ARTICLE

Check for updates

Dual RNA-seq of *Orientia tsutsugamushi* informs on host-pathogen interactions for this neglected intracellular human pathogen

Bozena Mika-Gospodorz [1,11], Suparat Giengkam [2,11], Alexander J. Westermann [1,3], Jantana Wongsantichon [2], Willow Kion-Crosby [4], Suthida Chuenklin [2], Loo Chien Wang [5,6], Piyanate Sunyakumthorn [7], Radoslaw M. Sobota [5,6], Selvakumar Subbian [8], Jörg Vogel [1,3], Lars Barquist [1,9,11✉] & Jeanne Salje [2,8,10,11✉]

Studying emerging or neglected pathogens is often challenging due to insufficient information and absence of genetic tools. Dual RNA-seq provides insights into host-pathogen interactions, and is particularly informative for intracellular organisms. Here we apply dual RNA-seq to *Orientia tsutsugamushi* (Ot), an obligate intracellular bacterium that causes 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 to study the transcriptional architecture of Ot, we find evidence for wide-spread post-transcriptional antisense regulation. Comparing the host response to two clinical isolates, we identify distinct immune response networks for each strain, leading to predictions of relative virulence that are validated 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.

[1] Helmholtz Institute for RNA-based Infection Research (HIRI), Helmholtz Centre for Infection Research (HZI), Würzburg, Germany. [2] Mahidol-Oxford Tropical Medicine Research Unit, Faculty of Tropical Medicine, Mahidol University, Bangkok, Thailand. [3] Institute for Molecular Infection Biology (IMIB), University of Würzburg, Würzburg, Germany. [4] Rutgers, the State Univeristy of New Jersey, New Jersey, NJ, USA. [5] Functional Proteomics Laboratory, Institute of Molecular and Cell Biology, Agency for Science, Technology and Research (A*STAR), Singapore, Singapore. [6] SingMass - National Mass Spectrometry Laboratory, Institute of Molecular and Cell Biology, Agency for Science, Technology and Research (A*STAR), Singapore, Singapore. [7] Armed Forces Research Institute of Medical Sciences, Bangkok, Thailand. [8] Public Health Research Institute, Rutgers University, New Jersey, NJ, USA. [9] Faculty of Medicine, University of Würzburg, Würzburg, Germany. [10] Centre for Tropical Medicine and Global Health, Nuffield Department of Medicine, University of Oxford, Oxford, UK. [11] These authors contributed equally: Bozena Mika-Gospodorz, Suparat Giengkam, Lars Barquist, Jeanne Salje. ✉email: lars.barquist@helmholtz-hiri.de; js2522@njms.rutgers.edu

Improved surveillance and diagnostics have led to the recognition of previously neglected bacteria as serious pathogens, whereas 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-third 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 infects 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[6,7]. These strains are classified into seven geographically diverse genotype groups, named after the serotypes of strains within them and dominated by the Karp, Kato and Gilliam groups[8,9]. Different strains of Ot exhibit different levels of virulence[10–12], dependent on both bacterial and host genotype. For example, strain Karp (group Karp) causes lethal infection in BALBc and C3H/He mice at low doses, strain Gilliam (group Gilliam) causes lethal infection in C3H/He but not BALBc mice at similar doses, whereas strain TA716 (group TA716) does not cause lethal infection in either mouse model at similar doses[11,13]. 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[14,15], 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*[16] 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). 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. Together, this illustrates the value of using a dual RNA-seq approach to study the biology of obligate intracellular bacteria.

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[17], and UT176, closely related to Karp based on whole genome sequencing[8], taken from a patient in northern Thailand in 2004[18]. These strains share a sequence identity of 95% in their TSA56 gene (commonly 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 umbilical 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 32:1 (UT176) and 35:1 (Karp) and grown for 5 days (Fig. 1a), by which point host cells were heavily loaded with bacteria (Representative growth curves shown in Fig. 1b, c, immunofluorescence microscopy images in Supplementary 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[9] and, in parallel, the human genome. As the *Orientia* genome is repeat-rich, we additionally applied model-based quantification with Salmon[19], which uses uniquely mapping reads to assign multi-mapping reads to these transcriptomes to improve our estimates of transcript abundance (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; Supplementary Fig. 3). This likely reflects differences in both cell entry efficiency and growth rate between Karp and UT176, which have doubling times of 19 and 27 h in HUVEC, respectively (Fig. 1b). The distribution of reads to RNA classes (Fig. 1e) indicated efficient depletion of ribosomal transcripts in the host transcriptome (<0.001% human rRNA reads). In contrast, we found an average of 32% and 44% rRNA reads in Karp and UT176, respectively. Most of these remaining bacterial ribosomal reads were derived from 5S rRNA (Supplementary Data 1), 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/). Reads mapping to coding sequences (CDSs) were abundant in both the HUVEC data (54% of all host-mapped reads across all sample) and in the Ot-specific reads (35% of the Karp- and 38% 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[9] 599 were expressed and 369 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; Supplementary Data 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
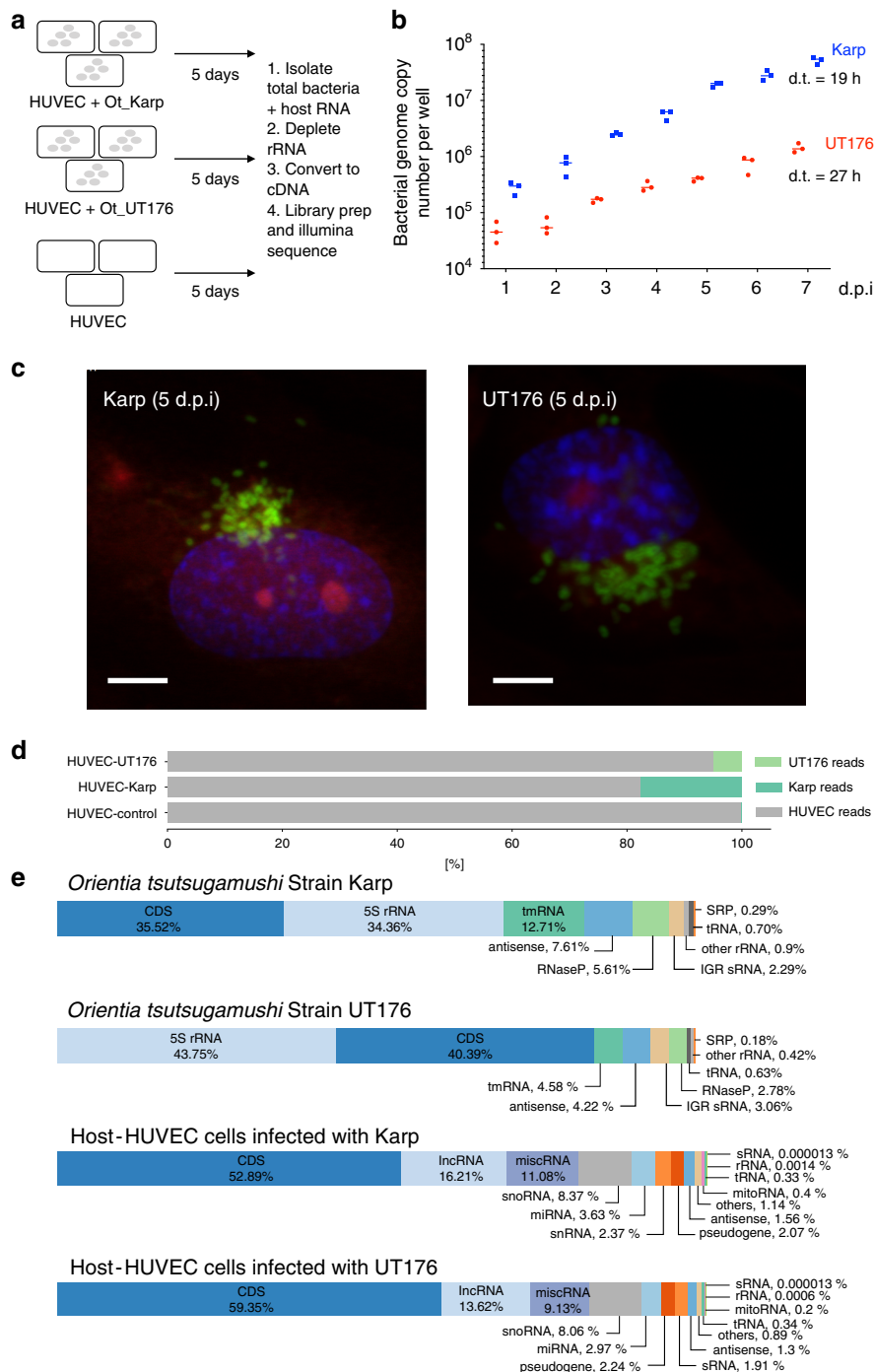
Fig. 1 Overview. a Schematic experimental overview. HUVEC = human umbilical vein endothelial cell. **b** Growth curve showing replication of Ot in cultured HUVEC cells. Bacteria were grown in 24-well plates and the total bacterial genome copy number per well as measured by qPCR is shown. Bacteria were added at an multiplicity of infection (MOI) of 8:1 (UT176) and 25:1 (Karp). 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 Supplementary Figs. 1 and 2. Blue = DAPI (DNA), Red = Evans blue (host cells), green = Ot labeled with Alexa488-click-methionine. Scale bar = 5 μm. **d** RNA mapping statistics showing the fraction of host and Ot RNA for each condition. The first replicate of the experiment is shown. Individual results for each replicate are shown in Supplementary Fig. 3. **e** Percentage of RNA-seq reads assigned to different classes of RNA in Karp, UT176 and HUVEC. Source data are provided as a Source Data file.

