxed ORF identification parameters pertaining to ORF size and start codon utilization. NCBI ORFfinder and the Sequence Manipulation Suite ORF Finder tools were used to find all ORFs more than 30 nt long using start codons ATG, CTG, TTG, and GTG[20]. Found ORFs were searched using BLASTX+ version 2.6.0 against lncRNA sequences that included their introns to find the final position of ORFS within the lncRNAs. UTRs were defined as 50 nucleotide bases directly following the CDS. TE and RRS were calculated according as previously described[3]. TE was the number of ribosomal reads covering the exons divided by the number of RNA reads in the exon, normalized for length. Briefly, RRS interrogates the known phenomenon of ribosome release from a transcript directly after recognition of the ORFs cognate stop codon[39] by enumerating the number of ribosome profiling reads before and after the in frame stop codon of a given mRNA. RRS is the result of ribosomal reads covering the exon divided by RNA reads covering the exon, divided by the result of ribosome reads covering the 3′ UTR divided by RNA reads covering the 3′ UTR. RRS ≥ 7 and TE ≥ 0.0001 values are considered that of the annotated protein coding genome. Exons had to have at least 1 ribosomal read for these metrics to be calculated. A fpkm cutoff of ≥ 4 in either treatment was used when identifying top significant differentially regulated features. Significance of $p<0.05$ was used. RibORF version 0.1 was used identify transcripts undergoing active translation measured by the correct alignment of the ribosome A-site, 3-nt periodicity of translating ribosomes and coverage uniformity across all codons of protein coding genes. RibORF was then used to assign percentage of maximum entropy (pme) scores to the OFRs found with ORFfinder. A PME above 0.6 was used to predict positive coding potential. RiboCode version 1.2.10 was used to identify

ORFs within the ribosome profiling data using start codons ATG, CTG, TTG, and GTG. A combined analysis p value of <0.05 was used to determine protein coding potential. The nucleotide sequences of all predicted coding ORFs found were searched for homology in the human CHESS 2.0 genome (http://ccb.jhu.edu/chess/) using BLASTX+ version 2.6.0. PhyloCSF was used on the top hit of ORFs for which homology was found to further assess coding potential[40]. All lncRNA RNA sequencing and analysis is provided in Supplementary Table 1.

Expression of Non-Canonical ORFs

Geneblocks corresponding to the open readings identified by ribosomal profiling were synthesized by IDT and inserted into pCMV6-Entry Tagged Cloning Vector (PS100001) using MLU I and Sgf I restriction digestion. Plasmids were verified by sanger sequencing and propagated by transformation in Top10 competent cells. HEK293 cells were transfected using Transit LT1 liposomal transfection reagent (Mirus) as per manufacture instructions. For confocal microscopy studies, HEK293 cells were transfected with 0.5 μg of plasmid DNA and incubated at 37°C for 24 hr. Cells were fixed in methanol. Anti-FLAG mAB (M2 clone), Phalloidin 647 (Santa Cruz) and DAPI (Sigma). Confocal imaging was conducted with a Nikon-Ti microscope combined with UltraVox spinning disk (PerkinElmer) and data was analyzed using the Volocity software (PerkinElmer).

Aw112010^{HA} Immunoblotting and Immunoprecipitation

WT and Aw112010^{HA} BMDMs were generated and treated with LPS (10 ng/ml) for the indicated times. Protein lysates and HA-immunoprecipitation fractions were generated using the Pierce™ HA-Tag Magnetic IP/Co-IP Kit as per manufactures instructions. Protein lysates and HA-IP samples were resolved on NuPAGE™ 4–12% Bis-Tris Protein Gels using MES running buffer (Invitrogen) and transferred to nitrocellulose membrane. Detection of endogenous Aw112010 was conducted using a monoclonal anti-HA antibody (HA1.1 clone) and β-tubulin (E7 clone) was used as a loading control.

Proteomics Methods

1×10^7 BMDMs stimulated with LPS (10 ng/ml) for 6 hrs were lysed and subjected to HA-immunoprecipitation as described above. An aliquot of HA-purified protein lysate was boiled for 15 min in 1% SDS followed by chloroform/methanol precipitation[41]. Reduction, alkylation, and Lys-C protease digestion were performed according to standard protocols[42],[43]. All resulting peptides were purified and desalted using a SepPak Classic SPE cartridge (Waters, Milford, MA, USA) according to the manufacturer's instructions and dried in a rotary vacuum centrifuge. Samples were resuspended in 0.1% trifluoroacetic acid and diluted to approximately 0.5 μg/μl. Peptide standards (JPT Peptide Technologies) with 13C(6)15N(4) R isotopic labels were added to a final concentration of 100 ng/mL[44]. A total volume of 5 μl of sample was injected onto an analytical column (75 μm × 50 cm PicoFrit column packed with 1.9 μm ReproSil-Pur 120Å C18-AQ resin) using ACQUITY UPLC M-Class (Waters) and a Q Exactive Plus (Thermo). Separation was performed on a 330 min non-linear gradient from 1% mobile phase B to 99% mobile phase B (mobile phase A: 1% ACN 0.1% FA in water, mobile phase B: 80% ACN 0.1% FA in water); MS: 70,000 resolution, 3e6 AGC target, 300–1700 m z−1 scan range; dd-MS2: top10 method, 17,500

resolution, 1e6 AGC target, 10 loop count, 1.6 m z−1 isolation window, 27 NCE). In all experiments, a full mass spectrum was followed by ten parallel reaction monitoring (PRM) scans at 17,500 resolution 2e5 AGC target, 4 m z−1 isolation window 100 ms maximum injection time) as triggered by an inclusion list. ProteoWizard MS Convert was used for peak picking and files were analyzed using Mascot (version 2.5.1). RNA-seq data from LPS and non-treated mouse macrophages were aligned to Ensemble GRCm38.p5 genome and translated in three reading frames using CLC SequenceViewer (Qiagen) and the resulting databases as well as a contaminant database were used for proteomics searches. Carbamidomethyl (C) was set as a fixed modification, and 13C(6)15N(4) R, oxidation (M), and acetyl (N-term) as variable modifications. The false discovery rate was set to 1%.

S. Typhimurium Infection

Prior to infection, 8–10 week-old mice were restricted from food and water for 4 hr followed by gavage of Streptomycin (20 mg). 20 hr later mice were fasted again for 4 hr and infected with streptomycin resistant *Salmonella enterica* subsp. *enterica* serovar Typhimurium (SL1344 strain, kindly provided by Dr. Jorge Galan, Yale University). *S.* Typhimurium was maintained as a glycerol stock at −80°C. Prior to infection, bacteria were propagated overnight in LB containing streptomycin (100 μg/ml). Bacteria was subcultured for 4 hr the following day in antibiotic free LB containing 0.3M NaCl to return it to log phase growth and increase virulence[45]. Using spectrophotometry, bacterial CFU was calculated with an infection dose ranging from 1×10^2 to 1×10^3 CFU per animal. To calculate fecal CFU, fecal pellets were resuspended in PBS at 50 mg/ml and vortexed for 20 min. Bacteria containing supernatants were clarified by centrifugation at 50 g for 10 min. Serial dilutions were conducted, and bacteria plated in triplicate on LB streptomycin (100 μg/ml) plates. For cecal, liver and spleen CFU enumeration, organs were isolated and weighted, and added to 2ml of PBS. Tissue was dissociated with GentleMacs C Tubes (Miltenyi Biotech) as per manufacturer's instructions. CFU counts were calculated using similar methodology as above. All CFU counts were preformed blinded. Mice which lost 30% body weight, or which were unresponsive, were euthanized.

Confocal Microscopy for Splenic Salmonella Dissemination

WT and Aw112010^{Stop} codon mice were infected with 1×10^2 CFUs as described above. Mice were euthanized 3 days post infection. For in-situ immunofluorescence, spleens were dissected, fixed in 4% paraformaldehyde for 1 hr at 4 °C, followed by incubation in 10% Sucrose/PBS for 1 hr, 20% Sucrose/PBS for 1hr and 30% Sucrose/PBS overnight at 4 °C, all under gentle agitation. Tissue samples were frozen in OCT on dry ice and kept in −80 °C until sectioning. Sections 8–10 μm thick were prepared with a cryostat (Leica). Spleen sections were permeabilized with Perm/Wash buffer (BD Biosciences) for 10 min and blocked with Protein Block (Dako) for 7 min. Primary antibodies included anti-F4/80 (BM8), anti-B220 (RA3–6B2) and anti-Salmonella (Abcam ab35156). Primary and secondary antibodies were incubated in Perm/Wash buffer for 1 h. After washing with Perm/Wash buffer, sections were mounted with ProLong Gold with DAPI (Invitrogen), covered and sealed with nail polish. Confocal imaging was conducted with a Nikon-Ti microscope combined with UltraVox spinning disk (PerkinElmer) and data was analyzed using the Volocity software (PerkinElmer).

Dextran Sulfate Sodium (DSS) Colitis Induction

DSS colitis was conducted as previously described[30]. Briefly, 10–12 week-old mice were administered 2.5% DSS (MP Bio) for 5 days and returned to regular drinking water and monitored daily. Weight loss was measured every day. On day 12, mice were euthanized and colons were extracted. Colonic shortening was used as a metric of colitis severity.

Phagocytosis Assay

WT and Aw112010^{Stop} BMDMs were generated and plated on non-tissue culture treated non-adherent plates and pretreated with the phagocytosis inhibitor Cytochalasin D (10 μM) for 1 hr, LPS (10 ng/ml) for 6 hr or left non-treated. Cells were then administered with 1 mg/ml pHrodo Red *E. coli* BioParticles (Invitrogen) conjugates for 1 hr. Cells were isolated, washed and stained with anti-CD11b (M1/70). Cd11b positive cells were assessed for pHrodo positivity which indicates cells which have phagocytosed the pH sensitive bio particles and initiated phagosome acidification using a LSRII flow cytometer.

