

# **Pathogenic bacteria experience pervasive RNA polymerase backtracking during infection**

Kaitlyn R. Browning<sup>1</sup> and Houra Merrikh<sup>1,\*</sup>

<sup>1</sup> Vanderbilt University School of Medicine, Department of Biochemistry, Nashville, TN 37232, USA

\* To whom correspondence should be addressed:

Email: [houra.merrikh@vanderbilt.edu](mailto:houra.merrikh@vanderbilt.edu)

# Abstract

Pathogenic bacteria and their eukaryotic hosts are in a constant arms race. Hosts have numerous defense mechanisms at their disposal that not only challenge the bacterial invaders, but have the potential to disrupt molecular transactions along the bacterial chromosome. However, it is unclear how the host impacts association of proteins with the bacterial chromosome at the molecular level during infection. This is partially due to the lack of a method that could detect these events in pathogens while they are within host cells. We developed and optimized a system capable of mapping and measuring levels of bacterial proteins associated with the chromosome while they are actively infecting the host (referred to as PIC-seq). Here, we focused on the dynamics of RNA polymerase (RNAP) movement and association with the chromosome in the pathogenic bacterium *Salmonella enterica* as a model system during infection. Using PIC-seq, we found that RNAP association patterns with the chromosome change during infection genome-wide, including at regions that encode for key virulence genes. Importantly, we found that infection of a host significantly increases RNAP backtracking on the bacterial chromosome. RNAP backtracking is the most common form of disruption to RNAP progress on the chromosome. Interestingly, we found that the resolution of backtracked RNAPs via the anti-backtracking factors GreA and GreB is critical for pathogenesis, revealing a new class of virulence genes. Altogether, our results strongly suggest that infection of a host significantly impacts transcription by disrupting RNAP movement on the chromosome within the bacterial pathogen. The increased backtracking events have important implications not only for efficient transcription, but also for mutation rates as stalled RNAPs increase the levels of mutagenesis.

## Introduction

Antimicrobial resistance is an urgent threat to human health. Bacterial pathogens continue to evolve at a rate that outpaces almost all old and new therapeutics. Resolving this problem requires a general understanding of how hosts and pathogens interact during infection. Numerous studies over the past few decades have demonstrated how different pathogens reach, invade, and proliferate within their target hosts. However, how the host impacts the pathogen during infection at the molecular level, specifically the complexes that function on the bacterial chromosome, has not been thoroughly investigated.

Bacterial pathogens face significant challenges from the moment of initial contact with mammalian hosts. They must successfully survive the harsh host environment and regulate expression of key virulence factors like pathogenicity island-encoded secretion systems, flagella, ion transporters, and stress response genes. For enteric intracellular pathogens, ingested cells must respond to the acidic pH<sup>1</sup> and the presence of reactive nitrogen species<sup>2</sup> within the host stomach. They must also overcome nutritional barriers caused by dense microbiota layers<sup>3</sup> and withstand antimicrobial peptides released by host cells<sup>4</sup>. The bacteria are also heavily bombarded by oxidative stress<sup>5-9</sup>. Although pathogenic bacteria are well equipped to respond to these host defense mechanisms, any number of these stresses during infection could lead to an accumulation of DNA damage, especially oxidative stress. Indeed, studies in *Salmonella enterica*<sup>10,11</sup>, *Staphylococcus aureus*<sup>12</sup>, and *Helicobacter pylori*<sup>13</sup> demonstrate the critical importance of DNA repair during infection.

DNA damage could have numerous consequences for bacterial cells during infection, including perturbations to DNA replication movement through regions of the chromosome that are packed with stalled RNA polymerases (RNAPs), which increase mutagenesis. Additionally, given that RNAPs are the most abundant proteins associated with the chromosome, DNA damage during infection could have a profound impact on transcription. It is well documented that obstacles on the DNA template, such as damaged template bases or DNA-protein adducts, stall transcription elongation complexes<sup>14-17</sup>. Gene expression programs that respond to the host environment are highly intricate and tightly regulated<sup>18,19</sup>. Thus, any unresolved disruption to RNAP progression during infection could be catastrophic to survival of the pathogen. However, the impact that infection of a host has on RNAP dynamics along the pathogen chromosome has not been investigated. This is partially due to the lack of a precise method capable of detecting changes at the molecular level on the chromosome of pathogens while they are within host cells.

In addition to perturbing transcriptional profiles of pathogenic species during infection, disruptions to RNAP movement could increase mutagenesis. Work from our lab and others' have shown that RNAP stalling activates transcription-coupled nucleotide repair, which is mutagenic specifically because of endogenous reactive oxygen species<sup>14,20-24</sup>. Therefore, potential disruptions to RNAP movement induced by the host are likely to not only abrogate the transcriptional profile of the pathogen, but also increase the chances of generating hypervirulent strains and facilitating AMR development due to increased mutagenesis.

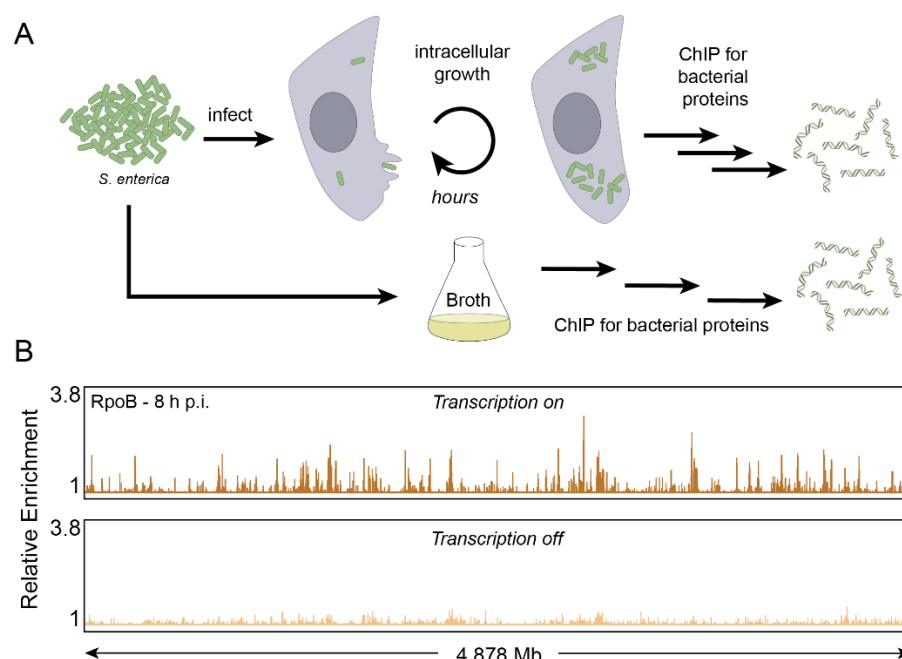
Here, we investigated how the host impacts RNAP movement on the chromosome of a bacterial pathogen. We used the facultative intracellular pathogen *Salmonella enterica* serovar Typhimurium (*S. Typhimurium*) as our model organism. We developed a novel method, post-infection ChIP-seq (PIC-seq), that can measure RNAP density and changes in RNAP movement on the bacteria chromosome while the pathogen is inside the host cell. We show that RNAP backtracking, the most common form of stalled RNAP complexes (described below), is significantly more prevalent during infection compared to cells grown in broth culture. Furthermore, we find that these disruptions to RNAP progression occur across the genome, including at regions encoding for key *S. Typhimurium* virulence genes. We also show that

resolution of backtracking is critical to pathogenesis, highlighting a key role for anti-backtracking factors GreA and GreB in infections. Our results suggest that RNAP movement is substantially impacted by the host, and that the resolution of these disruptions is critical for the survival of pathogens during infection.

## Results

### The development of a method to measure changes in bacterial protein association with the chromosome during infection

We hypothesized that RNAP backtracking is more prevalent during infection genome-wide, and that resolution is critical to pathogenesis. To test this hypothesis, we needed a system that would allow us to map RNAP occupancy on the chromosome of the pathogen while the bacteria still resided within host cells. To our knowledge, such a method did not exist. We developed and optimized a system whereby well-established infection protocols are paired with chromatin immunoprecipitation (ChIP) of bacterial proteins, followed by deep sequencing of the immunoprecipitating DNA (Fig 1A). We refer to this method as post-infection ChIP-seq (PIC-seq). PIC-seq enabled us to determine RNAP dynamics at the molecular level specifically during infection.



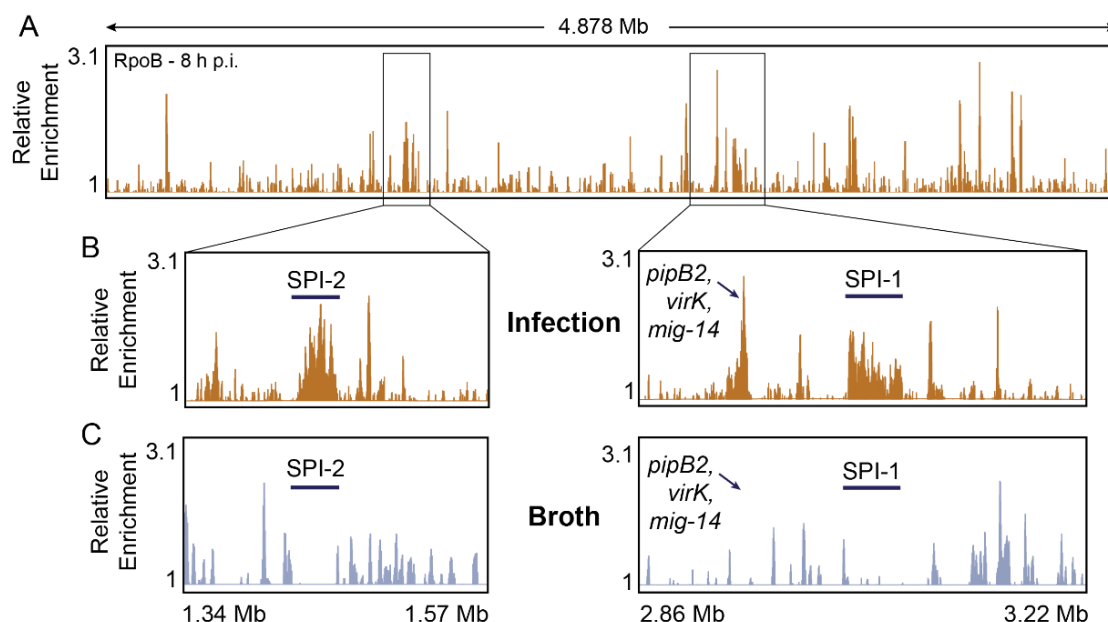
**Figure 1. PIC-seq method can be used to measure RNAP occupancy during infection.** (A) Schematic of the method to perform chromatin immunoprecipitation (ChIP) followed by deep sequencing from broth culture and from infected cells. Eukaryotic host cells are infected with pathogenic bacteria cells. Once the infection has progressed for a given time, ChIP experiments can be performed. (B) Representative example of *S. Typhimurium* RNAP occupancy 8 h post infection (p.i.) of HeLa cells as determined by PIC-seq of RpoB (the beta subunit of RNAP). Inhibiting transcription with rifampicin during the infection (“transcription off”) abrogates RpoB (RNAP) enrichment genome wide. Relative enrichment is defined as the ratio of IP and total read counts. Total genome size is 4.878 Mb.