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[20], and is known to contribute to virulence in pathogens as diverse as *Salmonella* Typhimurium[21] and *Francisella tularensis*[22]. In our data, tmRNA appears to be expressed at unusually high levels, contributing between 4.6 and 13% of total bacterial reads (Fig. 1e; Supplementary Data 1), suggesting an important role in Ot survival in mammalian cells. tmRNA generally consists of a tRNA-like (acceptor) domain encoded
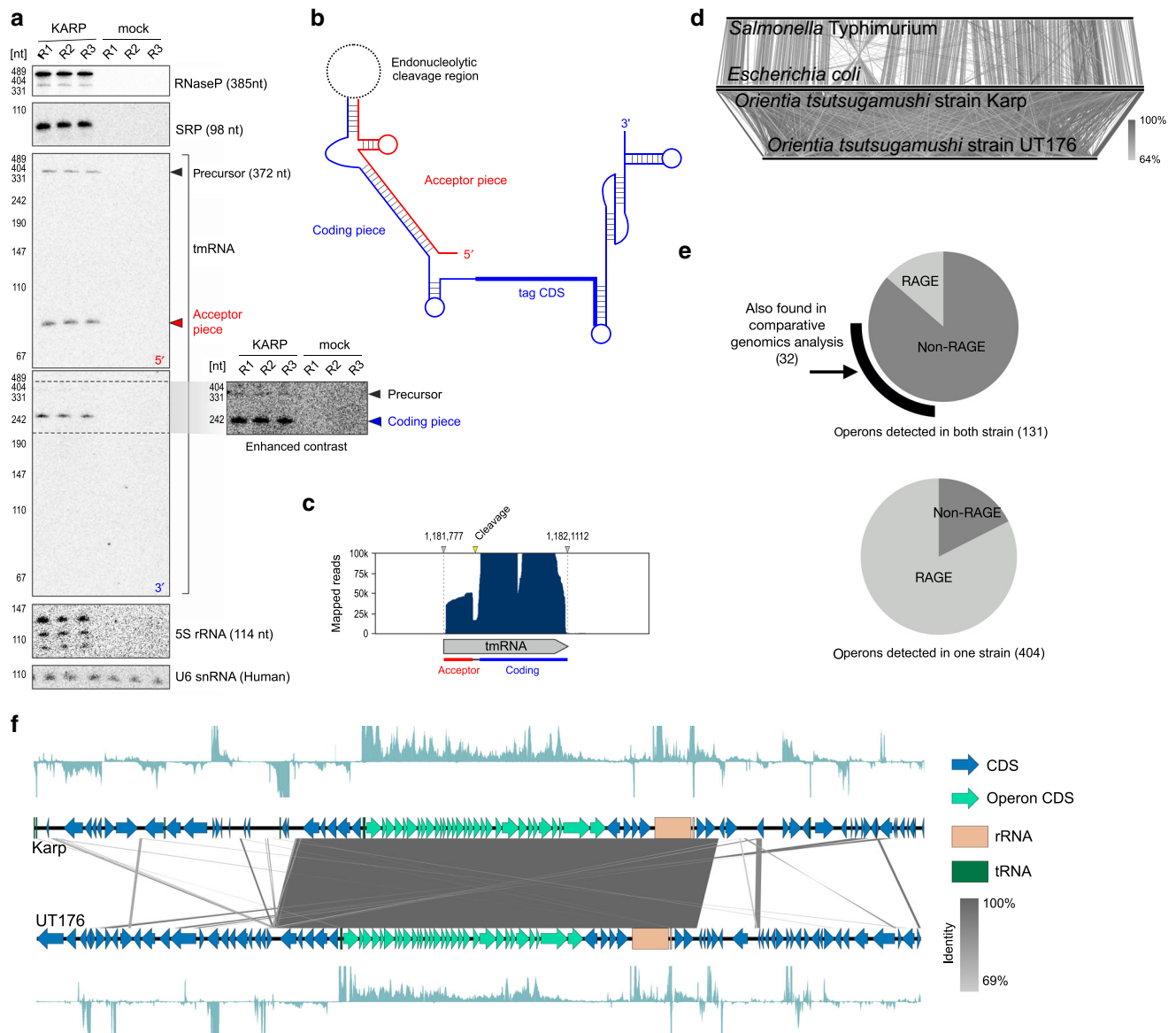
Fig. 2 RNA biology in Ot. a Northern blot analysis of core non-coding RNAs in Ot, showing results of three independent biological replicates. **b** Structure of the two-piece tmRNA observed in the Ot transcriptome. **c** RNA-seq read coverage over the tmRNA gene mirrors cleavage observed by Northern 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 (top) and strain-specific (bottom) operons. **f** Visualization of the largest conserved operon in Ot, encoding multiple ribosomal genes, showing RNA-seq coverage in both strains. Source data are provided as a Source Data file.

upstream of a short open reading frame (coding domain). However, the transcript has undergone a circular permutation in some clades of bacteria[23], including the Alphaproteobacteria[24], which requires processing of a precursor transcript into separate, base-pairing acceptor and coding RNA chains[25,26] (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[27,28]. Our RNA-seq data identified 55 intergenic sRNA candidates, between 77 and 803 nt, in the Karp transcriptome (Supplementary Data 1 and 2). When normalized to the genome size of Ot, this is consistent with the number of sRNAs reported in model bacterial pathogens[29–33].

Conserved operons in a dynamic genome. The genome of Ot is highly dynamic[9], and while the timescales and mechanisms of its rearrangements are unknown they are thought to be driven by an extreme proliferation of mobile elements[34,35], 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[36] 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) (Supplementary Data 2) 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[9], 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%, Supplementary 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[37]. In summary, the identification of co-transcribed 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[35]. 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 genes including ankyrin-repeat-containing proteins (*ank*), tetratricopeptide repeat-containing proteins (*TPR*), *spoT/relA* genes, DNA methyltransferases, and replicative DNA helicases. Many of these genes are truncated and most RAGE copies are highly degraded, containing only a subset of genes from the complete element. It is not known if this ICE is still active for transposition, nor whether Ot can express a functional type IV secretion apparatus. In our RNA-seq data set ~50% of the most highly expressed genes were repetitive genes encoded by the RAGE (defined throughout our analysis as integrase, transposase, conjugal transfer genes and hypothetical genes) in both strains. These same genes were also highly expressed in the antisense direction in both strains (Fig. 3a; Supplementary Fig. 6), leading us to hypothesize that the repetitive RAGE genes may be regulated by antisense gene expression.

Antisense transcription is widespread in bacteria[38], with between 5 and 75% of coding sequences exhibiting antisense transcription. Although functions for a number of specific antisense transcripts have been described, including regulation through occlusion of the ribosome binding-site or induction of RNase III mediated decay[39] their relevance as a general functional class remains unclear. Antisense promoters tend to be weakly conserved[40], arguing against specific functions, and mathematical modeling has suggested the majority of antisense transcripts are not expressed at sufficient levels to affect the regulation of their cognate coding sequence[41].

To explore the relationship between sense and antisense expression of core Ot genes and repetitive RAGE genes, we combined our Karp RNA-seq data set with a proteomics data set generated under the same experimental growth conditions. We chose to investigate Karp initially, as the higher bacterial load

makes detection of bacterial proteins more likely. We observed substantially fewer RAGE gene products detected by proteomics, compared with RNA-seq (Fig. 3a). Genes with detected protein products had higher transcript expression on average compared to those not detected by proteomics (Fig. 3b). However, many highly expressed transcripts appeared to produce no protein. Given our previous observations, we asked whether antisense transcription would correlate with protein expression. All genes with detected proteins had an antisense–sense read count ratio of <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 has a widespread regulatory role in Ot. This was confirmed by cross-validation (Methods, Supplementary 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-one core genes also exhibited an antisense–sense ratio of >1 (Supplementary Data 9) and these include the chromosomal replication initiator protein *dnaA*, DNA polymerase subunit III, an outer membrane autotransporter protein *scaD*, glutamine synthetase, two transporters, the protein export protein *secB* and 13 hypothetical proteins. None of these models achieved >65% 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 (Supplementary Discussion, Supplementary Figs. 8 and 9).