Intracellular Salmonella Survival Assay

WT and Aw112010^{Stop} BMDM generation and S. Typhimurium bacterial culturing was conducted as previously described. On day 7, the BMDM culture media was replaced with antibiotic free media. 1×10^6 BMDMs were infected in triplicate with 1×10^7 CFUs of *S.* Typhimurium (MOI 10) by spinfection at 800 × g at 37°C for 10 min and returned to the incubator for a further 20 min. Media was then removed and replaced with media containing Gentamycin (100 μg/ml) to kill extracellular bacteria for 1 hr. Media was replaced with fresh media including Gentamycin (25 μg/ml) for 4.5 hr. Cells were lysed in a 1% Triton and 0.1% SDS buffer for 5 min under gentle agitation. Cell lysates were plated on Streptomycin (100 μg/ml) containing LB plates and CFUs enumerated.

BMDM Cell Death Assay

WT and Aw112010^{Stop} BMDMS were generated and seeded at 50,000 cells per 96 well plate in triplicate. Cells were stimulated with LPS (100 ng/ml) for 5 hr. Media was replaced with fresh antibiotic free media containing LPS (100 ng/ml) and *S.* Typhimurium (MOI 100). Cells were centrifuged at 800 × g at 37°C for 10 min. Cells were returned to the incubator for a further 50 min. Supernatants were collected and clarified. The Pierce™ LDH Cytotoxicity Assay Kit was used to measure cell death, as per manufacturer's instructions.

In vitro BMDM LPS induced Cytokine Measurement

WT and Aw112010^{Stop} BMDMs were generated and stimulated with LPS (10 ng/ml) for the indicated times. Supernatants were collected and clarified by centrifugation (12,000 g, 10 min at 4°C) and analyzed for IL-12p40, IL-6 and IL-10 by ELISA (R&D Duosets). Similarly, BMDMs were generated and stimulated as above and were lysed in Trizol and RNA extracted as per manufacturer's instructions. RNA was equalized to 700 ng and converted to cDNA using Maxima Reverse transcriptase kit (Thermo). qPCR was conducted using Sigma KiCqStart predesigned SYBR green primers as indicated.

In vivo LPS induced Cytokine Measurement

8–10 week old WT and Aw112010^{Stop} mice were weighed and administered 100 μl of either PBS or LPS (Serotype O111:B4, Enzo LifeSciences) (10 mg/kg) via intraperitoneal injection. After 6 hr, mice were euthanized and serum collected and analyzed for IL-6 and IL-12p40 cytokines by ELISA (R&D Duosets).

Non-Sense Mediated Decay Studies

WT and Aw112010^{Stop} BMDMs were generated and stimulated with cycloheximide (50 μg/ml) or a non-sense mediated decay inhibitor (Calbiochem) (iNMD) (50 μM) for 6 hr. RNA was extracted and qPCR conducted for *Aw112010* mRNA expression.

RNA and Protein Structural Prediction tools

Two Aw112010 rescue expression plasmids were generated. The WT Aw112010 mRNA sequence was cloned into the pCMV6 expression vector:

CTGAGCTGCAAGATGTCTCCCATCCCTCTGATATTTATCTTTGGTGGTGTGCTCATC
ATCTGCCTGATGCAACAATACCTGGCGTATAAGTCTTCTAAGAACGTCGTTAAAGT
CTTCTGCCATCAAGCCAATGATGTACATATATACCAGACCCAGGTCGTCATGACAA
ACACACTGGAAACCTCAAGTGGAAAAAGCCACCCACTGGGTCGTTCAGGAGAGA
TCCAGTCTTTAAAGAAGCAAAAC.

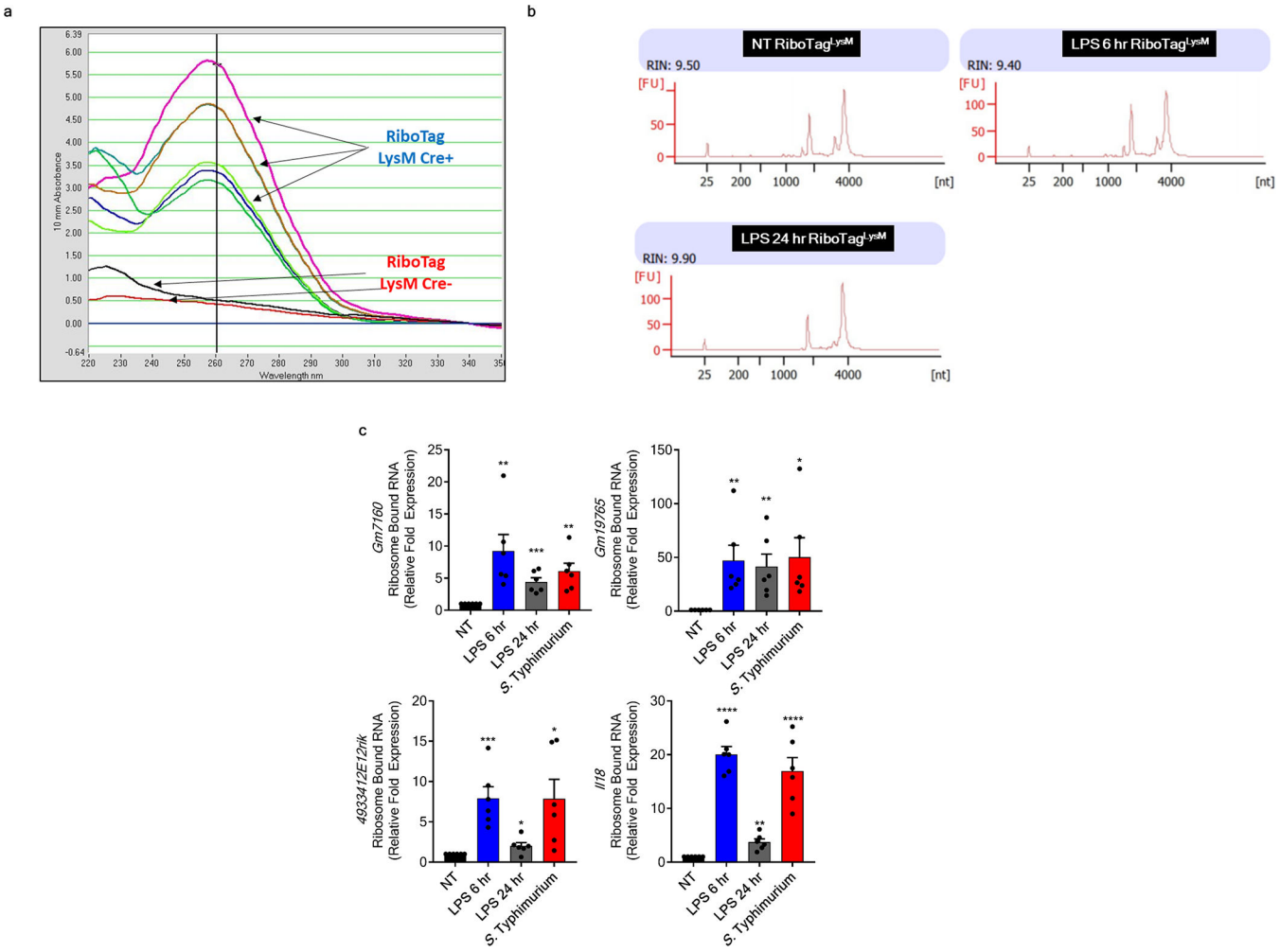
The Mut Aw112010 mRNA sequence contains extensive synonymous mutation:
CTGTCGTGTAAAATGTCACCTATTCCCCTAATCTTCATTTTCGGCGGGGTCCTTATT
ATTTGTTTAATGCAGCAGTATCTAGCATACAAATCATCGAAAAATGTAGTAAAGGT
ATTTTGTCACCAGGCGAACGACGTTCACATTTATCAAACTCAAGTGGTAATGACGA
ATACTCTCGAGACAAGCAGCGGGAAGTCACATCCGTTAGGCCGGTCCGGCGAAAT
ACAATCGTTGAAAAAACAGAAT.

The Vienna RNA package was used to calculate minimum free energy structures of each RNA[46]. Both mRNA sequences encode the same protein product: LSCKMSPIPLIFIFGGVLIICLMQQYLAYKSSKNVVKVFCHQANDVHIYQTQVVM TNTLETSSGKSHPLGRSGEIQSLKKQN. In order to predict the structure of this protein we used the Quark software package[47].