We used *Salmonella enterica* serovar Typhimurium (*S. Typhimurium*) to infect HeLa cells based on highly standardized protocols<sup>9,25–27</sup>. At the end of the experiment, bacterial proteins were chemically crosslinked to the DNA while the bacteria still resided within the host cell. Crosslinking the entire infection system in this way ensured that any transient protein interactions within the pathogen that may be unique to infection were not lost during sample processing. Both bacterial and host cells were then lysed and their DNA was fragmented by sonication. RpoB, the  $\beta$  subunit of RNAP, was immunoprecipitated from the mixed lysate using a native antibody. Using Western blot analyses, we determined that the antibody is highly specific to RpoB and does not cross react with proteins specific to HeLa cells (S1 Fig). The associated DNA was then extracted and deep sequenced. Any contaminating DNA from the host is naturally excluded from further analysis, as it does not map to the *S. Typhimurium* genome. On average, 12.4% of total reads mapped to the *S. Typhimurium* genome in our experiments, providing an average of approximately 80x coverage (approximately 2.6 million aligned reads per sample). Though we used this system to immunoprecipitate RNAP from *Salmonella*, PIC-seq can, in principle, be adapted to examine any protein of interest with a specific antibody in any host-pathogen combination.

We measured RNAP association with the pathogen genome with PIC-seq. In parallel, as a control to determine if our signal was specific to the bacterial pathogen, we treated cells with the antibiotic rifampicin, which is a transcription inhibitor, during the infection. Upon rifampicin treatment, the signal for RNAP occupancy was abrogated genome-wide, demonstrating the specificity of PIC-seq (Fig 1B).

### **RNAP association with virulence genes can be specifically detected during infection**

To further confirm that our method can differentiate between cells grown in broth culture and infection, we analyzed the levels of RNAP occupancy with regions encoding virulence genes (Fig 2A, S2A Fig). As expected, RNAP was significantly enriched at regions of the chromosome containing multiple, successive virulence genes (known as *Salmonella* pathogenicity islands, “SPIs”) during infection (Fig 2B). Importantly, no significant RNAP signal was found at these genes in cells grown in broth (Fig 2C). Similar trends were observed at a trio of virulence genes, *pipB2*, *virK*, and *mig-14*, where there is no RNAP enrichment in cells grown in broth, but clear enrichment during infection. Unsurprisingly, almost all genes with significant RNAP enrichment during infection (but not in broth culture) have a role (confirmed or putative) in virulence (S1 Table). These results are consistent with previous RNA-seq datasets of *S. Typhimurium* during infection<sup>9,28,29</sup>, as well as the well-defined roles of SPI-1 and SPI-2 proteins in facilitating host cell invasion and promoting intracellular survival, respectively<sup>30</sup>. In contrast, we observed similar levels of RNAP at housekeeping gene regions, such as at rDNA genes, in both conditions (S2B Fig, S2 Table). Our results demonstrate that PIC-seq can specifically map and measure the levels of RNAP occupancy during infection. Furthermore, the high RNAP occupancy at virulence genes we observed further supports prior work and the conclusion that these genes are vital during infection.



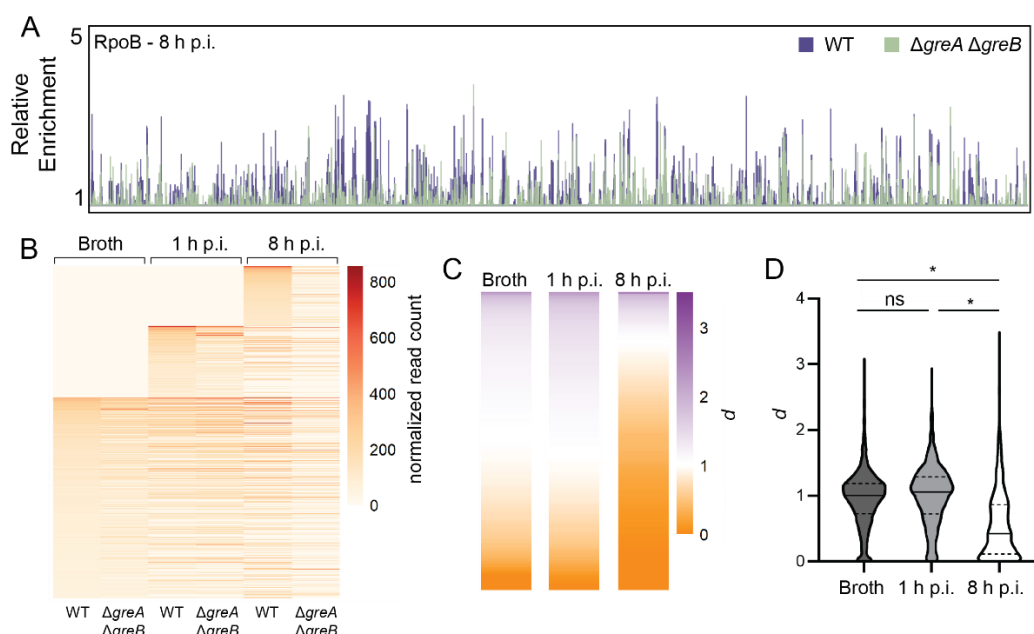
**Figure 2. Specific detection of RNAP occupancy at *S. Typhimurium* virulence genes.** (A) Representative example of *S. Typhimurium* RNAP occupancy 8 h post infection (h p.i.) of HeLa cells as determined by ChIP-seq of RpoB. (B) Reads mapped to the *S. Typhimurium* genome display areas of high RNAP enrichment during infection, notably within *Salmonella* pathogenicity islands (SPIs) and at a trio of virulence genes (*pipB2*, *virK*, and *mig-14*). (C) No significant RNAP signal was found at these genes in cells grown in broth culture. Relative enrichment is defined as the ratio of IP and total read counts.

## RNAP backtracking increases significantly during infection

We hypothesized that RNAP backtracking is more prevalent during infection genome wide. We postulated this based on the fact that the host can damage DNA (such as through oxidative stress), potentially inhibiting RNAP movement. The most common type of RNAP stalling is backtracking, which is a stable state of the complex<sup>31</sup>. When backtracked, the active site of RNAP becomes disengaged, and the 3' end of the nascent RNA is extruded through the secondary channel of the enzyme<sup>32-34</sup>. In *S. Typhimurium*, the anti-backtracking factors GreA and GreB rescue backtracked RNAPs by stimulating the intrinsic endonucleolytic activity of RNAP to restore the active site<sup>35,36</sup>.

To test the hypothesis that there is increased RNAP backtracking on the genome of the pathogen during infection, we constructed strains that lack GreA and GreB<sup>37</sup>. If our hypothesis is correct, then cells lacking the Gre factors would have increased levels of stalled RNAPs during infection compared to growth in broth cultures. To detect this, we measured the differences in RNAP occupancy levels across the genome in the presence (wild-type, WT) and absence ( $\Delta greA \Delta greB$ ) of these anti-backtracking factors. Since both Gre factors are directly and specifically involved in the resolution of backtracked RNAPs<sup>35,36</sup>, we deduced that sites where RNAP occupancy changes in the absence of Gre factors are sites where backtracking is occurring. The prevalence of backtracking can then be measured by determining how these RNAP occupancy differences in the presence and absence of Gre factors change in cells grown in broth versus during infection.

We measured RNAP occupancy in the presence and absence of Gre factors in cells grown in broth and at 1 or 8 h p.i. using PIC-seq. Indeed, deletion of both Gre factors led to changes in RNAP occupancy genome-wide (Fig 3A, S3A-B Fig). We then identified genes that exhibited differential RNAP occupancy in the presence and absence of Gre factors in each condition. However, because cells grown in these conditions have widely different gene expression profiles, we could not directly compare backtracking in the same gene across each condition. In other words, we could not compare backtracking at a gene that is transcriptionally active in one condition but not in the other. Therefore, we categorized the genes by transcription level, which was determined for each gene by normalizing the number of reads mapping to that gene in the IP sample versus the total sample in the WT strain. Genes were categorized independently for each condition. This resulted in 631 top transcribed genes in broth, 645 at 1 h p.i., and 590 at 8 h p.i. (S3 Table, S3C Fig). When comparing the 283 top transcribed genes that are shared between all three conditions, it is clear that RNAP occupancy differs in the presence and absence of Gre factors, not only in each independent condition, but also across conditions (Fig 3B).



**Figure 3. RNAP occupancy decreases in the absence of Gre factors.** (A) Representative RNAP occupancy profile for wild-type (WT) cells or cells lacking Gre factors ( $\Delta greA \Delta greB$ ) 8 h post infection (p.i.) of HeLa cells as determined by ChIP-seq of RpoB. Relative enrichment is defined as the ratio of IP and total read counts. (B) RNAP occupancy changes visualized as a heatmap, where every horizontal line represents the normalized read count for the same gene across each condition. All top transcribed genes for each condition are plotted (1043 genes total, see S3 Table). Each value represents the average of three biological replicates. Zeros are plotted for genes that do not arise in one condition but do arise in one or both of the other conditions. (C) Ratio  $d$  of normalized read count in the absence of Gre factors versus in the presence of Gre factors (WT) for top-transcribed genes in each condition averaged across three biological replicates. (D) Quantification of (C) showing ratio  $d$  calculated for every top-transcribed gene, where the median is shown by a solid line and quartiles by dashed lines. \* $p < 0.0001$ , ns: not significant, one-way ANOVA.

We found that RNAP occupancy generally decreases in the absence of Gre factors. This is consistent with the activity of previously described accessory factors, such as helicases, that can remove stalled RNAPs from DNA<sup>16,21,38–42</sup>. Backtracked RNAPs that are not rescued by the Gre factors are expected to eventually become a target for removal by such helicases and other transcription terminators, leading to the decrease in RNAP occupancy that we observe.