Differential expression of genes in Karp and UT176. Due to a lack of genetic tools, identification of virulence mechanisms in Ot has been difficult, with only a small number of antigenic surface proteins and effectors known. As pan-genome diversity appears to primarily be the result of gene duplication and decay, differences in virulence between strains are likely due to differences in expression. To investigate this hypothesis, we performed differential expression analysis between Karp and UT176 at 5 days after infection of HUVEC cells. Pathway and gene ontology (GO) analyses of differentially expressed genes (Fig. 4a; Supplementary Data 16 and 17) indicated that most pathways were upregulated in Karp compared with UT176, including those involved in DNA replication and metabolism, consistent with Karp's higher growth rate (Fig. 1b). At the gene level (Supplementary Data 18; Fig. 4b) we found a number of surface and effector proteins (Anks) were differentially regulated between the two strains. Ot encodes five autotransporter domain-containing proteins (ScaA-ScaE) and three immunogenic type surface antigens (TSA22, TSA47, and TSA56). All these surface proteins are immunogenic, based on their reactivity to patient sera[42], with TSA56 being the most abundant Ot surface protein. TSA56 has four variable domains
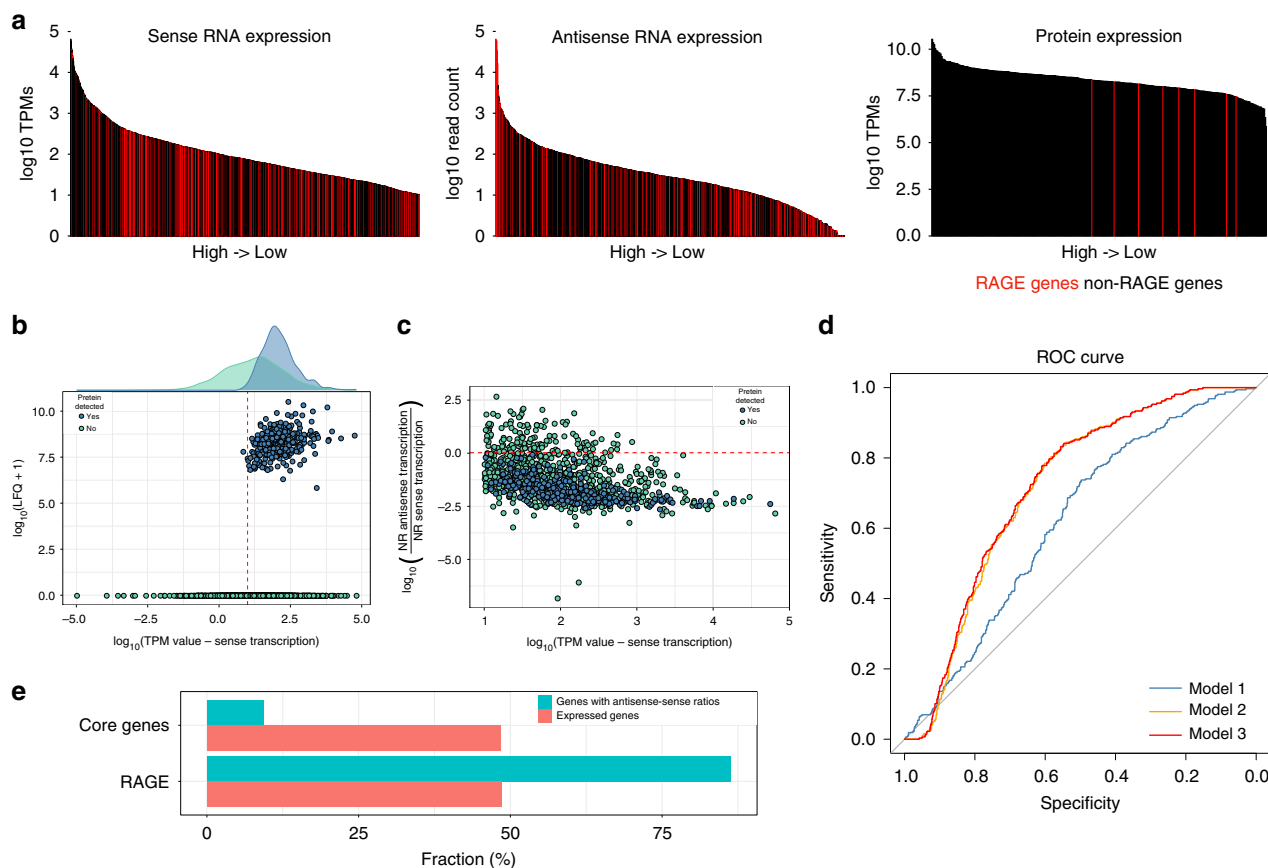
Fig. 3 Antisense transcription in Ot. a Sense RNA expression, antisense RNA expression, and protein expression over genes, ranked from high to low; RAGE (rickettsial-amplified genetic element) genes are marked in red. **b** Plot showing the relationship between protein expression, defined by LFQs (label-free quantitations), and transcript expression, defined by TPMs (transcripts per million). 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 (1.06) above which translation was not detected by mass spectrometry. **d** ROC (receiver operating characteristic) 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.
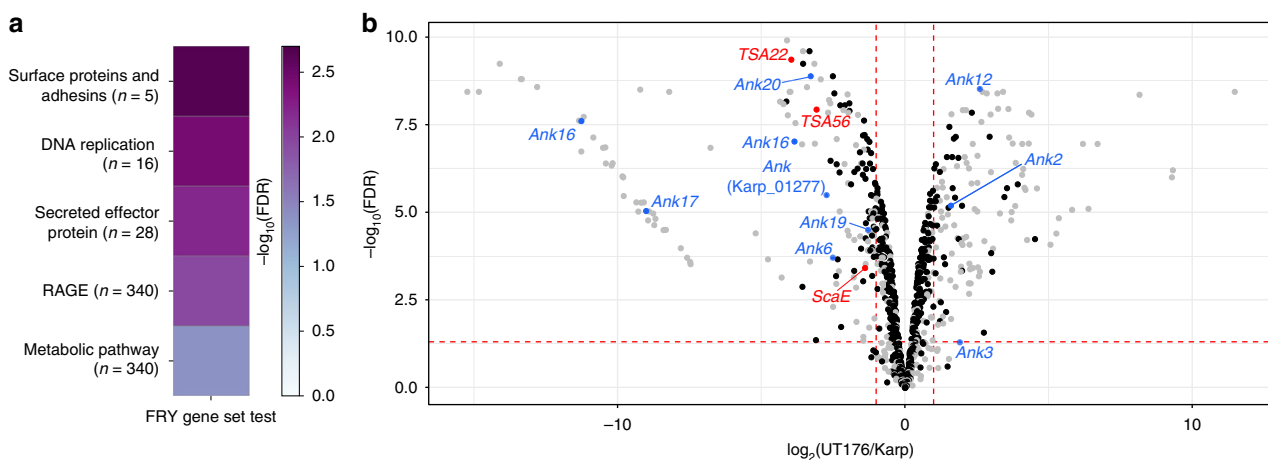

Fig. 4 Differential bacterial gene expression. a Heatmap illustrating pathways enriched in differentially expressed genes. All illustrated categories are more highly expressed in Karp. FDR-corrected *p*-values were calculated using the fry gene set enrichment test in the edgeR R package. **b** Volcano plot showing the 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. Gray dots represent RAGE (rickettsial-amplified genetic element) genes. FDR-corrected two-sided *p*-values were calculated using the quasi-likelihood *F*-test in the edgeR R package.

and these lead to strain-specific antibody responses in patients. TSA47, TSA56, and ScaA have been evaluated as possible vaccine candidates[43,44]. Of the core Ot genes, among those most differentially expressed between Karp and UT176 were *scaE*, *tsa56*, and *tsa22* (1.40, 3.08, and 3.96 logFC in Karp over UT176, respectively), and differential expression was confirmed by qRT-PCR in an independent infection experiment (Supplementary Fig. 12A). In contrast, *scaD* levels were increased in UT176 but to a lesser degree (0.99 logFC in UT176 over Karp). It is likely that different levels of expression of these bacterial surface proteins will affect interactions with host cells, for example through stronger binding of host cell receptors or activation of innate immune receptors. In the context of animal infection, differential expression of these immunogenic proteins may affect the induced adaptive immune response.

Genes for Ank and tetratricopeptide repeat-containing proteins (TPR) are present in 33 (Ank)/29 (TPR) and 21 (Ank)/22 (TPR) copies in Karp and UT176, respectively[9]. Some *ank* genes function as effectors in eukaryotic cells while others are uncharacterized. We compared the expression of Ank and TPR genes in Karp and UT176, using annotations derived from protein similarity to strain Ikeda for which the Anks have been best characterized[45]. *ank2*, *ank3*, *ank12*, and two copies of *tpr8* were upregulated with a logFC >1.5 in UT176, whereas six Anks including *ank6* and *tpr1*, *tpr3*, and *tpr5* were upregulated with a logFC >1.5 in Karp. Most of these proteins were not detected in the Karp proteomics data set 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 localize to the endoplasmic reticulum or host cell cytoplasm when ectopically expressed[46]. Ank6 interferes with NFkB translocation to the nucleus and inhibits its transcriptional activation[47]. 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 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 (Supplementary Data 7 and 8), seen previously in cultured endothelial cells and monocytes, as well as patient-derived macrophages[45,48–51]. 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 upregulated 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 upregulated 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 upregulation of proinflammatory chemokine genes including *CXCL10*, *CXCL11*, and for cytokine receptors *IL13RA2*, *IL7R*, *IL15RA*, and *IL3RA* (Supplementary Data 7 and 8).

The upstream signals leading to activation of these signaling pathways are unknown but Ot has been shown to activate host cells by signaling through the NOD1-IL32[52] and TLR2[53] pathways. Our data showed that *TLR3* is upregulated in cultured HUVEC cells in response to both Karp and UT176 (Supplementary Table 1). TLR3 recognizes viral double-stranded (ds)RNA in the cytoplasm[54], and

it is possible that it responds to Ot dsRNA. The upregulation of the mRNA for transcription factor IRF7, which is known to respond to stimulation from membrane-bound TLRs, further supports a role for TLR2 and TLR3 in the detection of Ot.

