Dual RNA-seq provides insight into the biology of the neglected intracellular human pathogen *Orientia tsutsugamushi*

Bozena Mika-Gospodorz[#a], Suparat Giengkam[#b], Alexander J. Westermann[a,c], Jantana Wongsantichon[b], Willow Kion-Crosby[d], Suthida Chuenklin[b], Loo Chien Wang[e], Piyanate Sunyakumthorn[f], Radoslaw M. Sobota[j], Selvakumar Subbian[h], Jörg Vogel[a,c], Lars Barquist[#*a,g] and Jeanne Salje[#*b,h,i]

Running title: RNA sequencing of *Orientia tsutsugamushi*

Helmholtz Institute for RNA-based Infection Research (HIRI), Helmholtz Centre for Infection Research (HZI), Würzburg, Germany[a]

Mahidol-Oxford Tropical Medicine Research Unit, Faculty of Tropical Medicine, Mahidol University, Bangkok, Thailand[b];

Institute for Molecular Infection Biology (IMIB), University of Würzburg, Würzburg, Germany[c]

Rutgers, the State Univeristy of New Jersey, New Jersey, USA[d]

Functional Proteomics Laboratory, Institute of Molecular and Cell Biology, Agency for Science, Technology and Research (A*STAR), Singapore[e].

Armed Forces Research Institute of Medical Sciences, Bangkok, Thailand[f];

Faculty of Medicine, University of Würzburg, Würzburg, Germany[g];

Public Health Research Institute, Rutgers University, New Jersey, USA[h];

Centre for Tropical Medicine and Global Health, Nuffield Department of Medicine, University of Oxford, Oxford, United Kingdom[i];

SingMass - National Mass Spectrometry Laboratory, Institute of Molecular and Cell Biology, Agency for

24 Science, Technology and Research (A*STAR), Singapore[j]

25

26

27 [#] these authors contributed equally

28 *address correspondence to js2522@njms.rutgers.edu; lars.barquist@helmholtz-hiri.de

Summary

Emerging and neglected diseases pose challenges as their biology is frequently poorly understood, and genetic tools often do not exist to manipulate the responsible pathogen. Organism agnostic sequencing technologies offer a promising approach to understand the molecular processes underlying these diseases. Here we apply dual RNA-seq to *Orientia tsutsugamushi* (Ot), an obligate intracellular bacterium and the causative agent of the vector-borne human disease scrub typhus. Half the Ot genome is composed of repetitive DNA, and there is minimal collinearity in gene order between strains. Integrating RNA-seq, comparative genomics, proteomics, and machine learning, we investigated the transcriptional architecture of Ot, including operon structure and non-coding RNAs, and found evidence for wide-spread post-transcriptional antisense regulation. We compared the host response to two clinical isolates and identified distinct immune response networks that are up-regulated in response to each strain, leading to predictions of relative virulence which were confirmed in a mouse infection model. Thus, dual RNA-seq can provide insight into the biology and host-pathogen interactions of a poorly characterized and genetically intractable organism such as Ot.

Keywords: neglected and emerging pathogens, intracellular bacteria, dual RNA-seq, transcriptomics, host-pathogen cell biology, bacterial virulence, antisense transcription.

Main Text

Introduction

Improved surveillance and diagnostics have led to the recognition of previously neglected bacteria as serious pathogens, whilst human population growth, globalization and increased travel have contributed to the emergence of new pathogens and changing patterns of infectious disease. The biology of neglected and emerging pathogens is often poorly understood, but is essential to developing therapeutic and preventative strategies. Obligate intracellular pathogens present additional challenges, as many cause diseases that are difficult to diagnose and are difficult to manipulate experimentally.

Obligate intracellular bacteria include the Rickettsiales, an order which includes the arthropod and nematode symbiont *Wolbachia* as well as a number of human and veterinary pathogens. *Orientia tsutsugamushi* (Ot, Class Alphaproteobacteria, Order Rickettsiales, Family Rickettsiaceae) causes the mite-borne human disease scrub typhus, a leading cause of severe febrile illness in the Asia Pacific region[1], home to roughly two thirds of the world's population. Locally acquired cases in the Middle East and Latin America suggest that this disease may be more widespread than previously appreciated[2,3]. Under-recognition and under-reporting are a major problem in scrub typhus because unambiguous diagnosis is difficult, and awareness is low amongst many clinicians. Symptoms are non-specific and include headache, fever, rash, and lymphadenopathy beginning 7-14 days after inoculation via a feeding larval stage mite. If untreated, this can progress to cause multiple organ failure and death. In the mite vector Ot colonizes the ovaries and salivary glands. During acute infection of its mammalian host, the bacteria infect endothelial cells, dendritic cells and monocytes/macrophages at the mite bite site[4], and then disseminate via blood and lymphatic vessels to multiple organs including lung, liver, kidney, spleen and brain[5].

Ot strains are highly variable in terms of antigenicity and virulence. Hundreds of strains have been described based on differences in the sequence of the surface protein TSA56[9–11]. These strains are classified into 7 geographically diverse serotype groups, dominated by the Karp, Kato and Gilliam groups[12]. As with many other pathogenic bacteria, whole genome sequencing has revealed that serotype-based groupings do not necessarily reflect phylogenetic relationships[8]. Different strains of Ot exhibit different levels of virulence[19–21], dependent on both bacterial and host genotype. For example, strain Karp (serotype group Karp) causes lethal infection in BALBc and C3H/He mice at low doses, strain Gilliam (serotype group Gilliam) causes lethal infection in C3H/He but not BALBc mice at similar doses, whilst strain TA716 (serotype group TA716) does not cause lethal infection in either mouse model at similar doses[20,22]. The underlying causes of this variation in infection outcomes remain obscure.

Dual RNA-seq quantifies RNA transcripts of intracellular pathogens and host cells in a single experiment[24,25], and can provide insight into both the host and pathogen response to infection. For example, dual RNA-seq has been used to study obligate intracellular *Chlamydia trachomatis*[26] revealing the rewiring of *Chlamydia* metabolism during the onset of an infection of human epithelial cells, together with the corresponding host responses. Here we apply dual RNA-seq to deepen our understanding of the RNA biology of Ot and its consequences for virulence. We survey the transcriptome of Ot strain Karp, identifying non-coding RNAs and transcribed operons in a genome broken by frequent recombination and transposition of the rickettsial amplified genetic element (RAGE) integrative and conjugative element (ICE)[6,7]. Integrating proteomic measurements, we further provide evidence that RAGE genes are regulated through prevalent antisense transcription. Finally, we compare infection between strain Karp and strain UT176 identifying a core host response to Ot dominated by type-I interferon signaling, as well as distinct

immune responses to each strain. We show that this in turn leads to different outcomes in a mouse model of scrub typhus.

Results

Dual RNA-seq of *Orientia tsutsugamushi* infecting endothelial cells

We focused on two Ot clinical isolates: Karp, taken from a patient in New Guinea in 1943[27], and UT176, a separate strain in the same serotype group as Karp, taken from a patient in northern Thailand in 2004[28]. These strains are closely related, with a sequence identity of 95% in their TSA56 gene (used to classify strains)[8]. Consistent with a closed pan-genome for Ot, the gene content of Karp and UT176 are similar, with differences primarily in gene copy number, pseudogenes, and gene order along the genome. Human vein endothelial cells (HUVEC) were selected as host cells due to their similarity to cell types involved in both early and advanced infection. HUVEC cells were infected with bacteria at an MOI of 100:1 and grown for 5 days (Fig. 1A), by which point host cells were heavily loaded with bacteria (Fig. 1C, Supp. Figs 1, 2). Uninfected HUVEC cells were grown in parallel. After 5 days total RNA was isolated, depleted for rRNA, converted to cDNA and sequenced to ~35 million reads per library using Illumina technology. Reads were mapped to the completed genomes of Karp, UT176[8] and, in parallel, the human genome. As the *Orientia* genome is repeat-rich, we additionally applied model-based quantification with Salmon[29] which uses uniquely mapping reads to assign multi-mapping reads to these transcriptomes to improve our estimates of transcript abundance (see Methods).

We observed 17.1-17.5% bacterial reads in HUVECs infected with Karp and 2.8-4.9% bacterial reads in HUVECs infected with UT176 (Fig. 1D, Supp. Fig. 3). This difference likely reflects growth rate differences between Karp and UT176, which have doubling times of 19 and 27 hours in HUVEC, respectively (Fig. 1B). The distribution of reads to RNA classes (Fig. 1D) indicated efficient depletion of ribosomal transcripts in the host transcriptome (<0.001% human rRNA reads). In contrast, we found an average of 16% and 30% rRNA reads in Karp and UT176, respectively. Most of these remaining bacterial

ribosomal reads were derived from 5S rRNA (Supp. Table), likely reflecting the divergence of 5S rRNA sequences between Ot and bacterial model organisms used for optimization of the Ribo-Zero approach (https://emea.illumina.com/products/selection-tools/ribo-zero-kit-species-compatibility.html?langsel=/de/). This notwithstanding, reads mapping to coding sequences (CDSs) were not only abundant in the HUVEC data subset (53% of all host-mapped reads), but also in the Ot-specific reads (38% of the Karp- and 49% of the UT176-mapped reads), allowing differential expression analysis. Dual RNA-seq also readily detected the various non-coding RNA classes from both host and bacteria (Fig. 1E). Of 657 predicted core Ot genes[8] 491 were expressed and 212 were highly expressed (see Methods for definitions).

Ot ncRNAs and evidence for tmRNA processing

Bacterial genomes encode many non-coding (nc)RNAs. Among the most conserved are several specialized, abundant housekeeping ncRNAs, including the RNA components of ribonuclease P (RNase P), the signal recognition particle (SRP), and transfer-messenger RNA (tmRNA), all of which were detected in the Karp transcriptome data (Fig. 1E, Supp. Table 1). To validate the RNA-seq data, we performed Northern blot analysis for conserved housekeeping ncRNAs (Fig. 2A). These include the M1 RNA component of RNase P, a ribozyme responsible for tRNA processing, and 4.5S, the RNA component of the SRP involved in translocation of membrane proteins. Both ran at their expected lengths of ~385 and ~100 nt, respectively. However, a second stronger band for the M1 transcript ran slightly higher, indicative of a length of ~450 nt, suggesting the existence of a precursor-M1.

We also found evidence of tmRNA processing in Ot. tmRNA has both mRNA-like and tRNA-like features, rescues stalled ribosomes[30], and is known to contribute to virulence in pathogens as diverse as

Salmonella Typhimurium[31] and *Francisella tularensis*[32]. In our data, tmRNA appears to be expressed at unusually high levels, contributing between 3 and 7% of total bacterial reads (Fig 1E), suggesting an important role in Ot survival in mammalian cells. tmRNA generally consists of a tRNA-like (acceptor) domain encoded upstream of a short open reading frame (coding domain). However, the transcript has undergone a circular permutation in some clades of bacteria[33], including the Alphaproteobacteria[34], which requires processing of a precursor transcript into separate, base-pairing acceptor and coding RNA chains[35,36] (Fig. 2B). We detected three Ot tmRNA forms using Northern blot: (i) a long precursor tmRNA (372 nt); (ii) a 5' fragment of ~80 nt, the acceptor domain; (iii) and the 3' coding domain of ~240 nt (Fig. 2B). Read coverage over the tmRNA locus in the Karp genome supported a cleavage event within the loop region that connects the tRNA- and mRNA-like domains in the full-length precursor (Fig. 2C).

In addition to these universally conserved housekeeping ncRNAs, bacterial genomes encode family-, genus-, species-, or strain-specific small ncRNAs (sRNAs) to adapt their gene expression to specific intrinsic and environmental cues[37,38]. Our RNA-seq data identified 55 intergenic sRNA candidates, between 77-803 nt, in the Karp transcriptome (Supp. Table 1). When normalized to the genome size of Ot, this is consistent with the number of sRNAs reported in model bacterial pathogens[39–44].

Conserved operons in a dynamic genome

The genome of Ot is highly dynamic[8], and while the timescales and mechanisms of its rearrangements are unknown they are thought to be driven by an extreme proliferation of mobile elements[6,7], in particular the RAGE. The consequences of this are evident when comparing the high degree of synteny in bacteria from two related 'normal' genera (*Escherichia* and *Salmonella*) to the complete shuffling we observe between the two Ot strains studied here (Fig. 2D). As bacterial genomes are normally organized into co-transcribed

operons of functionally related genes, we wondered how this macroscale loss of synteny would affect conservation of these transcripts. Using Rockhopper[45] and manual curation, we identified adjacent genes expressed in a continuous transcript, classifying these as operons. We identified 131 operons fully conserved between Karp and UT176 (all genes expressed in both strains) and seven partly conserved (some genes expressed in both strains). Our previous analysis of 8 Ot genomes identified 51 universally conserved genomic islands, including 35 potential collinear gene clusters containing two to thirteen genes[8], and we found evidence for operonic transcripts originating from 24 of these. We also identified 212 and 192 transcribed operons present only in Karp or UT176, respectively, and these were generally associated with the RAGE mobile element (73% in Karp and 93% in UT176) in contrast to conserved operons (14% of fully conserved operons, Fig 2E).

The majority (84%, Supp. Fig. 6) of conserved operons consisted of only two or three genes. Longer operons tended to encode for core cellular processes, the longest being a 30 gene operon encoding almost half of Ot ribosomal proteins proximal to the ribosomal RNA operon itself (Fig. 2F). Others included an 8 gene operon involved in iron-sulfur cluster assembly, and 6 and 5 gene operons in distinct loci each encoding for portions of the NADH-ubiquinone oxidoreductase complex in an organization similar to that observed in *Rickettsia prowazekii* and eukaryotic mitochondria[46]. In summary, the identification of cotranscribed gene clusters in a genome as highly dynamic as that of Ot indicates strong selection for those genes to remain coupled, indicating involvement in the same pathways and likely shared regulation.