We performed a more quantitative analysis to get an accurate picture of RNAP occupancy changes for cells grown during infection versus growth in broth culture. We defined a ratio,  $d$ , as the normalized read counts mapping to a gene in the backtracking-prone state ( $\Delta greA \Delta greB$ ) versus those in WT. An equivalent ratio ( $d \approx 1$ ) indicates that there is little difference in RNAP occupancy independent of the presence of the Gre factors. We interpreted these regions to be those that are not experiencing backtracking. A ratio that deviates from one ( $d < 1$  or  $d > 1$ ) indicates that there is a difference in RNAP occupancy at the gene that is dependent on the Gre factors, suggesting that these factors are playing an important role in maintaining RNAP progression at that gene. We interpreted these regions to be those that are experiencing significant RNAP backtracking. This ratio was calculated for every top-transcribed gene in each condition (Fig 3C, S3D Fig). For cells grown in broth and at 1 h p.i., the average ratio of RNAP occupancy in the absence of Gre factors versus in the presence is equivalent (Table 1, Fig 3D), suggesting that, on average, cells do not depend on Gre factors to maintain proper RNAP occupancy under these conditions. In contrast, cells at 8 h p.i. do rely on the Gre factors to maintain RNAP occupancy (Table 1, Fig 3D), suggesting that backtracking becomes pervasive as the infection progresses.

**Table 1. Gre factor-dependent RNAP occupancy changes in broth versus infection.**

Transcription level	$d_{broth}$	$d_{1 \text{ h p.i.}}$	$d_{8 \text{ h p.i.}}$	$d'$ 1 h p.i. vs. broth	$d'$ 8 h p.i. vs. broth
++	0.96	1.01	0.56	1.06	0.59
+++	0.91	0.99	0.52	1.09	0.57
++++	0.91	0.94	0.48	1.03	0.53
All genes	0.94	1.00	0.55	1.06	0.59

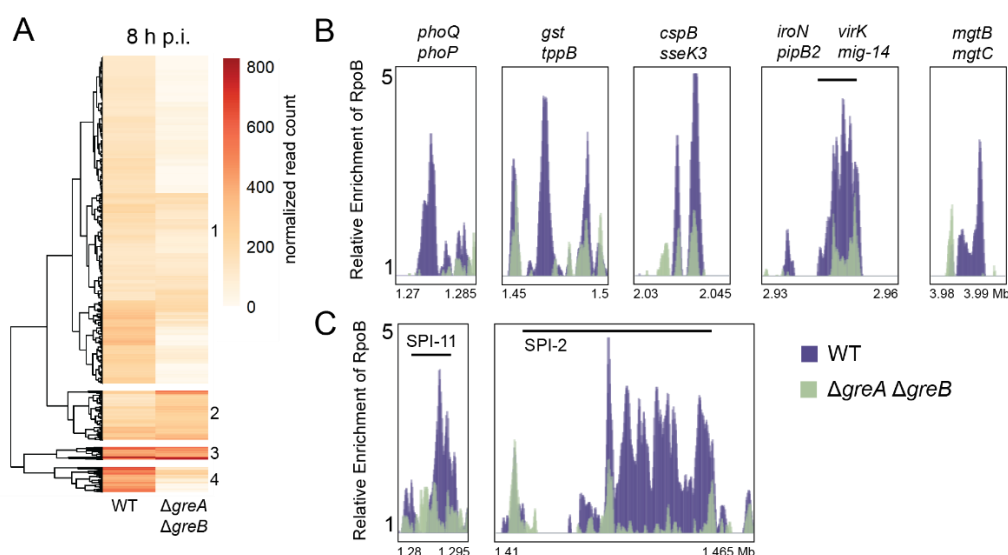
Ratio  $d$  was calculated as the normalized read counts mapping to a gene in the backtracking-prone state ( $\Delta greA \Delta greB$ ) versus in WT for each condition and each transcription level (++, +++, ++++). Ratio  $d'$  comparing ratio  $d$  for 1 or 8 h p.i. to broth for the groups of genes with equivalent transcription levels.

We determined the extent to which backtracking is more prevalent during infection than broth. We accomplished this by quantifying the average differences in RNAP occupancy changes in broth versus infection. We defined a new ratio,  $d'$ , as the ratio of  $d$  for infection versus  $d$  for broth,  $d' = d_{inf} / d_{broth}$ . Backtracking early in infection, at 1 h p.i., is not more prevalent than in broth, but it is more prevalent at 8 h p.i.,  $d' = 0.59$  (Table 1).

### Backtracking occurs within key genes necessary for *Salmonella* pathogenesis

Having established that backtracking is more prevalent during infection genome-wide, specifically at later time points, we endeavored to identify those regions where backtracking occurs during infection. Thus, we looked at the genes where the greatest changes in RNAP occupancy in the absence of Gre factors

occur. These were identified by plotting RNAP occupancy as heat maps and using hierarchical clustering (Fig 4A, S4 Fig, S4 Table).



**Figure 4. Backtracking is prevalent at key virulence genes during infection.** (A) RNAP occupancy changes at 8 h post infection (p.i.) in the absence of Gre factors visualized as a heat map. Each value represented is the average of three biological replicates. Hierarchical clustering (numbered 1-4) was performed using the pheatmap function in RStudio (see S4 Table). (B) Key genes related to virulence in cluster 4 in (A) exhibit significant differences in RNAP occupancy ( $p < 0.05$  for every gene listed, unpaired t test). (C) Genes belonging to SPIs that fall into cluster 1 in (A) and exhibit significant differences in RNAP occupancy.

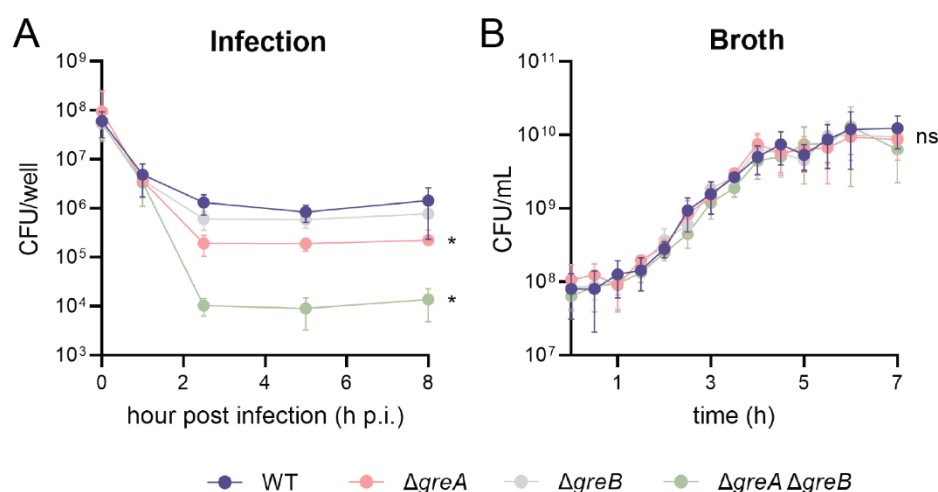
As backtracking is minimal throughout all clusters in broth and at 1 h p.i. (Fig 3, S4 Fig), we focused on characterizing gene functions for each cluster at 8 h p.i. (Fig 4A). Genes at 8 h p.i. fall into four clusters: cluster one, 466 genes; cluster two, 70 genes; cluster three, 18 genes; and cluster four, 36 genes (S4 Table). In clusters four and one, RNAP occupancy is generally higher in the presence of Gre factors (WT) than in the absence ( $\Delta greA \Delta greB$ ). Genes in cluster four exhibit dramatic differences in RNAP occupancy depending on the presence of Gre factors, with an average  $d$  of 0.23. There was not enough statistical power to perform gene ontology (GO) functional enrichment analysis for genes in cluster four. However, manual annotation of these gene functions reveals that out of 36 genes, at least 15 have been previously implicated in *S. Typhimurium* virulence (Fig 4B, S4 Table). Although RNAP occupancy does depend on the presence of Gre factors in cluster one ( $d = 0.43$ ), these genes have lower RNAP occupancy levels in both genotypes (Fig 4A). GO functional enrichment analysis revealed that genes in cluster one are enriched in functions related to cellular and macromolecular localization, membrane biogenesis, regulation of metabolic processes, secretion, and translation. This cluster also consists of numerous SPI genes that exhibit large differences in RNAP occupancy when Gre factors are absent (Fig 4C, S4 Table). Of the 25 top transcribed SPI-2 genes in this dataset, 22 fall within cluster one. Similarly, all five of the top transcribed SPI-11 genes in this dataset fall within cluster one. Efficient RNAP progression at these genes would be critical to ensuring cell survival during infection, as these genes are required throughout *S. Typhimurium* infection.

Genes in cluster two exhibit an average  $d$  ratio of 1.41, or greater RNAP occupancy in the absence than in the presence of Gre factors. Genes in cluster two are enriched in functions related to cellular nitrogen compound biosynthetic processes (GO analysis). Cluster three includes genes where RNAP occupancy is high and does not change in the absence of Gre factors ( $d = 0.99$ ). Manual annotation of genes in cluster three reveal functions related to essential processes such as translation, carbon metabolism, and cell wall maintenance (S4 Table). Overall, our results suggest that backtracking occurs at key virulence genes during infection, and that the Gre factors are critical for RNAP progression at these regions.

We found that the prevalence of backtracking was not uniform genome-wide. We reasoned that longer genes are more likely to accumulate DNA damage and therefore more likely to accumulate backtracked RNAPs. Accordingly, we found that the average length of genes in cluster four, which exhibited the greatest differences in RNAP occupancy in a Gre factor-dependent manner, was significantly longer than the average length of all top-transcribed genes in this dataset (S5 Fig).

### Resolution of backtracking is critical to pathogenesis

Knowing that backtracking is prevalent at key virulence genes during infection, we tested whether disruption to RNAP progression is problematic for pathogenesis in general. To elucidate the impact of prevalent backtracking on *S. Typhimurium* pathogenicity, we infected HeLa cells with backtracking-prone *S. Typhimurium* lacking one or both Gre factors. We measured the number of viable intracellular *S. Typhimurium* from the infection over time using a standard gentamicin protection assay (Fig 5). Deletion of *greB* alone did not significantly affect the growth of *S. Typhimurium* inside HeLa cells over the course of the infection, while deletion of *greA* led to a modest, yet statistically significant, loss in intracellular growth. However, deletion of both Gre factors significantly abrogated intracellular growth of *S. Typhimurium* throughout the duration of the infection (Fig 5A). This effect is specific to infection, as there were no significant differences in growth of these strains compared to wild type when grown in broth (Fig 5B). The infection-specific effect of depleting cells of both Gre factors suggests that pervasive backtracking during infection disrupts key virulence gene expression, and resolution is critical to intracellular survival.



**Figure 5. Resolution of backtracking during infection is key to pathogenesis.** Average number of CFUs of backtracking-prone cells (A) isolated during infection or (B) grown in broth culture. Error bars indicate standard deviation of at least six biological replicates. \* $p < 0.005$ , ns: not significant, one-way ANOVA with Bonferroni correction.

## Discussion

During infection, bacterial pathogens must survive numerous host defense mechanisms that are not only damaging to the cell itself, but can also lead to DNA damage capable of stalling RNAP<sup>14–17</sup>. Pathogens must quickly sense and respond to the host environment, and any disruption to these virulence gene programs at the transcriptional level would be significant. Our work suggests that infection of a host significantly impacts bacterial transcription, causing pervasive backtracking that must be resolved to ensure successful pathogenesis.