Differential host responses to Karp and UT176. Although Karp and UT176 both induced a type-I interferon proinflammatory response compared to uninfected HUVEC cells (Fig. 5b, c), each strain also induced its own unique response. Some of these expression changes were validated by qRT-PCR (Supplementary Fig. 12). The mRNA levels of multiple cytokines, chemokines, and cytokine receptors were higher in HUVEC cells infected with UT176 compared with Karp (Fig. 6a; Supplementary Figs. 13, 14, and 16). A network map of proinflammatory chemokines and cytokines, and their differential induction in response to UT176 and Karp is shown in Fig. 6a and Supplementary Fig. 13A. Most of the genes for cytokines, chemokines, and cytokine receptors were differentially upregulated by infection with UT176 compared with Karp, including *CXCL8*, *CXCL1*, *CXCL2*, *CXCL10*, *IL6*, *IL1RL1*, and *IL18R1*. The mRNA levels of surface adhesion molecules associated with activation of the endothelium, VCAM1 and ICAM1, were also upregulated in UT176-infected cells compared with Karp (Supplementary Data 7 and 8). Although *TLR3* was upregulated in both strains, TLR3 activation in UT176-infected cells was 1.5 logFC higher than in response to Karp. Comparison of NFkB pathway genes and genes associated with NOS2 production revealed that genes in both pathways were upregulated in UT176 but they were not upregulated or were significantly less upregulated in response to Karp infection (Supplementary Fig. 14). Expression of host genes associated with leukocyte proliferation and mononuclear leukocyte differentiation was strongly induced in HUVECs infected with UT176 but significantly less so when infected with Karp (Supplementary Fig. 15). Thus, UT176 seems to induce a stronger proinflammatory response and this may lead to more effective pathogen clearance (Fig. 1b).

In contrast to the multiple chemokines and cytokines upregulated in UT176-infected HUVEC cells, only *IL33* was specifically upregulated in Karp-infected HUVEC cells (5 logFC difference; Supplementary Data 7 and 8). IL33 is a proinflammatory cytokine that is involved in pathogenicity in a mouse model of scrub typhus[55]. To investigate Karp-mediated activation of *IL33*, we analyzed gene induction in the IL33-FAS network (Fig. 6b; Supplementary Fig. 13B). Most genes in the network were differentially induced in Karp-infected HUVEC cells compared to in UT176-infected HUVECs. Upregulation of IL33-NOS-mediated signaling contributes to tissue inflammation. We analyzed networks of genes involved in (i) organismal growth failure (ii) organismal morbidity and mortality and (iii) organismal death. In all cases, Karp induced these networks while UT176 dampened them (Supplementary Fig. 16).

Two Ot strains differ in virulence in a mouse model. To investigate how differences in Karp and UT176 extend to behavior in a host, we tested the relative virulence of the two strains in an intravenous mouse infection model. 1.25×10^6 bacteria were intravenously inoculated into female C57BL/6NJcl mice (6–8 weeks, 8 mice per group) and monitored for disease symptoms for 12 days prior to euthanasia. Both the more severe clinical symptoms (Fig. 7a; Supplementary Fig. 18) and lower weight gain over 12 days (Fig. 7a; Supplementary Fig. 17) of Karp-infected compared to UT176-infected mice support Karp being the more virulent Ot strain.

Blood and tissue from lung, liver, spleen, and kidneys were isolated and the bacterial load measured by qPCR. The bacterial
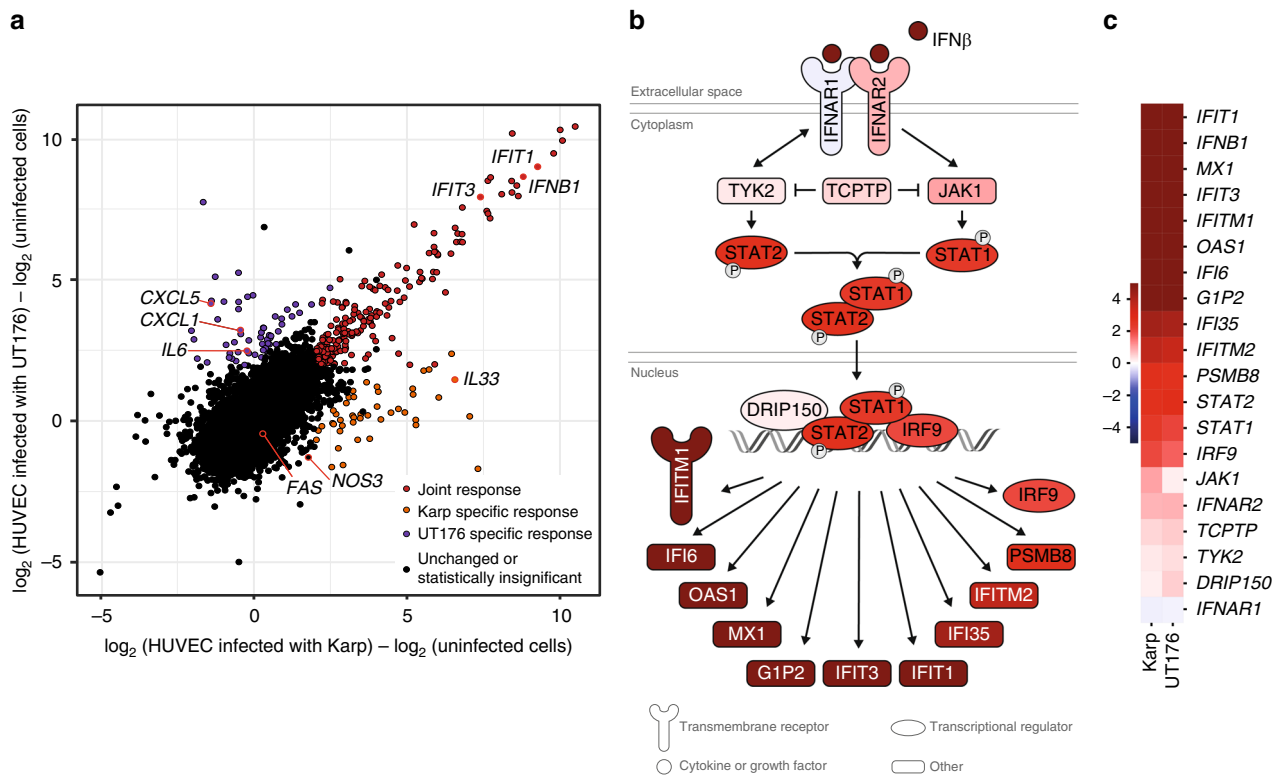
Fig. 5 Ot induces an antiviral interferon response in HUVECs. a Summary of the host response showing joint and strain-specific responses. The joint response is defined as genes with a $\log_2$ fold change (logFC) > 2 and FDR-corrected p-value < 0.01 for infection with both Karp and UT176. Strain-specific responses are genes with a logFC > 2 and FDR-corrected p-value < 0.01 for infection with either Karp or UT176, excluding genes already included in the joint response. FDR-corrected two-sided p-values were calculated using the quasi-likelihood F-test in the edgeR R package. **b** Activation of multiple genes in the canonical interferon signaling pathway in Karp-infected HUVECs compared with uninfected HUVEC cells. **c** Heatmap showing upregulation of genes in the interferon signaling pathway in HUVEC cells infected with Karp and UT176 compared with uninfected cells. The color scale represents the logFC in gene expression.

copy number in the blood (Fig. 7c) and tissues (Fig. 7d) was significantly higher in Karp-infected mice. Tissues were stained by hematoxylin and eosin, and the extent of tissue damage scored by histopathological scoring (Figs. 7e, f; Supplementary Fig. 19). Lesion scoring was significantly more severe in lung, liver, kidney and spleen of Karp-infected mice than UT176-infected mice. Although lungs of all Ot-infected mice showed diffuse thickening of alveolar septa, and infiltration of macrophages and lymphocytes, this was more pronounced in Karp-infected mice. Tissues were only analyzed at a single time point (12 days) and therefore it is possible that the disease dynamics differed between the two strains, and that different results would be observed at different times after infection. It is also worth noting that these differences in virulence in a mouse model may not translate to equivalent differences in human pathogenicity. Together these data showed that Karp exhibited higher virulence in a mouse model of Ot infection than UT176. This is consistent with our observations in HUVEC cells, though investigation of the differential host response to Karp- and UT176-infected mice, including in particular the roles of the adaptive immune response and dissemination kinetics within the host, should be a focus of future work.

