Genome-wide association and function studies identify Mfd as a critical RNA polymerase co-factor at hard-to-transcribe regions

Mark N. Ragheb[1] **and Houra Merrikh**[2,3,*]

[1] Molecular and Cellular Biology Graduate Program and Medical Scientist Training Program, University of Washington, Seattle, WA, USA.
[2] Department of Biochemistry, Vanderbilt University, Nashville, TN 37205, USA
[3] Department of Pathology, Microbiology, and Immunology, Vanderbilt University Medical Center, Nashville, TN 37232, USA

*Corresponding (Lead) author
Houra Merrikh
Email: houra.merrikh@vanderbilt.edu
Tel: 615-343-3846

Abstract

RNA polymerase (RNAP) encounters various roadblocks during transcription. Given that these obstacles can change the dynamics of RNAP movement, they are likely to influence transcription either directly or through RNAP associated factors. One such factor is Mfd; a highly conserved DNA translocase that is thought to primarily function in repair of DNA lesions that have stalled RNAP. However, the interaction between Mfd and RNAP may also be important for transcription regulation at generally hard-to-transcribe regions where RNAP frequently stalls in living cells, even in the absence of DNA lesions. If so, then Mfd may function as a critical RNAP co-factor and a transcription regulator, at least for some genes. This model has not been directly tested.

Here, we assessed the function of Mfd *in vivo* and determined its impact on RNAP association and transcription regulation. We performed genome-wide studies, and identified chromosomal loci bound by Mfd. We found many genomic regions where Mfd modulates RNAP association and represses transcription. Additionally, we found that almost all loci where Mfd associates and regulates transcription contain highly structured regulatory RNAs. The RNAs in these regions regulate a myriad of biological processes, ranging from metabolism, to tRNA regulation, to toxin-antitoxin functions. We found that transcription regulation by Mfd, at least at the toxin-antitoxin loci, is essential for cell survival. Based on these data, we propose that Mfd is a critical RNAP co-factor that is essential for transcription regulation at difficult-to-transcribe regions, especially those that express structured regulatory RNAs.

Keywords: Mfd, RNA polymerase, transcription termination, regulatory RNAs, RNA secondary structure, toxin-antitoxins

Significance

The Mfd translocase recognizes stalled RNAPs. This recognition is generally thought to facilitate transcription-coupled DNA repair, based largely on data from biochemical studies. Little is known about Mfd's function in living cells, especially in the absence of exogenous DNA damage. Our data show that Mfd is a critical RNAP co-factor that modulates RNAP association and regulates transcription at various loci, especially those containing highly structured, regulatory RNAs. This work improves our understanding of Mfd's function in living cells and assigns a new function to Mfd as a regulator of transcription at hard-to-transcribe regions where maintaining transcriptional equilibrium (e.g. at toxin-antitoxin loci) is essential for viability. Altogether, this work also expands our understanding of how transcription is regulated at difficult-to-transcribe loci.

Introduction

Timely and efficient transcription is a fundamental requirement for maintaining cellular homeostasis. The process of transcription elongation is discontinuous, with RNA polymerase (RNAP) processivity altered by a wide range of obstacles. These obstacles vary in severity, from pause sites that slow the rate of RNAP(1–3) to more severe obstacles, such as protein roadblocks and the replication fork, which induce the reverse translocation of RNAP with respect to both DNA and nascent RNA (RNAP backtracking) (4–7). Roadblocks to RNAP processivity are both prevented and resolved through various mechanisms in bacteria, including the coupling of transcription and translation, as well as various cellular factors that help re-establish transcription elongation (i.e. anti-backtracking factors)(8).

In vitro work shows that the DNA translocase Mfd utilizes its RNAP binding properties and forward translocase activity to rescue arrested RNAP, and in doing so, can restore RNAP to its active elongation(9) as well as promote transcription termination state(10). However, despite decades of research on the biochemical characteristics of Mfd, the endogenous contexts in which its translocase and anti-backtracking functions become critical for transcription remains elusive.

Mfd was initially described as a critical DNA repair factor *in vivo* that promotes transcription-coupled repair (TCR)(11–13). In the TCR pathway, Mfd removes stalled RNAP at bulky DNA lesions to expose the offending lesion to the nucleotide excision repair (NER) pathway via its UvrA binding capacity (for a review of the pathway, see (14)). However, cells lacking Mfd show little to no sensitivity to DNA damaging agents that promote RNAP stalling(15, 16), especially relative to other TCR factors(17, 18). Such data suggests that Mfd may have a broader cellular function outside of DNA repair. Indeed, further research has implicated Mfd in other cellular roles, including increased mutagenesis at specific loci where collisions between the replication

and transcription machineries occur(8), as well as regulating catabolite repression in *Bacillus subtilis*(19, 20). Additionally, recent *in vitro* experiments have shown that Mfd is capable of autonomously translocating on DNA in the absence of a lesion, but whether this is the *case in vivo* is unclear (21).

Despite our limited understanding of Mfd's central cellular functions, its high level of conservation in bacteria implies a fundamental role for Mfd that may be separate from TCR. Mfd is highly structurally conserved, and complementation experiments across divergent species shows that at least some of its functions are also highly conserved(15). In addition, the expression of Mfd is thought to be constitutive, suggesting a possible homeostatic role in regulating transcription. Indeed, recent work suggests that Mfd plays a housekeeping function in cells by associating with RNAP in the absence of exogenous stressors(22). However, the role of Mfd in association with RNAP and what particular genomic sites (if any) Mfd may be acting upon, particularly in the absence of exogenous stress, remains unclear.

In this work, we define the *in vivo* role of Mfd in regulating transcription by identifying specific regions in the genome where RNAP and transcription are altered in the absence of Mfd. We found that Mfd promotes RNAP release and transcription termination at regions containing highly structured regulatory RNAs. We also determined that Mfd's activity is important for regulating transcription at various critical loci in *B. subtilis*, including toxin-antitoxin systems, and that misregulation in the absence of Mfd can be lethal. Our work suggests that RNA secondary structure is a major impediment to transcription *in vivo* and that transcription regulation by Mfd is especially important at sites transcribing highly structured RNAs. Based on these and other results, we propose that Mfd is an essential RNAP co-factor that regulates transcription, especially at regions harboring critical regulatory RNAs.

Results

Mapping the genomic loci where Mfd's associates

We began by assessing Mfd's genome-wide association using chromatin immunoprecipitation sequencing (ChIP-seq). We constructed a *B. subtilis* strain where Mfd is C-terminally Myc-tagged (Fig. S1). To identify the chromosomal regions where Mfd associates, we harvested exponentially growing cells and performed ChIP-seq analysis. We controlled for potential ChIP artifacts by comparing this signal to ChIP-seq performed using Myc antibody in *B. subtilis* lacking a Myc-tagged Mfd (Fig. 1*A*). Under these conditions, we found 489 genes out of 5755 genes analyzed (a total of 8%) of the *B. subtilis* genome exhibit preferential Mfd association (defined as one standard deviation greater than the average Mfd ChIP signal across all genes) (Dataset S1).