Aw112010 Rescue Experiments

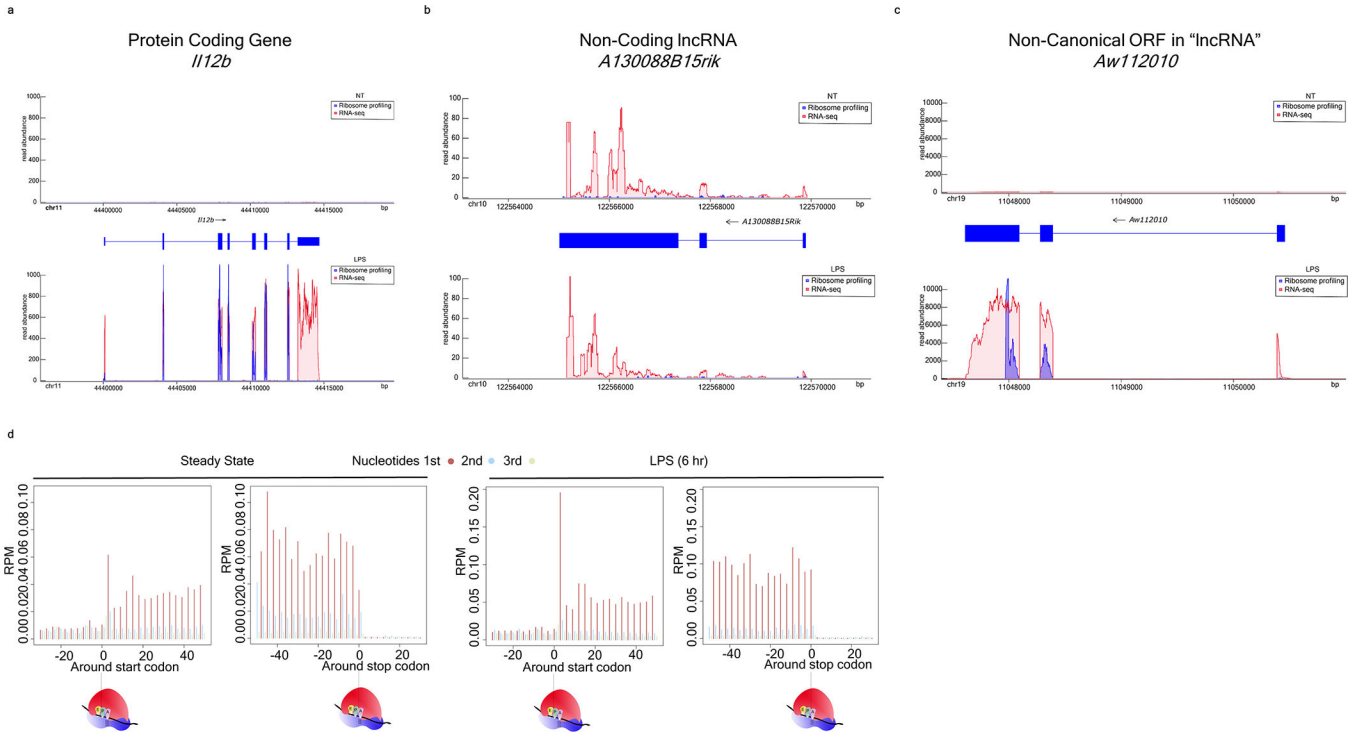
WT and Aw112010^{Stop} BMDMs were generated and cultured on non-tissue cultured treated non-adherent plates for 7 days. 1 ×10^6 BMDMs were electroporated with 1 μg of rescue expression vector as indicated using the Amaxa Mouse Macrophage Nucleofector Kit using a Nucleofector 2b Device as per manufactures instructions (Lonza). Cells were rested for 6 hr and stimulated as indicated. Expression of Aw112010 was confirmed by anti-FLAG tag immunoblotting. IL-12p40 protein and *Il12b* mRNA gene expression was measured as previously described.

Extended Data



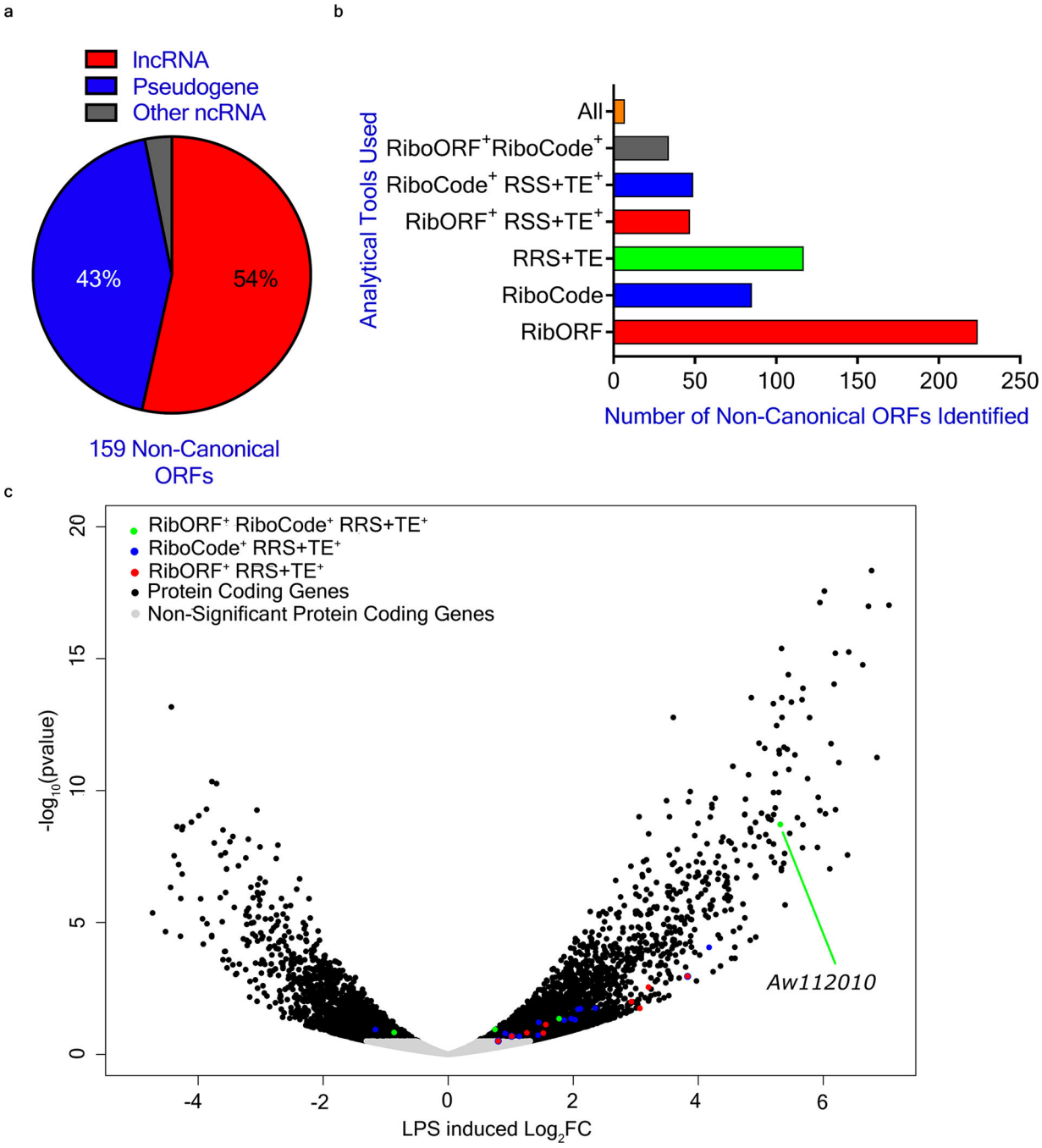
Extended Data Figure 1: RiboTag RNA Isolation and mRNA Expression

a) Nanodrop analysis of Ribosome associated RNA isolated from RiboTag and RiboTag^{Lysm} mice showing no detected RNA isolated from WT BMDMs**. b)** Bioanalyzer traces and RNA Integrity Number (RIN) of ribosome associated RNA isolated from BMDMs from RiboTag^{LysM} mice non-treated (NT) or stimulated with LPS for 6 and 24 hours. **c)** qPCR analysis of ribosome associated transcripts of non-treated (NT) BMDMs, stimulated with LPS (10 ng/ml) or infected with *S.* Typhimurium at an MOI of 1 for 6 hr. Data is presented as SEM from 6 biological replicates.



Extended Data Figure 2: Ribosome Profiling, RNAseq Read Tracing and RibORF Analysis

a-d) WT BMDMs were non-treated (NT) or stimulated with LPS (10 ng/ml) for 6 hr and ribosome profiling conducted. Data is representative of 2 biological replicates **a)** Pattern of RNAseq transcriptional reads (red) and Riboprofiling translational reads (blue) for *IL12b* from NT (upper trace) and LPS stimulated (lower trace) BMDMs. The gene structure of *Il12b* is located in the center, with a very thin blue line representing the introns and wide blue rectangles indicating exonic structure. Thinner exonic structures represent annotated 5′ and 3' UTRs**. b)** Pattern of RNAseq transcriptional reads (red) and Riboprofiling translational reads (blue) for a Non-RiboTag identified lncRNA, *A130088B15rik*, from NT (upper trace) and LPS stimulated (lower trace) BMDMs. The gene structure of *A130088B15rik* is located in the center, with a thin blue line representing the introns and wide blue rectangles indicating exonic structure. **c)** Pattern of RNAseq transcriptional reads (red) and Riboprofiling translational reads (blue) for a RiboTag identified lncRNA, *Aw112010*, from NT (upper trace) and LPS stimulated (lower trace) BMDMs. The gene structure of Aw112010 is located in the center, with a thin blue line representing the introns and wide blue rectangles indicating exonic structure. **d)** RibORF analysis of read distribution (reads/million mappable reads; RPM) around start and stop codons of known, annotated protein coding genes in steady state and LPS stimulated samples.

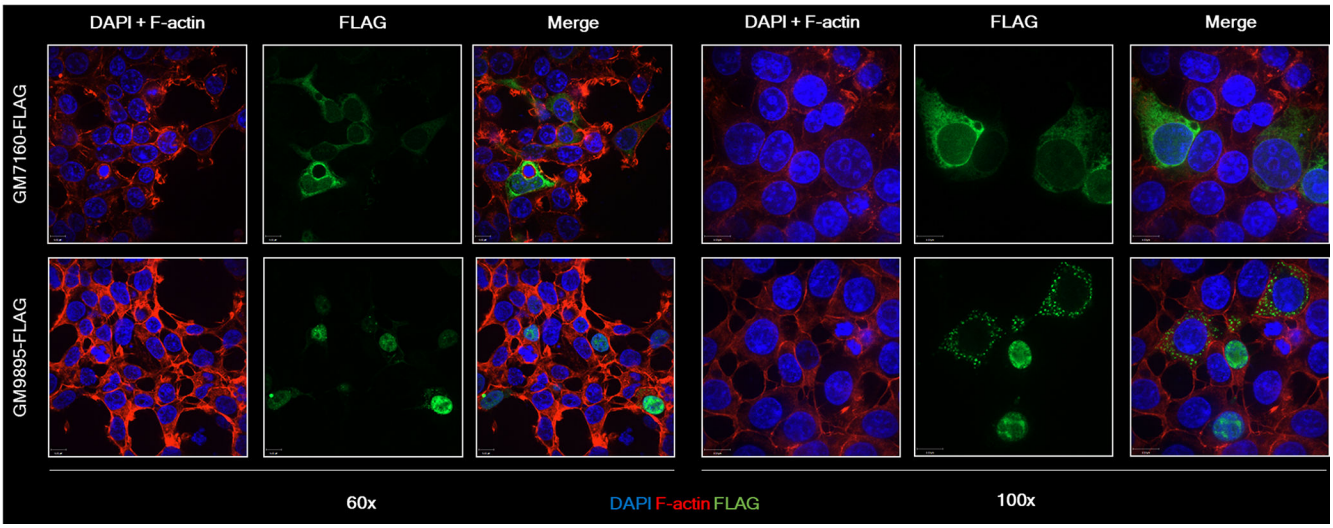


Extended Data Figure 3: Breakdown of Different Analytical approaches to Predict Protein Coding lncRNAs

a) RiboCode analysis of ribosome profiling data identifies 85 ORFs within lncRNAs with protein coding potential **b**) Comparison of non-canonical ORFs identified by RibORF, Ribosome Release Score and Translation Efficiency (RRS+TE+) and RiboCode analytical strategies from BMDM expressing lncRNA using ribosome profiling. **c)** WT BMDMs were non-treated or stimulated with LPS (10 ng/ml) for 6 hr and ribosome profiling conducted.