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 >1000 bp in length[9]. This is in contrast to the most closely related rickettsial species, whose genomes are

typically around 1.1–1.3 Mbp[56]. 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 two or three 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 in strain Karp. Although it is unclear whether the same phenomenon occurs across Ot strains, our observation that similar genes are enriched for antisense transcription in strain UT176 (Supplementary Fig. 6) suggests that it may be. 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[57], 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[58]. It is unclear whether these untranslated transcripts have some function in Ot, or
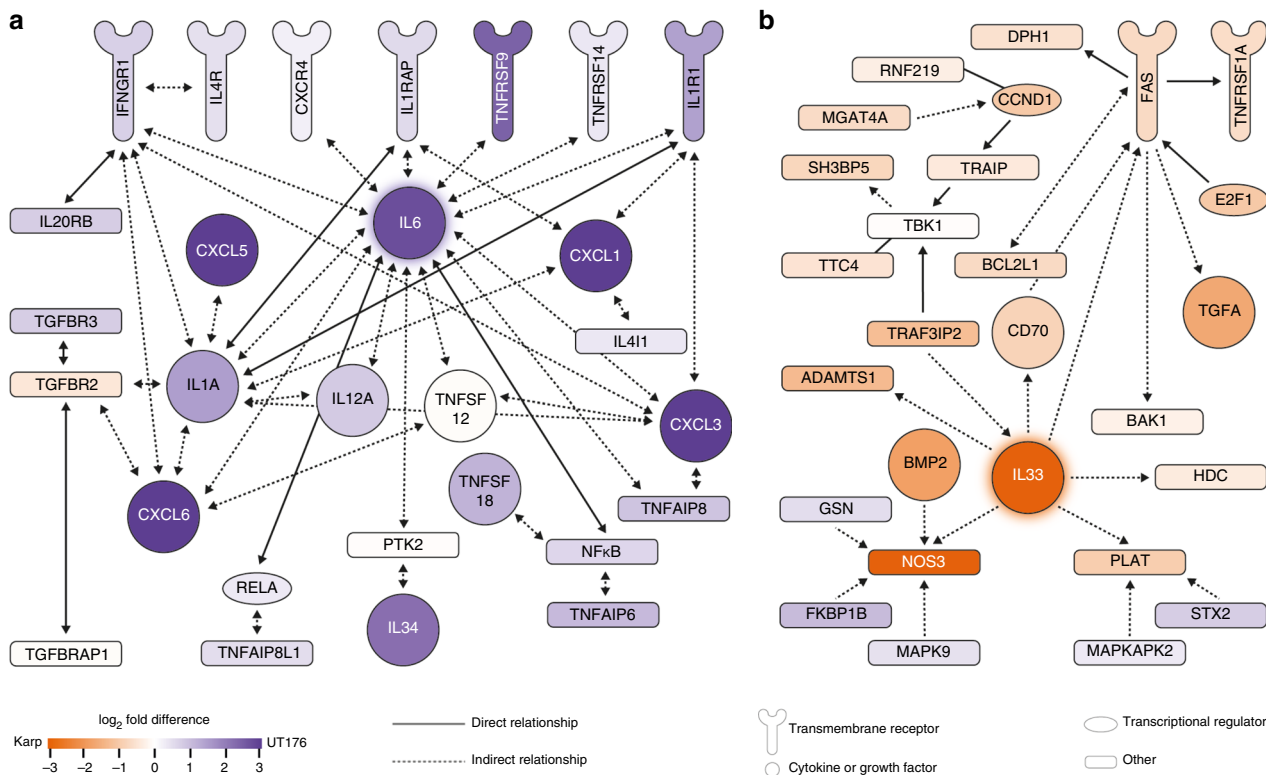
Fig. 6 Karp and UT176 lead to the upregulation of distinct networks in HUVECs. a Upregulation of multiple proinflammatory chemokines and cytokines in HUVECs infected with UT176. **b** Induction of the IL33-FAS-mediated anoikis network in Karp-infected HUVECs.

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 although 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 upregulation 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. Although both Karp and UT176 induced an antiviral proinflammatory response, as shown previously[45,48–51], UT176 strongly induced an IL6-mediated proinflammatory response, whereas Karp induced an IL33-NOS3-FAS response, differences likely to influence the relative virulence of these strains.

Our study has a number of limitations. First, we cannot distinguish between differential host responses due to actively replicating bacteria as compared to non-replicating bacteria, nor between Ot-specific responses as compared to non-specific uptake responses. A second limitation in the interpretation of our data is that Karp has a higher growth rate than UT176

(Fig. 1b) and produced a higher number of reads (Fig. 1d). This is unlikely to affect the differential bacterial gene expression measurements, which were normalized between samples, but it does make it difficult to separate the effect of differences in bacterial growth from differences in bacterial virulence on the host response. Finally, in order to obtain a sufficient read count, a relatively high MOI (~30:1) was used in the RNA-seq experiment. Only a subset of these bacteria will be viable and the exact number entering host cells is unknown. However, the final infectious dose per host cell was likely higher than that encountered under physiological conditions and this is likely to affect the immune response of host cells.

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 have 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 increased severity and mortality[55]. Our observations of reduced induction of *IL33* by the less virulent UT176 strain further support a role for this cytokine in the pathogenesis of scrub typhus. Future studies could investigate the causal links between Ot strain variability and the host immune response, for instance by screening panels of Ot strains and applying genome- or transcriptome-wide association studies on the bacterial side, or through genetic manipulation or the application of immunomodulating agents on the host side.

In summary, we have used dual RNA-seq to gain insights into the transcriptome structure and mechanisms of gene regulation in the neglected intracellular pathogen Ot during infection. We provide evidence for widespread antisense regulation, in particular for the RAGE genes. We identified a relationship between the relative induction of IL33- and IL6-based gene networks in the host and disease severity. These findings will lay the
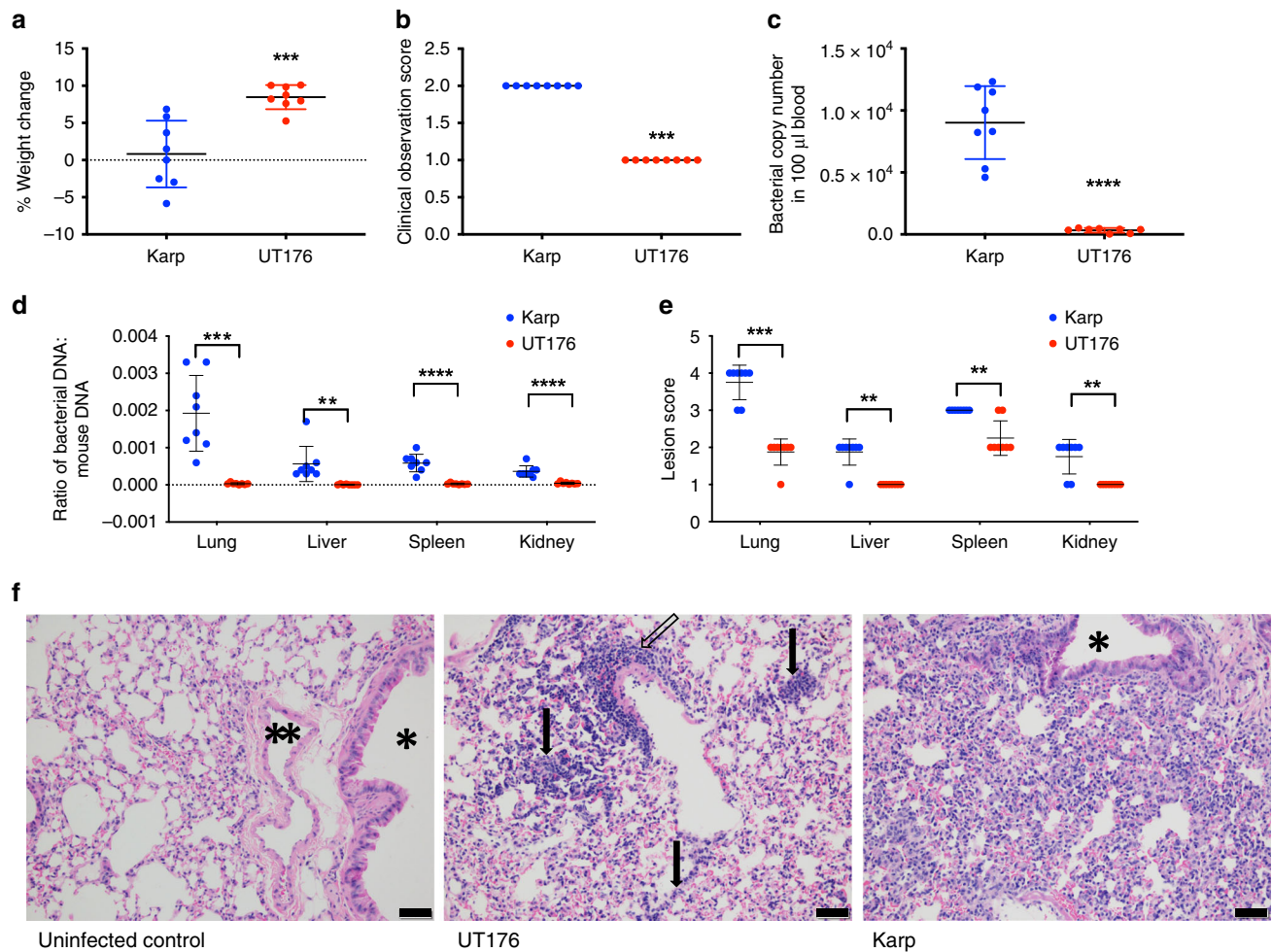
Fig. 7 Karp is more virulent than UT176 in a mouse infection model. a Weight change over 12 days of infection. **b** Clinical observation score of mice 12 days post infection. This number is a composite score based on appetite, activity, and hair coat with higher numbers representing low appetite, low activity, and ruffled fur. Details provided in Supplementary Fig. 19. **c** Bacterial genome copy number in 100 µl blood taken from euthanized mice 12 days post infection, measured by qPCR. **d** The ratio of bacterial genome copy number to mouse genome 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. **$p \leq 0.01$ ***$p \leq 0.001$ ****$p \leq 0.0001$. **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 Supplementary Fig. 19. Mean and SD from eight individual mice is shown. Source data are provided as a Source Data file.

groundwork for subsequent studies on the regulation of gene expression in Ot and mechanisms of pathogenesis. More generally, the present study may serve as a blueprint for the characterization of further obligate intracellular, genetically intractable bacterial pathogens.