Our data show that disruptions to RNAP movement increase as the infection progresses. As evidenced by the lack of difference in RNAP occupancy in the presence and absence of Gre factors, it appears that after 1 h of infection, RNAPs are not yet perturbed by interactions with the host. This is further emphasized by growth of cells at 1 h p.i.: cells at this time point do not require Gre factors to survive during infection. It is possible, however, that RNAP stalling at 1 h p.i. is underestimated in our experiments due to detection thresholds of PIC-seq as a population assay and the heterogeneity of the cell population at this early point post infection<sup>43</sup>. Cells at 8 h p.i., however, have endured stress from the host environment longer than those at 1 h p.i., potentially increasing the likelihood of DNA damage accumulation that will stall RNAPs and lead to backtracking. Cells at this later time point exhibit large differences in RNAP occupancy and, based on infection experiments in the presence and absence of Gre factors, require backtracking resolution for intracellular survival.

In addition to perturbing transcription, an increase in the prevalence of backtracking during infection could have implications for DNA replication. Transcription is an obstacle to DNA replication due to the lack of spatiotemporal separation between the two processes. These conflicts pose significant threats to genome stability regardless of directionality. Interestingly, it has been demonstrated that co-directional conflicts between a backtracked RNAP and an oncoming replication fork lead to double-stranded breaks (DSBs) in the DNA<sup>37</sup>. Our data show that *S. Typhimurium* cells lacking previously described key conflict resolution factors, such as UvrD or Rep helicases<sup>44</sup>, exhibit significant growth defects during infection of HeLa cells (S6 Fig). However, further work is required to determine the extent to which replication-transcription conflicts occur in pathogens during infection. Whether increased backtracking during infection exacerbates conflicts, thus compromising genome integrity, is untested. Previous work has shown that cells that are prone to backtracking (lacking either Gre factor) accumulate mutations faster<sup>37</sup> and have higher rates of recombination than cells with both Gre factors<sup>45</sup>.

Previous literature shows that stalled RNAPs trigger downstream DNA repair pathways that can be mutagenic<sup>14,20–24</sup>. Although pathogenic cells do have the Gre factors and will thus experience less backtracking under regular conditions, the fact that these RNAP stalling events are much more prevalent during infection inherently increases the chances of mutagenic events occurring. Therefore, we postulate that, though the host is fighting the bacteria with DNA damaging mechanisms such as oxidative stress, they are also enabling bacteria to attain mutations that could lead to adaptive evolution. This idea is outside the scope of this work; however, it is certainly interesting to investigate in future studies as this dichotomy could accelerate the development of AMR and/or hypervirulence.

Our results also highlight an underappreciated role for Gre factors during infection and suggest the existence of a new class of so-called virulence genes in bacterial pathogens. This class of proteins promote survival and guarantee proper gene expression during infection by regulating and modulating RNAP through transcription initiation, elongation, and termination. Consistent with our findings, in *S. Typhimurium*, Gre factors have been associated with promoting intracellular fitness during oxidative stress by facilitating transcriptional fidelity and elongation of central metabolic genes<sup>46</sup>. They have also been demonstrated to prevent backtracking within one gene, *hilD*, which encodes a master regulator of SPI genes to promote pathogenesis<sup>47</sup>, and to facilitate biofilm formation through mediating transcription elongation at key effector genes<sup>48</sup>. Similar impact of Gre factors on virulence phenotypes have also been shown in other, highly divergent bacterial species which highlights the conservation of this phenomenon<sup>49-51</sup>.

Precise regulation of gene expression during infection is critical to the survival of bacterial pathogens. Our data strongly suggest that DNA damage to the pathogen resulting from host defense mechanisms threatens this precision. Disruptions to RNAP progression must be resolved. Altogether, our results support the model that backtracking is prevalent during infection and that Gre factors are required to ensure proper gene regulation and survival during infection.

## Methods

### Bacterial strains and growth conditions

*Salmonella enterica* serovar Typhimurium SL1344 was the wild-type (WT) strain used in these studies. Derivative mutant strains are listed in S5 Table. Bacteria were grown on LB-Lennox agar plates at 37°C with the following antibiotics when appropriate: 25 mg/mL chloramphenicol or 50 mg/mL kanamycin. Single colonies were used to inoculate liquid cultures in LB-Lennox and grown at 37°C with aeration (260 rpm).

### Construction of chromosomal deletion mutants

The bacteriophage  $\lambda$  Red recombination system was used for construction of chromosomal deletion mutants<sup>52,53</sup>. The plasmid pSIM27 was transformed into wild-type SL1344 for expression of the Red recombinase system. Kanamycin and chloramphenicol resistance cassettes were amplified with Phusion High-Fidelity polymerase (Thermo) from T-SACK gDNA<sup>54</sup> using primers that contained 40 nucleotides homologous to the regions flanking the *greA* or *greB* (including the ATG start codon, S5 Table). Amplicons were electroporated separately into SL1344 and cells were grown on selective LB agar to yield single knockout strains  $\Delta greA$  (*greA*::Kan, HM4527) and  $\Delta greB$  (*greB*::Cat, HM4525). To make the double knockout strain  $\Delta greA \Delta greB$  (HM4529), the *greA* gene was replaced with the kanamycin resistance cassette in the  $\Delta greB$  strain as above. Every deletion was confirmed by Sanger sequencing.

### Mammalian cell culture

HeLa cells (ATCC) were grown in high glucose Dulbecco's minimal essential medium (DMEM, Gibco – 11995065) supplemented with 10% heat inactivated fetal bovine serum (FBS, R&D Systems), 4 mM L-glutamine, and 1X Penicillin/Streptomycin (Gibco – 15140122). Antibiotic-free media of the same formulation was used for bacterial infections. Cells were maintained at 37°C with 5% CO<sub>2</sub> and passaged following ATCC guidelines.

## Seeding and bacterial invasion of mammalian cells

HeLa cells at low passage number were seeded at the following densities 16-18 hours before infection:  $1.5 \times 10^5$  cells per well (24-well dish) or  $1.14 \times 10^7$  (15-cm plate). Immediately prior to infection, HeLa cells were washed with 1X PBS (Gibco), and antibiotic-free media was added to the cells. Cultures of *S. Typhimurium* were inoculated in LB from single colonies and grown overnight at 37°C while shaking. The next day, the precultures were diluted back to OD<sub>600</sub> = 0.05 in LB and grown at 37°C while shaking until OD<sub>600</sub> reached 0.6 (approximately 2 h). Bacteria were collected by centrifugation, washed in 1X PBS, resuspended in antibiotic-free media, and used immediately to infect HeLa cells at an MOI of ~100:1. Bacteria were allowed to invade for 1 h at 37°C with 5% CO<sub>2</sub>. Bacteria were removed after invasion, and HeLa cells were washed once with 1X PBS and fresh antibiotic-free media was added. Thirty minutes later (1.5 h post infection [p.i.]), gentamicin was added to a final concentration of 50 ug/mL for the duration of the experiment. Infected HeLa cells were maintained at 37°C with 5% CO<sub>2</sub> until times indicated below.

## Gentamicin protection assays

HeLa cells were seeded in 24-well dishes and infected with bacteria as described above. Extracellular growth of *S. Typhimurium* was inhibited by gentamicin as described above. At indicated time points, infected HeLa cells were washed once with 1X PBS and lysed with ice-cold 1% Triton X-100 in H<sub>2</sub>O. Viable bacteria were enumerated by plating on LB agar and grown at 37°C overnight. A one-way ANOVA was performed at 2.5, 5 and 8 h p.i. timepoints with Bonferroni correction to determine statistical significance.

## Chromatin immunoprecipitation from infection

HeLa cells were seeded in 15-cm plates and infected with bacteria as described above. For treatment with rifampicin, 100 mg/mL was added to the media ten minutes prior to each time point. At indicated time points, infected HeLa cells were washed once with 1X PBS and crosslinked in 1X PBS + 1% methanol-free formaldehyde (Thermo) in H<sub>2</sub>O for 10 min at room temperature and subsequently quenched with 0.5 M glycine. Crosslinked infected HeLa cells were washed twice with 1X PBS, then dislodged in 1X PBS by scraping. Cells from two 15-cm plates were combined for each replicate and collected by centrifugation. Cell pellets were stored at -80°C for future processing. Pellets were thawed on ice and resuspended in 2.5 mL ice-cold NPT lysis buffer<sup>55</sup> [(50 mM Tris-HCl pH 7.5, 150 mM NaCl, 5 mM EDTA, 0.5 % NP-40, 0.1 % Triton X-100, complete protease inhibitor cocktail [Roche] added fresh) for 10 min on ice. Lysozyme was added to 10 mg/mL and the lysate was incubated at 37°C for 30 min. Lysates were sonicated for 5 cycles of 30 s on/off (2.5 min total sonication time) at 4°C in a Bioruptor Plus sonication system (Diagenode) and pelleted by centrifugation at 8000 RPM for 15 min at 4°C. A 40 µL aliquot was taken from the lysate supernatant as the input control. For the immunoprecipitation, 6 µL RpoB monoclonal antibody (8RB13, Thermo) was added to the lysate supernatant and rotated overnight at 4°C. The next day, 90 µL of a 50% protein A Sepharose bead slurry (GE) was added, and IPs were incubated for 1 h at room temperature with gentle rotation. Beads were pelleted by centrifugation at 2000 RPM for 1 min. The supernatant was discarded, and the beads were washed six times for 3 min each in wash buffer (50 mM Tris-HCl pH 7.0, 150 mM NaCl, 5 mM EDTA, 1% Triton X-100), followed by one wash with TE pH 8.0. The elution was carried out at 65°C for 10 min in 200 µL elution buffer I (50 mM Tris pH 8.0, 10 mM EDTA, 1% SDS). Beads were pelleted at 5000 RPM for 1 min and the supernatant saved. The beads were washed with 150 µL elution buffer II (10 mM Tris-HCl pH 8.0, 1 mM EDTA, 0.67% SDS) and pelleted at 7000 RPM for 1 min. The second supernatant was combined with the first eluate. The combined eluates and the accompanying input controls were de-crosslinked overnight by

incubation at 65°C. The following day, the eluates and input controls were treated with proteinase K (0.4 mg/mL) at 37°C for 2 h. Sodium acetate was added and the DNA purified by phenol:chloroform:isoamyl alcohol extraction. The DNA was precipitated in 100% ethanol at -20°C and pelleted at 14,000 RPM for 15 min at 4°C, before being resuspended in resuspension buffer (10 mM Tris-HCl pH 8.0, 1 mM EDTA).