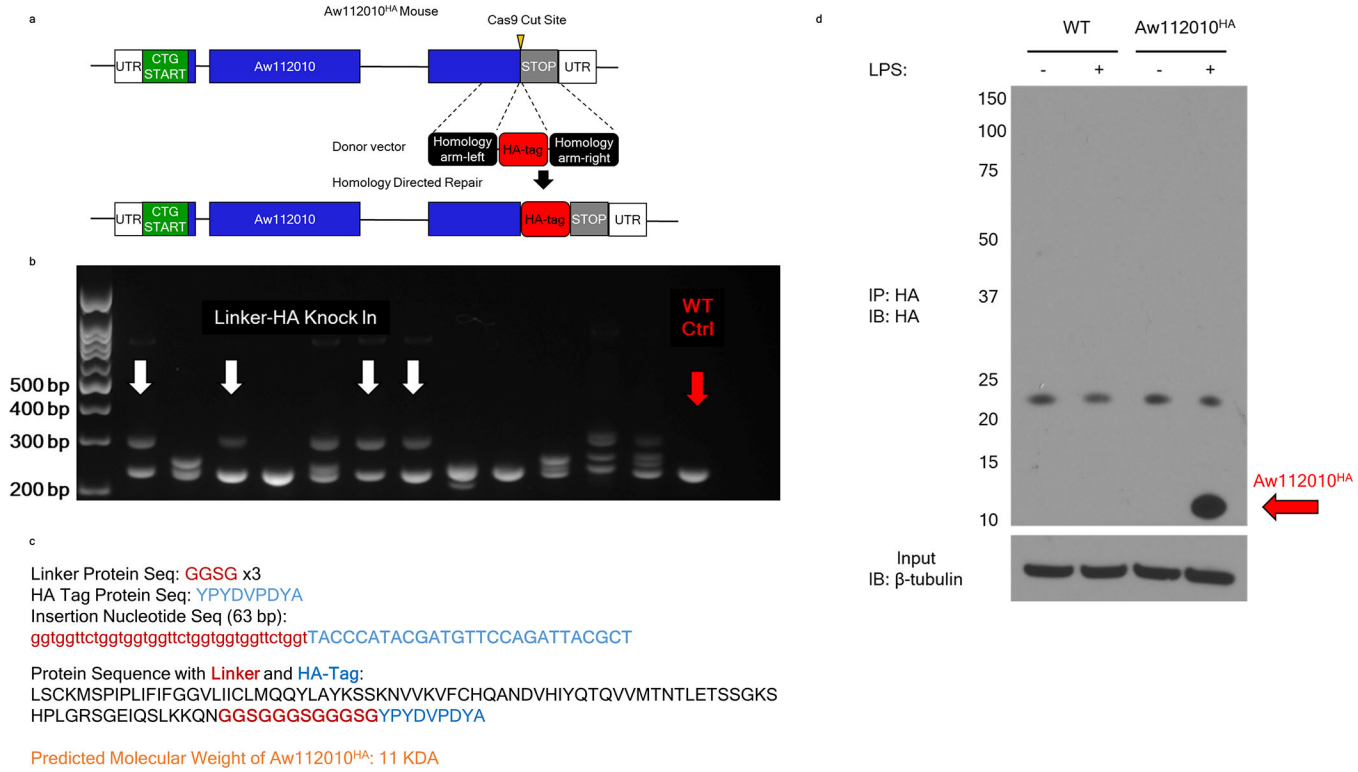
Data is representative of 2 biological replicates. Volcano plot of LPS induced differentially regulated genes identified by RibORF, RiboCode and RRS+TE analysis.

a



Extended Data Figure 4: Overexpression of Non-Canonical ORFs Reveals Distinct Subcellular Localization

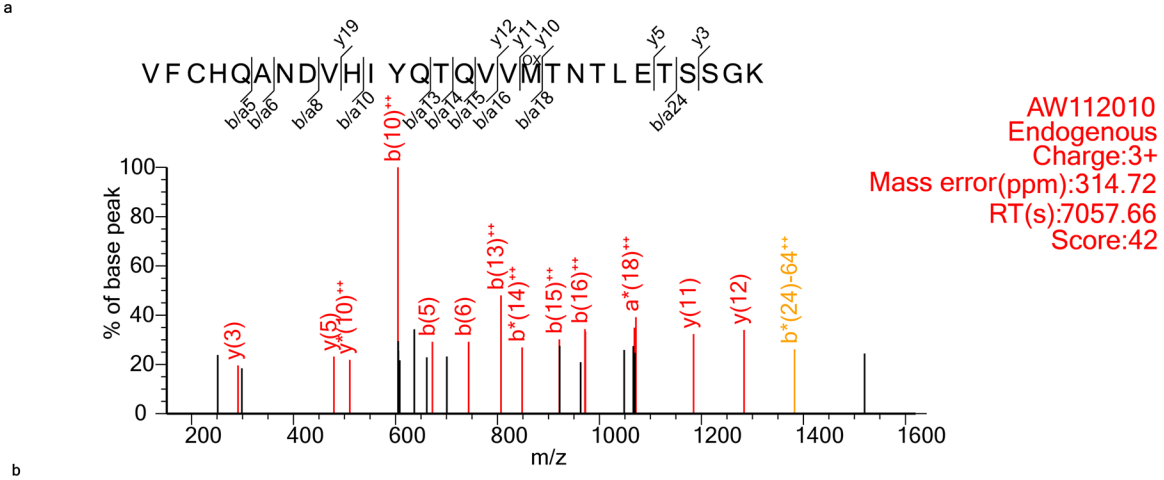
a) HEK293 cells were transfected with 500 ng of flag-tagged plasmids encoding the non-canonical ORFs for GM7160 and GM9895. Cells were fixed and stained with DAPI (blue, nucleus), Phalloidin (red, cytoskeletal F-actin) and anti-flag (green, ORF of interest). Confocal microscopy was conducted at 60x and 100x objectives, as indicated. Data are representative of at least 3 independent experiments.



Extended Data Figure 5: Aw112010^{HA} Mouse Characterization

a) Schematic representation of Aw112010^{HA} knock in mice. **b)** Genotyping for Aw112010^{HA} mice from CRISPR/Cas9 injections **c)** Sequence information for GGSG(x3)-HA Tag insertion used to generate Aw112010^{HA} mice. **d)** WT and AW112010^{HA} BMDMs were left untreated or stimulated with LPS (10 ng/ml) for 6hrs. Protein lysates were generated and were incubated overnight with anti-HA magnetic beads. Purified lysates were probed for HA by western blot. Whole cell lysates were used as a loading control and probed for β-tubulin. Data is representative of 3 independent experiments.

a



b

Predicted Sequence 51.3% coverage:

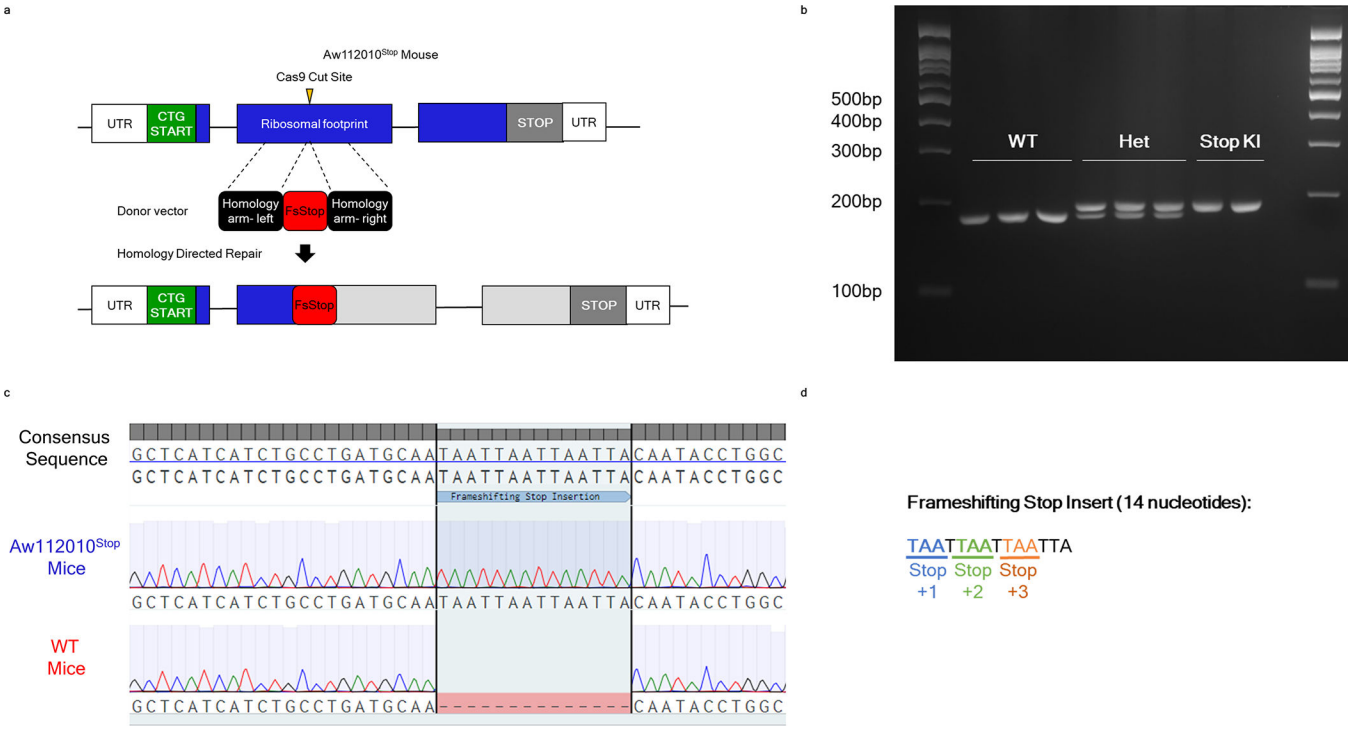
LSCKMSPIPLIFIFGGVLIICLMQQYLAYKSSKNVVKVFCHQANDVHIYQTQVVMTNTLETSSGKSHPLGRSGEIQSLKKQN