Mfd's genomic association pattern correlates with that of RNAP

Given the known physical interaction between Mfd and RNAP(23), we hypothesized that Mfd binding peaks would be correlated with RNAP association on the genome. We performed ChIP-seq of RpoB, the β subunit of RNAP using a native antibody. We indeed found that Mfd association genome-wide is largely correlated with RpoB occupancy (Pearson coefficient 0.68) (Fig. 1*B* and Fig. S2), suggesting that Mfd functions as an RNAP co-factor in *B. subtilis*.

Mfd requires interaction with RNAP for its association with all genomic loci

Mfd is a multimodular protein, consisting of eight domains connected by flexible linkers (23). Of these domains, the RNAP interacting domain (RID) and the translocase module (composed of domains D5 and D6) are critical for Mfd's ability to rescue stalled transcription complexes (24, 25). Mfd is thought to be recruited to the identified genomic regions through its interaction with RpoB. We therefore tested whether the interaction of Mfd with RpoB is critical for its recruitment to the genomic loci we identified. Prior *in vitro* work suggested RID mutations abrogate the

interaction between Mfd and RNAP(23). We therefore constructed a Mfd-myc strain with a mutation at the L522 residue to disrupt Mfd's binding to RpoB, without disrupting the stability or folding of Mfd, as described in *Escherichia coli* (23). Upon confirming that the *B. subtilis* L522A mutation disrupted Mfd's interaction with RNAP via a bacterial 2-hybrid assay (Fig. S3), we performed ChIP-seq experiments with this mutant. The ChIP signal we detected in WT strains was abrogated in the strain expressing the L522A allele (Fig. 1*C*). These results strongly suggested that Mfd's interaction with RNAP is essential for its recruitment to all genomic loci identified in the previous experiments. These findings suggest that Mfd functions as a genome-wide RNAP co-factor *in vivo*.

Mfd's association with DNA requires transcription elongation

In vitro, Mfd helps promote the rescue of arrested transcription elongation complexes (TECs), yet how Mfd recognizes stalled RNAP *in vivo* remains unclear. We therefore sought to discriminate whether Mfd association with various genomic loci was facilitated via loading during the transcription initiation or elongation phase. To distinguish between transcription initiation and elongation, we utilized the antimicrobial rifampicin, which directly blocks transcription initiation(26, 27), subsequently eliminating the formation of transcription elongation complexes (TECs). Rifampicin treatment largely eliminated Mfd ChIP-seq binding signal (Fig. 1*D*), suggesting that active transcription elongation (and not initiation) is a requirement for Mfd association with RNAP, and subsequently, genomic regions.

Mfd promotes release of RNAP at some genomic loci

Mfd can function to promote both transcription elongation as well as transcription termination, at least under *in vitro* conditions (9, 10). However, the importance of these functions *in vivo* remained elusive. We wondered if Mfd's close association with RNAP occupancy *in vivo* was relevant for transcription elongation, transcription termination, or both. We therefore performed

ChIP-seq of RpoB in WT and *Δmfd* strains to look for genes where RpoB occupancy was altered in the absence of Mfd. ChIP-seq experiments did not detect alterations in RpoB occupancy at the majority of genes where Mfd associates based on our Mfd ChIP-seq studies. This may be a result of various factors, such as the existence of redundant transcription-associated factors, the specific growth conditions of the experiment, or potential limitations of detection thresholds which are possible in an ensemble assay such as ChIP-seq. However, we did find a number of genes where RpoB occupancy either increased or decreased in the *Δmfd* strain compared to WT (Fig. 2*A*). We specifically noticed a bias towards a greater number of genes that exhibited increased (rather than decreased) RpoB occupancy in the *Δmfd* strain. Quantification of these results revealed that a total of 116 genes exhibited at least a two-fold increase, while 53 genes exhibited at least a two-fold decrease in RpoB occupancy without Mfd. Many of these genes are within the same operon, and therefore are expressed as single transcripts. Thus, rather than single genes, we grouped and analyzed the identified genes as transcription units (TUs)(28). Our analysis revealed that *Δmfd* strains contain 71 TUs with at least one gene containing a minimum of two-fold increase and 31 TUs with at least one gene containing a minimum of two-fold decreased in RpoB occupancy compared to WT (Table 1, Datasets S2 and S3).

We next wanted to determine whether the changes in RpoB occupancy observed in *Δmfd* were directly due to Mfd's activity at those regions or whether we were detecting indirect effects. To address this, we looked for a correlation between the regions where Mfd associates (ChIP-seq experiments) and where there are also changes in RpoB association levels in the absence of Mfd. We observed Mfd association at many of the genes with increased RpoB occupancy in the *Δmfd* strain, but not at any of the genes with decreased RpoB occupancy in the *Δmfd* strain (Fig. 2*B*). More specifically, 52 of the 116 genes (35 of the 71 TUs) with increased RpoB occupancy in *Δmfd* compared to WT were regions with Mfd association, while we did not detect

Mfd association at any of the 53 genes (31 TUs) with decreased RpoB occupancy. From these findings we concluded that increased RNAP occupancy (but not decreased) is a direct result of RNAP release/termination by Mfd. This conclusion is in line with prior observations made in biochemical experiments (10, 29)

Regions where Mfd promotes RNAP release are enriched for regulatory RNAs

In vitro, Mfd's translocase activity can help release RNAP exposed to various obstacles. However, whether there are endogenous hotspots of RNAP stalling that requires Mfd function remained unknown. Furthermore, if such hotspots exist, the nature of the potential obstacles remained unclear. Intriguingly, 92% of the TUs that showed both an increase in RNAP density in the *Δmfd* strain and direct Mfd association at those loci express a minimum of one regulatory RNA (Table S1). These regulatory RNAs are a subset of the 1583 regulatory RNAs in *B. subtilis*, which encompass a wide variety of RNAs, including independent, non-coding transcripts, antisense RNAs, and multiple riboswitches(28). In comparison, only 39% of the TUs with decreased RNAP density in *Δmfd* strains contain regulatory RNAs (Table S2), which is consistent with the average percentage of TUs with predicted regulatory RNAs in *B. subtilis* (28, 30).

Regions of Mfd function are enriched for high secondary structure in RNA

We hypothesized that the mechanism underlying the increase in RNAP association at the identified TUs in the *Δmfd* strain was related to RNA secondary structure impeding RNAP processivity. This hypothesis is consistent with changes in RNAP dynamics due to secondary RNA structures, such as hairpins in the context of intrinsic transcription termination, and other transcription regulatory processes(31–33). Previous work characterized the predicted secondary structure for each regulatory RNA in *B. subtilis*(30). We sought to test whether the regions with increased RNAP in *Δmfd* strains are more prone to transcribing RNAs with more stable