Evidence for Ot RAGE regulation by antisense RNA

The RAGE of Ot is present in at least 185 remnant copies[7]. It encodes an integrase (*int*) and transposase gene *(tra),* multiple genes from the VirB type IV secretion system (*vir*) and a number of potential effector

185 genes including ankyrin-repeat containing proteins (*ank),* tetratricopeptide repeat containing proteins
186 (*TPR*), SpoT/RelA genes, DNA methyltransferases and replicative DNA helicases. Many of these genes
187 are truncated and most RAGE copies are highly degraded, containing only a subset of genes from the
188 complete element. It is not known if this ICE is still active for transposition, nor whether Ot can express
189 a functional type IV secretion apparatus. In our RNA-seq dataset ~50% of the most highly expressed genes
190 were repetitive genes encoded by the RAGE (defined throughout our analysis as integrase, transposase,
191 conjugal transfer genes and hypothetical genes) in both strains. These same genes were also highly
192 expressed in the antisense direction (Fig. 3A, Supp. Fig. 6), leading us to hypothesize that the repetitive
193 RAGE genes may be regulated by antisense gene expression.

194 Antisense transcription is widespread in bacteria[47], with between 5% and 75% of coding sequences
195 exhibiting antisense transcription. While functions for a number of specific antisense transcripts have been
196 described, including regulation through occlusion of the ribosome binding-site or induction of RNase III
197 mediated decay[48] their relevance as a general functional class remains unclear. Antisense promoters tend
198 to be weakly conserved[49], arguing against specific functions, and mathematical modeling has suggested
199 the majority of antisense transcripts are not expressed at sufficient levels to affect the regulation of their
200 cognate coding sequence[50].

201 To explore the relationship between sense and antisense expression of core Ot genes and repetitive RAGE
202 genes, we combined our Karp RNA-seq dataset with a proteomics dataset generated under the same
203 experimental growth conditions. We observed substantially fewer RAGE gene products detected by
204 proteomics, compared with RNA-seq (Fig. 3A). Genes with detected protein products had higher
205 transcript expression on average compared to those not detected by proteomics (Fig. 3B). However, many
206 highly expressed transcripts appeared to produce no protein. Given our previous observations, we asked
207 whether antisense transcription would correlate with protein expression. All genes with detected proteins

had an antisense-sense read count ratio of less than 1, in contrast to genes with no detected protein product, which had an antisense-sense read count spanning several orders of magnitude (Fig. 3C) suggesting antisense RNA expression may be a factor in inhibiting translation.

To test this hypothesis more rigorously, we constructed three logistic regression models to predict protein detection from our transcriptomic data. The first used only transcripts per million (TPMs) derived from the sense strand as a predictor; the second used only the antisense-sense read count ratio as a predictor; the third used both features. Comparisons of the predictive power of these three models showed that antisense transcription is predictive of protein expression (Fig. 3D). Model 1, relying only on sense expression, did little better than chance at predicting protein detection. Models 2 and 3, which incorporate the antisense-sense ratio, led to large improvements in predictive power, suggesting that antisense transcription plays a wide-spread regulatory role in Ot. This was confirmed by cross-validation (Methods, Supp. Fig. 7). We found significant enrichment for RAGE genes among those with high antisense-sense ratios (Fig. 3E), suggesting antisense transcription may work to control the expression of selfish genetic elements at the protein level. Thirty core genes also exhibited an antisense-sense ratio of >1 (Supp. Table core genes) and these include the chromosomal replication initiator protein DnaA, DNA polymerase subunit III, an outer membrane autotransporter protein ScaD, glutamine synthetase, three transporters, the protein export protein SecB and 12 hypothetical proteins. None of these models achieved greater than 67% balanced accuracy, which may be due to both the existence of other modes of post-transcriptional regulation and the lack of sensitivity in our proteomics. For instance, we have also performed a preliminary investigation of codon bias and found some evidence for differential codon usage in genes expressed at the RNA, but not protein, level (Supp. Text, Supp. Fig. 8-9).

Differential expression of genes in Karp and UT176

231 Due to a lack of genetic tools, identification of virulence mechanisms in Ot has been difficult, with only
232 a small number of antigenic surface proteins and effectors known. As pangenome diversity appears to
233 primarily be the result of gene duplication and decay, differences in virulence between strains are likely
234 due to differences in expression. To investigate this hypothesis, we performed differential expression
235 analysis between Karp and UT176 at 5 d after infection of HUVEC cells. Pathway and gene ontology
236 (GO) analyses of differentially expressed genes (Fig. 4A, Supp. Table 1) indicated that most pathways
237 were up-regulated in Karp compared with UT176, including those involved in DNA replication and
238 metabolism, consistent with Karp's higher growth rate (Fig. 1B). At the gene level (Supp. Table 1, Fig.
239 4B) we found a number of surface and effector proteins (Anks) were differentially regulated between the
240 two strains. Ot encodes five autotransporter domain-containing proteins (ScaA-ScaE) and three
241 immunogenic-type surface antigens (TSA22, TSA47, TSA56). All these surface proteins are
242 immunogenic, based on their reactivity to patient sera[60], with TSA56 being the most abundant Ot surface
243 protein. TSA56 has four variable domains and these lead to strain-specific antibody responses in patients.
244 TSA47, TSA56, and ScaA have been evaluated as possible vaccine candidates[61,62]. Of the core Ot genes,
245 those most differentially expressed between Karp and UT176 were *scaE*, *tsa56,* and *tsa22* (1.39, 3.06 and
246 3.94 logFC in Karp over UT176, respectively). In contrast, *scaD* levels were increased in UT176 but to a
247 lesser degree (0.80 logFC in UT176 over Karp). Given their high immunogenicity it is likely that their
248 differential expression results in differential host responses.
249
250 Genes for Ank and tetratricopeptide repeat-containing proteins (TPR) are present in 33 (Ank)/29 (TPR)
251 and 21 (Ank)/22 (TPR) copies in Karp and UT176, respectively [8]. Some *ank* genes function as effectors
252 in eukaryotic cells whilst others are uncharacterized. We compared the expression of Ank and TPR genes
253 in Karp and UT176. *ank2*, *ank3*, *ank12*, *tpr6,* and *tpr8* were up-regulated with a logFC >1.5 in UT176,

while *ank16*, *ank17*, *ak20,* and *tpr1* were up-regulated with a logFC >1.5 in Karp. Most of these proteins were not detected in the Karp proteomics dataset suggesting that either the mRNAs were not translated, or that the proteins were secreted and lost during purification. The protein products of all of these *ank* genes localise to the endoplasmic reticulum or host cell cytoplasm when ectopically expressed[63]. Ank6 interferes with NFkB translocation to the nucleus and inhibits its transcriptional activation[64]. The activity of the other differentially expressed Anks is not known. Given that these effector proteins interact directly with host cell proteins we expect that this differential expression will lead to downstream differences in host response.

Karp and UT176 induce a type-I interferon proinflammatory response

The transcriptional profile of HUVEC cells infected with Karp or UT176 showed a clear core response to Ot (Fig. 5A, red), with smaller gene sets responding specifically to a single strain (Fig. 5A purple and orange). The core response was dominated by a type-I interferon proinflammatory response (Supp. Table 1), seen previously in cultured endothelial cells and monocytes, as well as patient-derived macrophages[13–17]. This is further illustrated by activation of the canonical interferon signaling pathway in response to Karp (Fig. 5B), with a similar response observed for UT176 (Fig. 5C).

Host genes commonly up-regulated upon infection with either Ot strain include *IFNB1* (interferon beta) and genes involved in regulating the type-I interferon response: *IRF9* (interferon-regulated factor 9) and *STAT1*/*2*. Interferon-stimulated genes were also up-regulated upon Ot infection, including various interferon induced proteins with tetratricopeptide repeats (*IFIT*) genes and 2'-5'-oligoadenylate synthase 1 (*OAS1*). In addition to the type-I interferon pathway, the joint Ot response led to up-regulation of

276 proinflammatory chemokine genes including *CXCL10*, *CXCL11*, and for cytokine receptors *IL13RA2*,
277 *IL7R*, *IL15RA*, and *IL3RA* (Supp. Table 1).
278
279 The upstream signals leading to activation of these signaling pathways are unknown but Ot has been
280 shown to activate host cells by signaling through the NOD1-IL32[65] and TLR2[66] pathways. Our data
281 showed that *TLR3* is up-regulated in cultured HUVEC cells in response to both Karp and UT176 (Supp.
282 Table 1). TLR3 recognizes viral double-stranded (ds)RNA in the cytoplasm[67] and it is possible that it
283 responds to Ot dsRNA. The up-regulation of the mRNA for transcription factor IRF7, which is known to
284 respond to stimulation from membrane-bound TLRs, further supports a role for TLR2 and TLR3 in the
285 detection of Ot.
286
287 **Differential host responses to Karp and UT176**
288 Although Karp and UT176 both induced a type-I interferon proinflammatory response compared to
289 uninfected HUVEC cells (Fig. 5B,C), each strain also induced its own unique response. Some of these
290 expression changes were validated by qRT-PCR (Supp. Fig. 12). The mRNA levels of multiple cytokines,
291 chemokines, and cytokine receptors were higher in HUVEC cells infected with UT176 compared with
292 Karp (Fig. 6A, Supp Fig 14, 15, 17). A network map of proinflammatory chemokines and cytokines, and
293 their differential induction in response to UT176 and Karp is shown in Fig 6A and Supp. Fig. 14A. Most
294 of the genes for cytokines, chemokines, and cytokine receptors were differentially up-regulated by
295 infection with UT176 compared with Karp, including *CXCL8*, *CXCL1*, *CXCL2*, *CXCL10*, *IL6*, *IL1RL1*,
296 *IL18R1*. The mRNA levels of surface adhesion molecules associated with activation of the endothelium,
297 VCAM1 and ICAM1, were also up-regulated in UT176 infected cells compared with Karp (Supp. Table
298 1). Whilst *TLR3* was up-regulated in both strains, TLR3 activation in UT176 infected cells was 1.5 logFC

299 higher than in response to Karp. Comparison of NFkB pathway genes and genes associated with NOS2
300 production revealed that genes in both pathways were up-regulated in UT176 but they were not up-
301 regulated or were significantly less up-regulated in response to Karp infection (Supp. Fig. 15). Expression
302 of host genes associated with leukocyte proliferation and mononuclear leukocyte differentiation was
303 strongly induced in HUVECs infected with UT176 but significantly less so when infected with Karp
304 (Supp. Fig. 16). Thus, UT176 seems to induce a stronger proinflammatory response and this may lead to
305 more effective pathogen clearance (see Fig. 1B).
306
307 In contrast to the multiple chemokines and cytokines up-regulated in UT176-infected HUVEC cells, only
308 *IL33* was specifically up-regulated in Karp-infected HUVEC cells (5 logFC difference; Supp. Table 1).
309 IL33 is a proinflammatory cytokine that is involved in pathogenicity in a mouse model of scrub typhus[68].
310 To investigate Karp-mediated activation of *IL33* we analyzed gene induction in the IL33-FAS network
311 (Fig. 6B/Supp. Fig. 14B). Most genes in the network were differentially induced in Karp-infected HUVEC
312 cells compared to in UT176-infected HUVECs. Up-regulation of IL33-NOS-mediated signaling
313 contributes to tissue inflammation. We analyzed networks of genes involved in (i) organismal growth
314 failure (ii) organismal morbidity and mortality and (iii) organismal death. In all cases Karp induced these
315 networks whilst UT176 dampened them (Supp. Fig. 17).
316
317 **Two Ot strains differ in virulence in a mouse model**
318 To assay the effects of the differential inflammatory responses induced by Karp and UT176 we tested the
319 relative virulence of the two strains in an intravenous mouse infection model. $1.25x10^6$ bacteria were
320 inoculated into outbred ICR mice and monitored for disease symptoms for 12 days prior to euthanasia.
321 Blood and tissue from lung, liver, spleen, and kidneys were isolated and the bacterial load measured by

322 qPCR. We found that Karp was significantly more virulent than UT176 as determined by disease outcome,
323 weight loss, bacterial load, and histopathological analysis (Fig. 7, Supp. Fig. 19-21). This difference is
324 likely to result from a combination of differential bacterial growth rate (Fig. 1B), differential expression
325 of bacterial virulence related genes (Fig. 4), and differential induction of host immune networks (Fig. 6).

326
327

Discussion

Both its obligate intracellular lifestyle and the complexity of rearrangements in the Ot genome make it difficult to study. Ot has a genome of 1.9-2.5 Mbp, almost half of which is composed of repetitive regions of >1,000 bp in length[8]. This is in contrast to the most closely related rickettsial species, whose genomes are typically around 1.1-1.3 Mbp[69]. The Ot genome is remarkably unstable, which makes inference of its transcriptional architecture particularly difficult. Using RNA-seq, we were able to identify core ncRNAs, putative sRNAs, and operonic transcripts. In sharp contrast to most bacteria, only a handful of operons containing more than 2 or 3 genes were conserved between Karp and UT176, and these primarily encode for proteins involved in core cellular processes like respiration and translation. Given that Karp encodes only 12 predicted transcription factors and 3 sigma factors, in contrast to 300 and 7, respectively, in *E. coli*, this raises the question of how transcription in Ot is coordinated.

One possible explanation is that much Ot transcription is not stringently controlled, and alternative mechanisms have arisen in Ot to control protein expression. This is supported in part by our observation that protein expression is partially predicted by antisense transcription. This mode of regulation seems to be particularly prevalent for genes encoded by the RAGE, a transposable element of the integrative and conjugative element group. Transposable element regulation by antisense transcripts was one of the earliest discovered examples of riboregulation[71], though it has not previously been observed at the scale implied by our RNA-seq analysis. Such antisense regulation could arise spontaneously through capture of transcriptional noise, providing a parsimonious alternative to transcriptional control[72]. It is unclear whether these untranslated transcripts have some function in Ot, or whether they are purely selfish DNA elements that Ot has been unable to dispose of due to its small population size. One intriguing possibility

is that this regulatory mechanism would provide a large pool of double stranded RNA upon intracellular bacterial lysis, which may explain Ot induction of TLR3 and an antiviral immune response.