c

AW112010 Labeled Standard	AW112010 Endogenous
Charge:2+	Charge:2+
Mass error (ppm):3.30	Mass error (ppm):3.45
RT(s):3797.43	RT(s):3797.57
Score:71	Score:46

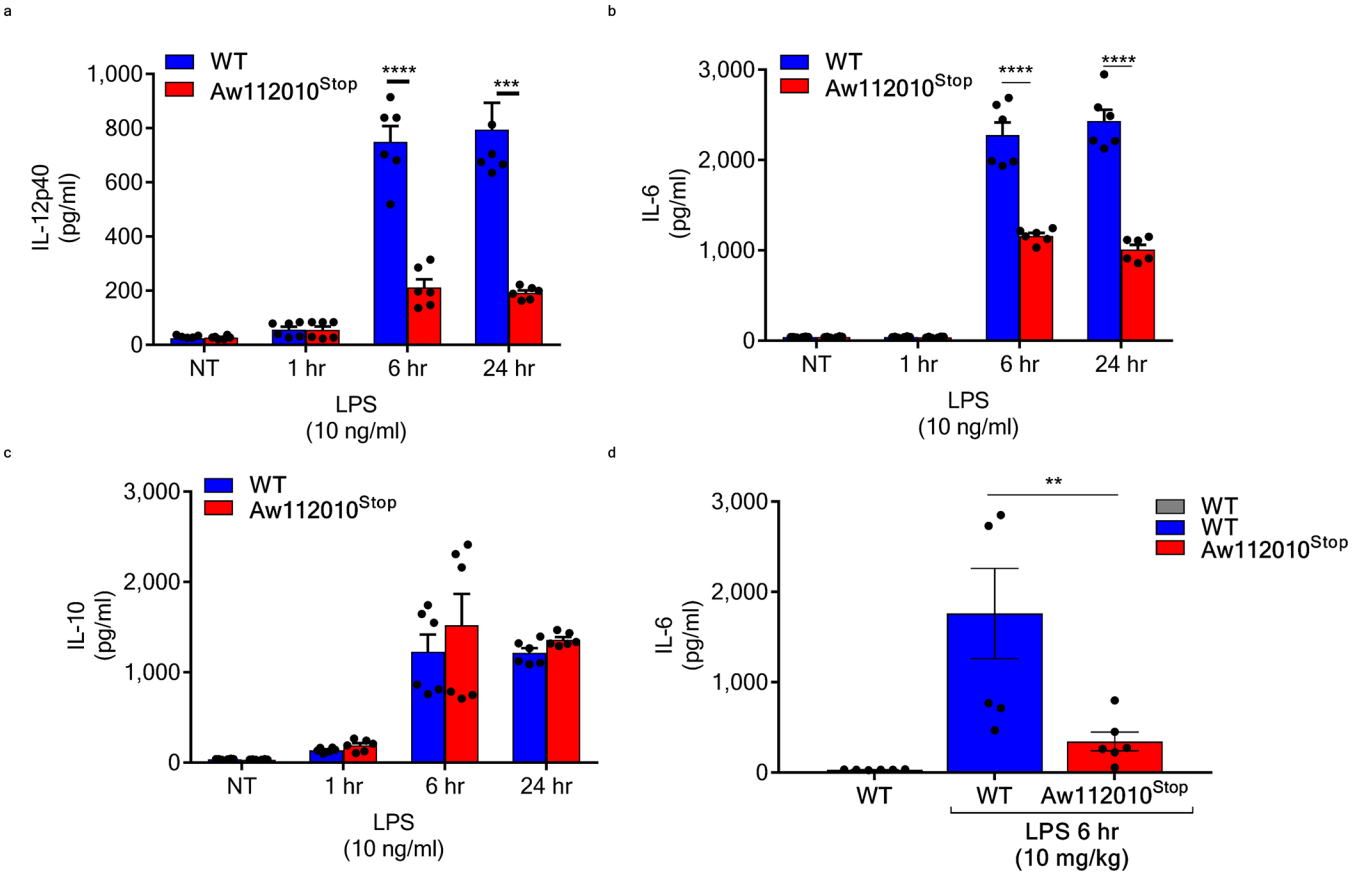
Extended Data Figure 6: Mass Spec Validation of the Aw112010 ORF

a) MS/MS fragmentation of an endogenous peptide from Aw112010 found in LPS stimulated macrophages after HA IP purification. Identified fragment ions (b and y-ions, red) are indicated above and below the peptide sequence. **b)** Aw112010 predicted protein sequence. Peptides detected by mass spec are highlighted in red and blue. **c)** Mass Spec information for the identified fragments displayed in Fig 2h.



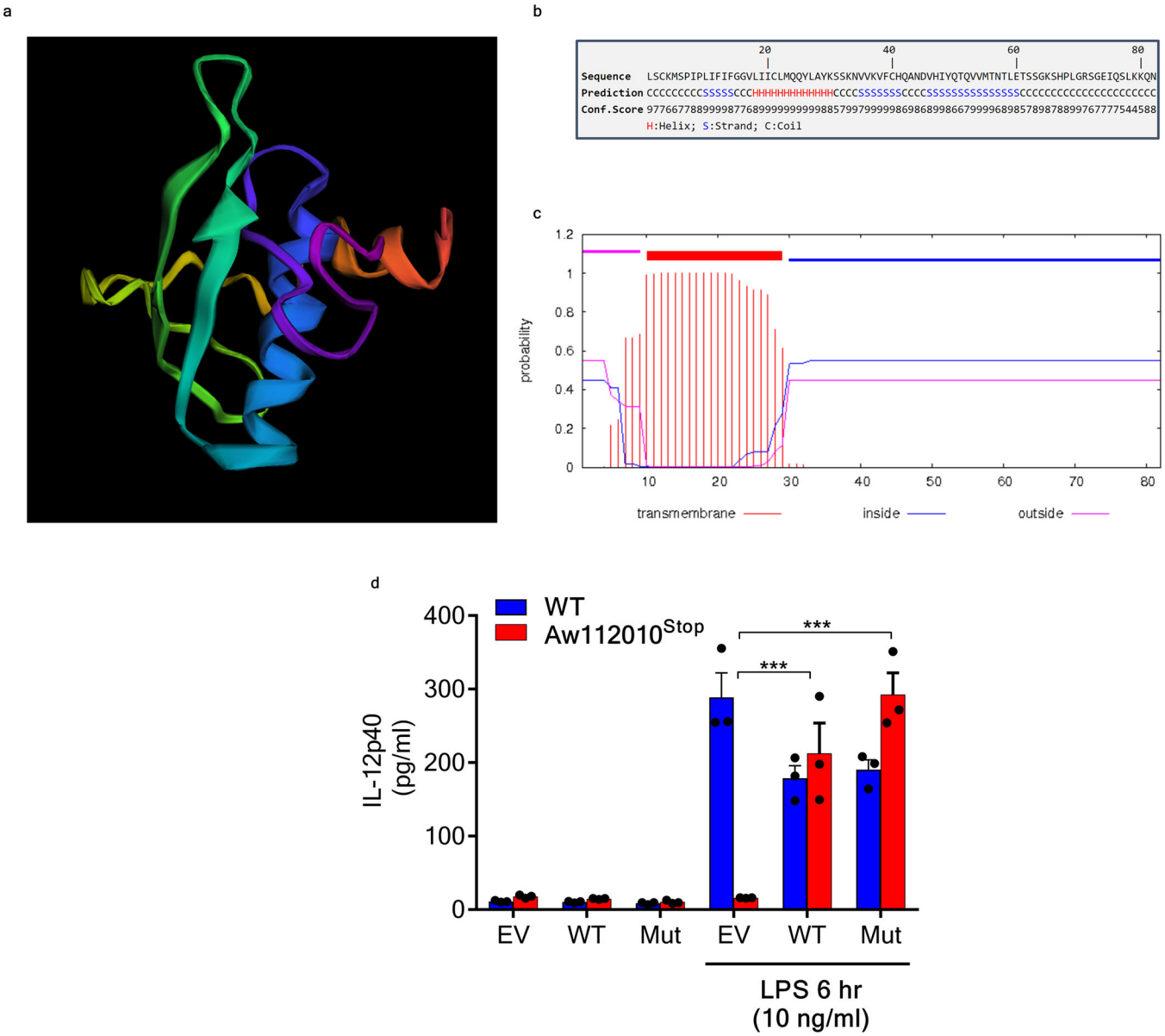
Extended Data Figure 7: Characterization of Aw112010^{Stop} Mice

a) Schematic representation of Aw112010^{HA} knock in mice. **b)** Genotyping for Aw112010^{Stop} mice generated using CRISPR/Cas9. **c)** Sanger sequencing of the frameshifting Stop Codon insertion in Aw112010^{Stop} Mice and WT controls. **d)** Sequence of frameshifting stop insertion. Stop codons and frame positions are indicated below the sequence.