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; Gibco C0035C) for 5 days at MOI 100:1. Cells were cultured using Media200 (ThermoFisher, Catalog number M200-500) supplemented with LVES media (ThermoFisher, Catalog number A14608-01) at 35 °C and 5% CO_2. The infectivity was determined by qPCR of the single copy Ot gene 47 kDa at days 5–7[59]. Primer sequences are given in Supplementary Table 1.

For growing bacteria for RNA isolation, bacteria from frozen stocks were first pregrown in HUVEC cells in a T25 culture flask. After 5 days they were harvested

and immediately inoculated onto a fresh lawn of HUVEC cells, with each condition filling 2 × 6-well plates (12 wells), for a second round of growth. These bacteria were used for RNA isolation. Because it is not possible to rapidly quantify the number of purified bacteria, the MOI of infections for RNA isolation was estimated by measuring the number of bacteria in the pregrowth supernatant one day before bacteria were harvested. The exact inoculum that had been used was subsequently confirmed by using qPCR of a sample of the inoculum, and it was determined that the MOI for infection was 35:1 bacteria:host (Karp) and 32:1 bacteria:host (UT176). Note that the actual number of bacteria that entered into host cells is likely to be less than this, as not all bacteria are viable for infection. Following infection bacteria that did not enter host cells were washed away with fresh media 3 h post infection. Both uninfected cells and infected cells were harvested by incubating the cells on ice and quickly resuspending in RNAprotect Bacteria Reagent (Qiagen, catalog number 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[60].

Bacteria prepared for growth curve measurements were prepared in the same way, with 5 days pregrowth in HUVEC cells, except the bacteria were then grown

in 24-well plates. The MOI was subsequently determined to be 8:1 (UT176) and 25:1 (Karp). At each time point bacterial DNA was isolated using alkaline lysis extraction and the bacterial genome copy number determined by qPCR[59].

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-MLV reverse transcriptase (NEB) wherein the 3′ adapter served as a primer. The first-strand cDNA was purified, and the 5′ Illumina TruSeq sequencing adapter was ligated to the 3′ end of the antisense cDNA. The resulting cDNA was PCR-amplified to about 10–20 ng/μl using a high-fidelity DNA polymerase. The TruSeq barcode sequences were part of the 5′ and 3′ TruSeq sequencing adapters. The cDNA library was purified using the Agencourt AMPure XP kit (Beckman Coulter Genomics) and analyzed by capillary electrophoresis (Shimadzu MultiNA microchip).

For sequencing, cDNA libraries were pooled in approximately equimolar amounts. The cDNA pool was size fractionated in the size range of 200–600 bp using a differential cleanup with the Agencourt AMPure kit (Beckman Coulter Genomics). Aliquots of the cDNA pools were analyzed by capillary electrophoresis (Shimadzu MultiNA microchip). Sequencing was performed on a NextSeq 500 platform (Illumina) at Vertis Biotechnologie AG, Freising-Weihenstephan, Germany (single-end mode; 75 cycles).

Northern blots. Each 15 μg of total RNA (i.e. a mixture of human and Ot RNA) prepared as above were loaded per lane and separated in 6% (vol/vol) polyacrylamide–7 M urea gels. RNA was transferred onto Hybond XL membranes (Amersham) by electro-blotting (1 h, 50 V, 4 °C) in a tank blotter (Peqlab), cross-linked with UV light, and hybridized at 42 °C with gene-specific 32P-end-labeled DNA oligonucleotides (Supplementary Fig. 12) in Hybri-Quick buffer (Carl Roth AG). After exposure, the screens were read out on a Typhoon FLA 7000 phosphorimager (GE Healthcare).

qRT-PCR. qRT-PCR was performed with the Power SYBR Green RNA-to-CT1-Step kit (Applied Biosystems) according to the manufacturer's instructions and a CFX96 Touch real-time PCR detection system (Bio-Rad). Human U6 snRNA served as reference transcripts. Fold changes in expression were determined using the $2^{(-\Delta\Delta Ct)}$ method[61]. Primer sequences are given in Supplementary Table 1, and their specificity had been confirmed using Primer-BLAST (NCBI).

RNA-seq read processing and quantification. The raw reads were initially processed according to our established dual RNA-seq pipeline[14]. Briefly, raw reads were trimmed for adaptor sequences and a minimum read quality of 20 using cutadapt[62]. Reads were then mapped against the human (GRCh38) and Ot (UT176 accession: LS398547.1; Karp accession: LS398548.1) reference sequences using the READemption pipeline (v0.4.3[63]) and segemehl with the lack remapper (v0.2.0[64]), removing reads that mapped equally well to the bacterial and host genomes. For downstream analysis of human gene expression, only uniquely mapping reads were retained for quantification.

To improve quantification of repetitive sequences, reads mapped to the Ot genomes were used for quantification of bacterial transcript expression using Salmon (v0.9.1)[19]. Salmon is a quasi-mapping based gene expression quantification tool that consists of two steps, indexing and quantification.

Transcript fasta files were created from the Genbank annotations using the gene coordinates. The indexing step was performed in quasi-mapping mode (–type quasi). Expression of the transcripts was quantified using both stranded forward library type (-lSF) and removing incompatible mappings (–incompatPrior 0.0). Salmon identified identical gene repeats that are collected in 218 and 127 groups for Karp and UT176, respectively (Supplementary Data 10 and 11). For quantification purposes, we retained a single gene from each group. Antisense reads were quantified in the same way using reverse complemented transcript sequences.

For the purposes of summarizing gene expression, we calculated mean TPM values from three replicates for each strain. Genes with a mean TPM >10 were classified as expressed, and those with a mean TPM value >50 highly expressed.

Gene annotation. For each gene, we retrieved the gene name, gene product, and amino acid sequence from the Genbank annotation. In addition, using eggNOG-mapper[65] we predicted gene names and both KEGG pathways[66] and GO terms. We manually identified surface antigen encoding proteins using BLAST. The

KEGGREST (Tenenbaum, D (2019) KEGGREST: Client-side REST access to KEGG. R package version 1.18.1.) and GO.db (Carlson M (2019). GO.db: A set of annotation maps describing the entire Gene Ontology. R package version 3.5.0.) R packages were used to retrieve KEGG and GO terms, respectively. We additionally added specific annotations for ankyrin and tetrapeptide repeat proteins through manual comparison using BLAST search to annotations in the Ot Ikeda strain annotation [Genbank assembly number GCA_000010205.1].

Non-coding RNA prediction. Non-coding RNAs were annotated using Rock-hopper (v 2.03)[67], ANNOgesic[68] (v0.7.17), and Infernal[69] (v1.1.2) searching sequences against the Rfam database[70]. These provided inconsistent predictions of intergenic sRNAs. Intergenic sRNAs were manually curated by visual comparison of the predicted sRNA coordinates with the read coverage in the Integrative Genomics Viewer[71] (v2.5.2). Infernal predicted the core housekeeping ncRNAs tmRNA, RNase P, SRP, and 5S rRNA. The quantification of the bacterial transcriptomes complemented with predicted ncRNAs was performed using Salmon.

Genomic alignment. Genomic comparisons in Fig. 1d, f were performed using Easyfig[72]. *Escherichia coli* K-12 MG1655 (accession number U00096) and *Salmonella enterica* serovar *Typhimurium* SL1344 (accession number FQ312003) were used as comparators for synteny analysis.

Orthology and conserved operon prediction. We predicted orthologous genes between the two *Orientia* strains using Poff (included in ProteinOrtho v 5.16)[73] with default parameters in synteny mode. To identify conserved operons, we used operon structures predicted in each strain by Rockhopper[36]. Based on visual analysis of read coverage in the Integrative Genomics Viewer, some of the operons were manually extended by addition of genes or merging two operons into one. We also identified partially conserved operons missing some genes in one strain.

Differential gene expression. For the bacteria, differential gene expression analysis was performed between orthologous genes identified by Poff. Genes that were predicted as an orthologous group (more than two genes) were removed from the analysis. In addition, we removed duplicates (transcripts with perfectly identical sequence) that were identified by Salmon in either strain. For both human and bacterial RNA-seq data, we performed differential gene expression analysis with the edgeR package[74] (v3.20.9) using robust quasi-likelihood estimation[75], including genes with CPM (counts per million) > 10 (for Ot) or CPM > 1 (for HUVEC) in at least three libraries. To identify biological processes that differ between two *Orientia* strains, we have performed gene set analysis using KEGG and GO terms that contain at least four expressed genes using the fry test in the edgeR package.

Proteomic sample preparation. Bacteria were propagated in HUVEC cell line at an MOI 70:1 (Karp) and 159:1 (UT176), and harvested at 5 dpi. Ot was isolated, washed with 0.3 M sucrose, and lysed with 1% Triton-X prior to acetone precipitation of protein. Total protein was then alkylated, reduced, and subsequently treated with Lys-C/Trypsin. Digested peptides were desalted using Oasis® HLB reversed-phase cartridges, vacuum dried, and stored for MS runs.