### **Chromatin immunoprecipitation from broth**

Cultures of *S. Typhimurium* were inoculated in LB from single colonies and grown overnight at 37°C while shaking. The next day, the precultures were diluted back to OD<sub>600</sub> = 0.05 in LB and grown at 37°C while shaking until OD<sub>600</sub> reached 0.6 (approximately 2 h). Bacteria were crosslinked with 1% formaldehyde for 20 min at room temperature before being quenched with 0.5 M glycine. Cells were collected by centrifugation and washed once in cold 1X PBS. Cell pellets were resuspended in 1.5 mL Solution A (10 mM Tris-HCl pH 8.0, 20% w/v sucrose, 50 mM NaCl, 10 mM EDTA, 10 mg/ml lysozyme, 1 mM AEBSF) and incubated at 37°C for 30 min. After incubation, 1.5 mL of 2X IP buffer (100 mM Tris pH 7.0, 10 mM EDTA, 2% Triton X-100, 300 mM NaCl and 1 mM AEBSF, added fresh) was added and lysates were chilled on ice for 30 min. Lysates were sonicated four times for 10 s (40 s total sonication time) at 30% amplitude and pelleted by centrifugation at 8000 RPM for 15 min at 4°C. A 40 µL aliquot was taken from the lysate supernatant as the input control. For the immunoprecipitation, 2 µL of RpoB monoclonal antibody (clone 8RB13, Thermo) was added to 1 mL of the lysate supernatant and rotated overnight at 4°C. The next day, 30 µL of 50% protein A Sepharose bead slurry (GE) was added, and IPs were incubated for 1 h at room temperature with gentle rotation. Beads were pelleted by centrifugation at 2000 RPM for 1 min. The supernatant was discarded, and the beads were washed six times for 3 min each in wash buffer (50 mM Tris-HCl pH 7.0, 150 mM NaCl, 5 mM EDTA, 1% Triton X-100), followed by one wash with TE pH 8.0. The elution was carried out at 65°C for 10 min in 100 µL elution buffer I (50 mM Tris pH 8.0, 10 mM EDTA, 1% SDS). Beads were pelleted at 5000 RPM for 1 min and the supernatant saved. The beads were washed with 150 µL elution buffer II (10 mM Tris-HCl pH 8.0, 1 mM EDTA, 0.67% SDS) and pelleted at 7000 RPM for 1 min. The second supernatant was combined with the first elution. The combined eluates and the accompanying input controls were de-crosslinked overnight by incubation at 65°C. The following day, the eluates and input controls were treated with proteinase K (0.4 mg/mL) at 37°C for 2 hr. Sodium acetate was added and the DNA purified by phenol:chloroform:isoamyl alcohol extraction. The DNA was precipitated in 100% ethanol at -20°C for 1 hr and pelleted at 14,000 RPM for 15 min at 4°C, before being resuspended in resuspension buffer (10 mM Tris-HCl pH 8.0, 1 mM EDTA).

### **Deep sequencing and data processing**

Sequencing libraries were prepared using a Nextera XT DNA Library Preparation kit (Illumina). Libraries were deep sequenced by the Vanderbilt Technology for Advanced Genomics (VANTAGE) sequencing core on an Illumina NovaSeq platform, resulting in approximately 24M x 150 bp paired-end reads per sample. Raw reads were trimmed (Trimmomatic v0.39<sup>56</sup>) then mapped to the *S. enterica* serovar Typhimurium strain SL1344 genome (GenBank: FQ312003.1) using Bowtie2 v2.2.5<sup>57</sup>. Both PCR and optical duplicates were removed using Picard v1.3 (Broad Institute) and bam files were sorted and indexed using SAMtools v1.13<sup>58</sup>.

The number of reads mapping to each gene in both the IP and input samples was quantified using featureCounts v2.0.3<sup>59</sup> and then normalized to total mapped reads. For each gene, the normalized number of reads in the input sample was subtracted from the normalized number of reads in the IP sample and averaged across three biological replicates to calculate normalized read count. For any gene where this calculation resulted in a negative number, the normalized read count was redefined as 0 for that gene. At

this point, rDNA genes were excluded from further analysis. Genes were then categorized into five groups based on transcription level using K-means clustering for each individual condition. Transcription level was defined by the normalized read count in the wild-type sample for each condition. The three groups containing the most transcribed genes for each condition were used for further analysis. Ratio  $d$  was calculated for these genes by dividing the normalized read count for the  $\Delta greA \Delta greB$  strain by the normalized read count in the wild-type strain. Ratio  $d'$  was calculated by dividing  $d$  from the 1 h p.i. or 8 h p.i. condition by the  $d$  from the broth condition. Heatmaps were plotted and hierarchical clustering was performed using the pheatmap package in RStudio. Gene ontology functional enrichment analysis was performed using the PANTHER v.14 pipeline<sup>60</sup> through The Gene Ontology Resource<sup>61,62</sup>.

For enrichment analysis, peaks were called from processed bam files using macs2 v.2.2.7.1<sup>63</sup> and assigned to genome features using the “closest” tool from the BEDTools suite v.2.30.0<sup>64</sup>.

Bedgraph files were prepared from the processed bam files using the bamCompare tool from the deepTools suite v3.5.1<sup>65</sup> and visualized on the Integrated Genomics Viewer (igv) platform v2.12.1<sup>66</sup>.

All statistical tests were performed in GraphPad Prism v9.

## Acknowledgments

We would like to thank all past and present members of the Merrikh lab for their suggestions and advice on this project, especially Kevin Lang and Anna Johnson. We are also indebted to the Cortez and Hodges labs for providing their equipment and expertise throughout this project. This project was funded by T32ES007028 to K.R.B. NIH National Institute grant AI127422 to H.M.

## Contributions

H.M. and K.R.B. designed the experiments, K.R.B. performed the experiments, both H.M. and K.R.B. analyzed the data and wrote the paper.

## References

1. Foster, J. W. & Hall, H. K. Inducible pH homeostasis and the acid tolerance response of *Salmonella typhimurium*. *J. Bacteriol.* **173**, 5129–5135 (1991).
2. Henard, C. A. & Vázquez-Torres, A. Nitric Oxide and *Salmonella* Pathogenesis. *Front. Microbiol.* **2**, 84 (2011).
3. Rogers, A. W. L., Tsolis, R. M. & Bäumlér, A. J. *Salmonella* versus the Microbiome. *Microbiol. Mol. Biol. Rev. MMBR* **85**, e00027-19 (2021).

4. Ouellette, A. J. Paneth Cell  $\alpha$ -Defensins in Enteric Innate Immunity. *Cell. Mol. Life Sci. CMLS* **68**, 2215–2229 (2011).
5. Fang, F. C., Frawley, E. R., Tapscott, T. & Vazquez-Torres, A. Bacterial Stress Responses during Host Infection. *Cell Host Microbe* **20**, 133–143 (2016).
6. Fu, J. *et al.* Salmonella proteomics under oxidative stress reveals coordinated regulation of antioxidant defense with iron metabolism and bacterial virulence. *J. Proteomics* **157**, 52–58 (2017).
7. Garai, P., Gnanadhas, D. P. & Chakravorty, D. Salmonella enterica serovars Typhimurium and Typhi as model organisms. *Virulence* **3**, 377–388 (2012).
8. Imlay, J. A. The molecular mechanisms and physiological consequences of oxidative stress: lessons from a model bacterium. *Nat. Rev. Microbiol.* **11**, 443–454 (2013).
9. Hautefort, I. *et al.* During infection of epithelial cells Salmonella enterica serovar Typhimurium undergoes a time-dependent transcriptional adaptation that results in simultaneous expression of three type 3 secretion systems. *Cell. Microbiol.* **10**, 958–984 (2008).
10. Buchmeier, N. A. *et al.* DNA repair is more important than catalase for Salmonella virulence in mice. *J. Clin. Invest.* **95**, 1047–1053 (1995).
11. Suvarnapunya, A. E., Lagassé, H. A. D. & Stein, M. A. The role of DNA base excision repair in the pathogenesis of Salmonella enterica serovar Typhimurium. *Mol. Microbiol.* **48**, 549–559 (2003).
12. Ha, K. P. *et al.* Staphylococcal DNA Repair Is Required for Infection. *mBio* **11**, e02288-20 (2020).
13. O’Rourke, E. J. *et al.* Pathogen DNA as target for host-generated oxidative stress: Role for repair of bacterial DNA damage in Helicobacter pylori colonization. *Proc. Natl. Acad. Sci.* **100**, 2789–2794 (2003).
14. Agapov, A., Olina, A. & Kulbachinskiy, A. RNA polymerase pausing, stalling and bypass during transcription of damaged DNA: from molecular basis to functional consequences. *Nucleic Acids Res.* **50**, 3018–3041 (2022).
15. Lass-Napiorkowska, A. & Heyduk, T. Real-time observation of backtracking by bacterial RNA polymerase. *Biochemistry* **55**, 647–658 (2016).

16. Epshtein, V., Toulmé, F., Rahmouni, A. R., Borukhov, S. & Nudler, E. Transcription through the roadblocks: the role of RNA polymerase cooperation. *EMBO J.* **22**, 4719–4727 (2003).
17. Walter, W., Kireeva, M. L., Studitsky, V. M. & Kashlev, M. Bacterial polymerase and yeast polymerase II use similar mechanisms for transcription through nucleosomes. *J. Biol. Chem.* **278**, 36148–36156 (2003).
18. Colgan, A. M. *et al.* The Impact of 18 Ancestral and Horizontally-Acquired Regulatory Proteins upon the Transcriptome and sRNA Landscape of *Salmonella enterica* serovar Typhimurium. *PLOS Genet.* **12**, e1006258 (2016).
19. Rhen, M. & Dorman, C. J. Hierarchical gene regulators adapt *Salmonella enterica* to its host milieu. *Int. J. Med. Microbiol. IJMM* **294**, 487–502 (2005).
20. Million-Weaver, S. *et al.* An underlying mechanism for the increased mutagenesis of lagging-strand genes in *Bacillus subtilis*. *Proc. Natl. Acad. Sci. U. S. A.* **112**, E1096-1105 (2015).
21. Ragheb, M. N., Merrikh, C., Browning, K. & Merrikh, H. Mfd regulates RNA polymerase association with hard-to-transcribe regions in vivo, especially those with structured RNAs. *Proc. Natl. Acad. Sci. U. S. A.* **118**, e2008498118 (2021).
22. Carvajal-Garcia, J., Samadpour, A. N., Viera, A. J. H. & Merrikh, H. Oxidative stress drives mutagenesis through transcription coupled repair in bacteria. 2022.06.28.497968 Preprint at <https://doi.org/10.1101/2022.06.28.497968> (2023).
23. Gómez-Marroquín, M. *et al.* Stationary-Phase Mutagenesis in Stressed *Bacillus subtilis* Cells Operates by Mfd-Dependent Mutagenic Pathways. *Genes* **7**, 33 (2016).
24. Ross, C. *et al.* Novel Role of mfd: Effects on Stationary-Phase Mutagenesis in *Bacillus subtilis*. *J. Bacteriol.* **188**, 7512–7520 (2006).
25. Klein, J. A., Powers, T. R. & Knodler, L. A. Measurement of *Salmonella enterica* Internalization and Vacuole Lysis in Epithelial Cells. *Methods Mol. Biol. Clifton NJ* **1519**, 285–296 (2017).