Extended Data Figure 8: Cytokine Production in WT and Aw112010^{Stop} macrophages and mice

a-c) WT and Aw112010^{Stop} BMDMs were stimulated with LPS for indicated times and supernatants were analyzed for IL-12p40, IL-6 and IL-10 by ELISA. Data is from 6 biological replicates conducted over 2 independent experiments. **d)** Mice were administered PBS (n=5) or LPS (n=6, WT and Aw112010^{Stop}) (10 mg/kg) for 6 hr via intraperitoneal injection. Serum was analyzed for IL-6 by ELISA. SEM of replicates is presented for ELISA and qPCR data and significance determined by unpaired two-tailed t-tests, (*p<0.05; **p<0.01; ***p<0.001; ****p<0.0001).



Extended Data Figure 9: The Aw112010 Protein is required for IL-12-40 Production

a-c) Predicted protein structure of the Aw112010 non-canonical ORF encoded protein generated using Quark Package. Aw112010 is predicted to contain a single transmembrane domain. **d)** BMDMs were subjected to electroporation with indicated plasmids. BMDMs were stimulated with LPS (10 ng/ml) for 6 hr. Supernatants were analyzed for IL-12p40 production by ELISA. Data is of 3 biological replicates. Statistical significance was determined by unpaired two-tailed t tests. Significance indicated by ***p<0.001.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We would like to thank Jon Alderman, Caroline Lieber, Cynthia Hughes, Linda Evangelisti, Elisabeth Hughes-Picard, Erica Ryke, Lauren Machado, Christopher Castaldi for help in facilitating this work. We would like to thank Dr. Jorge Galan and Dr. Hui Sun for providing the *S.* Typhimurium and for extremely helpful discussion on the infection model. We would like to offer a special thank you to Drs. Roni Nowarski, Marc Healy and Noah Palm for insightful comments and discussion on the manuscript. This work was supported by the Howard Hughes Medical Institute and the Blavatnik Family Foundation (R.A.F.). This work was supported in part by the Searle Scholars Program, the Leukemia Research Foundation, an American Cancer Society Institutional Research Grant Individual Award for New Investigators (IRG-58-012-57), the NIH (R01GM122984), and Yale University West Campus start-up funds (to S.A.S.). A.K. was in part supported by an NIH Predoctoral Training Grant (5T32GM06754 3-12) (S.S).

References

1. Couso JP & Patraquim P Classification and function of small open reading frames. Nat Rev Mol Cell Biol 18, 575–589, doi:10.1038/nrm.2017.58 (2017). [PubMed: 28698598]
2. Kozak M Regulation of translation in eukaryotic systems. Annu Rev Cell Biol 8, 197–225, doi: 10.1146/annurev.cb.08.110192.001213 (1992). [PubMed: 1335743]
3. Guttman M, Russell P, Ingolia NT, Weissman JS & Lander ES Ribosome profiling provides evidence that large noncoding RNAs do not encode proteins. Cell 154, 240–251, doi:10.1016/j.cell. 2013.06.009 (2013). [PubMed: 23810193]
4. Ingolia NT, Lareau LF & Weissman JS Ribosome profiling of mouse embryonic stem cells reveals the complexity and dynamics of mammalian proteomes. Cell 147, 789–802, doi:10.1016/j.cell. 2011.10.002 (2011). [PubMed: 22056041]
5. Hoagland MB, Stephenson ML, Scott JF, Hecht LI & Zamecnik PC A SOLUBLE RIBONUCLEIC ACID INTERMEDIATE IN PROTEIN SYNTHESIS. Journal of Biological Chemistry 231, 241–257 (1958). [PubMed: 13538965]
6. Sanz E et al. Cell-type-specific isolation of ribosome-associated mRNA from complex tissues. Proc Natl Acad Sci U S A 106, 13939–13944, doi:10.1073/pnas.0907143106 (2009). [PubMed: 19666516]
7. Clausen BE, Burkhardt C, Reith W, Renkawitz R & Forster I Conditional gene targeting in macrophages and granulocytes using LysMcre mice. Transgenic Res 8, 265–277 (1999). [PubMed: 10621974]
8. Mudge JM & Harrow J Creating reference gene annotation for the mouse C57BL6/J genome assembly. Mamm Genome 26, 366–378, doi:10.1007/s00335-015-9583-x (2015). [PubMed: 26187010]
9. Pruitt KD, Tatusova T, Brown GR & Maglott DR NCBI Reference Sequences (RefSeq): current status, new features and genome annotation policy. Nucleic Acids Res 40, D130–135, doi: 10.1093/nar/gkr1079 (2012). [PubMed: 22121212]
10. Kozomara A & Griffiths-Jones S miRBase: annotating high confidence microRNAs using deep sequencing data. Nucleic Acids Research 42, D68–D73, doi:10.1093/nar/gkt1181 (2014). [PubMed: 24275495]
11. Carpenter S et al. A long noncoding RNA mediates both activation and repression of immune response genes. Science 341, 789–792, doi:10.1126/science.1240925 (2013). [PubMed: 23907535]
12. Kotzin JJ et al. The long non-coding RNA Morrbid regulates Bim and short-lived myeloid cell lifespan. Nature 537, 239–243, doi:10.1038/nature19346 (2016). [PubMed: 27525555]
13. Osuna BA, Howard CJ, Kc S, Frost A & Weinberg DE In vitro analysis of RQC activities provides insights into the mechanism and function of CAT tailing. Elife 6, doi:10.7554/eLife.27949 (2017).
14. Ingolia NT, Ghaemmaghami S, Newman JR & Weissman JS Genome-wide analysis in vivo of translation with nucleotide resolution using ribosome profiling. Science 324, 218–223, doi: 10.1126/science.1168978 (2009). [PubMed: 19213877]
15. Ji Z, Song R, Regev A & Struhl K Many lncRNAs, 5'UTRs, and pseudogenes are translated and some are likely to express functional proteins. eLife 4, e08890, doi:10.7554/eLife.08890 (2015). [PubMed: 26687005]

16. Lander ES et al. Initial sequencing and analysis of the human genome. Nature 409, 860–921, doi:10.1038/35057062 (2001). [PubMed: 11237011]
17. Venter JC et al. The Sequence of the Human Genome. Science 291, 1304 (2001). [PubMed: 11181995]
18. Kondo T et al. Small peptides switch the transcriptional activity of Shavenbaby during Drosophila embryogenesis. Science 329, 336–339, doi:10.1126/science.1188158 (2010). [PubMed: 20647469]
19. Kearse MG & Wilusz JE Non-AUG translation: a new start for protein synthesis in eukaryotes. Genes Dev 31, 1717–1731, doi:10.1101/gad.305250.117 (2017). [PubMed: 28982758]
20. Stothard P The sequence manipulation suite: JavaScript programs for analyzing and formatting protein and DNA sequences. Biotechniques 28, 1102, 1104 (2000). [PubMed: 10868275]
21. Wang H, Wang Y, Xie S, Liu Y & Xie Z Global and cell-type specific properties of lincRNAs with ribosome occupancy. Nucleic Acids Research 45, 2786–2796, doi:10.1093/nar/gkw909 (2017). [PubMed: 27738133]
22. Xiao Z et al. De novo annotation and characterization of the translatome with ribosome profiling data. Nucleic Acids Research 46, e61–e61, doi:10.1093/nar/gky179 (2018). [PubMed: 29538776]
23. D'Lima NG et al. A human microprotein that interacts with the mRNA decapping complex. Nat Chem Biol 13, 174–180, doi:10.1038/nchembio.2249 (2017). [PubMed: 27918561]
24. de Jong R et al. Severe mycobacterial and Salmonella infections in interleukin-12 receptor-deficient patients. Science 280, 1435–1438 (1998). [PubMed: 9603733]
25. Lehmann J et al. IL-12p40-Dependent Agonistic Effects on the Development of Protective Innate and Adaptive Immunity Against Salmonella Enteritidis. The Journal of Immunology 167, 5304 (2001). [PubMed: 11673546]
26. Mannon PJ et al. Anti-interleukin-12 antibody for active Crohn's disease. N Engl J Med 351, 2069–2079, doi:10.1056/NEJMoa033402 (2004). [PubMed: 15537905]
27. Neurath MF, Fuss I, Kelsall BL, Stüber E & Strober W Antibodies to interleukin 12 abrogate established experimental colitis in mice. The Journal of Experimental Medicine 182, 1281 (1995). [PubMed: 7595199]
28. Pulak R & Anderson P mRNA surveillance by the Caenorhabditis elegans smg genes. Genes Dev 7, 1885–1897 (1993). [PubMed: 8104846]
29. Martin L et al. Identification and characterization of small molecules that inhibit nonsense mediated RNA decay and suppress nonsense p53 mutations. Cancer research 74, 3104–3113, doi:10.1158/0008-5472.CAN-13-2235 (2014). [PubMed: 24662918]
30. Nowarski R et al. Epithelial IL-18 Equilibrium Controls Barrier Function in Colitis. Cell 163, 1444–1456, doi:10.1016/j.cell.2015.10.072 (2015). [PubMed: 26638073]
31. Gabanyi I et al. Neuro-immune interactions drive tissue programming in intestinal macrophages. Cell 164, 378–391, doi:10.1016/j.cell.2015.12.023 (2016). [PubMed: 26777404]
32. Obrig TG, Culp WJ, McKeehan WL & Hardesty B The mechanism by which cycloheximide and related glutarimide antibiotics inhibit peptide synthesis on reticulocyte ribosomes. J Biol Chem 246, 174–181 (1971). [PubMed: 5541758]
33. Schneider-Poetsch T et al. Inhibition of Eukaryotic Translation Elongation by Cycloheximide and Lactimidomycin. Nature chemical biology 6, 209–217, doi:10.1038/nchembio.304 (2010). [PubMed: 20118940]
34. Liao Y, Smyth GK & Shi W featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. Bioinformatics 30, 923–930, doi:10.1093/bioinformatics/btt656 (2014). [PubMed: 24227677]
35. Love MI, Huber W & Anders S Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol 15, 550, doi:10.1186/s13059-014-0550-8 (2014). [PubMed: 25516281]
36. Zhang H, Meltzer P & Davis S RCircos: an R package for Circos 2D track plots. BMC Bioinformatics 14, 244, doi:10.1186/1471-2105-14-244 (2013). [PubMed: 23937229]
37. Phanstiel DH, Boyle AP, Araya CL & Snyder MP Sushi.R: flexible, quantitative and integrative genomic visualizations for publication-quality multi-panel figures. Bioinformatics 30, 2808–2810, doi:10.1093/bioinformatics/btu379 (2014). [PubMed: 24903420]