In the absence of genetic tools it is difficult to identify specific genes that drive virulence differences between UT176 and Karp. However, comparative genomics has revealed that whilst the pan genome of Ot is open, it is largely composed of gene duplications rather than newly acquired genes. This lack of gene novelty likely reflects the environmental isolation associated with an obligate intracellular lifestyle. Consequently, strain-specific differences in virulence are likely to be driven largely by differences in relative gene expression rather than the presence or absence of virulence genes. Consistent with this, we observed an up-regulation of virulence-associated surface proteins in Karp compared with UT176.

The inflammatory response triggered by Ot infection is a key driver of virulence in scrub typhus. We compared the response of endothelial cells to the two strains of Ot and found that differential activation of the immune response correlated with differential outcomes in a scrub typhus mouse model. Whilst both Karp and UT176 induced an antiviral proinflammatory response, as shown previously[13–17], UT176 strongly induced an IL6-mediated pro-inflammatory response whilst Karp induced an IL33-NOS3-FAS response, differences likely to influence the relative virulence of these strains.

IL33 was one of the most strongly differentially regulated genes between UT176 and Karp infections (5.1 logFC higher in Karp-infected HUVECs). IL33 has previously been shown to play a role in pathogenesis in a scrub typhus murine model, using the Karp strain, where it was shown that IL33 levels were increased during Ot infection, that *IL33*-/- mice showed less severe disease symptoms, and that addition of rIL33

372 increased severity and mortality[68]. Our observations of reduced induction of *IL33* by the less virulent
373 UT176 strain further support a role for this cytokine in the pathogenesis of scrub typhus.
374
375 In summary, we have used dual RNA-seq to gain insights into the transcriptome structure and mechanisms
376 of gene regulation in the neglected intracellular pathogen Ot during infection. We provide evidence for
377 widespread antisense regulation, in particular for the RAGE genes. We identified a relationship between
378 the relative induction of IL33- and IL6-based gene networks in the host and disease severity. These
379 findings will lay the groundwork for subsequent studies on the regulation of gene expression in Ot and
380 mechanisms of pathogenesis. More generally, the present study may serve as a blueprint for the
381 characterization of further obligate intracellular, genetically intractable bacterial pathogens.
382

383 **Figure legends**

384 **Figure 1. Overview. A.** Schematic experimental overview. HUVEC = human umbilical vein endothelial
385 cell. **B.** Growth curve showing replication of Ot in cultured HUVEC cells. Bacteria were grown in 24 well
386 plates and the total bacteria per well is shown. Mean and SD from three independent replicates are shown.
387 **C.** Confocal microscopy images of Ot in HUVEC cells 5 days post infection. Additional images and time
388 points are shown in Supp. Fig. 1 and 2. Blue = DAPI (DNA), Red = Evans blue (host cells), green = Ot
389 labelled with Alexa488-click-methionine. **D.** RNA mapping statistics showing the fraction of host and Ot
390 RNA for each condition. The first replicate of the experiment is shown. Individual results for each replicate
391 are shown in Supp. Fig. 3. **E.** Percentage of RNA-seq reads assigned to different classes of RNA in Karp,
392 UT176 and HUVEC.
393
394 **Figure 2. RNA biology in Ot. A.** Northern blot analysis of core non-coding RNAs in Ot. **B.** Structure of
395 the two-piece tmRNA observed in the Ot transcriptome **C.** RNA-seq read coverage over the tmRNA gene
396 mirrors cleavage observed by Northern blot. **D.** A comparison of genomic synteny of two species within
397 the enterobacteriaceae (*Escherichia coli* MG1655 and *Salmonella enterica* serovar Typhimurium SL1344,
398 top), with synteny between the two Orientia strains from this study (bottom). **E.** Pie charts illustrating the
399 relative abundance of RAGE genes in conserved (top) and strain-specific (bottom) operons. **F.**
400 Visualization of the largest conserved operon in Ot, encoding multiple ribosomal genes, showing RNA-
401 seq coverage in both strains.
402
403 **Figure 3. Antisense transcription is enriched on mobile genetic elements and is predictive of an**
404 **absence of protein expression. A. B.** Plot showing the relationship between protein expression, defined
405 by LFQs, and transcript expression, defined by TPMs. Genes cluster into two groups based on their protein

406 expression. The red line indicates the threshold for expressed genes (TPM value equal to 10). **C.** Sense
407 transcription and the ratio of reads assigned to the antisense and sense strands, showing classification
408 based on proteomics detection. The red line indicates the sense-antisense ratio (1.059182) above which
409 translation was not detected by mass spectrometry. **D.** ROC curves evaluating the performance of logistic
410 regression models to predict protein expression from RNA-seq read counts. Model 1 strictly uses sense
411 expression, Model 2 the antisense-sense ratio, and Model 3 uses both. Incorporating antisense expression
412 clearly improves model performance. **E.** Fraction of core genes and RAGE genes in the set of genes with
413 high antisense-sense ratios, compared to all expressed genes.
414
415 **Figure 4. Differential bacterial gene expression. A.** Heatmap illustrating pathways enriched in
416 differentially expressed genes. All illustrated categories are more highly expressed in Karp. **B.** Volcano
417 plot showing the differential expression of bacterial genes in Karp and UT176. Bacterial surface genes
418 (red) and ankyrin repeat-containing effector proteins (blue) with log fold change ≥1 are highlighted.
419
420 **Figure 5. Ot induces an antiviral interferon response in HUVECs. A.** Summary of the host response
421 showing joint and strain specific responses. The joint response is defined as genes with a logFC > 2 and
422 FDR-corrected p-value < 0.01 for infection with both Karp and UT176. Strain-specific responses are genes
423 with a logFC > 2 and FDR-corrected p-value < 0.01 for infection with either Karp or UT176, excluding
424 genes already included in the joint response. **B.** Activation of multiple genes in the canonical interferon
425 signaling pathway in Karp infected HUVECs compared with uninfected HUVEC cells. **C.** Heat map
426 showing up-regulation of genes in the interferon signaling pathway in HUVEC cells infected with Karp
427 and UT176 compared with uninfected cells.
428

Figure 6. Karp and UT176 lead to the up-regulation of distinct networks in HUVECs. A. Up-regulation of multiple proinflammatory chemokines and cytokines in HUVECs infected with UT176. **B.** Induction of the IL33-FAS-mediated anoikis network in Karp infected HUVECs.

Figure 7. Karp is more virulent than UT176 in a mouse infection model. A. Weight change over 12 days of infection. **B.** Clinical observation score of mice 12 days post-infection. This number is a composite score based on appetite, activity, and hair coat with higher numbers representing low appetite, low activity, and ruffled fur. Details provided in Supp. Fig. 20. **C.** Bacterial copy number in 100 µl blood taken from euthanized mice 12 days post-infection, measured by qPCR. **D.** The ratio of bacterial DNA copy number to mouse DNA copy number in lung, liver, spleen, and kidney of euthanized mice 12 days post-infection, measured by qPCR. **E.** Lesion scores of Hematoxylin and Eosin stained lung, liver, spleen, and kidneys of euthanized mice 12 days post infection. Scores range from 0 to 5 with 0 representing normal tissue and 5 representing severe lesion damage. All graphs show mean and standard deviation. Statistical significance is calculated using unpaired Student t-test in GraphPad Prism software. **p≤0.01 ***p≤0.001 ****p≤0.0001 **F.** Images of Hematoxylin and Eosin stained lung tissue of mice infected with buffer, UT176 or Karp. Scale bars = 50 µm. * indicates airway and ** indicates blood vessel. Uninfected control: airway, blood vessel and alveoli all appear normal. UT176-infected lungs: There are diffuse thickening and infiltration of alveolar septa with a mixed population of macrophages and lymphocytes (arrows). There is also mild perivascular lymphohistiocytic inflammation (open arrow). Karp-infected lungs: There is diffuse moderate thickening and infiltration of alveolar septa with a mixed population of macrophages and lymphocytes. The airway (*) is unaffected and normal. Additional figures are shown in Supp. Fig. 20.

452

Materials and Methods

Growth of Ot and isolation of RNA

The clinical isolate strains (Karp and UT176) of *Orientia tsutsugamushi* were propagated in a confluent monolayer of host cells (HUVEC, Human Umbilical Vein Endothelial Cells; ATCC PCS-100-010) for 5 days at MOI 100:1. Cells were cultured using Media200 (Thermo Fisher, Catalog no. M200-500) supplemented with LVES media (Thermo Fisher, Catalog no. A14608-01) at 35 °C and 5% CO_2. The infectivity was determined by qPCR of the single copy Ot gene *47kDa* at day 5-7. Both uninfected cells and infected cells were harvested by incubating the cells on ice and quickly resuspending in RNAprotect Bacteria Reagent (Qiagen, catalog no. 76506), then storing at -80 °C until use. RNA extraction was performed using the Qiagen RNeasy Plus kit (Qiagen, catalog number 74136) according to manufacturer's instructions and as described previously (Atwal, 2017).

RNA processing and sequencing

The integrity of the DNase-treated RNA samples was assessed in a Bioanalyzer (Agilent). All samples had RIN (RNA integrity number) values ≥8.0. Ribosomal transcripts were removed using the Ribo-Zero Gold (epidemiology) kit (Illumina). Following the manufacturer's instructions, 500 ng of total, DNase-treated RNA was used as an input to the ribo-depletion procedure. rRNA-depleted RNA was precipitated in ethanol for 3 h at -20°C.

cDNA libraries for Illumina sequencing were generated by Vertis Biotechnologie AG, Freising-Weihenstephan, Germany. rRNA-free RNA samples were first sheared via ultrasound sonication (four 30-s pulses at 4°C) to generate on average 200- to 400-nt fragments. Fragments of 20 nt were removed using the Agencourt RNAClean XP kit (Beckman Coulter Genomics) and the Illumina TruSeq adapter was ligated to the 3' ends of the remaining fragments. First-strand cDNA synthesis was performed using M-

476 MLV reverse transcriptase (NEB) wherein the 3' adapter served as a primer. The first-strand cDNA was
477 purified, and the 5' Illumina TruSeq sequencing adapter was ligated to the 3' end of the antisense cDNA.
478 The resulting cDNA was PCR-amplified to about 10 to 20 ng/µl using a high fidelity DNA polymerase.
479 The TruSeq barcode sequences were part of the 5' and 3' TruSeq sequencing adapters. The cDNA library
480 was purified using the Agencourt AMPure XP kit (Beckman Coulter Genomics) and analyzed by capillary
481 electrophoresis (Shimadzu MultiNA microchip).
482 For sequencing, cDNA libraries were pooled in approximately equimolar amounts. The cDNA pool was
483 size fractionated in the size range of 200 to 600 bp using a differential cleanup with the Agencourt AMPure
484 kit (Beckman Coulter Genomics). Aliquots of the cDNA pools were analyzed by capillary electrophoresis
485 (Shimadzu MultiNA microchip). Sequencing was performed on a NextSeq 500 platform (Illumina) at
486 Vertis Biotechnologie AG, Freising-Weihenstephan, Germany (single-end mode; 75 cycles).
487
488 **Northern blots**
489 Each 15 µg of total RNA (i.e. a mixture of human and Ot RNA) prepared as above were loaded per lane
490 and separated in 6% (vol/vol) polyacrylamide–7 M urea gels. Blotting was performed as previously
491 described[24]. After the transfer onto Hybond XL membranes (Amersham), RNA was cross-linked with UV
492 light and hybridized at 42°C with gene-specific 32P-end-labeled DNA oligonucleotides (Supp. Fig. 13)
493 in Hybri-Quick buffer (Carl Roth AG). After exposure, the screens were read out on a Typhoon FLA 7000
494 phosphorimager (GE Healthcare).
495
496 **qRT-PCR**
497 qRT-PCR was performed with the Power SYBR Green RNA-to-CT1-Step kit (Applied Biosystems)
498 according to the manufacturer's instructions and a CFX96 Touch real-time PCR detection system (Bio-

499 Rad). Human U6 snRNA served as reference transcripts. Fold changes in expression were determined
500 using the $2^{(-\Delta\Delta Ct)}$ method [76]. Primer sequences are given in Supp. Fig. 13, and their specificity had been
501 confirmed using Primer-BLAST (NCBI).
502
503 **RNA seq read processing and quantification**
504 The raw reads were initially processed according to our established dual RNA-seq pipeline[24]. Briefly, raw
505 reads were trimmed for adaptor sequences and a minimum read quality of 20 using cutadapt[77]. Reads were
506 then mapped against the human (GRCh38) and Ot (UT176 accession: LS398547.1; Karp accession:
507 LS398548.1) reference sequences using the READemption pipeline (v0.4.3, [78]) and segemehl with the
508 lack remapper (v0.2.0 [79]), removing reads that mapped equally well to the bacterial and host genomes. For
509 downstream analysis of human gene expression, only uniquely mapping reads were retained for
510 quantification.
511 To improve quantification of repetitive sequences, reads mapped to the Ot genomes were used for
512 quantification of bacterial transcript expression using Salmon (v0.9.1) [29]. Salmon is a quasi-mapping
513 based gene expression quantification tool that consists of two steps, indexing and quantification.
514 Transcript fasta files were created from the Genbank annotations using the gene coordinates. The indexing
515 step was performed in quasi-mapping mode (--type quasi). Expression of the transcripts was quantified
516 using both stranded forward library type (-lSF) and removing incompatible mappings (--incompatPrior
517 0.0). Salmon identified identical gene repeats that are collected in 218 groups (see supplementary table,
518 Karp groups of duplicates). For quantification purposes we retained a single gene from each group.
519 For the purposes of summarizing gene expression, we calculated mean TPM values from three replicates
520 for each strain. Genes with a mean TPM greater than 10 were classified as expressed, and those with a
521 mean TPM value greater than 50 highly expressed.