247 secondary structures. We determined the average minimum free energy (MFE) z-score, a proxy
248 of RNA structure stability(34, 35), for the 42 regulatory RNAs in the TUs that had both increased
249 RpoB density in the *Δmfd* strain and Mfd association. We then compared the results to TUs that
250 showed no difference in RpoB density between WT and *Δmfd* strains. We found that those
251 associated with Mfd binding and increased RpoB density in the *Δmfd* strain have significantly
252 higher predicted RNA secondary structure relative to all regulatory RNAs (Fig. 3). This finding
253 suggests that Mfd is critical for regulating RNAP at regions containing highly structured RNAs.
254
255 **RNAP termination at structured regulatory RNAs is specific to Mfd**
256 Various factors are known to help rescue arrested RNAP through different mechanisms. One of
257 the most well-known and highly conserved factors is GreA, which functions as an RNAP
258 antibacktracking factor and functions by cleaving the nascent 3' RNA that has extruded from the
259 RNAP catalytic channel during backtracking(8, 36).To test whether the antibacktracking activity
260 of GreA also contributed to RNAP release at the loci transcribing structured RNAs, we
261 performed RpoB ChIP-seq of *B. subtilis* WT and a *ΔgreA* strain. We found that the *ΔgreA* strain
262 only had 12 genes (and six TUs) with increased RpoB occupancy (Dataset S4, Fig. S4). Two of
263 the six TUs transcribe regulatory RNAs and neither contained significant predicted secondary
264 structure. These results suggest that unlike Mfd, GreA does not function in releasing RNAP from
265 sites containing secondary structure. We also observed that in the *ΔgreA* strain, a total 469
266 genes exhibited less than two-fold RpoB occupancy (Dataset S4, Fig. S4). The high number of
267 genes with decreased RpoB occupancy suggests that GreA functions *in vivo* largely to promote
268 transcription elongation, which is consistent with prior biochemical studies (36).
269
270 **Mfd promotes transcription repression at sites expressing structured, regulatory RNAs**
271 We next tested the effect of Mfd on transcription at sites containing structured RNAs. We began
272 by performing RNA-seq of WT and *Δmfd* strains in *B. subtilis*. Consistent with our WT and *Δmfd*

RpoB ChIP-seq results, we find more genes were upregulated than downregulated in the *Δmfd* strain (240 genes upregulated compared to 138 genes downregulated) (Fig. 4*A* and Dataset S5). When comparing our RpoB ChIP-seq finding to RNA-seq, we found that of the 116 genes with greater than 2-fold RpoB ChIP-seq signal in the *Δmfd* strain, 30 of them directly showed increased expression, while none show decreased expression in the absence of Mfd. Because standard RNA sequencing protocols are often not suitable for accurate measurement of small RNAs(37), we wondered if perhaps there were additional genes with increased RpoB density in the *Δmfd* strain that have corresponding increases in transcription, but were not accurately detected in our RNA-seq analysis. We therefore directly measured RNA levels using qRT-PCR at three loci containing non-coding RNAs (the *trnY* locus, *txpa-ratA*, and *bsrH-asBsrH*), all of which show Mfd binding and increased RpoB signal in the *Δmfd* strain. We found that all three of these loci have increased gene expression in the *Δmfd* strain compared to WT (Fig. 4*B*). These findings suggest that at many sites, Mfd's *in vivo* RNAP release activity directly represses transcription.

Transcriptional regulation by Mfd at toxin-antitoxin loci is essential for cell survival

Aside from promoting mutagenesis and evolution, there are not many phenotypic defects that have been detected in the absence of Mfd, even upon exposure to DNA damage(15, 16, 38–40). We wondered whether the transcriptional regulation activity of Mfd at regions we detected were physiologically relevant. We began by focusing on the highest structured regulatory RNAs which had altered RpoB density in the *Δmfd* strain and were directly bound by Mfd. These regulatory RNAs were present in two pairs of type I toxin-antitoxin (TA) loci in *B. subtilis*: the *txpA/ratA* locus and the bsrH/as-bsrH locus. Type I TA loci are characterized by the expression of a small toxic peptide and a noncoding RNA that neutralizes toxin expression by direct binding and either inhibiting translation or promoting degradation of the toxin mRNA(41). The cellular functions of type I TA loci remain unclear, but they have been proposed to be important for

diverse aspects of physiology, including persister formation(42), biofilm formation(43), and prophage maintenance(44). Five type I TA loci have been identified in *B. subtilis*(41)- we found that three of these loci have both Mfd binding and a minimum two-fold increase in RpoB density in the *Δmfd* strain, while a fourth locus, *yonT/as-YonT*, showed a significant increase in RpoB density in the *Δmfd* strain (Table S3).

Transcription regulation at type I TA loci is essential for cell survival, as overexpression of type I toxins can be lethal(45). We hypothesized that Mfd's regulation of transcription at TA loci was important for survival. We therefore overexpressed the TxpA toxin under an IPTG (isopropyl b-D-1-thiogalacto-pyranoside) inducible promoter in both WT and *Δmfd* strains and performed cellular viability assays. We found that cells lacking Mfd are highly sensitized to both chronic (Fig. 5*A*) and acute (Fig. 5*C*) overexpression of TxpA. Cells lacking Mfd show up to five orders of magnitude sensitivity to overexpression of this toxin. Similarly, we overexpressed the BsrH toxin in WT and *Δmfd* strains to test whether Mfd's effect was conserved. Indeed, we see roughly 4 orders of magnitude sensitization of the *Δmfd* strain with both chronic (Fig. 5*B*) and acute (Fig. 5*D*) overexpression of BsrH. To confirm that this effect was directly due to overexpression of BsrH or TxpA, we performed qRT-PCR analysis in our WT and *Δmfd* strains containing the overexpression constructs. We confirmed that in the TxpA and BsrH overexpression strains, toxin overexpression was increased by ~3-fold and ~2-fold in the *Δmfd* strain, respectively (Fig. 5*E-F*). This finding suggests that Mfd's sensitivity is directly due to toxin overexpression.

Discussion

In this work, we unraveled the fundamental importance of Mfd on regulating transcription and RNAP association at sites of endogenous regions. We found that most sites which share both Mfd association as well as changes in RpoB levels in the absence of Mfd contain regulatory

RNAs. Many of these regulatory RNAs are highly structured. We additionally found that Mfd plays an essential regulatory role in cellular physiology, at least at toxin-antitoxin loci, which contain highly structured RNAs. From these findings, we propose that Mfd is a critical RNAP co-factor, important for regulation of transcription at many chromosomal regions. Regulation by Mfd at sites containing highly structured RNAs are especially pronounced in our datasets and are essential at least for homeostasis of TA systems (Fig. 6). Given the known structural and functional conservation of Mfd(15, 46), it is likely that its role in regulating transcription at highly structured RNAs is also conserved in other species.

Prior work has shown that Mfd can promote transcription termination at significant protein roadblocks(8, 19, 20, 29, 47, 48). Our findings suggest that Mfd's transcription termination activity is particularly critical *in vivo* at regions of high RNA secondary structure. While what induces Mfd association and Mfd's influence on RNAP association at these sites remains unclear, we do know that RNA secondary structure can regulate RNAP dynamics *in vivo*. For example, RNAP pausing can be promoted by the formation of stable RNA hairpin structures in the exit channel of RNAP and inhibit its movement(3, 49–51). Additionally, pausing via RNA secondary structure can regulate gene expression at riboswitches, can promote coupling of transcription and translation(52, 53), and is critical for the process of intrinsic transcription termination in bacteria(10, 54). At certain sites, such pausing may induce RNAP backtracking, but mechanistic studies suggest that more commonly pausing induces a "half translocated" state of RNAP(50), which inhibits its immediate processivity (55). It therefore seems plausible that Mfd is critical for helping promote RNAP release at sites with significant RNAP pausing. Indeed, other RNAP associated factors, specifically NusA and NusG, are known to alter the dynamics between RNAP pausing and transcription elongation(56, 57). Although our experiments do not allow us to identify the precise state of RNAP at sites of Mfd activity *in vivo*, current models regarding RNAP activity hint at the possibility that Mfd can recognize and act not

only on backtracked RNAPs but also on RNAPs that are paused and not fully backtracked, or perhaps simply decreased in its elongation rate. These mechanisms however are not mutually exclusive. It is possible that the specific regions where Mfd modulates transcription have increased RNAP backtracking, however, the results of our GreA experiments are inconsistent with this model.