38. Quinlan AR & Hall IM BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics 26, 841–842, doi:10.1093/bioinformatics/btq033 (2010). [PubMed: 20110278]

39. Frolova L et al. A highly conserved eukaryotic protein family possessing properties of polypeptide chain release factor. Nature 372, 701–703 (1994). [PubMed: 7990965]

40. Pertea M et al. Thousands of large-scale RNA sequencing experiments yield a comprehensive new human gene list and reveal extensive transcriptional noise. bioRxiv (2018).

41. Wessel D & Flugge UI A method for the quantitative recovery of protein in dilute solution in the presence of detergents and lipids. Anal Biochem 138, 141–143 (1984). [PubMed: 6731838]

42. Slavoff SA et al. Peptidomic discovery of short open reading frame-encoded peptides in human cells. Nat Chem Biol 9, 59–64, doi:10.1038/nchembio.1120 (2013). [PubMed: 23160002]

43. Gundry RL et al. Preparation of proteins and peptides for mass spectrometry analysis in a bottom-up proteomics workflow. Curr Protoc Mol Biol Chapter 10, Unit10 25, doi: 10.1002/0471142727.mb1025s88 (2009).

44. Gerber SA, Rush J, Stemman O, Kirschner MW & Gygi SP Absolute quantification of proteins and phosphoproteins from cell lysates by tandem MS. Proc Natl Acad Sci U S A 100, 6940–6945, doi: 10.1073/pnas.0832254100 (2003). [PubMed: 12771378]

45. Chen LM, Kaniga K & Galán JE Salmonella spp. are cytotoxic for cultured macrophages. Molecular Microbiology 21, 1101–1115, doi:doi:10.1046/j.1365-2958.1996.471410.x (1996). [PubMed: 8885278]

46. Gruber AR, Lorenz R, Bernhart SH, Neubock R & Hofacker IL The Vienna RNA websuite. Nucleic Acids Res 36, W70–74, doi:10.1093/nar/gkn188 (2008). [PubMed: 18424795]

47. Xu D & Zhang Y Ab initio protein structure assembly using continuous structure fragments and optimized knowledge-based force field. Proteins 80, 1715–1735, doi:10.1002/prot.24065 (2012). [PubMed: 22411565]

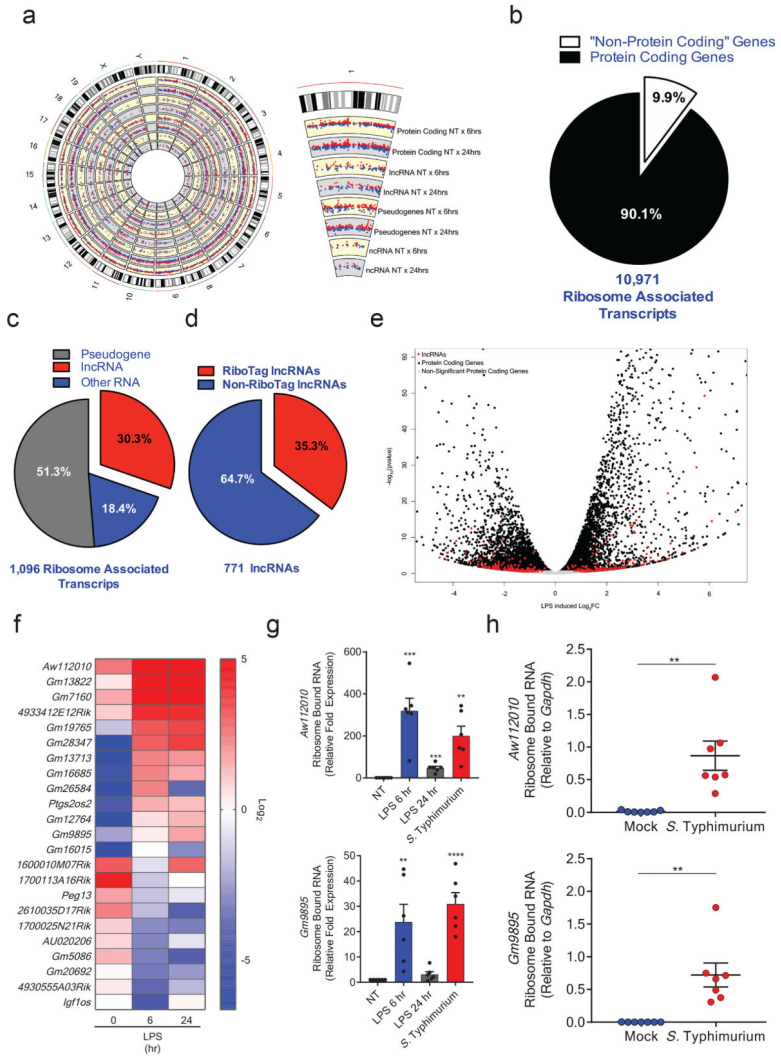


Figure 1: Bacterial Infection Drives Widespread Ribosomal Association with "Non-Coding RNAs"

a-f) BMDMs from RiboTag^{LysM} mice were non-treated (NT) or stimulated with LPS (1 ng/ml). RNA was subjected to RNAseq. Data is presented as a combination of 2 independent biological replicates. **a)** Circos plot shows differentially expressed (Log_2FC) ribosome associated transcripts upon LPS 6 and 24 hr stimulation with red depicting upregulation and blue downregulation. Each track from the periphery to the core represents: chromosomes location; 12,820 known protein coding transcripts; 1,176 lncRNAs; 1,107 pseudogenes; and 413 other non-coding RNA. **b)** Pie chart of percentage breakdown of protein coding gene annotated from RiboTag RNAseq (fpkm ≥ 1). **c)** The exploded "non-protein coding" are further classified. **d)** Stratification of detectable BMDM lncRNAs based upon ribosome association. Ribosome associated lncRNA with an fpkm of ≥1 in RiboTag RNAseq are represented in the red exploded section. Blue section depicts lncRNAs not found in RiboTag RNAseq, but with an fpkm of ≥ 0.01 in conventional RNAseq. **e)** Volcano plot and **f)** heatmap analysis of lncRNAs associated with ribosomes after LPS stimulated in BMDMs. **g)** qPCR analysis of ribosome associated transcripts of non-treated BMDMs or stimulated

with LPS (10 ng/ml) or infected with *S.* Typhimurium at an MOI of 1 for 6 hr. Data is presented as 6 biological replicates and fold expression calculated from each individual NT sample **h)** RiboTag^{LysM} mice were gavaged with 2×10^8 CFU of *S.* Typhimurium. After 24 hr, colonic tissue was extracted and lysed. Macrophage ribosome associated RNA was isolated and qPCR analysis conducted. Data is presented as 7 biological replicates. **g-h)** SEM of replicates is present for qPCR data and significance determined by unpaired two-tailed t-tests, (**p<0.01; ***p<0.001; ****p<0.0001).

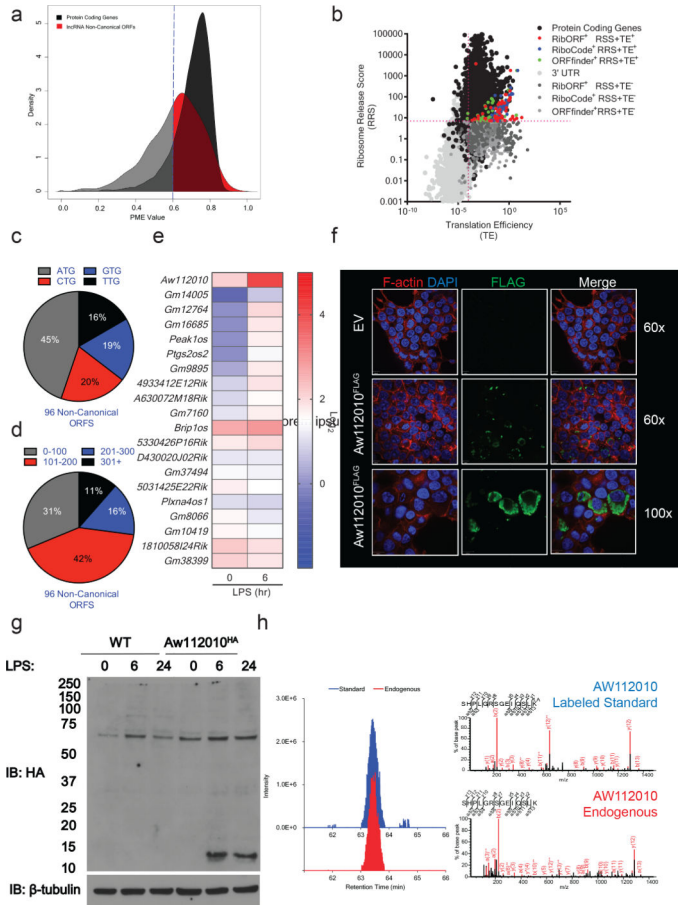


Figure 2: LPS Triggers Genome Wide Differential Translation of Non-Canonical ORFs in lncRNAs