522

523 **Gene annotation**

524 For each gene we retrieved the gene name, gene product, and amino acid sequence from the Genbank
525 annotation. In addition, using eggNOG-mapper[10] we predicted gene names and both KEGG pathways [11]
526 and GO terms. We manually identified surface antigen encoding proteins using BLAST. The KEGGREST
527 (Tenenbaum, D (2019) KEGGREST: Client-side REST access to KEGG. R package version 1.18.1.) and
528 GO.db (Carlson M (2019). GO.db: A set of annotation maps describing the entire Gene Ontology. R
529 package version 3.5.0.) R packages were used to retrieve KEGG and GO terms, respectively.

530

531 **Non-coding RNA prediction**

532 Noncoding RNAs were annotated using Rockhopper[76], ANNOgesic [77] (v0.7.17) and Infernal[78] (v1.1.2)
533 searching sequences against the Rfam database[79]. These provided inconsistent predictions of intergenic
534 sRNAs. Intergenic sRNAs were manually curated by visual comparison of the predicted sRNA
535 coordinates with the read coverage in the Integrative Genomics Viewer [80] (v2.5.2). Infernal predicted the
536 core housekeeping ncRNAs tmRNA, RNaseP, SRP and 5S rRNA. The quantification of the bacterial
537 transcriptomes complemented with predicted sRNAs was performed using Salmon.

538

539 **Genomic alignment**

540 Genomic comparisons in Figure panels D and F were performed using Easyfig[81]. Escherichia coli K-12
541 MG1655 (Accession number U00096) and Salmonella enterica serovar Typhimurium SL1344 (Accession
542 number FQ312003) were used as comparators for synteny analysis.

543

544 **Orthology and conserved operon prediction**

545 We predicted orthologous genes between the two Orientia strains using Poff[82] with default parameters in
546 synteny mode. To identify conserved operons we used operon structures predicted in each strain by
547 Rockhopper [41]. Based on visual analysis of read coverage in the Integrative Genomics Viewer, some of
548 the operons were manually extended by addition of genes or merging two operons into one. We also
549 identified partially conserved operons missing some genes in one strain.

550

551 **Differential gene expression**

552 For the bacteria, differential gene expression analysis was performed between orthologous genes
553 identified by Poff. Genes that were predicted as an orthologous group (more than two genes) were
554 removed from the analysis. Additionally, we removed duplicates (transcripts with perfectly identical
555 sequence) that were identified by Salmon in either strain. For both human and bacterial RNA-seq data, we
556 performed differential gene expression analysis with the edgeR package [83] (v3.20.9) using robust quasi-
557 likelihood estimation [84], including genes with CPM (Counts Per Million) > 10 (for Ot) or CPM > 1 (for
558 HUVEC) in at least 3 libraries. To identify biological processes that differ between two Orientia strains,
559 we have performed gene set analysis using KEGG and GO terms that contain at least 4 expressed genes
560 using the fry test in the edgeR package.

561

562 **Proteomic sample preparation**

563 Bacteria were propagated in HUVEC cell line at MOI 100:1 and harvested at 5 dpi. Ot was isolated,
564 washed with 0.3 M sucrose, and lysed with 1% triton-X prior to acetone precipitation of protein. Total

565 protein was then alkylated, reduced, and subsequently treated with Lys-C/Trypsin. Digested peptides were
566 desalted using Oasis® HLB reversed-phase cartridges, vacuum dried, and stored for MS runs.
567
568 **Mass spectrometry**
569 The dried samples were resuspended of 2% (v/v) acetonitrile solution containing 0.06% (v/v)
570 trifluoroacetic acid and 0.5% (v/v) acetic acid and loaded onto an autosampler plate. Online
571 chromatography was performed using EASY-nLC 1000 (Thermo Scientific) in single-column setup using
572 0.1% formic acid in water and 0.1% formic acid in acetonitrile as mobile phases. using reversed-phase
573 C18 column (EASY-Spray LC Column, 75 µm inner diameter x 50 cm, 2 µm particle size) (Thermo
574 Scientific). The samples were injected and separated on the analytical column maintained at 50 °C using
575 a 2–23% (v/v) acetonitrile gradient over 60 min, then ramped to 50% over the next 20 min, and finally to
576 90% within 5 min. The final mixture was maintained for 5 min to elute all remaining peptides. Total run
577 duration for each sample was 90 min at a constant flow rate of 300 nl/min.
578 Data was acquired using an Orbitrap Fusion mass spectrometer (Thermo Scientific) in data-dependent
579 mode. Samples were ionized 2.5 kV and 300 °C at the nanospray source and positively-charged precursor
580 MS1 signals were detected using an Orbitrap analyzer set to 60,000 resolution, automatic gain control
581 (AGC) target of 400,000 ions, and maximum injection time (IT) of 50 ms. Precursors with charges 2–7
582 and having the highest ion counts in each MS1 scan were further fragmented using collision-induced
583 dissociation (CID) at 35% normalized collision energy and their MS2 signals were analyzed by ion trap
584 at an AGC of 10,000 and maximum IT of 35 ms. Precursors used for MS2 scans were excluded for 90 s
585 in order to avoid re-sampling of high abundance peptides. The MS1-MS2 cycles were repeated every 3 s
586 until completion of the run.

587 Identification of proteins within each sample was performed using MaxQuant (v1.5.5.1). Raw mass
588 spectra were searched against Orientia tsutsugamushi primary protein sequences derived from complete
589 genome data. Human whole proteome sequences were obtained from Uniprot and included as background.
590 Carbamidomethylation on Cys was set as the fixed modification and acetylation on protein N-terminus
591 and oxidation of Met were set as dynamic modifications for the search. Trypsin was set as the digestion
592 enzyme and was allowed up to 3 missed cleavage sites. Precursors and fragments were accepted if they
593 had a mass error within 20 ppm. Peptides were matched to spectra at a false discovery rate (FDR) of 1%
594 against the decoy database.
595
596 **Proteomic data analysis**
597 Protein expression was measured by label free quantification values (LFQs). A protein was classified as
598 detected if at least two peptides were detected in at least 2 biological replicates, and the mean LFQ across
599 the three replicates was used for further analysis. Otherwise the protein was classified as undetected, and
600 the LFQ value was set to zero. The proteomic data includes 19 protein groups that couldn't be resolved,
601 consisting of 93 proteins. In our analysis we discarded these proteins to simplify the analysis.
602
603 **Transcript classification**
604 Sense transcript expression was defined by mean TPM value across replicates. The antisense/sense ratio
605 was calculated as the ratio of mean read counts assigned to the antisense and sense strand of coding
606 annotations. The duplicated sequences identified by Salmon (Supp. Table 1. Karp groups of duplicates)
607 and non-coding RNAs were removed from the analysis.
608

We divided the data set into two classes, detected and undetected in proteomics. Within our analyzed data set, 322 genes were detected, whereas 1608 genes were not detected by mass spectrometry. We found a weak positive correlation between TPMs and LFQs for genes with detected proteins (Spearman's correlation coefficient equal to 0.32), but it was not a linear association (Pearson's correlation coefficient equal to 0.04). For the further analysis we selected transcripts with sense expression greater than 10 TPMs, previously defined as our expression threshold.

Logistic Regression model

To test whether antisense-sense ratios are predictive of protein expression, we have applied logistic regression, which models the probability of a binary response, that is, whether a protein is expressed or not. We have built 3 competing models. Model 1 makes predictions of the protein expression based solely on sense transcription:

$\beta_0 + \beta_1$* (TPM sense)

Model 2 makes predictions solely on the antisense-sense ratio:

$\beta_0 + \beta_1$* ((number of antisense reads) / (number of sense reads))

Model 3 uses both sense transcription and the antisense-sense ratio to make predictions:

$\beta_0 + \beta_1$* (TPM sense) + β_2* ((number of antisense reads) / (number of sense reads))

Since the data is highly imbalanced, 295 transcripts with detected proteins, and 814 without, we used a downsampling procedure (downSample function) implemented in the caret R package[85] to create a balanced data set for model training purposes. Next, the function glm() with a logit link function from the caret package was used to fit models to the reduced data set. For a first indication as to whether any of these models are predictive, we trained all three models on a downsampled data set consisting of 590 genes, then tested them on the complete data set. To more rigorously assess this result, we have applied 500-fold cross validation. For each fold the data was split randomly into 2 data sets, training and testing which included 1055 and 54 genes, respectively. Each time the new training data set was reduced to 562 genes, which were used to estimate the model parameters, and then the model was evaluated on the testing data set. The model performance was evaluated using a variety of measures, i.e. sensitivity, specificity, accuracy (caret R package) as well as with ROC curves [86] (pROC 1.14.0) and the area under the ROC curve (AUC).

Immunofluorescence microscopy

The protocol for L-Homopropargylglycine (HPG) incorporation, click chemistry and fluorescence detection were based on recommendations from Click-iT® HPG Alexa Fluor® Protein Synthesis Assay Kits (Molecular probe by Life Technologies). HUVECs were grown on chambered coverslip slides (Ibidi, USA), for 2 days before infection with bacteria at MOI 100:1. To incorporate HPG at times indicated, medium was removed and replaced with L-methionine-free medium (Dulbecco's Modified Eagle Medium, DMEM, Cat. no. 21013) containing 25 µM HPG for 30 min at 37 ºC. Labeled bacteria were washed twice in 1X PBS+ 1mg/ml BSA, pH 7.4 before fixing with 4% formaldehyde and subsequently being permeabilized with 0.5% TritonX for 20 min on ice. After washing with PBS + 1 mg/ml BSA, the Click-iT® reaction cocktail (Click-iT® HPG Alexa Fluor® Protein Synthesis Assay Kits cat. C10428)

was incubated with cells for 30 min at room temperature in the dark. The Azide dye (Alexa Fluor®488) was used at a final concentration of 5 μM. After the click reaction, cells were labeled with the actin probe Alexa Fluor® 594 phalloidin at a dilution of 1:40 and the nuclear stain Hoechst diluted to 1:1000 for 30 min at 37 °C. Cells were washed 3X with PBS which was replaced with mounting media after the final wash. Imaging was performed using a Zeiss LSM 7000 equipped with a 63 × 1.4 NA objective lens (Carl Zeiss, USA) and also a Leica SP8 laser scanning confocal microscope.

Analysis of codon bias

We calculated the RSCU (relative synonymous codon usage) for each codon to quantify genome-wide or gene-specific codon usage bias following[52]. To determine the genomic codon counts for each species and gene set, we parsed nucleotide sequence data and annotation in the GenBank file format, downloaded from the NCBI database. We also obtained tRNA gene copy numbers from the GtRNAdb database[87,88], and integrated protein abundance for E. coli K-12 MG1655 data from PaxDB 89.

Host Network/Pathway analysis

To identify pathways that are affected in Karp and/or UT176 infected host cells, genes differentially expressed with an adjusted p-value of < 0.05 were analyzed using Ingenuity Pathway Analysis (IPA) software (Ingenuity® Systems, Inc. Redwood City, CA)[81] as described previously[82]. Selected pathways were chosen based on enrichment p-values and activation Z-scores, and served as the basis for Figs 5, 6, and Supp. Fig. 11, 14, 15, 16, and 17.

Mice and Ethics statement

All animal research was performed strictly under approved Institutional Animal Care and Use Committee (IACUC) protocol by the IACUC and Biosafety Review Committee at the Armed Forces Research Institute of Medical Sciences (AFRIMS) Bangkok, Thailand, an AAALAC International-accredited facility. The protocol number was PN16-05. The animal research was conducted in compliance with Thai laws, the Animal Welfare Act, and all applicable U.S. Department of Agriculture, Office of Laboratory Animal Welfare and U.S. Department of Defense guidelines. All animal research adhered to the Guide for the Care and Use of Laboratory Animals, NRC Publication (8^{th} Edition).

C57BL/6NJcl mice were purchased from Nomura Siam International, Bangkok, Thailand. Mice were housed in an animal biosafety level 2 facility and moved to an animal biosafety level 3 containment 2 days before the inoculation. Mice at 6- 8 weeks of age were used in these experiments. Two group of mice (n =8 per group) were intravenously injected in the tail vein with $1.25x10^6$ genome copies of *O. tsutsugamushi* of either Karp strain or UT176 strain. The *O. tsutsugamushi* inoculum was derived from *O. tsutsugamushi*-infected L929 cells. Clinical signs and body weight were evaluated daily. After 12 days post inoculation, all mice were euthanized. Blood and tissue samples including lungs, liver, spleen, and kidneys were collected for bacteria quantification and histopathology.

Acknowledgements

JS is funded by a Royal Society Dorothy Hodgkin Research Fellowship (DH140154) and this project was additionally funded by a grant from the University of Oxford Medical Sciences Division Medical Research Fund. The authors are grateful to Guy Riddihough from Life Science Editors for editorial support on this manuscript. The authors would like to thank Sandy Pernitzsch/Scigraphix for assistance with Figures 5 and 6.

Author contributions

Conceptualization, J.V., L.B., J.S.; Investigation, B.M.G., S.G., A.J.W., J.W., S. C., L.C.W., P.S., L.B.; Formal Analysis, B.M.G., W.K.C., S.S., R.S., L.B.; Resources, R.S., J.V., L.B., J.S.; Writing – Original Draft, L.B., J.S.; Writing – Review and Editing, A.J.W., L.B., J.S.; Supervision, L.B., J.S.; Funding Acquisition, J.V., J.S.