Given the wide range of biological functions inherent to structured RNAs, the consequences of Mfd's activity at these sites are likely to be broad ranging. For example, we identified multiple riboswitches containing TUs which Mfd binds and promotes RNAP release. These TUs are involved in many critical metabolic processes, ranging from beta-glucoside metabolism (*bglP-bglH-yxiE*)(58, 59), to the utilization of glycerol (*glpF-glpK* and *glpT-glpQ*)(60) to purine metabolism (*purEKBCSQLFMNHD*)(61, 62). We also identified Mfd binding and RNAP release at a locus containing a long cis-acting antisense RNA (*YabE*/S25) that is thought to play a role cell wall maintenance (63). Lastly, amongst many other sites, we also identified Mfd binding and RNAP termination at tRNA loci (the *trnY* locus in *B. subtilis* containing a highly structured RNA of unknown function). The physiological relevance of Mfd's activity at many of these sites requires further investigation that is outside of the scope of the current study.

Various mechanisms of transcription-associated mutagenesis (TAM) exist(64, 65). Based on our findings, we propose that the inherent structure of RNA may be an additional novel mechanism by which transcription promotes mutagenesis, specifically through Mfd. Interestingly, RNA secondary structure has been reported to enhance mutation rates in replicating retroviruses(66), suggesting that evolution via secondary structure may be a universal mechanism. Moreover, a recent study by Thornlow, et al.(67), using computational analyses, revealed that tRNAs have higher mutation rates relative to other parts of the genome(67). They also suggest that this phenomenon is linked to transcription. It is therefore quite possible that, at least in bacteria, the

377 evolution of tRNA structures is mediated by Mfd. By promoting DNA mutagenesis at sites of
378 highly structured RNAs, Mfd may inherently alter the secondary structure encoded at the site of
379 its activity, leading to novel or altered functions of the RNA.
380
381 Interestingly, Mfd also functions as an evolvability factor in diverse bacterial species, promoting
382 rapid evolution of antibiotic resistance development(15) as well as stationary-phase
383 mutagenesis(38, 39). Additionally, non-coding RNAs are well-known to evolve very quickly (68),
384 however, the mechanisms by which this occurs is unknown. Our results suggest that RNAP
385 release contributes to the evolution of these regions. While outside the scope of our studies,
386 addressing this possibility requires further investigations.
387
388 **Materials and Methods**
389 **Strain constructions**
390 All strains and plasmids used and constructed in this study are listed in Table S4 and primers
391 used are listed in Table S5. *B. subtilis* stains used in this study were derivates of the HM1
392 (JH642) parent strain(69). Transformations into *B. subtilis* HM1 were performed under standard
393 conditions as previously described(70). Plasmids used in this study were grown in *E. coli* DH5α.
394 Plasmids were cloned using chemical transformations of competent *E. coli*. All plasmid
395 purification was performed by growth of appropriate *E. coli* strain overnight at 37° C in Luria-
396 Bertani (LB) medium supplement with the appropriate antibiotic and plasmids were
397 subsequently purified using the GeneJet Plasmid Miniprep Kit (Thermo). Further details on
398 strain construction can be found in the SI appendix.
399
400 **Growth conditions**
401 For experiments in *B. subtilis*, cultures were grown as described unless otherwise indicated.
402 Cells were plated on LB supplemented with the appropriate antibiotic for isolation of single

403 colonies. Overnight cultures from single colonies were grown at 37° C in LB at 260 RPMs and
404 the following day cells were diluted back to OD600 0.05 and grown to exponential phase
405 (OD600 0.3-0.5) before harvesting. For acute rifampicin ChIP experiments, cultures were grown
406 in identical fashion until they reach OD600 0.3-0.5 and rifampicin was added at a concentration
407 of 100µg/mL for 5 minutes before harvesting.
408
409 **ChIP-seq and ChIP-qPCR experiments**
410 For *B. subtilis* Mfd ChIP experiments, c-Myc mouse monoclonal antibody (clone 9E10) was
411 used (Thermo). For RpoB experiments, RNA polymerase beta mouse monoclonal antibody
412 (clone 8RB13) was used (Thermo).
413
414 ChIP experiments were performed as previously described(71, 72). Briefly, Cells were grown to
415 exponential phase as previously described and crosslinked with 1% formaldehyde v/v. After 20
416 minutes at room temperature, .5M final concentration of glycine was added and cells were
417 pelleted, washed in cold 1x PBS and pelleted again. Cells were resuspended in solution A (10
418 mM Tris pH 8.0, 10 mM EDTA, 50 mM NaCl, 20% sucrose) supplemented with 1 mg/ml
419 lysozyme and 1 mM AEBSF at 37° C for 30 minutes. 2x IP buffer (100 mM Tris pH 7.0, 10 mM
420 EDTA, 300 mM NaCl, 20% triton x-100), supplemented with 1mM AEBSF, was then added and
421 lysates were incubated on ice for 30 minutes. Cell lysates were sonicated four times as 30%
422 amplitude for ten seconds using a Fisher sonic dismembrator (Fisher FB120). Lysates were
423 centrifuged at 8000 RPMs for 15 minutes at 4° C. The supernatant was transferred into new
424 microfuge tubes.
425
426 ChIP lysates were split into a total DNA input control (40µl of lysate) and immunoprecipitation
427 (IP) (1mL of lysate). For Mfd ChIP experiments 12µl anti-c-Myc antibody were added to the IP
428 samples and 2µl of anti-RpoB antibody was added for RpoB ChIPs. IP lysates were rotated

overnight at 4° C. The following day, 30µl Protein A sepharose beads (GE) were added to the IP samples and rotated for one hour at room temperature. Beads were then pelleted with centrifugation at 2000 RPMs for one minute. Supernatant was decanted and beads were subsequently washed six times with 1x IP buffer and one time with 1x TE pH 8.0. Beads were then pelleted and resuspended in 100µl of elution buffer (50mM Tris pH 8.0, 10mM EDTA), and 1% SDS and incubated at 65° C for 10 minutes. Beads were pelleted by centrifugation and supernatant was transferred to a new microfuge tube. A second round of elution was performed by resuspension of beads in 150µl of elution buffer II (10mM Tris pH 8.0, 1 mM EDTA, 0.67% SDS). Beads were pelleted and supernatant was transferred to microfuge tube containing eluate from the first elution. IP samples were then incubated overnight at 65° C. The following day, proteinase K was added at a final concentration 0.4 mg/mL and samples were incubated for two hours at 37° C. Purification was performed by using the GeneJet PCR Purification Kit (Thermo).

Library preparation for ChIP-seq was performed using the Nextera XT DNA Library Prep Kit (Illumina) according to manufacturer's instructions. For ChIP-quantitative PCR (qPCR), Sso Advanced Universal SYBR Green Supermix (BioRad) was used according to manufacturer's instructions.