Mass spectrometry. The dried samples were resuspended in 2% (v/v) acetonitrile solution containing 0.06% (v/v) trifluoroacetic acid and 0.5% (v/v) acetic acid, and loaded onto an autosampler plate. Online chromatography was performed using EASY-nLC 1000 (ThermoScientific) in single-column setup using 0.1% formic acid in water and 0.1% formic acid in acetonitrile as mobile phases using reversed-phase C18 column (EASY-Spray LC Column, 75 μm inner diameter × 50 cm, 2 μm particle size) (ThermoScientific). The samples were injected and separated on the analytical column maintained at 50 °C using a 2–23% (v/v) acetonitrile gradient over 60 min, then ramped to 50% over the next 20 min, and finally to 90% within 5 min. The final mixture was maintained for 5 min to elute all remaining peptides. Total run duration for each sample was 90 min at a constant flow rate of 300 nl/min.

Data were acquired using an Orbitrap Fusion mass spectrometer (ThermoScientific) in data-dependent mode. Samples were ionized 2.5 kV and 300 °C at the nanospray source and positively-charged precursor MS1 signals were detected using an Orbitrap analyzer set to 60,000 resolution, automatic gain control (AGC) target of 400,000 ions, and maximum injection time (IT) of 50 ms. Precursors with charges 2–7 and having the highest ion counts in each MS1 scan were further fragmented using collision-induced dissociation (CID) at 35% normalized collision energy and their MS2 signals were analyzed by ion trap at an AGC of 10,000 and maximum IT of 35 ms. Precursors used for MS2 scans were excluded for 90 s in order to avoid re-sampling of high abundance peptides. The MS1–MS2 cycles were repeated every 3 s until completion of the run.

Identification of proteins within each sample was performed using MaxQuant (v1.5.5.1). Raw mass spectra were searched against *Orientia tsutsugamushi* primary protein sequences derived from complete genome data for the Karp and UT176 strains. Human whole proteome sequences were obtained from Uniprot and included as background. Carbamidomethylation on Cys was set as the fixed modification and acetylation on protein N terminus and oxidation of Met were set

as dynamic modifications for the search. Trypsin was set as the digestion enzyme and was allowed up to three missed cleavage sites. Precursors and fragments were accepted if they had a mass error within 20 ppm. Peptides were matched to spectra at a false discovery rate (FDR) of 1% against the decoy database.

Proteomic data analysis. Protein expression was measured by label-free quantification values (LFQs). A protein was classified as detected if at least two peptides were detected in at least two biological replicates, and the mean LFQ across the three replicates was used for further analysis. Otherwise, the protein was classified as undetected, and the LFQ value was set to zero. The proteomic data includes 23 protein groups that could not be resolved, consisting of 97 proteins. In our analysis, we discarded these proteins to simplify the analysis.

Transcript classification. Sense transcript expression was defined by mean TPM value across replicates. The antisense/sense ratio was calculated as the ratio of mean read counts assigned to the antisense and sense strand of coding annotations. The duplicated sequences identified by Salmon (Supplementary Data 10) and non-coding RNAs were removed from the analysis.

We divided the data set into two classes, detected and undetected in proteomics. Within our analyzed data set, 318 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 0.33), but it was not a linear association (Pearson's correlation coefficient = 0.04). For the further analysis, we selected transcripts with sense expression >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 three competing models. Model 1 makes predictions of the protein expression based solely on sense transcription:

$$\beta_0 + \beta_1 * (\text{TPM sense}).$$

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

$$\beta_0 + \beta_1 * ((\text{number of antisense reads})/(\text{number of sense reads})).$$

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

$$\beta_0 + \beta_1 * (\text{TPM sense}) + \beta_2 * ((\text{number of antisense reads})/(\text{number of sense reads})).$$

As data are highly imbalanced, 316 transcripts with detected proteins, and 915 without, we used a downsampling procedure (downSample function) implemented in the caret R package[76] 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 632 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, data were split randomly into two data sets, training and testing, which included 1171 and 60 genes, respectively. Each time the new training data set was reduced to 602 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. precision, recall, and balanced accuracy (caret R package) as well as with ROC curves[77] (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. number 21013) containing 25 μM HPG for 30 min at 37 °C. Labeled bacteria were washed twice in 1× PBS + 1 mg/ml BSA, pH 7.4 before fixing with 4% formaldehyde and subsequently being permeabilized with 0.5% Triton-X 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 3× 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 Plotkin et al.[78]. 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[79,80], and integrated protein abundance for *E. coli* K-12 MG1655 data from PaxDB[81].

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)[82,83]. Selected pathways were chosen based on enrichment *p*-values and activation *Z*-scores, and served as the basis for Figs. 5, 6, and Supplementary Figs. 11, 13, 14, 15, and 16.

Mice and ethics statement. All animal research was performed strictly under protocol approved by the Armed Forces Research Institute of Medical Sciences (AFRIMS) Animal Care and Use Committee and carried out in accordance with the 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. The protocol number was PN16-05. The animal research was conducted in compliance with All animal research adhered to the Guide for the Care and Use of Laboratory Animals, NRC Publication (8th Edition). AFRIMS is an AAALAC International-accredited facility located in Bangkok, Thailand. Mice were co-housed (4 mice/case, 2 cases/group) in standard polycarbonate microisolator cages with filter tops and natural ventilation and stainless steel metal feeding hoppers and water bottle holders at 21 °C, and relative humidity was maintained within the range of 30–70%. The acceptable range is 21 °C ± 1 °C or 20–22 °C.

Female C57BL/6NJcl mice (Inbred) at age of 6–8 weeks (lot numbers 2-37, 2-41, and 2-45) were purchased from Nomura Siam International, Bangkok, Thailand. Mice were housed under specific pathogen-free (SPF) in an animal biosafety level 2 facility, AFRIMS and moved to an animal biosafety level 3 containment, AFRIMS 2 days before the inoculation. Female mice at 6–8 weeks of age were used in these experiments. Two group of female mice ($n = 8$ per group) were intravenously injected in the tail vein with 1.25×10^6 genome copies of *O. tsutsugamushi* of either Karp strain or UT176 strain. The *O. tsutsugamushi* inoculum was derived from *O. tsutsugamushi*-infected L929 cells (kind gift from Stuart Blacksell, Mahidol Oxford Tropical Medicine Research Unit, Bangkok, Thailand). Clinical signs and body weight were evaluated daily. After 12 days post inoculation, all mice were killed. Blood and tissue samples including lungs, liver, spleen, and kidneys were collected for bacteria quantification and histopathology. Adult mice were humanely euthanized with CO_2 inhalation. Gas flow at 2 l/min (at 15 psi CO_2) were maintained in the euthanasia chamber at least 5 min after the animals stop breathing. Death were confirmed by physical examination (the absence of a heartbeat) and ensured by an adjunctive physical method such as cervical dislocation or exsanguination.

Reporting summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability
Sequencing data have been deposited in GEO with accession number GSE139498. Proteomics data have been deposited in jPOSTrepo with accession number PXD017956. Source data are provided with this paper.

References
1. Luce-Fedrow, A. et al. A review of scrub typhus (*Orientia tsutsugamushi* and related organisms): then, now, and tomorrow. *Trop. Med. Infect. Dis.* **3**, 8 (2018).

2. Izzard, L. et al. 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. et al. Endemic scrub typhus in South America. *N. Engl. J. Med.* **375**, 954–961 (2016).

4. Paris, D. H. et al. *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. 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–S230 (2009).

7. Varghese, G. M. et al. 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).

8.   James, S. L. et al. Antigenic relationships among human pathogenic *Orientia tsutsugamushi* isolates from Thailand. *PLoS Negl. Trop. Dis.* **10**, e0004723 (2016).

9.   Batty, E. M, et al. Long-read whole genome sequencing and comparative analysis of six strains of the human pathogen *Orientia tsutsugamushi*. *PLoS Negl. Trop. Dis.* **12**, e0006566 (2018).

10. 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. *Micro. Pathog.* **39**, 177–187 (2005).

11. 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).

12. Sunyakumthorn, P. et al. 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).

13. Groves, M. G. & Kelly, D. J. Characterization of factors determining *Rickettsia tsutsugamushi* pathogenicity for mice. *Infect. Immun.* **57**, 1476–1482 (1989).

14. Westermann, A. J. et al. Dual RNA-seq unveils noncoding RNA functions in host–pathogen interactions. *Nature* **529**, 496–501(2016).

15. Westermann, A. J., Barquist, L. & Vogel, J. Resolving host-pathogen interactions by dual RNA-seq. *PLoS Pathog.* **13**, e1006033 (2017).

16. Humphrys, M. S. et al. Simultaneous transcriptional profiling of bacteria and their host cells. *PLoS ONE* **8**, e80597 (2013).

17. 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).

18. 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).

19. 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).

20. Withey, J. H. & Friedman, D. I. A salvage pathway for protein structures: tmRNA and trans-translation. *Annu. Rev. Microbiol.* **57**, 101–123 (2003).

21. 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).

22. 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).

23. Mao, C. et al. Variations on the tmRNA gene. *RNA Biol.* **6**, 355–361 (2009).

24. 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. USA* **97**, 7778–7783 (2000).

25. 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).

26. Sharkady, S. M. & Williams, K. P. A third lineage with two-piece tmRNA. *Nucleic Acids Res.* **32**, 4531–4538 (2004).

27. Storz, G., Vogel, J. & Wassarman, K. M. Regulation by small RNAs in bacteria: expanding frontiers. *Mol. Cell* **43**, 880–891 (2011).

28. 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).

29. Albrecht, M. et al. The transcriptional landscape of *Chlamydia pneumoniae*. *Genome Biol.* **12**, R98 (2011).

30. Kröger, C. et al. The transcriptional landscape and small RNAs of *Salmonella enterica* serovar *Typhimurium*. *Proc. Natl Acad. Sci. USA* **109**, E1277–E1286 (2012).