Competing interests

The authors declare no competing interests.

References

1. Luce-Fedrow, A., Lehman, M., Kelly, D., Mullins, K., Maina, A., Stewart, R., Ge, H., John, H., Jiang, J. & Richards, A. A Review of Scrub Typhus (Orientia tsutsugamushi and Related Organisms): Then, Now, and Tomorrow. *Tropical Medicine and Infectious Disease* **3**, 8 (2018).
2. Izzard, L., Fuller, A., Blacksell, S. D., Paris, D. H., Richards, A. L., Aukkanit, N., Nguyen, C., Jiang, J., Fenwick, S., Day, N. P., Graves, S. & Stenos, J. Isolation of a novel Orientia species (O. chuto sp. nov.) from a patient infected in Dubai. *J Clin Microbiol* **48**, 4404-4409 (2010).
3. Weitzel, T., Dittrich, S., López, J., Phuklia, W., Martinez-Valdebenito, C., Velásquez, K., Blacksell, S. D., Paris, D. H. & Abarca, K. Endemic Scrub Typhus in South America. *N Engl J Med* **375**, 954-961 (2016).
4. Paris, D. H., Phetsouvanh, R., Tanganuchitcharnchai, A., Jones, M., Jenjaroen, K., Vongsouvath, M., Ferguson, D. P., Blacksell, S. D., Newton, P. N., Day, N. P. & Turner, G. D. Orientia tsutsugamushi in human scrub typhus eschars shows tropism for dendritic cells and monocytes rather than endothelium. *PLoS Negl Trop Dis* **6**, e1466 (2012).
5. Moron, C. G., Popov, V. L., Feng, H. M., Wear, D. & Walker, D. H. Identification of the target cells of Orientia tsutsugamushi in human cases of scrub typhus. *Mod Pathol* **14**, 752-759 (2001).
6. Cho, N. H., Kim, H. R., Lee, J. H. & Kim, I. S. The Orientia tsutsugamushi genome reveals massive proliferation of conjugative type IV secretion system and host– cell interaction genes. *PNAS* **104**, 7981-7986 (2007).
7. Nakayama, K., Yamashita, A., Kurokawa, K., Morimoto, T., Ogawa, M., Fukuhara, M., Urakami, H., Ohnishi, M., Uchiyama, I., Ogura, Y., Ooka, T., Oshima, K., Tamura, A., Hattori, M. & Hayashi, T. The Whole-genome sequencing of the obligate intracellular bacterium Orientia tsutsugamushi revealed massive gene amplification during reductive genome evolution. *DNA Res* **15**, 185-199 (2008).
8. Batty EM, C. S., Blacksell SB, Richards A, Paris D, Bowden R, Chan C, Lachumanan R, Day N, Donnelly P, Chen SL, Salje J. Long-read whole genome sequencing and comparative analysis of six strains of the human pathogen *Orientia tsutsugamushi*. *Plos Negl Trop Dis* (2018).
9. Kelly, D. J., Fuerst, P. A., Ching, W. M. & Richards, A. L. Scrub typhus: the geographic distribution of phenotypic and genotypic variants of Orientia tsutsugamushi. *Clin Infect Dis* **48 Suppl 3**, S203-30 (2009).
10. Premaratna, R., Blanton, L. S., Samaraweera, D. N., de Silva, G. N., Chandrasena, N. T., Walker, D. H. & de Silva, H. J. Genotypic characterization of Orientia tsutsugamushi from patients in two geographical locations in Sri Lanka. *BMC Infect Dis* **17**, 67 (2017).
11. Varghese, G. M., Janardhanan, J., Mahajan, S. K., Tariang, D., Trowbridge, P., Prakash, J. A., David, T., Sathendra, S. & Abraham, O. C. Molecular Epidemiology and Genetic Diversity of Orientia tsutsugamushi from Patients with Scrub Typhus in 3 Regions of India. *Emerg Infect Dis* **21**, 64-69 (2015).
12. James, S. L., Blacksell, S. D., Nawtaisong, P., Tanganuchitcharnchai, A., Smith, D. J., Day, N. P. & Paris, D. H. Antigenic Relationships among Human Pathogenic Orientia tsutsugamushi Isolates from Thailand. *PLoS Negl Trop Dis* **10**, e0004723 (2016).
13. Cho, N., others & Kim, I. S. Expression of Chemokine Genes in Murine Macrophages Infected with Orientia tsutsugamushi. *infection and Immunity* (2000).
14. Cho, N. H., Seong, S. Y., Choi, M. S. & Kim, I. S. Expression of chemokine genes in human dermal microvascular endothelial cell lines infected with Orientia tsutsugamushi. *Infect Immun* **69**, 1265-1272 (2001).

15. Koh, Y. S., Yun, J. H., Seong, S. Y., Choi, M. S. & Kim, I. S. Chemokine and cytokine production during Orientia tsutsugamushi infection in mice. *Microbial pathogenesis* (2004).
16. Tantibhedhyangkul, W., Prachason, T., Waywa, D., El Filali, A., Ghigo, E., Thongnoppakhun, W., Raoult, D., Suputtamongkol, Y., Capo, C., Limwongse, C. & Mege, J. L. Orientia tsutsugamushi stimulates an original gene expression program in monocytes: relationship with gene expression in patients with scrub typhus. *PLoS Negl Trop Dis* **5**, e1028 (2011).
17. Tantibhedhyangkul, W., Ben Amara, A., Textoris, J., Gorvel, L., Ghigo, E., Capo, C. & Mege, J. L. Orientia tsutsugamushi, the causative agent of scrub typhus, induces an inflammatory program in human macrophages. *Microb Pathog* **55**, 55-63 (2013).
18. Astrup, E., Janardhanan, J., Otterdal, K., Ueland, T., Prakash, J. A., Lekva, T., Strand, O. A., Abraham, O. C., Thomas, K., Damas, J. K., Mathews, P., Mathai, D., Aukrust, P. & Varghese, G. M. Cytokine network in scrub typhus: high levels of interleukin-8 are associated with disease severity and mortality. *PLoS Negl Trop Dis* **8**, e2648 (2014).
19. Fukuhara, M., Fukazawa, M., Tamura, A., Nakamura, T. & Urakami, H. Survival of two Orientia tsutsugamushi bacterial strains that infect mouse macrophages with varying degrees of virulence. *Microb Pathog* **39**, 177-187 (2005).
20. Hanson, B. Comparative susceptibility to mouse interferons of Rickettsia tsutsugamushi strains with different virulence in mice and of Rickettsia rickettsii. *Infect Immun* **59**, 4134-4141 (1991).
21. Sunyakumthorn, P., Paris, D. H., Chan, T. C., Jones, M., Luce-Fedrow, A., Chattopadhyay, S., Jiang, J., Anantatat, T., Turner, G. D., Day, N. P. & Richards, A. L. An intradermal inoculation model of scrub typhus in Swiss CD-1 mice demonstrates more rapid dissemination of virulent strains of Orientia tsutsugamushi. *PLoS One* **8**, e54570 (2013).
22. Groves, M. G. & Kelly, D. J. Characterization of factors determining Rickettsia tsutsugamushi pathogenicity for mice. *Infect Immun* **57**, 1476-1482 (1989).
23. Saliba, A. E., C Santos, S. & Vogel, J. New RNA-seq approaches for the study of bacterial pathogens. *Curr Opin Microbiol* **35**, 78-87 (2017).
24. Westermann, A. J., Förstner, K. U., Amman, F., Barquist, L., Chao, Y., Schulte, L. N., Müller, L., Reinhardt, R., Stadler, P. F. & Vogel, J. Dual RNA-seq unveils noncoding RNA functions in host-pathogen interactions. *Nature* (2016).
25. Westermann, A. J., Barquist, L. & Vogel, J. Resolving host-pathogen interactions by dual RNA-seq. *PLOS Pathog* **13**, (2017).
26. Humphrys, M. S., Creasy, T., Sun, Y., Shetty, A. C., Chibucos, M. C., Drabek, E. F., Fraser, C. M., Farooq, U., Sengamalay, N., Ott, S., Shou, H., Bavoil, P. M., Mahurkar, A. & Myers, G. S. Simultaneous transcriptional profiling of bacteria and their host cells. *PLoS One* **8**, e80597 (2013).
27. Enatsu, T., Urakami, H. & Tamura, A. Phylogenetic analysis of Orientia tsutsugamushi strains based on the sequence homologies of 56-kDa type-specific antigen genes. *FEMS Microbiol Lett* **180**, 163-169 (1999).
28. Paris, D. H., Aukkanit, N., Jenjaroen, K., Blacksell, S. D. & Day, N. P. A highly sensitive quantitative real-time PCR assay based on the groEL gene of contemporary Thai strains of Orientia tsutsugamushi. *Clin Microbiol Infect* **15**, 488-495 (2009).
29. Patro, R., Duggal, G., Love, M. I., Irizarry, R. A. & Kingsford, C. Salmon provides fast and bias-aware quantification of transcript expression. *Nat Methods* **14**, 417-419 (2017).
30. Withey, J. H. & Friedman, D. I. A salvage pathway for protein structures: tmRNA and trans-translation. *Annu Rev Microbiol* **57**, 101-123 (2003).
31. Julio, S. M., Heithoff, D. M. & Mahan, M. J. ssrA (tmRNA) plays a role in Salmonella enterica serovar Typhimurium pathogenesis. *J Bacteriol* **182**, 1558-1563 (2000).

32. Svetlanov, A., Puri, N., Mena, P., Koller, A. & Karzai, A. W. Francisella tularensis tmRNA system mutants are vulnerable to stress, avirulent in mice, and provide effective immune protection. *Mol Microbiol* **85**, 122-141 (2012).
33. Mao, C., Bhardwaj, K., Sharkady, S. M., Fish, R. I., Driscoll, T., Wower, J., Zwieb, C., Sobral, B. W. & Williams, K. P. Variations on the tmRNA gene. *RNA Biol* **6**, 355-361 (2009).
34. Keiler, K. C., Shapiro, L. & Williams, K. P. tmRNAs that encode proteolysis-inducing tags are found in all known bacterial genomes: A two-piece tmRNA functions in Caulobacter. *Proc Natl Acad Sci U S A* **97**, 7778-7783 (2000).
35. Gaudin, C., Zhou, X., Williams, K. P. & Felden, B. Two-piece tmRNA in cyanobacteria and its structural analysis. *Nucleic Acids Res* **30**, 2018-2024 (2002).
36. Sharkady, S. M. & Williams, K. P. A third lineage with two-piece tmRNA. *Nucleic Acids Res* **32**, 4531-4538 (2004).
37. Storz, G., Vogel, J. & Wassarman, K. M. Regulation by small RNAs in bacteria: expanding frontiers. *Mol Cell* **43**, 880-891 (2011).
38. Wagner, E. G. H. & Romby, P. Small RNAs in bacteria and archaea: who they are, what they do, and how they do it. *Adv Genet* **90**, 133-208 (2015).
39. Albrecht, M., Sharma, C. M., Dittrich, M. T., Müller, T., Reinhardt, R., Vogel, J. & Rudel, T. The transcriptional landscape of Chlamydia pneumoniae. *Genome Biol* **12**, R98 (2011).
40. Dugar, G., Herbig, A., Förstner, K. U., Heidrich, N., Reinhardt, R., Nieselt, K. & Sharma, C. M. High-resolution transcriptome maps reveal strain-specific regulatory features of multiple Campylobacter jejuni isolates. *PLoS Genet* **9**, e1003495 (2013).
41. Kröger, C., Dillon, S. C., Cameron, A. D., Papenfort, K., Sivasankaran, S. K., Hokamp, K., Chao, Y., Sittka, A., Hébrard, M., Händler, K., Colgan, A., Leekitcharoenphon, P., Langridge, G. C., Lohan, A. J., Loftus, B., Lucchini, S., Ussery, D. W., Dorman, C. J., Thomson, N. R., Vogel, J. & Hinton, J. C. The transcriptional landscape and small RNAs of Salmonella enterica serovar Typhimurium. *Proc Natl Acad Sci U S A* **109**, E1277-86 (2012).
42. Sharma, C. M., Hoffmann, S., Darfeuille, F., Reignier, J., Findeiss, S., Sittka, A., Chabas, S., Reiche, K., Hackermüller, J., Reinhardt, R., Stadler, P. F. & Vogel, J. The primary transcriptome of the major human pathogen Helicobacter pylori. *Nature* **464**, 250-255 (2010).
43. Toledo-Arana, A., Dussurget, O., Nikitas, G., Sesto, N., Guet-Revillet, H., Balestrino, D., Loh, E., Gripenland, J., Tiensuu, T., Vaitkevicius, K., Barthelemy, M., Vergassola, M., Nahori, M. A., Soubigou, G., Régnault, B., Coppée, J. Y., Lecuit, M., Johansson, J. & Cossart, P. The Listeria transcriptional landscape from saprophytism to virulence. *Nature* **459**, 950-956 (2009).
44. Vogel, J., Bartels, V., Tang, T. H., Churakov, G., Slagter-Jäger, J. G., Hüttenhofer, A. & Wagner, E. G. RNomics in Escherichia coli detects new sRNA species and indicates parallel transcriptional output in bacteria. *Nucleic Acids Res* **31**, 6435-6443 (2003).
45. Tjaden, B. De novo assembly of bacterial transcriptomes from RNA-seq data. *Genome Biol* **16**, 1 (2015).
46. Andersson, S. G., Zomorodipour, A., Andersson, J. O., Sicheritz-Ponten, T., Alsmark, U. C., Podowski, R. M., Naslund, A. K., Eriksson, A. S., Winkler, H. H. & Kurland, C. G. The genome sequence of Rickettsia prowazekii and the origin of mitochondria. *Nature* **396**, 133-140 (1998).
47. Georg, J. & Hess, W. R. Widespread Antisense Transcription in Prokaryotes. *Microbiol Spectr* **6**, (2018).
48. Wade, J. T. & Grainger, D. C. Pervasive transcription: illuminating the dark matter of bacterial transcriptomes. *Nat Rev Microbiol* **12**, 647-653 (2014).