RNA-seq experiments

B. subtilis cultures were grown to exponential growth as previously described and harvested by addition of 1:1 volume 100% cold methanol and centrifugation at 5000 RPMs for five minutes. Cell pellets were subsequently lysed in TE and lysozyme (20mg/mL) and purified using the GeneJet RNA Purification Kit (Thermo). Library preparation for RNA-seq was performed using the Scriptseq Complete Kit (Bacteria) from Illumina, according to manufacturer's instructions.

Whole-genome sequencing analysis

455 ChIP-seq and RNA-seq samples were sequenced using the Illumina Nextseq 500/550
456 Sequencing system at the University of Washington Northwest Genomics Center and the
457 VANTAGE Sequencing Core at Vanderbilt University. After sequencing, sample reads from *B.*
458 *subtilis* were mapped to *B. subtilis* 168 genome (accession number: NC_000964.3) using
459 Bowtie2(73). For data visualization, SAMtools was used to process SAM files(74) to produce
460 wiggle plots(75). Wiggle files from all ChIP samples were normalized to input samples (total
461 input DNA subtracted from the ChIP signal). For quantification of ChIP-seq and RNA-seq
462 samples, BAM files were processed by the featureCounts program to determine read counts per
463 gene(76). To determine differential RNA-seq expression and differential ChIP-seq binding, read
464 counts were analyzed by DEseq2 software(77). In order to determine correlations between
465 RpoB ChIP binding and Mfd ChIP, read counts generated by featureCounts were divided by the
466 total number of sequencing reads per sample. ChIP samples were then divided by input
467 samples and $\log_2$ normalized.
468
469 **Quantitative RT-PCR assays**
470 For quantification of RNA using qRT-PCR, RNA was extracted as described in the RNA-seq
471 experiments section. Subsequently, 1µg of total RNA was treated with DNaseI (Thermo) for one
472 hour at 37° C. DNase denaturation was performed with addition of 10mM EDTA and incubation
473 at 65° C for 10 minutes. cDNA generation was performed using the iScript Supermix (BioRad),
474 according to the manufacturer's instructions. Quantitative PCR (qPCR) was performed using the
475 Sso Advanced Universal SYBR Green Supermix (BioRad), according to manufacturer's
476 instructions. For normalization of qRT-PCR, primers to *B. subtilis* rRNA was used.
477
478 **Cell survival assays**
479 For chronic survival assays, strains were struck out on LB agar plates and *B. subtilis* cultures
480 were grown in 2mL LB until they reached an OD600 of 0.5-1.0. All cultures were normalized to

OD600 0.3 and serial dilutions were performed in 1x Spizizen's salts. 5µl of cells were plated on control plates containing LB agar only and LB agar plates containing the designated concentration of IPTG (see figure legends for concentrations). Plates were grown at 30° C overnight and CFUs were enumerated the following day.

For acute survival assays, cultures were grown in 2ml until they reached an OD600 of 0.5-1.0 and then diluted back to OD600 0.05. Either was 1mM or .1 mM IPTG was then added and cells were grown for approximately 60 minutes (OD600~.3). Cells were subsequently washed two times with 1x Spizizen's salts to remove residual IPTG and were serially diluted. 5µl of cells were plated on LB agar and plates were grown at 30° C overnight for CFU enumeration. For both chronic and acute survival assays, images were taken using the BioRad Gel Doc XR+ Molecular Imager.

Acknowledgements

We thank Patrick Nugent and Ankunda Kariisa for assistance with strain construction and bacterial 2-hybrid assays. We also thank Charlie Lee from the Shendure Lab at the University of Washington as well as the VANTAGE genomics core at Vanderbilt for assistance with genome sequencing experiments. We also thank Chris Merrikh and Kevin Lang for helpful scientific discussions. This work was supported by the National Institute of Allergy and Infectious Diseases R01 AI127422.

Footnotes

Author contributions: M.N.R. and H.M. conceived of and designed experiments. M.N.R. performed experiments. M.N.R. and H.M. analyzed data and wrote the paper.

To whom the correspondence should be addressed. Email: houra.merrikh@vanderbilt.edu

507

508 The authors declare no conflict of interest

509

510 **References**

511 1. I. Artsimovitch, R. Landick, Pausing by bacterial RNA polymerase is mediated by
512 mechanistically distinct classes of signals. *Proc. Natl. Acad. Sci.* **97**, 7090–7095 (2000).

513 2. A. Mayer, H. M. Landry, L. S. Churchman, Pause & go: from the discovery of RNA
514 polymerase pausing to its functional implications. *Curr. Opin. Cell Biol.* **46**, 72–80 (2017).

515 3. C. L. Chan, D. Wang, R. Landick, Multiple interactions stabilize a single paused
516 transcription intermediate in which hairpin to 3' end spacing distinguishes pause and
517 termination pathways. *J. Mol. Biol.* **268**, 54–68 (1997).

518 4. D. A. Erie, The many conformational states of RNA polymerase elongation complexes
519 and their roles in the regulation of transcription. *Biochim. Biophys. Acta* **1577**, 224–239
520 (2002).

521 5. A. C. M. Cheung, P. Cramer, Structural basis of RNA polymerase II backtracking, arrest
522 and reactivation. *Nature* **471**, 249–253 (2011).

523 6. E. Nudler, A. Mustaev, E. Lukhtanov, A. Goldfarb, The RNA–DNA Hybrid Maintains the
524 Register of Transcription by Preventing Backtracking of RNA Polymerase. *Cell* **89**, 33–41
525 (1997).

526 7. E. Nudler, RNA Polymerase Backtracking in Gene Regulation and Genome Instability.
527 *Cell* **149**, 1438–1445 (2012).

528 8. D. Dutta, K. Shatalin, V. Epshtein, M. E. Gottesman, E. Nudler, Linking RNA polymerase
529 backtracking to genome instability in E. coli. *Cell* **146**, 533–543 (2011).

9. J. Park, M. T. Marr, J. W. Roberts, E. coli Transcription Repair Coupling Factor (Mfd Protein) Rescues Arrested Complexes by Promoting Forward Translocation. *Cell* **109**, 757–767 (2002).

10. J.-S. Park, J. W. Roberts, Role of DNA bubble rewinding in enzymatic transcription termination. *Proc. Natl. Acad. Sci.* **103**, 4870–4875 (2006).

11. I. Mellon, G. Spivak, P. C. Hanawalt, Selective removal of transcription-blocking DNA damage from the transcribed strand of the mammalian DHFR gene. *Cell* **51**, 241–249 (1987).

12. C. Selby, A. Sancar, Molecular mechanism of transcription-repair coupling. *Science (80-.).* **260**, 53 (1993).

13. Christopher P. Selby, Mfd Protein and Transcription-Repair Coupling in E. coli. *Photochem Photobiol.* **93**, 280–295 (2017).

14. P. C. Hanawalt, G. Spivak, Transcription-coupled DNA repair: two decades of progress and surprises. *Nat. Rev. Mol. Cell Biol.* **9**, 958–70 (2008).

15. M. N. Ragheb, *et al.*, Inhibiting the Evolution of Antibiotic Resistance. *Mol. Cell* **73**, 157-165.e5 (2019).

16. V. Kamarthapu, *et al.*, ppGpp couples transcription to DNA repair in E.coli. *Science (80-.).* (2016) https:/doi.org/10.1126/science.aad6945.