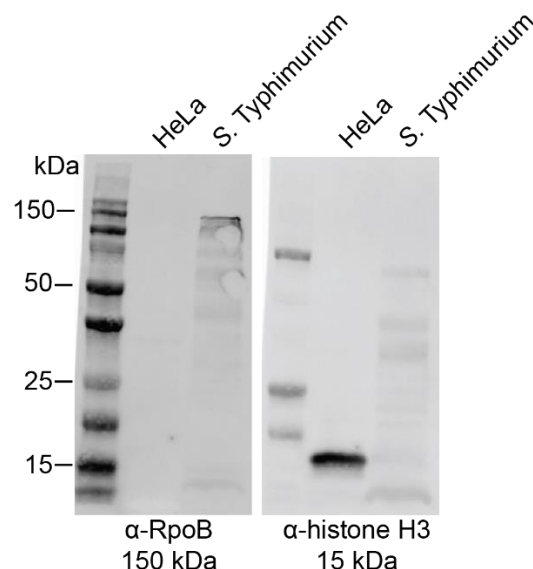
26. Steele-Mortimer, O., Méresse, S., Gorvel, J.-P., Toh, B.-H. & Finlay, B. B. Biogenesis of Salmonella typhimurium-containing vacuoles in epithelial cells involves interactions with the early endocytic pathway. *Cell. Microbiol.* **1**, 33–49 (1999).
27. Steele-Mortimer, O. Infection of Epithelial Cells With Salmonella enterica. in *Bacterial Pathogenesis: Methods and Protocols* (eds. DeLeo, F. R. & Otto, M.) 201–211 (Humana Press, 2008). doi:10.1007/978-1-60327-032-8\_16.
28. Powers, T. R. *et al.* Intracellular niche-specific profiling reveals transcriptional adaptations required for the cytosolic lifestyle of Salmonella enterica. *PLoS Pathog.* **17**, e1009280 (2021).
29. Canals, R. *et al.* Adding function to the genome of African Salmonella Typhimurium ST313 strain D23580. *PLOS Biol.* **17**, e3000059 (2019).
30. Fàbrega, A. & Vila, J. Salmonella enterica serovar Typhimurium skills to succeed in the host: virulence and regulation. *Clin. Microbiol. Rev.* **26**, 308–341 (2013).
31. Nudler, E. RNA Polymerase Backtracking in Gene Regulation and Genome Instability. *Cell* **149**, 1438–1445 (2012).
32. Abdelkareem, M. *et al.* Structural Basis of Transcription: RNA Polymerase Backtracking and Its Reactivation. *Mol. Cell* **75**, 298-309.e4 (2019).
33. Komissarova, N. & Kashlev, M. RNA polymerase switches between inactivated and activated states By translocating back and forth along the DNA and the RNA. *J. Biol. Chem.* **272**, 15329–15338 (1997).
34. Nudler, E., Mustaev, A., Lukhtanov, E. & Goldfarb, A. The RNA-DNA hybrid maintains the register of transcription by preventing backtracking of RNA polymerase. *Cell* **89**, 33–41 (1997).
35. Laptenko, O., Lee, J., Lomakin, I. & Borukhov, S. Transcript cleavage factors GreA and GreB act as transient catalytic components of RNA polymerase. *EMBO J.* **22**, 6322–6334 (2003).
36. Toulmé, F. *et al.* GreA and GreB proteins revive backtracked RNA polymerase in vivo by promoting transcript trimming. *EMBO J.* **19**, 6853–6859 (2000).

37. Dutta, D., Shatalin, K., Epshtein, V., Gottesman, M. E. & Nudler, E. Linking RNA Polymerase Backtracking to Genome Instability. *Cell* **146**, 533–543 (2011).
38. Park, J.-S., Marr, M. T. & Roberts, J. W. E. coli Transcription Repair Coupling Factor (Mfd Protein) Rescues Arrested Complexes by Promoting Forward Translocation. *Cell* **109**, 757–767 (2002).
39. Selby, C. P. & Sancar, A. Molecular Mechanism of Transcription-Repair Coupling. *Science* **260**, 53–58 (1993).
40. Kamarthapu, V. & Nudler, E. Rethinking Transcription Coupled DNA Repair. *Curr. Opin. Microbiol.* **24**, 15–20 (2015).
41. Hawkins, M. *et al.* Direct removal of RNA polymerase barriers to replication by accessory replicative helicases. *Nucleic Acids Res.* **47**, 5100–5113 (2019).
42. Gwynn, E. J. *et al.* The Conserved C-Terminus of the PcrA/UvrD Helicase Interacts Directly with RNA Polymerase. *PLoS ONE* **8**, e78141 (2013).
43. Knodler, L. A., Nair, V. & Steele-Mortimer, O. Quantitative Assessment of Cytosolic Salmonella in Epithelial Cells. *PLOS ONE* **9**, e84681 (2014).
44. Boubakri, H., de Septenville, A. L., Viguera, E. & Michel, B. The helicases DinG, Rep and UvrD cooperate to promote replication across transcription units in vivo. *EMBO J.* **29**, 145–157 (2010).
45. Poteete, A. R. Recombination Phenotypes of Escherichia coli greA Mutants. *BMC Mol. Biol.* **12**, 12 (2011).
46. Kant, S. *et al.* Gre factors help Salmonella adapt to oxidative stress by improving transcription elongation and fidelity of metabolic genes. *PLOS Biol.* **21**, e3002051 (2023).
47. Gaviria-Cantin, T., El Mouali, Y., Le Guyon, S., Römling, U. & Balsalobre, C. Gre factors-mediated control of hilD transcription is essential for the invasion of epithelial cells by Salmonella enterica serovar Typhimurium. *PLoS Pathog.* **13**, e1006312 (2017).
48. Gaviria-Cantin, T. *et al.* Gre Factors Are Required for Biofilm Formation in Salmonella enterica Serovar Typhimurium by Targeting Transcription of the csgD Gene. *Microorganisms* **10**, 1921 (2022).

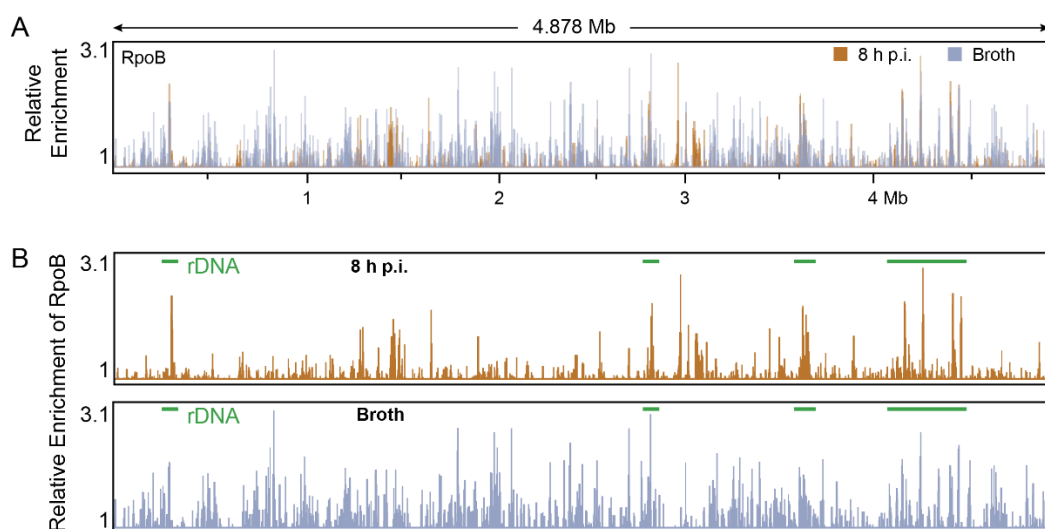
49. Cui, G., Wang, J., Qi, X. & Su, J. Transcription Elongation Factor GreA Plays a Key Role in Cellular Invasion and Virulence of *Francisella tularensis* subsp. *novicida*. *Sci. Rep.* **8**, 6895 (2018).
50. Feng, S. *et al.* Involvement of Transcription Elongation Factor GreA in *Mycobacterium* Viability, Antibiotic Susceptibility, and Intracellular Fitness. *Front. Microbiol.* **11**, 413 (2020).
51. Flashner, Y. *et al.* Generation of *Yersinia pestis* Attenuated Strains by Signature-Tagged Mutagenesis in Search of Novel Vaccine Candidates. *Infect. Immun.* **72**, 908–915 (2004).
52. Thomason, L. C., Costantino, N., Li, X. & Court, D. L. Recombineering: Genetic Engineering in *Escherichia coli* Using Homologous Recombination. *Curr. Protoc.* **3**, e656 (2023).
53. Yu, D. *et al.* An efficient recombination system for chromosome engineering in *Escherichia coli*. *Proc. Natl. Acad. Sci. U. S. A.* **97**, 5978–5983 (2000).
54. Li, X., Thomason, L. C., Sawitzke, J. A., Costantino, N. & Court, D. L. Positive and negative selection using the tetA-sacB cassette: recombineering and P1 transduction in *Escherichia coli*. *Nucleic Acids Res.* **41**, e204 (2013).
55. Leman, A. R. & Noguchi, E. Chromatin Immunoprecipitation to Investigate Origin Association of Replication Factors in Mammalian Cells. *Methods Mol. Biol. Clifton NJ* **1170**, 539–547 (2014).
56. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120 (2014).
57. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357.
58. Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).
59. Liao, Y., Smyth, G. K. & Shi, W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* **30**, 923–930 (2014).
60. Mi, H., Muruganujan, A., Ebert, D., Huang, X. & Thomas, P. D. PANTHER version 14: more genomes, a new PANTHER GO-slim and improvements in enrichment analysis tools. *Nucleic Acids Res.* **47**, D419–D426 (2019).

61. Ashburner, M. *et al.* Gene Ontology: tool for the unification of biology. *Nat. Genet.* **25**, 25–29 (2000).
62. The Gene Ontology Consortium *et al.* The Gene Ontology resource: enriching a GOld mine. *Nucleic Acids Res.* **49**, D325–D334 (2021).
63. Zhang, Y. *et al.* Model-based Analysis of ChIP-Seq (MACS). *Genome Biol.* **9**, R137 (2008).
64. Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841–842 (2010).
65. Ramírez, F. *et al.* deepTools2: a next generation web server for deep-sequencing data analysis. *Nucleic Acids Res.* **44**, W160–W165 (2016).
66. Robinson, J. T. *et al.* Integrative Genomics Viewer. *Nat. Biotechnol.* **29**, 24–26 (2011).