49. Raghavan, R., Sloan, D. B. & Ochman, H. Antisense transcription is pervasive but rarely conserved in enteric bacteria. *MBio* **3**, (2012).
50. Lloréns-Rico, V., Cano, J., Kamminga, T., Gil, R., Latorre, A., Chen, W. H., Bork, P., Glass, J. I., Serrano, L. & Lluch-Senar, M. Bacterial antisense RNAs are mainly the product of transcriptional noise. *Sci Adv* **2**, e1501363 (2016).
51. Andersson, S. G. & Sharp, P. M. Codon usage and base composition in Rickettsia prowazekii. *J Mol Evol* **42**, 525-536 (1996).
52. Botzman, M. & Margalit, H. Variation in global codon usage bias among prokaryotic organisms is associated with their lifestyles. *Genome Biol* **12**, R109 (2011).
53. Sharp, P. M., Bailes, E., Grocock, R. J., Peden, J. F. & Sockett, R. E. Variation in the strength of selected codon usage bias among bacteria. *Nucleic Acids Res* **33**, 1141-1153 (2005).
54. Duret, L. tRNA gene number and codon usage in the C. elegans genome are co-adapted for optimal translation of highly expressed genes. *Trends Genet* **16**, 287-289 (2000).
55. Ikemura, T. Correlation between the abundance of Escherichia coli transfer RNAs and the occurrence of the respective codons in its protein genes: a proposal for a synonymous codon choice that is optimal for the E. coli translational system. *J Mol Biol* **151**, 389-409 (1981).
56. Ikemura, T. Correlation between the abundance of yeast transfer RNAs and the occurrence of the respective codons in protein genes. Differences in synonymous codon choice patterns of yeast and Escherichia coli with reference to the abundance of isoaccepting transfer RNAs. *J Mol Biol* **158**, 573-597 (1982).
57. Plotkin, J. B. & Kudla, G. Synonymous but not the same: the causes and consequences of codon bias. *Nat Rev Genet* **12**, 32-42 (2011).
58. Frumkin, I., Lajoie, M. J., Gregg, C. J., Hornung, G., Church, G. M. & Pilpel, Y. Codon usage of highly expressed genes affects proteome-wide translation efficiency. *Proc Natl Acad Sci U S A* **115**, E4940-E4949 (2018).
59. Goodman, D. B., Church, G. M. & Kosuri, S. Causes and effects of N-terminal codon bias in bacterial genes. *Science* **342**, 475-479 (2013).
60. Ha, N. Y., Kim, Y., Choi, J. H., Choi, M. S., Kim, I. S., Kim, Y. S. & Cho, N. H. Detection of antibodies against Orientia tsutsugamushi Sca proteins in scrub typhus patients and genetic variation of sca genes of different strains. *Clin Vaccine Immunol* **19**, 1442-1451 (2012).
61. Ha, N. Y., Sharma, P., Kim, G., Kim, Y., Min, C. K., Choi, M. S., Kim, I. S. & Cho, N. H. Immunization with an Autotransporter Protein of Orientia tsutsugamushi Provides Protective Immunity against Scrub Typhus. *PLoS Negl Trop Dis* **9**, e0003585 (2015).
62. Ha, N. Y., Shin, H. M., Sharma, P., Cho, H. A., Min, C. K., Kim, H. I., Yen, N. T., Kang, J. S., Kim, I. S., Choi, M. S., Kim, Y. K. & Cho, N. H. Generation of protective immunity against Orientia tsutsugamushi infection by immunization with a zinc oxide nanoparticle combined with ScaA antigen. *J Nanobiotechnology* **14**, 76 (2016).
63. VieBrock, L., Evans, S. M., Beyer, A. R., Larson, C. L., Beare, P. A., Ge, H., Singh, S., Rodino, K. G., Heinzen, R. A., Richards, A. L. & Carlyon, J. A. Orientia tsutsugamushi ankyrin repeat-containing protein family members are Type 1 secretion system substrates that traffic to the host cell endoplasmic reticulum. *Front Cell Infect Microbiol* **4**, 186 (2014).
64. Evans, S. M., Rodino, K. G., Adcox, H. E. & Carlyon, J. A. Orientia tsutsugamushi uses two Ank effectors to modulate NF-κB p65 nuclear transport and inhibit NF-κB transcriptional activation. *PLoS Pathog* **14**, e1007023 (2018).

65. Cho, K. A., Jun, Y. H., Suh, J. W., Kang, J. S., Choi, H. J. & Woo, S. Y. Orientia tsutsugamushi induced endothelial cell activation via the NOD1-IL-32 pathway. *Microb Pathog* **49**, 95-104 (2010).
66. Gharaibeh, M., Hagedorn, M., Lilla, S., Hauptmann, M., Heine, H., Fleischer, B. & Keller, C. Toll-Like Receptor 2 Recognizes Orientia tsutsugamushi and Increases Susceptibility to Murine Experimental Scrub Typhus. *Infect Immun* **84**, 3379-3387 (2016).
67. Kawai, T. & Akira, S. Signaling to NF-kappaB by Toll-like receptors. *Trends Mol Med* **13**, 460-469 (2007).
68. Shelite, T. R., Liang, Y., Wang, H., Mendell, N. L., Trent, B. J., Sun, J., Gong, B., Xu, G., Hu, H., Bouyer, D. H. & Soong, L. IL-33-Dependent Endothelial Activation Contributes to Apoptosis and Renal Injury in Orientia tsutsugamushi-Infected Mice. *PLoS Negl Trop Dis* **10**, e0004467 (2016).
69. McLeod, M. P., Qin, X., Karpathy, S. E., Gioia, J., Highlander, S. K., Fox, G. E., McNeill, T. Z., Jiang, H., Muzny, D., Jacob, L. S., Hawes, A. C., Sodergren, E., Gill, R., Hume, J., Morgan, M., Fan, G., Amin, A. G., Gibbs, R. A., Hong, C., Yu, X. J., Walker, D. H. & Weinstock, G. M. Complete genome sequence of Rickettsia typhi and comparison with sequences of other rickettsiae. *J Bacteriol* **186**, 5842-5855 (2004).
70. Yildirim, A. & Feig, M. High-resolution 3D models of Caulobacter crescentus chromosome reveal genome structural variability and organization. *Nucleic Acids Res* **46**, 3937-3952 (2018).
71. Ellis, M. J. & Haniford, D. B. Riboregulation of bacterial and archaeal transposition. *Wiley Interdiscip Rev RNA* **7**, 382-398 (2016).
72. Jose, B. R., Gardner, P. P. & Barquist, L. Transcriptional noise and exaptation as sources for bacterial sRNAs. *Biochem Soc Trans* **47**, 527-539 (2019).
73. Narra, H. P., Schroeder, C. L., Sahni, A., Rojas, M., Khanipov, K., Fofanov, Y. & Sahni, S. K. Small Regulatory RNAs of Rickettsia conorii. *Sci Rep* **6**, 36728 (2016).
74. Schroeder, C. L., Narra, H. P., Rojas, M., Sahni, A., Patel, J., Khanipov, K., Wood, T. G., Fofanov, Y. & Sahni, S. K. Bacterial small RNAs in the Genus Rickettsia. *BMC Genomics* **16**, 1075 (2015).
75. Schroeder, C. L., Narra, H. P., Sahni, A., Rojas, M., Khanipov, K., Patel, J., Shah, R., Fofanov, Y. & Sahni, S. K. Identification and Characterization of Novel Small RNAs in Rickettsia prowazekii. *Front Microbiol* **7**, 859 (2016).
76. Livak, K. J. & Schmittgen, T. D. Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. *Methods* **25**, 402-408 (2001).
77. Martin, M. Cutadapt Removes Adapter Sequences From High-throughput Sequencing Reads. *EMBnet.journal* **17**, 10-12 (2011).
78. Förstner, K. U., Vogel, J. & Sharma, C. M. READemption-a tool for the computational analysis of deep-sequencing-based transcriptome data. *Bioinformatics* **30**, 3421-3423 (2014).
79. Otto, C., Stadler, P. F. & Hoffmann, S. Lacking alignments? The next-generation sequencing mapper segemehl revisited. *Bioinformatics* **30**, 1837-1843 (2014).
80. Huerta-Cepas, J., Szklarczyk, D., Forslund, K., Cook, H., Heller, D., Walter, M. C., Rattei, T., Mende, D. R., Sunagawa, S., Kuhn, M., Jensen, L. J., von Mering, C. & Bork, P. eggNOG 4.5: a hierarchical orthology framework with improved functional annotations for eukaryotic, prokaryotic and viral sequences. *Nucleic Acids Res* **44**, D286-93 (2016).
81. Krämer, A., Green, J., Pollard, J. & Tugendreich, S. Causal analysis approaches in Ingenuity Pathway Analysis. *Bioinformatics* **30**, 523-530 (2014).
82. Subbian, S., Bandyopadhyay, N., Tsenova, L., O'Brien, P., Khetani, V., Kushner, N. L., Peixoto, B., Soteropoulos, P., Bader, J. S., Karakousis, P. C., Fallows, D. & Kaplan, G. Early innate

immunity determines outcome of Mycobacterium tuberculosis pulmonary infection in rabbits. *Cell Commun Signal* **11**, 60 (2013).

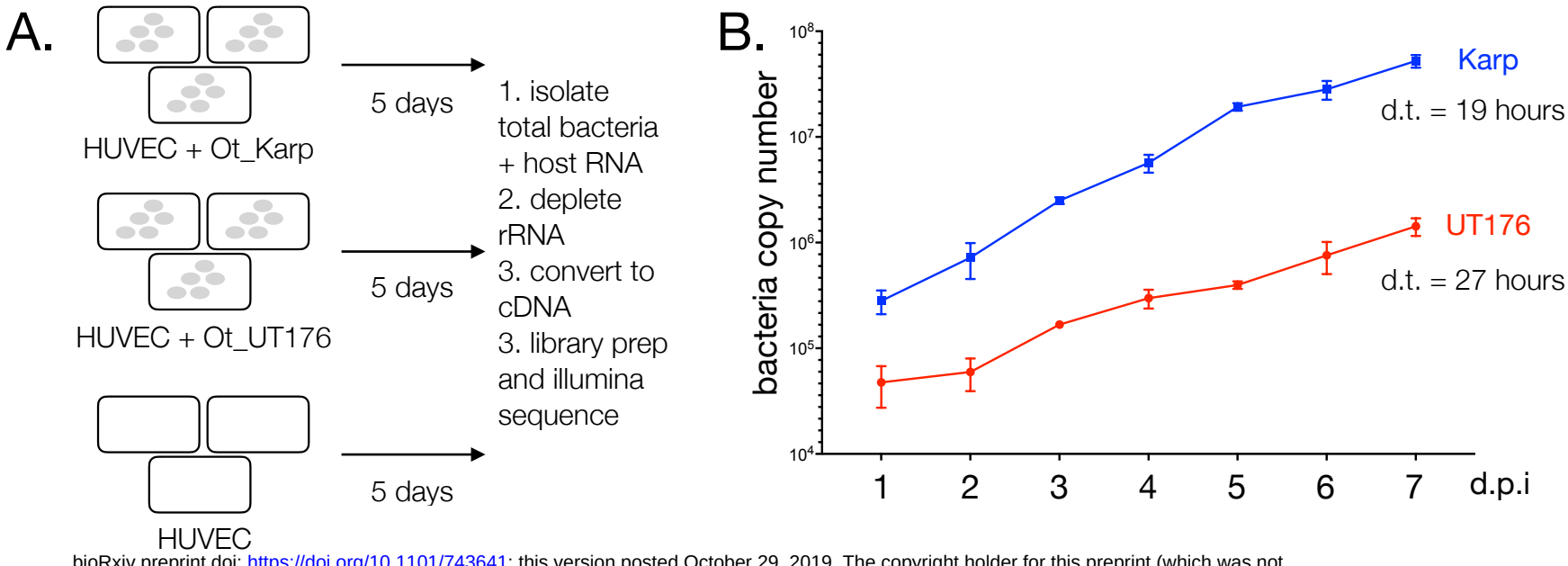


Figure 1. Overview. A. Schematic experimental overview. HUVEC = human umbilical vein endothelial cell. **B.** Growth curve showing replication of Ot in cultured HUVEC cells. Bacteria were grown in 24 well plates and the total bacteria per well is shown. Mean and SD from three independent replicates are shown. **C.** Confocal microscopy images of Ot in HUVEC cells 5 days post infection. Additional images and time points are shown in Supp. Fig. 1 and 2. Blue = DAPI (DNA), Red = Evans blue (host cells), green = Ot labelled with Alexa488-click-methionine. **D.** RNA mapping statistics showing the fraction of host and Ot RNA for each condition. The mean of three independent experiments is shown. Individual results for each replicate are shown in Supp. Fig. 3. **E.** Percentage of RNA-seq reads assigned to different classes of RNA in Karp, UT176 and HUVEC.