17. V. Epshtein, *et al.*, UvrD facilitates DNA repair by pulling RNA polymerase backwards. *Nature* **505**, 372–377 (2014).

18. S. E. Cohen, *et al.*, Roles for the transcription elongation factor NusA in both DNA repair and damage tolerance pathways in Escherichia coli. *Proc. Natl. Acad. Sci.* **107**, 15517–15522 (2010).

553 19. J. M. Zalieckas, L. V Wray, A. E. Ferson, S. H. Fisher, Transcription–repair coupling
554 factor is involved in carbon catabolite repression of the Bacillus subtilis hut and gnt
555 operons. *Mol. Microbiol.* **27**, 1031–1038 (1998).

556 20. J. M. Zalieckas, L. V Wray, S. H. Fisher, Expression of the Bacillus subtilis acsA gene:
557 position and sequence context affect cre-mediated carbon catabolite repression. *J.*
558 *Bacteriol.* **180**, 6649–54 (1998).

559 21. T. T. Le, *et al.*, Mfd Dynamically Regulates Transcription via a Release and Catch-Up
560 Mechanism. *Cell* **173**, 1823 (2018).

561 22. H. N. Ho, A. M. van Oijen, H. Ghodke, The transcription-repair coupling factor Mfd
562 associates with RNA polymerase in the absence of exogenous damage. *Nat. Commun.*
563 **9**, 1–12 (2018).

564 23. A. M. Deaconescu, *et al.*, Structural Basis for Bacterial Transcription-Coupled DNA
565 Repair. *Cell*, 507–520 (2006).

566 24. A. L. Chambers, A. J. Smith, N. J. Savery, A DNA translocation motif in the bacterial
567 transcription-repair coupling factor, Mfd. *Nucleic Acids Res.* **31** (2003).

568 25. A. J. Smith, N. J. Savery, RNA polymerase mutants defective in the initiation of
569 transcription-coupled DNA repair. *Nucleic Acids Res.* **33**, 755–764 (2005).

570 26. W. R. McClure, C. L. Cech, On the mechanism of rifampicin inhibition of RNA synthesis.
571 *J. Biol. Chem.* **253**, 8949–8956 (1978).

572 27. E. A. Campbell, *et al.*, Structural mechanism for rifampicin inhibition of bacterial RNA
573 polymerase. *Cell* **104**, 901–912 (2001).

574 28. P. Nicolas, *et al.*, Condition-dependent transcriptome reveals high-level regulatory
575 architecture in Bacillus subtilis. *Science* **335**, 1103–6 (2012).

576 29. R. S. Washburn, Y. Wang, M. E. Gottesman, Role of E .coli Transcription-Repair
577 Coupling Factor Mfd in Nun-mediated Transcription Termination. *J. Mol. Biol.* **2836**, 655–
578 662 (2003).

579 30. R. A. T. Mars, P. Nicolas, E. L. Denham, J. M. van Dijl, Regulatory RNAs in Bacillus
580 subtilis: a Gram-Positive Perspective on Bacterial RNA-Mediated Regulation of Gene
581 Expression. *Microbiol. Mol. Biol. Rev.* **80**, 1029–1057 (2016).

582 31. A. Ray-Soni, M. J. Bellecourt, R. Landick, Mechanisms of Bacterial Transcription
583 Termination: All Good Things Must End. *Annu. Rev. Biochem.* **85**, 319–347 (2016).

584 32. I. Gusarov, E. Nudler, The mechanism of intrinsic transcription termination. *Mol. Cell* **3**,
585 495–504 (1999).

586 33. A. Weixlbaumer, K. Leon, R. Landick, S. A. Darst, Structural basis of transcriptional
587 pausing in bacteria. *Cell* **152**, 431–441 (2013).

588 34. S. Y. Le, J. V Maizel, A method for assessing the statistical significance of RNA folding. *J.*
589 *Theor. Biol.* **138**, 495–510 (1989).

590 35. E. Freyhult, P. P. Gardner, V. Moulton, A comparison of RNA folding measures. *BMC*
591 *Bioinformatics* **6**, 241 (2005).

592 36. S. Borukhov, A. Polyakov, V. Nikiforov, A. Goldfarb, GreA protein: A transcription
593 elongation factor from Escherichia coli. *Proc. Natl. Acad. Sci.* **89**, 8899–8902 (1992).

594 37. F. Ozsolak, P. M. Milos, RNA sequencing: advances, challenges and opportunities. *Nat.*
595 *Rev. Genet.* **12**, 87–98 (2011).

596 38. H. A. Martin, M. Pedraza-reyes, R. E. Yasbin, E. A. Robleto, Transcriptional De-
597 Repression and Mfd Are Mutagenic in Stressed Bacillus subtilis Cells. *J Mol Microbiol*
598 *Biotechnol* **21**, 45–58 (2011).

599 39. C. Ross, *et al.*, Novel Role of mfd : Effects on Stationary-Phase Mutagenesis in Bacillus
600 subtilis. *J. Bacteriol.* **188**, 7512–7520 (2006).

601 40. S. Million-Weaver, *et al.*, An underlying mechanism for the increased mutagenesis of
602 lagging-strand genes in Bacillus subtilis. *Proc. Natl. Acad. Sci. U. S. A.* **112**, E1096-105
603 (2015).

604 41. S. Brantl, N. Jahn, SRNAs in bacterial type I and type III toxin-antitoxin systems. *FEMS*
605 *Microbiol. Rev.* **39**, 413–427 (2015).

606 42. T. Dörr, M. Vulić, K. Lewis, Ciprofloxacin Causes Persister Formation by Inducing the
607 TisB toxin in Escherichia coli. *PLoS Biol.* **8**, e1000317 (2010).

608 43. J. Domka, J. Lee, T. Bansal, T. K. Wood, Temporal gene-expression in Escherichia coli
609 K-12 biofilms. *Environ. Microbiol.* **9**, 332–346 (2007).

610 44. S. Durand, N. Jahn, C. Condon, S. Brantl, Type i toxin-antitoxin systems in Bacillus
611 subtilis. *RNA Biol.* **9**, 1491–1497 (2012).

612 45. J. M. Silvaggi, J. B. Perkins, R. Losick, Small Untranslated RNA Antitoxin in Bacillus
613 subtilis. *J. Bacteriol.* **187**, 6641–6650 (2005).

614 46. M. Ragheb, H. Merrikh, The enigmatic role of Mfd in replication-transcription conflicts in
615 bacteria. *DNA Repair (Amst).* **81** (2019).

616 47. C. P. Selby, A. Sancar, Transcription-repair coupling and mutation frequency decline.
617 *Microbiol. Rev.* **175**, 7509–7514 (1993).

618 48. S. C. Hung, M. E. Gottesman, Phage HK022 Nun Protein Arrests Transcription on Phage
619 lambda DNA in vitro and Competes with the Phage lambda N Antitermination Protein. *J.*
620 *Mol. Biol.*, 428–442 (1995).

621 49. J. Y. Kang, *et al.*, RNA Polymerase Accommodates a Pause RNA Hairpin by Global
622 Conformational Rearrangements that Prolong Pausing. *Mol. Cell* **69**, 802-815.e1 (2018).