## Supplementary Information



**S1 Figure. RpoB antibody is specific to bacterial protein.** Western blot showing that the native RpoB antibody used in these studies does not cross-react to proteins specific to HeLa cells. HeLa whole cell lysate was prepared in RIPA buffer and diluted into Laemmli sample buffer. *S. Typhimurium* whole cell lysate was prepared in lysis buffer (10 mM Tris-HCl pH = 7, 10 mM EDTA, 1X protease inhibitor, 0.1 mg/mL lysozyme) and diluted into Laemmli sample buffer. Lysates were boiled for 10 min and loaded onto a 12% Mini-PROTEAN TGX gel (Bio-Rad). Proteins were subsequently transferred to a PVDF nitrocellulose membrane (Bio-Rad). The membrane was blocked for 1 h at room temperature with Intercept PBS blocking buffer (LI-COR) before being immunoblotted with anti-RpoB (8RB13, Thermo) and anti-Histone H3 (PA5-16183, Thermo) antibodies overnight at 4°C. The blot was incubated with secondary antibodies IRDye-680RD and 800CW (LI-COR) for 30 min at room temperature before being imaged on an Odyssey Imager (LI-COR).



**S2 Figure. RpoB ChIP-seq signal from broth and infection, related to Figure 2.** (A) Representative overlay of *S. Typhimurium* RNAP occupancy as determined by ChIP-seq of RpoB (the beta subunit of RNAP) 8 h post infection (p.i.) of HeLa cells or from cells grown in broth culture. (B) Reads mapped to the *Salmonella* genome display areas of high RNAP enrichment at regions of rDNA in both conditions (green bars). Relative enrichment is defined as the ratio of IP and total read counts.

Locus	Fold Enrichment	Localization	Function	Reference
<i>mig-14</i>	2.82	Chromosome	antimicrobial peptide resistance	1
<i>virK</i>	2.79	Chromosome	antimicrobial peptide resistance	2
<i>sifB</i>	2.70	Chromosome	putatively contributes to the formation of the intracellular replication niche	3
<i>sitD</i>	2.62	SPI-1	resistance to nitrosative stress	4
<i>SL1344_1344</i>	2.57	Chromosome	putative component of the type III secretion system (T3SS-2) apparatus	5
<i>ssaK</i>	2.57	SPI-2	type III secretion system (T3SS-2) apparatus	6
<i>sitC</i>	2.53	SPI-1	resistance to nitrosative stress	4
<i>pdgL</i>	2.53	Chromosome	putative role in cell wall/membrane biogenesis	7
<i>SL1344_2902</i>	2.49	Chromosome	hypothetical protein	
<i>rplC</i>	2.49	Chromosome	ribosomal protein L3	
<i>ssaH</i>	2.46	SPI-2	putative role in type III secretion system apparatus	8
<i>ssaI</i>	2.46	SPI-2	putative role in type III secretion system apparatus	8
<i>ssaJ</i>	2.46	SPI-2	type III secretion system (T3SS-2) apparatus	6
<i>pipB2</i>	2.41	Chromosome	kinesin accumulation in the SCV	9
<i>ssaE</i>	2.40	SPI-2	chaperone	10
<i>SL1344_1530A</i>	2.37	Chromosome	hypothetical protein	
<i>ugtL</i>	2.37	Chromosome	PhoP activation	11
<i>sitA</i>	2.36	SPI-1	resistance to nitrosative stress	4
<i>sitB</i>	2.36	SPI-1	resistance to nitrosative stress	4
<i>sseB</i>	2.34	SPI-2	type III secretion system (T3SS-2) translocon	12
<i>sscA</i>	2.34	SPI-2	chaperone	13
<i>sseD</i>	2.33	SPI-2	type III secretion system (T3SS-2) translocon	12
<i>sseK/sseK3</i>	2.27	Chromosome	promote intracellular survival	14
<i>sseC</i>	2.20	SPI-2	type III secretion system (T3SS-2) translocon	12
<i>ssaL</i>	2.19	SPI-2	type III secretion system (T3SS-2) apparatus	6
<i>mgtC</i>	2.18	SPI-3	maintenance of ATP homeostasis	15
<i>ssaR</i>	2.18	SPI-2	type III secretion system (T3SS-2) apparatus	6
<i>sseG</i>	2.17	SPI-2	SIF biogenesis and SCV maintenance	16
<i>iscS</i>	2.15	Chromosome	tRNA synthesis	17
<i>prgJ</i>	2.15	SPI-1	type III secretion system (T3SS-1)	18
<i>prgI</i>	2.15	SPI-1	type III secretion system (T3SS-1)	18
<i>ssaQ</i>	2.10	SPI-2	type III secretion system (T3SS-2) apparatus	19

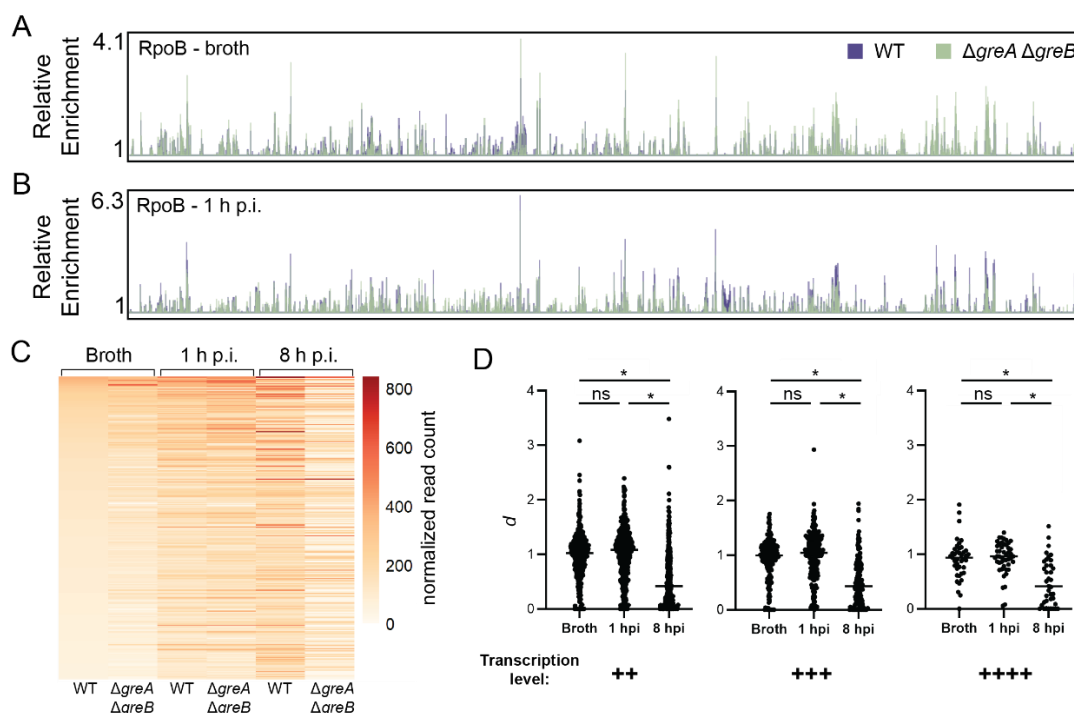
**S1 Table. RNAP is significantly enriched at virulence genes during infection.** Loci where RpoB was significantly enriched at least two-fold in *S. Typhimurium* 8 h post infection as determined by ChIP-seq. Each value represents the average fold enrichment ('signalValue' output of macs2) for at least two biological replicates.

# S1 Table References

1. Brodsky, I. E., Ernst, R. K., Miller, S. I. & Falkow, S. mig-14 Is a Salmonella Gene That Plays a Role in Bacterial Resistance to Antimicrobial Peptides. *J. Bacteriol.* **184**, 3203–3213 (2002).
2. Detweiler, C. S., Monack, D. M., Brodsky, I. E., Mathew, H. & Falkow, S. virK, somA and rcsC are important for systemic Salmonella enterica serovar Typhimurium infection and cationic peptide resistance. *Mol. Microbiol.* **48**, 385–400 (2003).
3. Freeman, J. A., Ohl, M. E. & Miller, S. I. The Salmonella enterica Serovar Typhimurium Translocated Effectors SseJ and SifB Are Targeted to the Salmonella-Containing Vacuole. *Infect. Immun.* **71**, 418–427 (2003).
4. Yousuf, S. *et al.* Manganese import protects Salmonella enterica serovar Typhimurium against nitrosative stress†. *Metallomics* **12**, 1791–1801 (2020).
5. Mazé, A., Glatter, T. & Bumann, D. The Central Metabolism Regulator EIIA Glc Switches Salmonella from Growth Arrest to Acute Virulence through Activation of Virulence Factor Secretion. *Cell Rep.* **7**, 1426–1433 (2014).
6. Hensel, M. *et al.* Functional analysis of ssaJ and the ssaK/U operon, 13 genes encoding components of the type III secretion apparatus of Salmonella Pathogenicity Island 2. *Mol. Microbiol.* **24**, 155–167 (1997).
7. Yoon, H., McDermott, J. E., Porwollik, S., McClelland, M. & Heffron, F. Coordinated Regulation of Virulence during Systemic Infection of Salmonella enterica Serovar Typhimurium. *PLoS Pathog.* **5**, e1000306 (2009).
8. Chakravorty, D., Rohde, M., Jäger, L., Deiwick, J. & Hensel, M. Formation of a novel surface structure encoded by Salmonella Pathogenicity Island 2. *EMBO J.* **24**, 2043–2052 (2005).
9. Fàbrega, A. & Vila, J. Salmonella enterica serovar Typhimurium skills to succeed in the host: virulence and regulation. *Clin. Microbiol. Rev.* **26**, 308–341 (2013).
10. Miki, T., Shibagaki, Y., Danbara, H. & Okada, N. Functional Characterization of SsaE, a Novel Chaperone Protein of the Type III Secretion System Encoded by Salmonella Pathogenicity Island 2. *J. Bacteriol.* **191**, 6843–6854 (2009).
11. Choi, J., Salvail, H. & Groisman, E. A. RNA chaperone activates Salmonella virulence program during infection. *Nucleic Acids Res.* **49**, 11614–11628 (2021).
12. Nikolaus, T. *et al.* SseBCD proteins are secreted by the type III secretion system of Salmonella pathogenicity island 2 and function as a translocon. *J. Bacteriol.* **183**, 6036–6045 (2001).
13. Zurawski, D. V. & Stein, M. A. The SPI2-encoded SseA chaperone has discrete domains required for SseB stabilization and export, and binds within the C-terminus of SseB and SseD. *Microbiology* **150**, 2055–2068 (2004).
14. Günster, R. A., Matthews, S. A., Holden, D. W. & Thurston, T. L. M. SseK1 and SseK3 Type III Secretion System Effectors Inhibit NF-κB Signaling and Necroptotic Cell Death in Salmonella-Infected Macrophages. *Infect. Immun.* **85**, e00010-17 (2017).
15. Lee, J.-W. & Lee, E.-J. Regulation and function of the Salmonella MgtC virulence protein. *J. Microbiol. Seoul Korea* **53**, 667–672 (2015).
16. Knuff-Janzen, K., Tupin, A., Yurist-Doutsch, S., Rowland, J. L. & Finlay, B. B. Multiple Salmonella-pathogenicity island 2 effectors are required to facilitate bacterial establishment of its intracellular niche and virulence. *PLoS ONE* **15**, e0235020 (2020).
17. Lundgren, H. K. & Björk, G. R. Structural alterations of the cysteine desulfurase IscS of Salmonella enterica serovar Typhimurium reveal substrate specificity of IscS in tRNA thiolation. *J. Bacteriol.* **188**, 3052–3062 (2006).
18. Lou, L., Zhang, P., Piao, R. & Wang, Y. Salmonella Pathogenicity Island 1 (SPI-1) and Its Complex Regulatory Network. *Front. Cell. Infect. Microbiol.* **9**, (2019).
19. Yu, X.-J., Liu, M., Matthews, S. & Holden, D. W. Tandem translation generates a chaperone for the Salmonella type III secretion system protein SsaQ. *J. Biol. Chem.* **286**, 36098–36107 (2011).