31. Sharma, C. M. et al. The primary transcriptome of the major human pathogen *Helicobacter pylori*. *Nature* **464**, 250–255 (2010).

32. Toledo-Arana, A. et al. The *Listeria* transcriptional landscape from saprophytism to virulence. *Nature* **459**, 950–956 (2009).

33. Vogel, J. et al. RNomics in *Escherichia coli* detects new sRNA species and indicates parallel transcriptional output in bacteria. *Nucleic Acids Res.* **31**, 6435–6443 (2003).

34. 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).

35. Nakayama, K. et al. 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).

36. Tjaden, B. De novo assembly of bacterial transcriptomes from RNA-seq data. *Genome Biol.* **16**, 1 (2015).

37. Andersson, S. G. et al. The genome sequence of *Rickettsia prowazekii* and the origin of mitochondria. *Nature* **396**, 133–140 (1998).

38. Georg, J. & Hess, W. R. Widespread antisense transcription in prokaryotes. *Microbiol. Spectr.* **6**, RWR-0029-2018 (2018).

39. Wade, J. T. & Grainger, D. C. Pervasive transcription: illuminating the dark matter of bacterial transcriptomes. *Nat. Rev. Microbiol.* **12**, 647–653 (2014).

40. Raghavan, R., Sloan, D. B. & Ochman, H. Antisense transcription is pervasive but rarely conserved in enteric bacteria. *MBio* **3**, e00156-12 (2012).

41. Lloréns-Rico, V. et al. Bacterial antisense RNAs are mainly the product of transcriptional noise. *Sci. Adv.* **2**, e1501363 (2016).

42. Ha, N. Y. et al. Detection of antibodies against *Orientia tsutsugamushi* Sca proteins in scrub typhus patients and genetic variation of sca genes of different strains. *Clin. Vaccin. Immunol.* **19**, 1442–1451 (2012).

43. Ha, N. Y. et al. Immunization with an autotransporter protein of *Orientia tsutsugamushi* provides protective immunity against scrub typhus. *PLoS Negl. Trop. Dis.* **9**, e0003585 (2015).

44. Ha, N. Y. et al. Generation of protective immunity against Orientia tsutsugamushi infection by immunization with a zinc oxide nanoparticle combined with ScaA antigen. *J. Nanobiotechnol.* **14**, 76 (2016).

45. Cho, N. et al. Expression of chemokine genes in murine macrophages infected with *Orientia tsutsugamushi*. *Infect. Immun.* **68**, 594–602 (2000).

46. VieBrock, L. et al. 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).

47. 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).

48. 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).

49. 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. *Microb. Pathog.* **36**, 51–57 (2004).

50. Tantibhedhyangkul, W. et al. *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).

51. Tantibhedhyangkul, W. et al. Orientia tsutsugamushi, the causative agent of scrub typhus, induces an inflammatory program in human macrophages. *Microb. Pathog.* **55**, 55–63 (2013).

52. Cho, K. A. et al. *Orientia tsutsugamushi* induced endothelial cell activation via the NOD1-IL-32 pathway. *Microb. Pathog.* **49**, 95–104 (2010).

53. Gharaibeh, M. et al. Toll-like receptor 2 recognizes *Orientia tsutsugamushi* and increases susceptibility to murine experimental scrub typhus. *Infect. Immun.* **84**, 3379–3387 (2016).

54. Kawai, T. & Akira, S. Signaling to NF-kappaB by Toll-like receptors. *Trends Mol. Med.* **13**, 460–469 (2007).

55. Shelite, T. R. et al. IL-33-dependent endothelial activation contributes to apoptosis and renal injury in *Orientia tsutsugamushi*-infected mice. *PLoS Negl. Trop. Dis.* **10**, e0004467 (2016).

56. McLeod, M. P. et al. Complete genome sequence of *Rickettsia typhi* and comparison with sequences of other rickettsiae. *J. Bacteriol.* **186**, 5842–5855 (2004).

57. Ellis, M. J. & Haniford, D. B. Riboregulation of bacterial and archaeal transposition. *Wiley Interdiscip. Rev. RNA* **7**, 382–398 (2016).

58. Jose, B. R., Gardner, P. P. & Barquist, L. Transcriptional noise and exaptation as sources for bacterial sRNAs. *Biochem. Soc. Trans.* **47**, 527–539 (2019).

59. Giengkam, S. et al. Improved quantification, propagation, purification and storage of the obligate intracellular human pathogen *Orientia tsutsugamushi*. *PLoS Negl. Trop. Dis.* **9**, e0004009 (2015).

60. Atwal, S. et al. Evidence for a peptidoglycan-like structure in Orientia tsutsugamushi. *Molecular microbiology*, **105**, 440–452 (2017).

61. 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).

62. Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet J.* **17**, 10–12 (2011).

63. 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).

64. Otto, C., Stadler, P. F. & Hoffmann, S. Lacking alignments? The next-generation sequencing mapper segemehl revisited. *Bioinformatics* **30**, 1837–1843 (2014).

65. Huerta-Cepas, J. et al. eggNOG 4.5: a hierarchical orthology framework with improved functional annotations for eukaryotic, prokaryotic and viral sequences. *Nucleic Acids Res.* **44**, D286–D293 (2016).

66. Kanehisa, M., Furumichi, M., Tanabe, M., Sato, Y. & Morishima, K. KEGG: new perspectives on genomes, pathways, diseases and drugs. *Nucleic Acids Res.* **45**, D353–D361 (2017).

67. McClure, R. et al. Computational analysis of bacterial RNA-Seq data. *Nucleic Acids Res.* **41**, e140 (2013).

68. Yu, S. H., Vogel, J. & Förstner, K. U. ANNOgesic: a Swiss army knife for the RNA-seq based annotation of bacterial/archaeal genomes. *Gigascience* **7**, giy096 (2018).

69. Nawrocki, E. P. & Eddy, S. R. Infernal 1.1: 100-fold faster RNA homology searches. *Bioinformatics* **29**, 2933–2935 (2013).

70. Kalvari, I. et al. Rfam 13.0: shifting to a genome-centric resource for non-coding RNA families. *Nucleic Acids Res.* **46**, D335–D342 (2018).

71. Robinson, J. T. et al. Integrative genomics viewer. *Nat. Biotechnol.* **29**, 24–26 (2011).

72. Sullivan, M. J., Petty, N. K. & Beatson, S. A. Easyfig: a genome comparison visualizer. *Bioinformatics* **27**, 1009–1010 (2011).

73. Lechner, M. et al. Orthology detection combining clustering and synteny for very large datasets. *PLoS ONE* **9**, e105015 (2014).

74. Robinson, M. D., McCarthy, D. J. & Smyth, G. K. edgeR: a bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**, 139–140 (2010).

75. Chen, Y., Lun, A. T. & Smyth, G. K. From reads to genes to pathways: differential expression analysis of RNA-Seq experiments using Rsubread and the edgeR quasi-likelihood pipeline. *F1000Res* **5**, 1438 (2016).

76. Kuhn, M. Building predictive models in R using the caret package. *J. Stat. Softw.* **28**, 1–26 (2008).

77. Robin, X. et al. pROC: an open-source package for R and S+ to analyze and compare ROC curves. *BMC Bioinformatics* **12**, 77 (2011).

78. Plotkin, J. B. & Kudla, G. Synonymous but not the same: the causes and consequences of codon bias. *Nat. Rev. Genet.* **12**, 32–42 (2011).

79. Chan, P. P. & Lowe, T. M. GtRNAdb: a database of transfer RNA genes detected in genomic sequence. *Nucleic Acids Res.* **37**, D93–D97 (2009).

80. Chan, P. P. & Lowe, T. M. GtRNAdb 2.0: an expanded database of transfer RNA genes identified in complete and draft genomes. *Nucleic Acids Res.* **44**, D184–D189 (2016).

81. Wang, M., Herrmann, C. J., Simonovic, M., Szklarczyk, D. & von Mering, C. Version 4.0 of PaxDb: protein abundance data, integrated across model organisms, tissues, and cell-lines. *Proteomics* **15**, 3163–3168 (2015).

82. Krämer, A., Green, J., Pollard Jr, J. & Tugendreich, S. Causal analysis approaches in ingenuity pathway analysis. *Bioinformatics*, **30**, 523–530 (2014).

83. Subbian, S. et al. Early innate immunity determines outcome of Mycobacterium tuberculosis pulmonary infection in rabbits. *Cell Communication and Signaling* **11**, 60 (2013).

Acknowledgements

Author contributions

Conceptualization, J.V., L.B., and J.S.; investigation, B.M.-G., S.G., A.J.W., J.W., S. C., L.C. W., P.S., and L.B.; formal analysis, B.M.-G., W.K.-C., S.S., R.M.S., and L.B.; resources, R.M.S., J.V., L.B., and J.S.; writing – original draft, L.B. and J.S.; writing—review and editing, A.J.W., L.B., and J.S.; supervision, L.B. and J.S.; funding acquisition, J.V., L.B., and J.S.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at https://doi.org/10.1038/s41467-020-17094-8.

Correspondence and requests for materials should be addressed to L.B. or J.S.

Peer review information *Nature Communications* thanks Christian Keller and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

Reprints and permission information is available at http://www.nature.com/reprints

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.