623 50. M. L. Kireeva, M. Kashlev, Mechanism of sequence-specific pausing of bacterial RNA
624 polymerase. *Proc. Natl. Acad. Sci.* **106**, 8900–8905 (2009).

625 51. V. R. Tadigotla, *et al.*, Thermodynamic and kinetic modeling of transcriptional pausing.
626 *Proc. Natl. Acad. Sci.* **103**, 4439–4444 (2006).

627 52. M. E. Winkler, C. Yanofsky, Pausing of RNA polymerase during in vitro transcription of
628 the tryptophan operon leader region. *Biochemistry* **20**, 3738–44 (1981).

629 53. A. V. Yakhnin, H. Yakhnin, P. Babitzke, RNA Polymerase Pausing Regulates Translation
630 Initiation by Providing Additional Time for TRAP-RNA Interaction. *Mol. Cell* **24**, 547–557
631 (2006).

632 54. A. Ray-Soni, R. A. Mooney, R. Landick, Trigger loop dynamics can explain stimulation of
633 intrinsic termination by bacterial RNA polymerase without terminator hairpin contact.
634 *Proc. Natl. Acad. Sci.* **114**, E9233–E9242 (2017).

635 55. J. Saba, *et al.*, The elemental mechanism of transcriptional pausing. *Elife* **8** (2019).

636 56. M. Turtola, G. A. Belogurov, NusG inhibits RNA polymerase backtracking by stabilizing
637 the minimal transcription bubble. *Elife* **5**, 1–27 (2016).

638 57. M. Strauß, *et al.*, Transcription is regulated by NusA:NusG interaction. *Nucleic Acids Res.*
639 **44**, 5971–5982 (2016).

640 58. S. Krüger, S. Gertz, M. Hecker, Transcriptional analysis of bglPH expression in Bacillus
641 subtilis: Evidence for two distinct pathways mediating carbon catabolite repression. *J.*
642 *Bacteriol.* **178**, 2637–2644 (1996).

643 59. I. Irnov, C. M. Sharma, J. Vogel, W. C. Winkler, Identification of regulatory RNAs in
644 Bacillus subtilis. *Nucleic Acids Res.* **38**, 6637–6651 (2010).

645 60. R.-P. Nilsson, L. Beijer, B. Rutberg, The glpT and glpQ genes of the glycerol regulon in
646 Bacillus subtilis. *Microbiology* **140**, 723–730 (1994).

647 61. M. Mandal, B. Boese, J. E. Barrick, W. C. Winkler, R. R. Breaker, Riboswitches Control
648 Fundamental Biochemical Pathways in Bacillus subtilis and Other Bacteria. *Cell* **113**,
649 577–586 (2003).

650 62. D. J. Ebbole, H. Zalkin, Cloning and characterization of a 12-gene cluster from Bacillus
651 subtilis encoding nine enzymes for de novo purine nucleotide synthesis. *J. Biol. Chem.*
652 **262**, 8274–8287 (1987).

653 63. W. Eiamphungporn, J. D. Helmann, Extracytoplasmic Function σ Factors Regulate
654 Expression of the Bacillus subtilis yabE Gene via a cis-Acting Antisense RNA. *J.*
655 *Bacteriol.* **191**, 1101–1105 (2009).

656 64. S. Jinks-Robertson, A. S. Bhagwat, Transcription-Associated Mutagenesis. *Annu. Rev.*
657 *Genet.* **48**, 341–359 (2014).

658 65. N. Kim, S. Jinks-Robertson, Transcription as a source of genome instability. *Nat. Rev.*
659 *Genet.* **13**, 204–214 (2012).

660 66. V. K. Pathakt, H. M. Temin, 5-Azacytidine and RNA Secondary Structure Increase the
661 Retrovirus Mutation Rate. *J. Virol.* **66**, 3093–3100 (1992).

662 67. T. M. Lowe, *et al.*, Transfer RNA genes experience exceptionally elevated mutation rates.
663 *Proc. Natl. Acad. Sci.* **115**, 8996–9001 (2018).

664 68. H. A. Dutcher, R. Raghavan, Origin, Evolution, and Loss of Bacterial Small RNAs.
665 *Microbiol. Spectr.* **6** (2018).

69. S. P. Brehm, S. P. Staal, J. A. Hoch, Phenotypes of pleiotropic-negative sporulation mutants of Bacillus subtilis. *J. Bacteriol.* **115**, 1063–70 (1973).

70. C. . Harwood, S. M. Cutting, J. Wiley, *Molecular Biological Methods for Bacillus* (1990) (April 11, 2019).

71. K. S. Lang, *et al.*, Replication-Transcription Conflicts Generate R-Loops that Orchestrate Bacterial Stress Survival. *Cell* **170**, 787-790.e18 (2017).

72. H. Merrikh, C. Machón, W. H. Grainger, A. D. Grossman, P. Soultanas, Co-directional replication-transcription conflicts lead to replication restart. *Nature* **470**, 554–7 (2011).

73. B. Langmead, S. L. Salzberg, Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 (2012).

74. H. Li, *et al.*, The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).

75. O. R. Homann, A. D. Johnson, MochiView: versatile software for genome browsing and DNA motif analysis. *BMC Biol.* **8**, 49 (2010).

76. Y. Liao, G. K. Smyth, W. Shi, featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* **30**, 923–930 (2014).

77. M. I. Love, W. Huber, S. Anders, Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).

78. A. Conesa, *et al.*, A survey of best practices for RNA-seq data analysis. *Genome Biol.* **17**, 13 (2016).

Figure Titles and Legends

Fig. 1. Mfd functions as an RNAP co-factor and requires transcription elongation for association with DNA

(A) ChIP-seq plot of *B. subtilis* Mfd tagged with 1x myc (red) and of WT *B. subtilis* (light green) using myc antibody. (B) ChIP-seq plot of WT *B. subtilis* RpoB. (C) ChIP-seq plot of *B. subtilis* MfdL522A-myc point mutant. (D) *B. subtilis* Mfd-myc treated with 50μg/mL of rifampicin for five minutes. Plots are normalized to total DNA input controls and are the average of at least two independent experiments.

Fig. 2. Mfd directly promotes release of RNAP *in vivo*

(A) RpoB ChIP-seq plots showing regions of RpoB enrichment in *Δmfd*. Top half of graph (read counts in red) reflects normalized RpoB ChIP-seq read counts where *B. subtilis Δmfd* had increased signal relative to WT. Bottom half of graph (read counts in green) reflects RpoB ChIP-seq read counts where *Δmfd* had decreased signal relative to WT *B. subtilis*. High background signal from ribosomal RNA was removed from plots for better visualization. Zoomed in plots are representative regions of high RpoB enrichment in *Δmfd*. (B) Scatter plot of WT and *Δmfd* RpoB ChIP-seq measuring signal at each gene in *B. subtilis*. For quantification of ChIP signal at each gene, read counts for each gene were normalized to total library counts and IP samples were normalized to total DNA input to calculate an IP/Total DNA ratio. Ratios were $\log_2$ normalized and averaged over at least two independent experiments. Data points above and below colored shading indicate greater than two-fold increase and decrease in RpoB signal in the *Δmfd* strain, respectively. Data points in red indicate genes bound by Mfd. Binding is defined as one standard deviation greater than the average ChIP signal across all genes in *B. subtilis*. Calculation of Mfd binding at each gene was determined as described for RpoB ChIP samples.