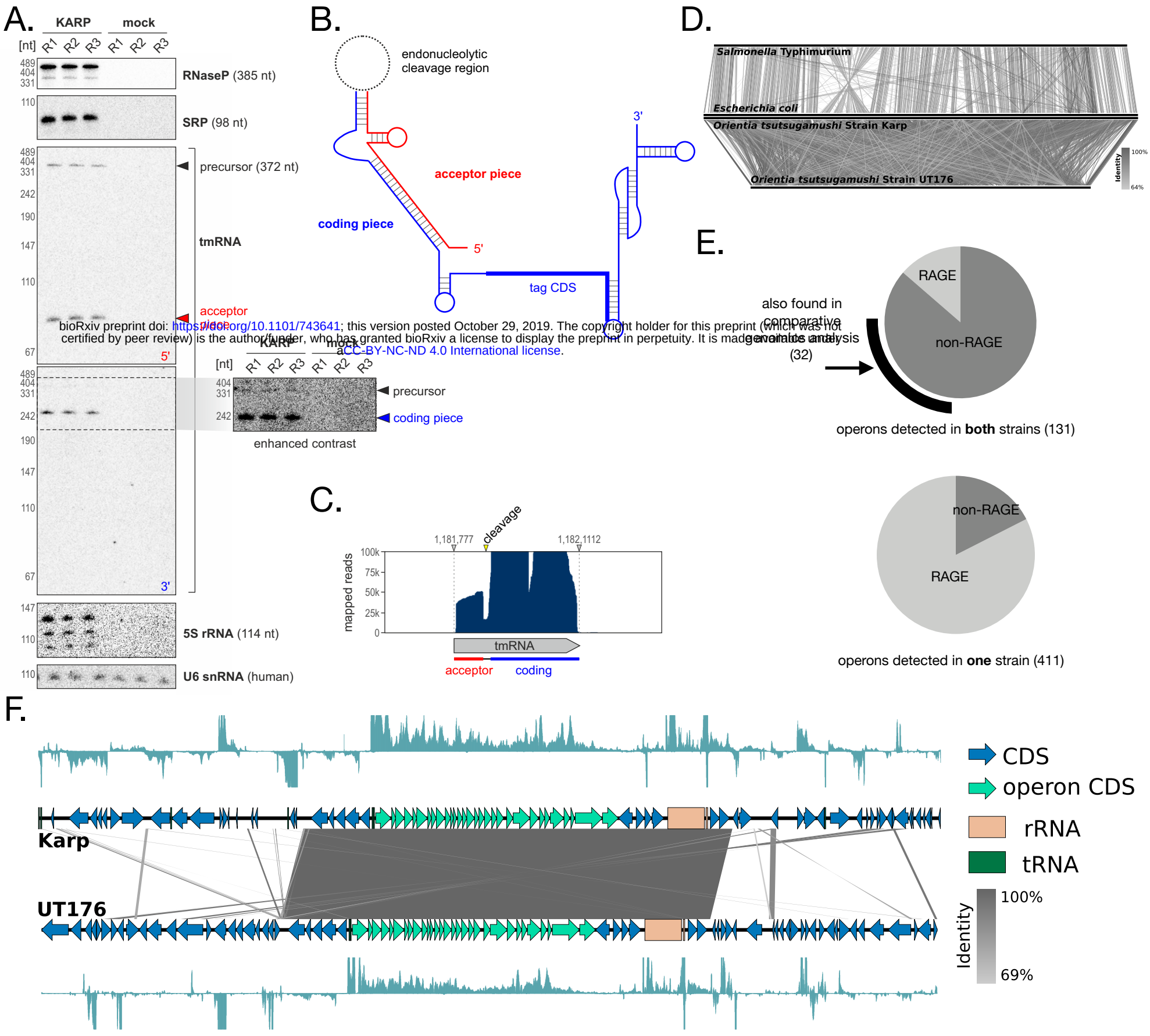


Figure 2. RNA biology in Ot. A. Northern blot analysis of core non-coding RNAs in Ot. **B.** Structure of the two-piece tmRNA observed in the Ot transcriptome **C.** RNA-seq read covereage over the tmRNA gene mirrors cleavage observed by Northern blot. **D.** A comparison of genomic synteny of two species within the enterobacteriaceae (*Escherichia coli* MG1655 and *Salmonella enterica* serovar Typhimurium SL1344, top), with synteny between the two Orientia strains from this study (bottom). **E.** Pie charts illustrating the relative abundance of RAGE genes in conserved (right) and strain-specific (left) operons. **F.** Visualization of the largest conserved operon in Ot, encoding multiple ribosomal genes, showing RNA-seq coverage in both strains.

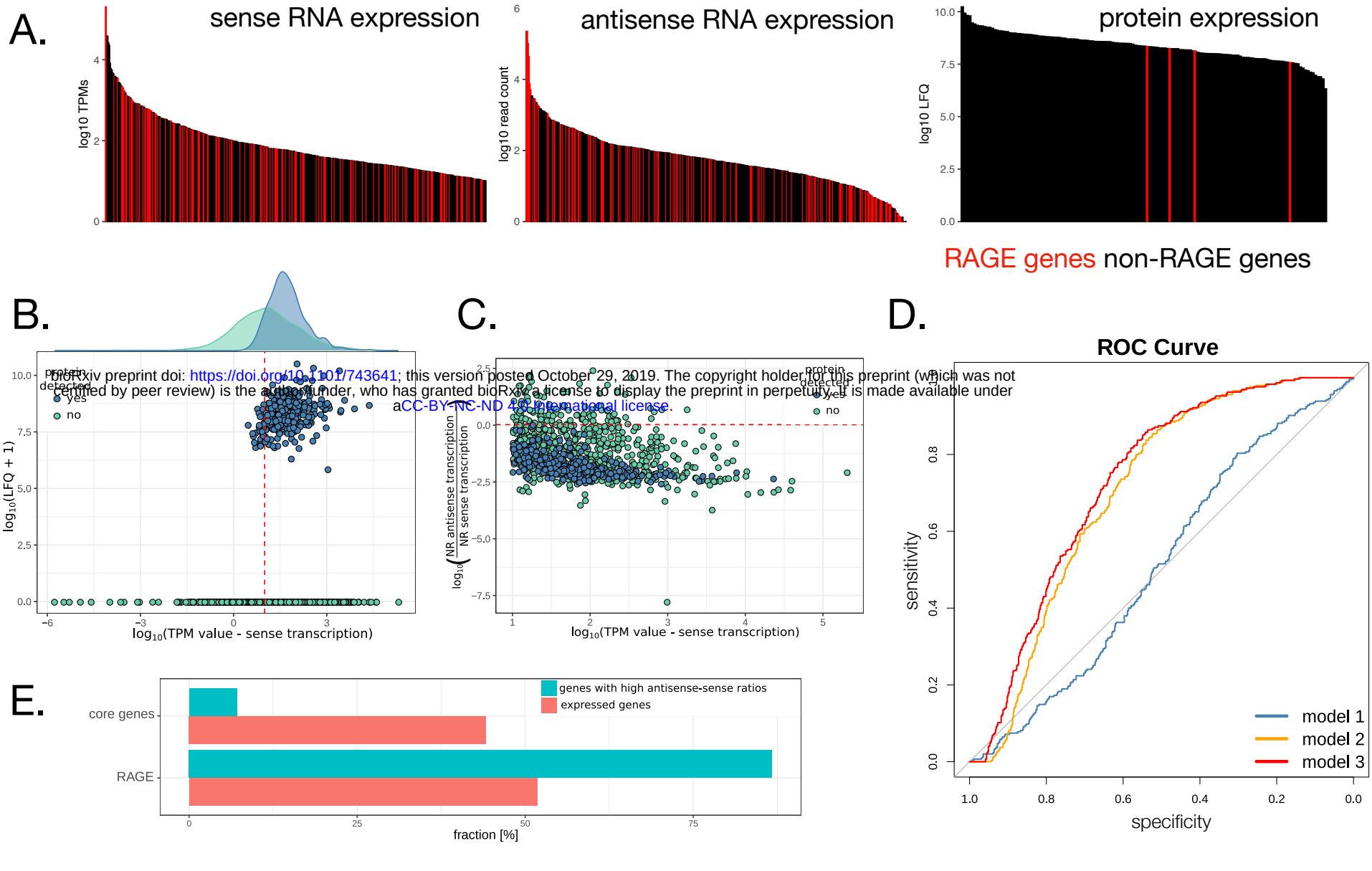


Figure 3. Antisense transcription is enriched on mobile genetic elements and is predictive of an absence of protein expression. **A. B.** Plot showing the relationship between protein expression, defined by LFQs, and transcript expression, defined by TPMs. Genes cluster into two groups based on their protein expression. The red line indicates the threshold for expressed genes (TPM value equal to 10). **C.** Sense transcription and the ratio of reads assigned to the antisense and sense strands, showing classification based on proteomics detection. The red line indicates the sense-antisense ratio (1.059182) above which translation was not detected by mass spectrometry. **D.** ROC curves evaluating the performance of logistic regression models to predict protein expression from RNA-seq read counts. Model 1 strictly uses sense expression, Model 2 the antisense-sense ratio, and Model 3 uses both. Incorporating antisense expression clearly improves model performance. **E.** Fraction of core genes and RAGE genes in the set of genes with high antisense-sense ratios, compared to all expressed genes.

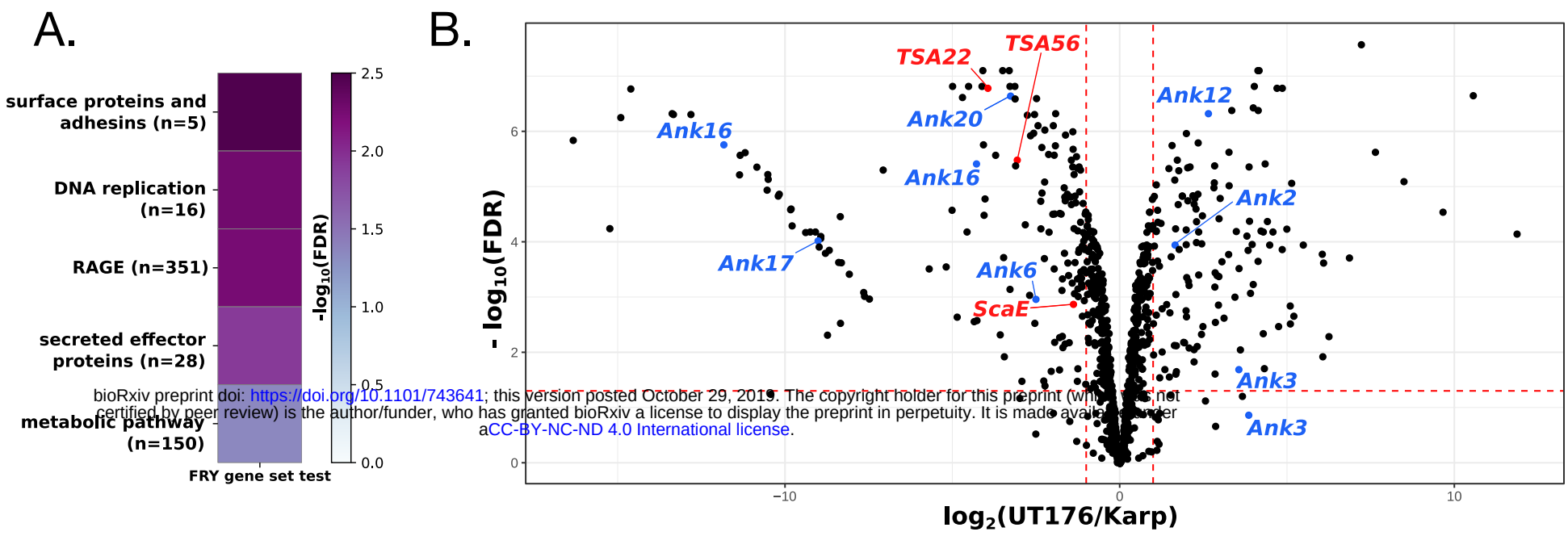


Figure 4. Differential bacterial gene expression. A. Heatmap illustrating pathways enriched in differentially expressed genes. All illustrated categories are more highly expressed in Karp. **B**. Volcano plot showing differential expression of bacterial genes in Karp and UT176. Bacterial surface genes (red) and ankyrin repeat containing effector proteins (blue) with log fold change ≥1 are highlighted.