a-e) WT BMDMs were non-treated or stimulated with LPS (10 ng/ml) for 6 hr and ribosome profiling conducted. Data is representative of 2 biological replicates. **a)** Percentage of Maximum Entropy (PME) values for protein coding genes and lncRNAs. ≥ 0.6% PME cutoff represents transcripts considered protein coding. **b)** Translation efficiency (TE) and ribosome release score (RRS) analysis was conducted on RiboProfiling identified transcripts. Purple broken lines represent the 95th percentile of the 3′ UTRs of known protein coding genes and discriminates coding and non-coding transcripts. **c)** Categorization of start codon usage and **d)** ORF size in RibORF and/or RiboCode identified lncRNAs with coding RRS+TE+ values. **e)** Heatmap of top significantly LPS differentially regulated lncRNA ORFs. **f)** HEK293 cells transfected with empty vector (EV) or Aw112010^{FLAG} ORF. Cells were stained with DAPI, Phalloidin and anti-FLAG. Confocal microscopy was conducted at 60x and 100x objectives. **g)** WT and Aw112010^{HA} BMDMs were untreated and stimulated with LPS (10 ng/ml), protein lysates generated and western blot conducted for HA and β-tubulin. Data is representative of 3 biological replicates. **h)** Aw112010^{HA} BMDMs were generated, stimulated with LPS for 6 hr and subjected to HA-immunoprecipitation. Purified lysates were subject to mass spec analysis. Precursor ion peaks in the MS1 extracted ion chromatogram corresponding to a spiked in synthetic isotopically labeled peptide standard (top) and co-elution of a peak consistent with the

endogenous Aw112010 peptide (bottom) in the same sample. Identified fragment ions (b and y-ions, red) are indicated above and below the peptide sequence. Data is representative of 2 biological replicates.

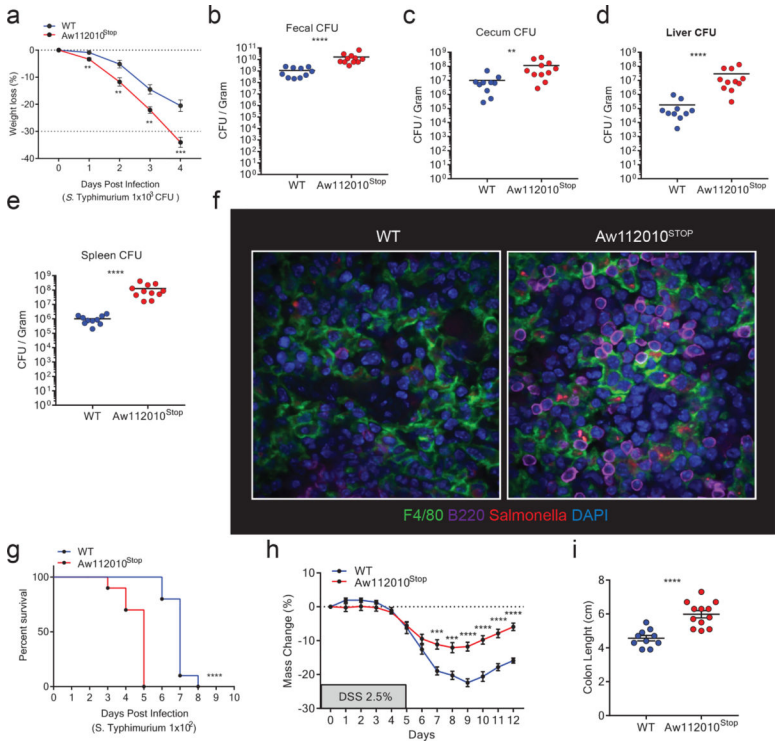


Figure 3: Translation of the Non-Canonical Aw112010 Encoded ORF is Essential for Mucosal Immunity

a-e) WT (n=10) and Aw112010^{Stop} (n=11) mice were administered Streptomycin (20mg) by oral gavage 24 hr prior to *S.* Typhimurium infection (1×10^3 Colony Forming Units (CFU). **a)** Weight loss was measured post infection. **b)** Enumeration of *S.* Typhimurium CFUs present in feces of WT and Aw112010^{Stop} mice 24 hr post infection. **c)** Enumeration of *S.* Typhimurium CFUs in the cecum of WT and Aw112010^{Stop} mice 96 hr post infection. **d)** Enumeration of *S.* Typhimurium CFUs in the liver and **e)** spleen of WT and Aw112010^{Stop} mice 96 hr post infection. **f)** Confocal immunostaining of macrophages (F4/80, Green), B cells (B220, Purple), Salmonella (anti-Salmonella, Red) in the spleens of WT and Aw112010^{Stop} mice infected with 1×10^2 CFUs of *S.* Typhimurium 72 hr post gavage. Representative of 3 independent biological replicates Original magnification 60x. **g)** Survival curve analysis of WT (n=10) and Aw112010^{Stop} (n=10) infected with 1×10^2 CFU via oral gavage. **h, i)** WT and Aw112010^{Stop} cohoused littermate mice were administered 2.5% DSS in their drinking water for 5 days. **h)** Weight loss from WT (n=11) and Aw112010^{Stop} (n=12) Weight loss was measured over 12 days. **j)** Colon length was measured from WT (n=10) and Aw112010^{Stop} (n=12). SEM of replicates is presented for weight loss and colon length data and statistical significance determined by unpaired two tailed t tests. Bacterial CFU data is presented SEM in log scale and statistical significance determined by nonparametric Mann-Whitney test. Survival curve statistical analysis was determined by a Log-rank test. Significance indicated by *p<0.05; **p<0.01; ***p<0.001; ****p<0.0001.

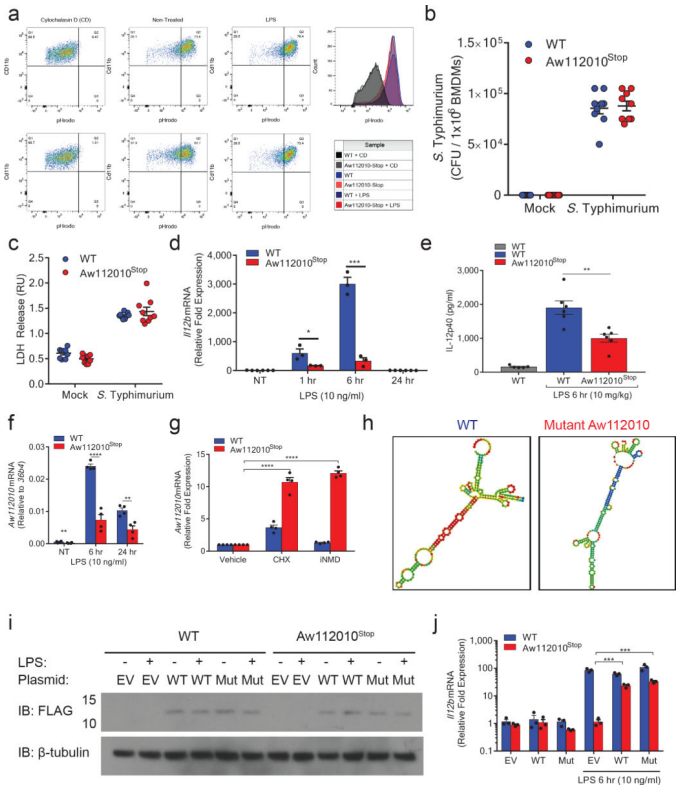


Figure 4: Translation of the Aw112010 Non-Canonical ORF Encoded Protein is required for IL-12 production

a) BMDMs were pretreated with CytochalasinD (CD) (10 μM) for 1 hr, LPS (10 ng/ml) for 6 hr or non-treated (NT). pHrodo BioParticles were administered for 1 hr and cells assessed for Cd11b and pHrodo. Plots are representative of 3 independent experiments. **b)** BMDMs were infected with *S.* Typhimurium for 6 hr. Cells were lysed and colony forming units enumerated (CFUs). **c)** BMDMs were pretreated with LPS (100 ng/ml) for 5 hr and infected with a *S.* Typhimurium for 1 hr and LDH release measured. **d)** BMDMs were stimulated with LPS (10 ng/ml). qPCR was conducted for *Il12b* expression. **e)** Mice were administered PBS (n=5) or LPS (n=6, WT and Aw112010^{Stop}) (10 mg/kg) for 6 hr via intraperitoneal injection. Serum was analyzed for IL-12p40 by ELISA. **f)** BMDMs were stimulated with LPS (10 ng/ml). qPCR conducted for *Aw112010* expression. **g)** WT and Aw112010^{Stop} BMDMs were treated with cycloheximide (50 μg/ml) or non-sense mediated decay inhibitor (iNMD) (50 μM) for 6 hr. qPCR was conducted for *Aw112010*. Fold expression was calculated from each individual replicates vehicle sample **h)** Predicted RNA folding of *Aw112010* mRNA (WT) and a mutant *Aw112010* transcript (Mut). **i, j)** BMDMs were subjected to electroporation with indicated plasmids. BMDMs were stimulated with LPS (10 ng/ml) for 6 hr. **i)** Western blot conducted for Aw112010-Flag and β-tubulin. **j)** qPCR was conducted for *Il12b* mRNA. Where applicable all data is presented as SEM. b, c) Data is from 3 independent experiments conducted with 3 biological and 3 technical replicates. d) Data is of 3 biological replicates and fold expression calculated from a WT NT sample. f,g**)** Data is from 4 independent experiments. i-j) Data is of 3 biological replicates. j) Fold expression is calculated from a single WT EV NT replicate for WT cells and a single

Aw112010^{Stop} EV NT replicate for Aw112010^{Stop} cells. Statistical significance was determined by unpaired two-tailed t tests. Significance indicated by *p<0.05; **p<0.01; ***p<0.001; ****p<0.0001.