Locus	Fold Enrichment	
	Broth	Infection
<i>yiaG</i>	2.01	3.44
<i>cspA</i>	2.01	3.44
<i>cueP</i>	2.01	3.44
<i>deaD</i>	2.47	2.93
<i>yrbN</i>	2.47	2.93
<i>nlpI</i>	2.47	2.93
<i>pnp</i>	2.47	2.90
<i>cspE (cspC?)</i>	2.23	2.84
<i>yobF</i>	2.23	2.84
<i>SL1344_RS09185</i>	2.23	2.82
<i>mgrB (yobG)</i>	2.23	2.82
<i>yqcC</i>	2.75	2.69
<i>hdfR</i>	2.33	2.50
<i>trxA</i>	2.07	2.35
<i>rhoL</i>	2.07	2.35
<i>rho</i>	2.07	2.35
<i>rpsD</i>	2.08	2.33
<i>rpsK</i>	2.08	2.33
<i>rpsM</i>	2.08	2.33
<i>rpmJ</i>	2.08	2.33
<i>rplN</i>	2.08	2.31
<i>rplO</i>	2.08	2.27
<i>secY/prlA</i>	2.08	2.26
<i>rpoA/pez</i>	2.08	2.18
<i>rpsC</i>	2.21	2.18
<i>rplV</i>	2.21	2.18
<i>rplB</i>	2.21	2.13
<i>rpoH</i>	2.07	2.12
<i>rluD</i>	2.37	2.08
<i>pgeF</i>	2.21	2.03

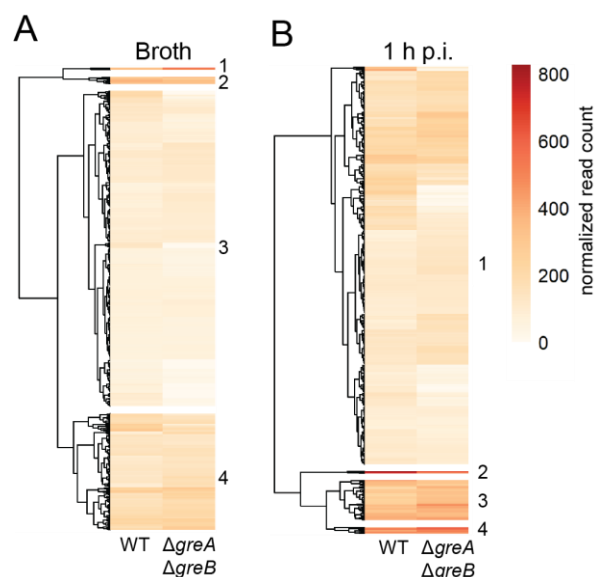
**S2 Table. RNAP is significantly enriched at some of the same genes in broth and infection.** Loci where RpoB was significantly enriched at least two-fold in *S. Typhimurium* in cells grown in broth and 8 h post infection, as determined by ChIP-seq. Each value represents the average fold enrichment ('signalValue' output of macs2) for at least two biological replicates.



**S3 Figure. RNAP occupancy changes in the presence and absence of Gre factors, related to Figure 3.** Representative RNAP occupancy profile for wild-type (WT) cells or cells lacking Gre factors ( $\Delta greA \Delta greB$ ) in cells (A) grown in broth and (B) at 1 h post infection (h p.i.) as determined by ChIP-seq of RpoB. Relative enrichment is defined as the ratio of IP and total read counts. (C) RNAP occupancy changes visualized as a heat map, where every horizontal line represents the normalized read count for the same gene across each condition. Only top-transcribed genes that arise in all three conditions are plotted (283 genes). Each value represented is the average of three biological replicates. (D) Quantification of  $d$  ratios calculated for each gene falling within each transcription level (see S3 Table). The number of genes belonging to each condition within each transcription level is as follows: ++, 354 (broth), 412 (1 h p.i.), 405 (8 h p.i.); +++, 160 (broth), 164 (1 h p.i.), 147 (8 h p.i.); +++++, 37 (broth), 42 (1 h p.i.), 38 (8 h p.i.). Each value represents the average of three biological replicates. \* $p < 0.0001$ , one-way ANOVA.

*Excel file*

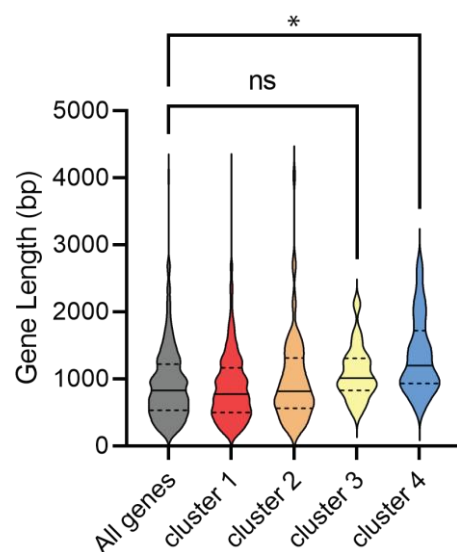
**S3 Table. Top transcribed genes in cells grown in broth, 1 h p.i., and 8 h p.i.** List of the top transcribed genes as determined by k-means clustering for each condition. Values represent the average normalized read count and the ratio of the normalized read counts of three biological replicates. This table also lists the 283 top transcribed genes that arise in all three conditions.



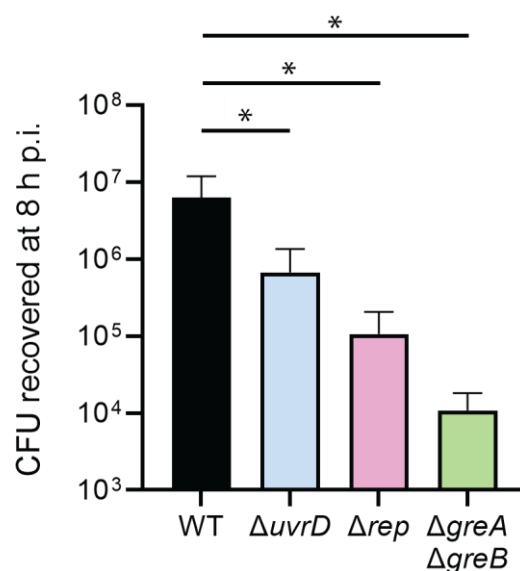
**S4 Figure. RNAP occupancy differences in broth and 1 h p.i., related to Figure 4.** RNAP occupancy changes in cells grown in broth and at 1 h post infection (h p.i.) in the absence of Gre factors visualized as a heat map. Each value represents the average of three biological replicates. Hierarchical clustering (numbered 1-4) was performed using the pheatmap function in RStudio (see Table S4).

*Excel file*

**S4 Table. RNAP occupancy changes for the top transcribed genes as categorized by hierarchical clustering.** Lists the top transcribed genes for each condition that call into each hierarchical cluster, as determined by pheatmap library in R. Values represent the average normalized read count and the ratio of the normalized read counts of three biological replicates. Also lists the manually annotated functions of genes in clusters three and four for the 8 h post infection (p.i.) condition.



**S5 Figure. Longer genes experience more prevalent backtracking, related to S4 Table.** All top transcribed genes at 8 h post infection were sorted by hierarchical clustering using the pheatmap function in RStudio (590 genes total). The average gene length was calculated for genes in each cluster. \* $p < 0.001$ , ns: not significant, one-way ANOVA.



**S6 Figure. Cells lacking key conflict resolution factors exhibit significant growth defects during infection.** HeLa cells were infected with *S. Typhimurium* cells lacking the indicated genes. Bacteria were harvested at 8 h post infection (p.i.) and plated for CFU enumeration. \* $p < 0.05$ , one-way ANOVA.

**S5 Table. Strains, plasmids, and primers used in this study.**

Strain	Genotype/Features	Identifier	Reference/Source
SL1344	Wild-type <i>S. Typhimurium</i> strain SL1344	HM4315	This study
$\Delta greB$	<i>greB::Cat</i>	HM4525	This study
$\Delta greA$	<i>greA::Kan</i>	HM4527	This study
$\Delta greA$ $\Delta greB$	<i>greA::Kan greB::Cat</i>	HM4529	This study
T-SACK	W3110 <i>araD</i> <math>\phi</math> <i>tetA-sacB amp fliC</i> <math>\phi</math> <i>cat argG::Tn5</i>	HM4177	<sup>1</sup>
Plasmid	Features		Reference/Source
pSIM27	For expression of Red recombinase system		Gift from Dr. Don Court
Primers	Sequence		Reference/Source
HM6042	CATTGCCCCCTACAGGAATGTTCAAGAGGT ATAACAAATGTATGGACAGCAAGCGAACCG		This study
HM6043	TTTACAATACACCAACAATTTGCGTATTGAG TACTGCTTATCAGAAGAACTCGTCAAGAAG		This study
HM6040	GTGTGCGCAATATCGACAGCAAAGGTAAAT CAACGAGATGTGTGACGGAAGATCACTTCG		This study
HM6041	TGCCAGCCATCAGCGGGGGCTTAGGATTCTT CTTGTCTTAACCAGCAATAGACATAAGCG		This study
HM5904	GAGAGCAAAATGCCACGCTC		This study
HM5905	GCTTCTGACCCTGGTATGGC		This study
HM5908	TCACGTGAGTCACCAGCAAAT		This study
HM5909	TGGCAAAAATGCCAGCCATC		This study

<sup>1</sup> Li, X., Thomason, L. C., Sawitzke, J. A., Costantino, N. & Court, D. L. Positive and negative selection using the *tetA-sacB* cassette: recombineering and P1 transduction in *Escherichia coli*. *Nucleic Acids Res.* **41**, e204 (2013).