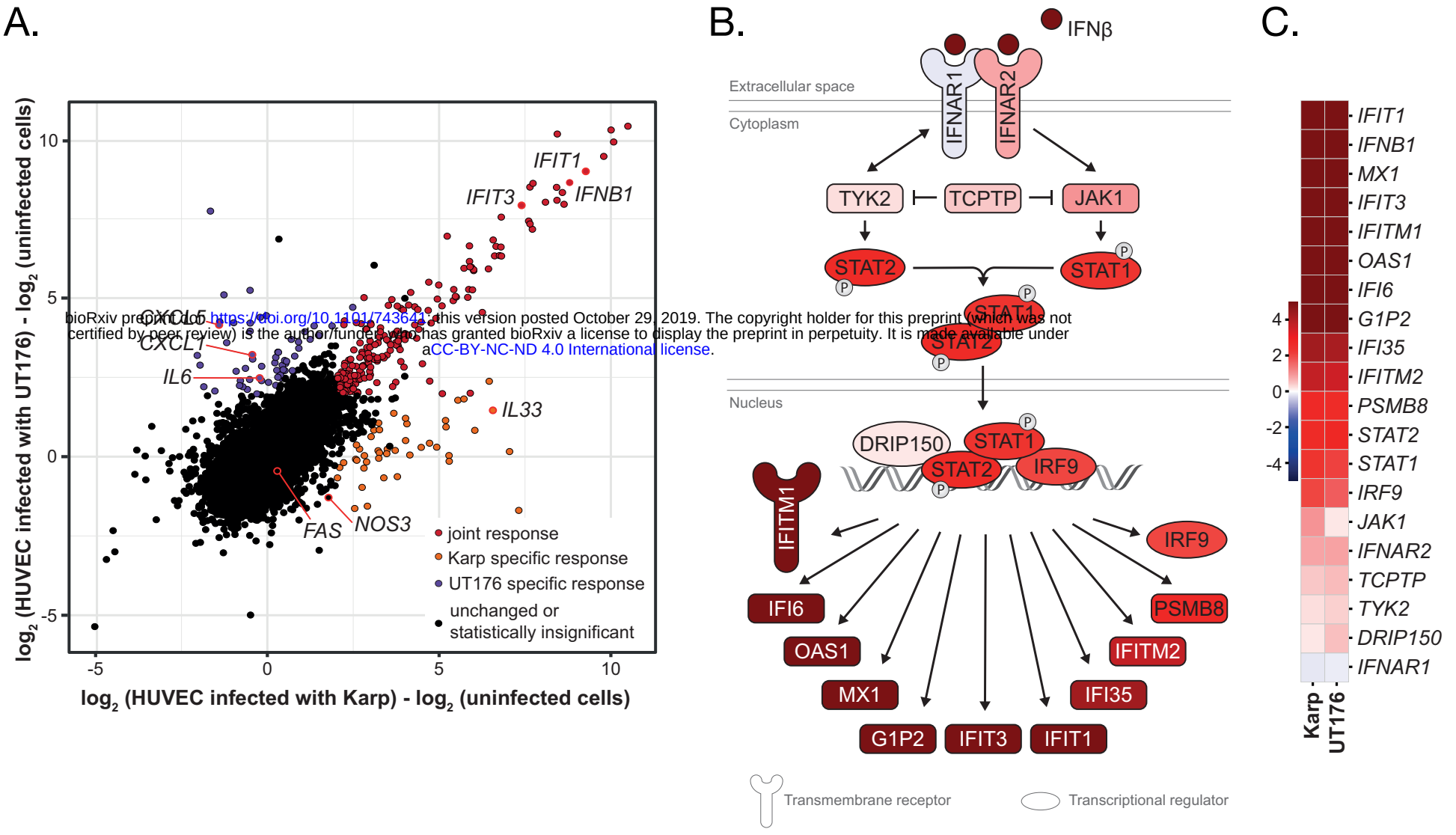


Figure 5. Ot induces an antiviral interferon response in HUVECs. A. Summary of the host response showing joint and strain specific responses. **B.** Activation of multiple genes in the canonical interferon signaling pathway in Karp infected HUVECs compared with uninfected HUVEC cells. **C.** Heat map showing up-regulation of genes in the interferon signaling pathway in HUVECs infected with Karp and UT176 compared with uninfected cells.

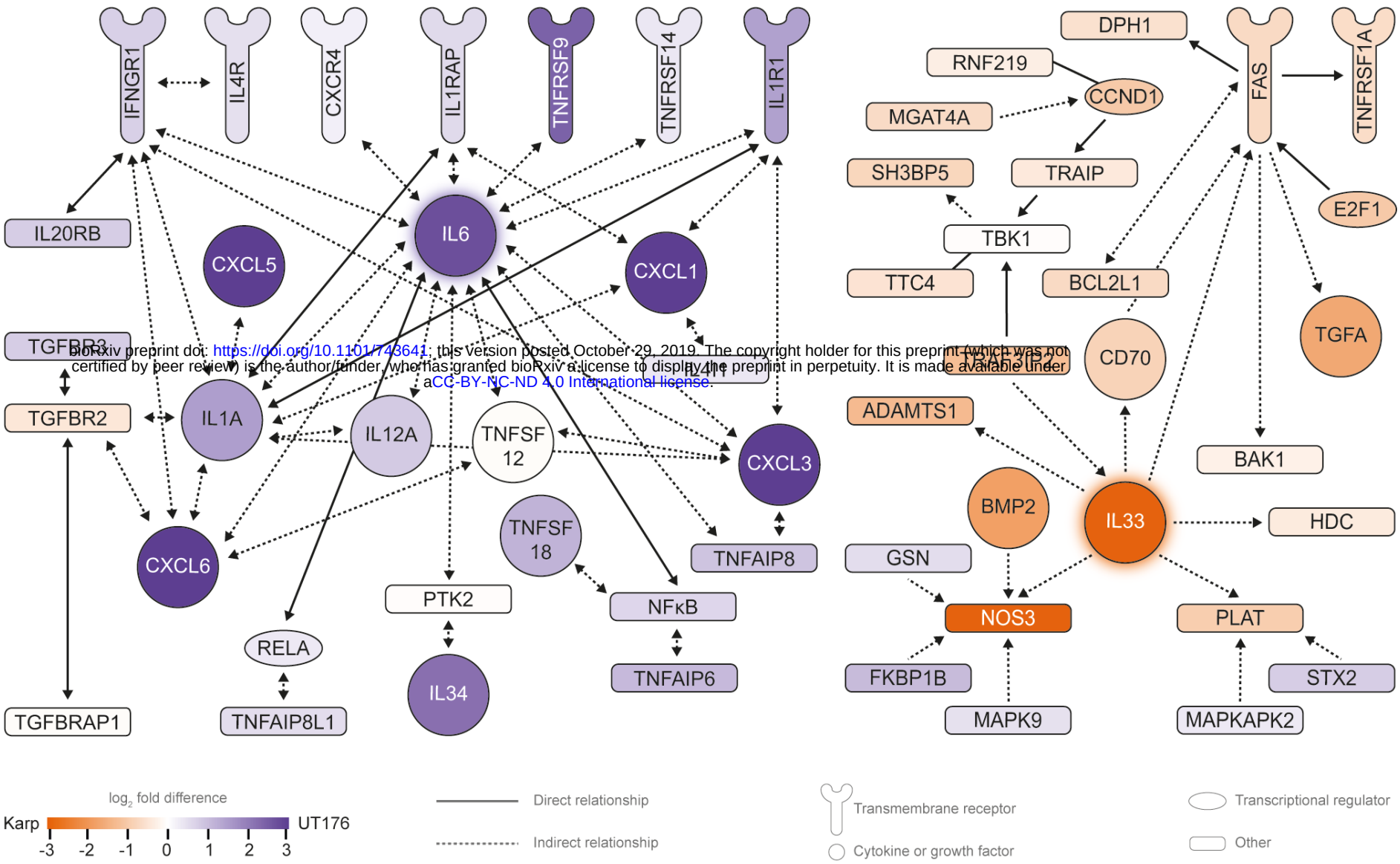
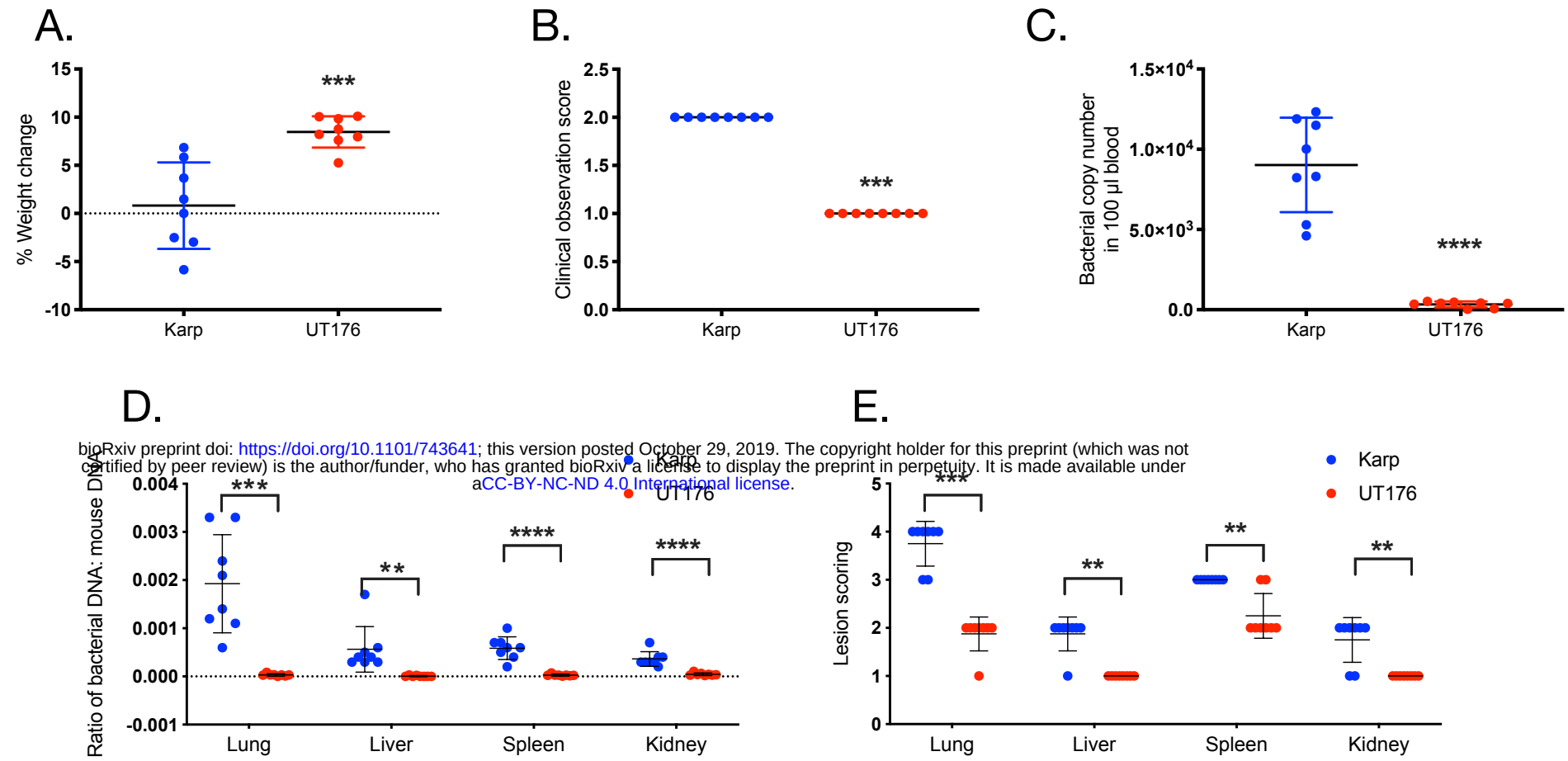


Figure 6. Karp and UT176 lead to up-regulation of distinct networks in HUVECs. A. Up-regulation of multiple proinflammatory chemokines and cytokines in HUVECs infected with UT176. **B.** Induction of the IL33-FAS-mediated anoikis network in Karp infected HUVEC cells.



F.

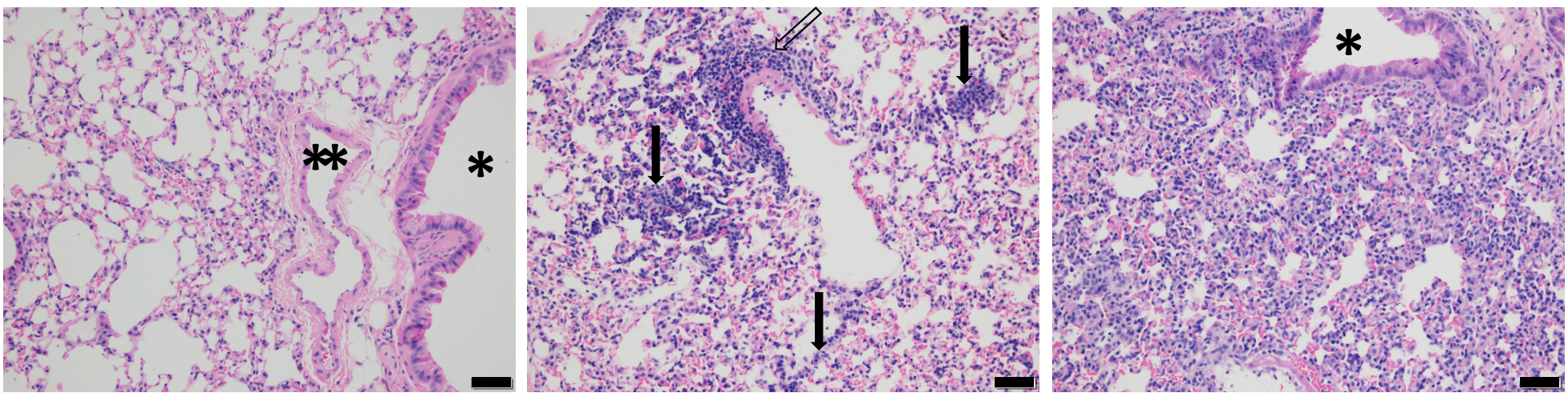


Figure 7. Karp is more virulent than UT176 in a mouse infection model. A. Weight change over 12 days infection. **B.** Clinical observation score of mice 12 days post infection. This number is a composite score based on appetite, activity, and hair coat with higher numbers representing low appetite, low activity, and ruffled fur. Details provided in Supp. Fig. 20. **C.** Bacterial copy number in 100 μl blood taken from euthanized mice 12 days post infection, measured by qPCR. **D.** Ratio of bacterial DNA copy number to mouse DNA copy number in lung, liver, spleen, and kidney of euthanized mice 12 days post infection, measured by qPCR. **E.** Lesion scores of Hematoxylin and Eosin stained lung, liver, spleen, and kidneys of euthanized mice 12 days post infection. Scores range from 0 to 5 with 0 representing normal tissue and 5 representing severe lesion damage. All graphs show mean and standard deviation. Statistical significance is calculated using unpaired t test using GraphPad Prism software. **p≤0.01 ***p≤0.001 ****p≤0.0001 **F.** Images of Hematoxylin and Eosin stained lung tissue of mice infected with buffer, UT176 or Karp. Scale bars = 50 μm. * indicates airway and ** indicates blood vessel. Uninfected control: airway, blood vessel and alveoli all appear normal. UT176-infected lungs: There is diffuse thickening and infiltration of alveolar septa with a mixed population of macrophages and lymphocytes (arrows). There is also mild perivascular lymphohistiocytic inflammation (open arrow). Karp-infected lungs: There is diffuse moderate thickening and infiltration of alveolar septa with a mixed population of macrophages and lymphocytes. The airway (*) is unaffected and normal. Additional figures are shown in Supp. Fig. 20.