Fig. 3. Transcription units with Mfd binding and increased RNAP density in *Δmfd* are enriched for structured regulatory RNAs

Scatter plot of the minimum free energy (MFE) Z-score for regulatory RNAs in *B. subtilis*. Data points represent regulatory RNAs within TUs that have no observed change in RpoB density between WT and *Δmfd* (grey data points), increased RpoB density in *Δmfd* and bound by Mfd (red data points), and decreased RpoB density in *Δmfd* (black data points). Error bars represent the standard error of the mean (SEM). Statistical significance was determined using two-tailed Z-test for two population means (****p<0.0001).

Fig. 4. Mfd promotes repression of transcription at sites of structured sRNAs

(A) Scatter plot of RNA-seq. Data points represent the expression level of each gene in *B. subtilis* in WT and *Δmfd* strains. Scatter plot represents expression level calculated using read per kilobase per million reads (RPKM)(78), from at least two independent experiments. Data points in red indicate genes with increased RpoB occupancy in *Δmfd*. (B) qRT-PCR analysis of three regions with increased RNAP occupancy in *Δmfd txpA/ratA* (left), *bsrH/as-bsrH* (middle), and the *trnY* locus (right). RNA values normalized to ribosomal RNA. Error bars represent the SEM from at least two different experiments. Statistical significance was determined using a two-tailed Student's T-test (****p<0.0001).

Fig. 5. Transcriptional regulation by Mfd at toxin-antitoxin loci is essential for cell survival

Survival assays under chronic (A) and acute (C) overexpression of TxpA toxin in WT and *Δmfd* and survival assays under chronic (B) and acute (D) overexpression of BsrH toxin in WT and *Δmfd*. For all figures, representative images shown above, and quantification of data shown below. Error bars represent the SEM from at least three independent experiments. (E and F) qRT-PCR analysis of txpA and bsrH overexpression in WT and *Δmfd* strains. RNA values normalized to ribosomal RNA. Error bars represent the SEM from at least two independent experiments. Statistical significance was determined using a two-tailed Student's T-test (*p<0.05, **p<0.01, ***p<0.001).

738 **Fig. 6. Model of Mfd activity at structured regulatory RNAs**

739 During transcription, elongating RNAP (shown in green) transcribed a highly structured RNA
740 sequence (shown in red). This can arrest RNAP (shown in maroon) on DNA. Bottom left- the
741 arrested complex is recognized by Mfd (shown in purple), which releases RNAP from the
742 template and in doing so represses transcription and promotes mutagenesis (shown as yellow
743 stars on DNA). Bottom right- in the absence of Mfd, RNAP resumes transcription, leading to
744 higher levels of RNA and decreased mutagenesis.

745 **Tables**

746 **Table 1.** Summary list of genes and previously defined TUs bound with changes in RpoB
747 density in *Δmfd* from ChIP-seq analysis. Changes in RpoB density and Mfd binding at TUs are
748 defined by changes in one or more genes corresponding to its associated TUs

749

	Decreased RpoB density in *Δmfd*	**Increased RpoB density in *Δmfd***
Genes (and TUs)	116 (71)	53 (31)
Genes (and TUs) bound by Mfd	52 (35)	0 (0)

750

Fig. 1

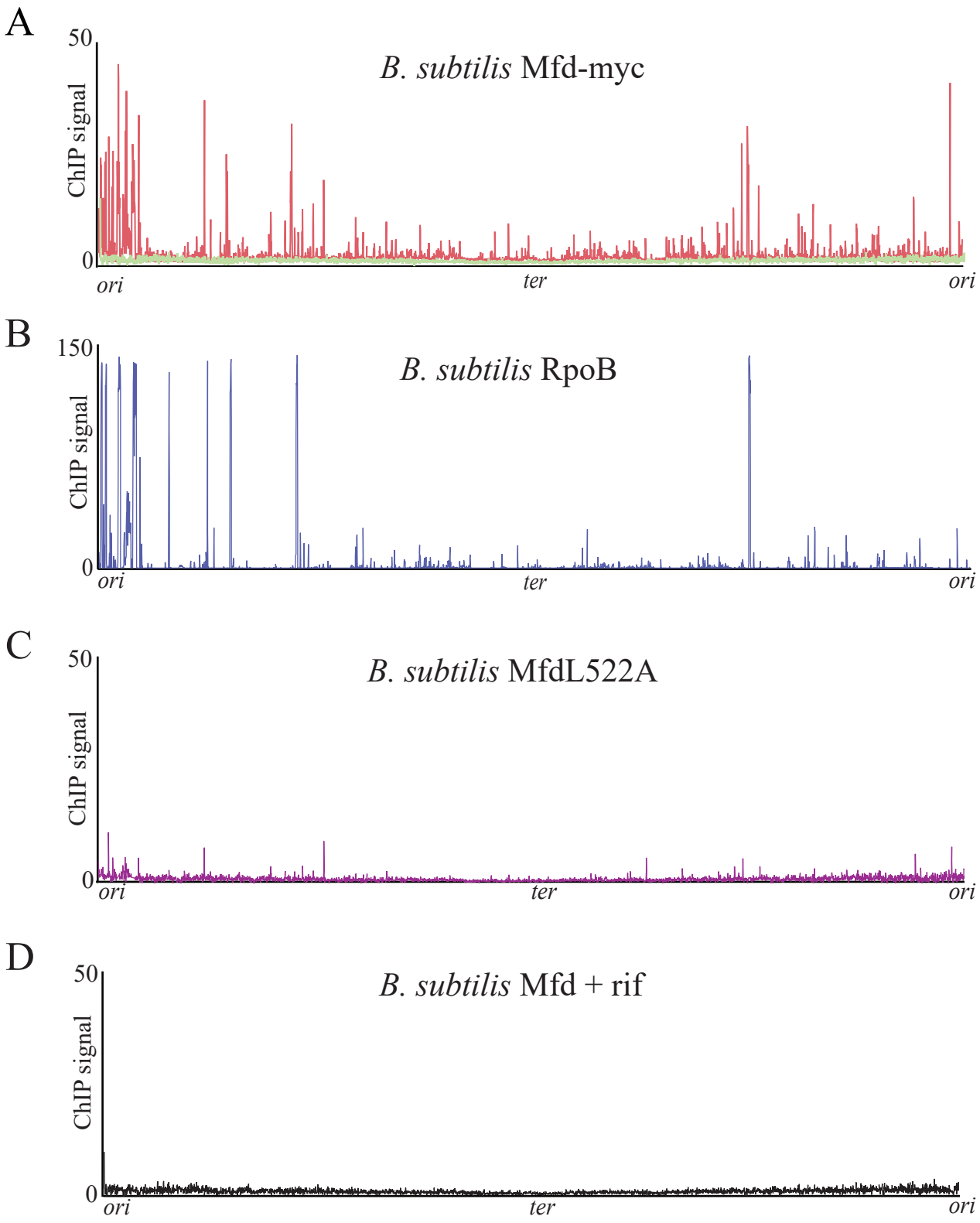
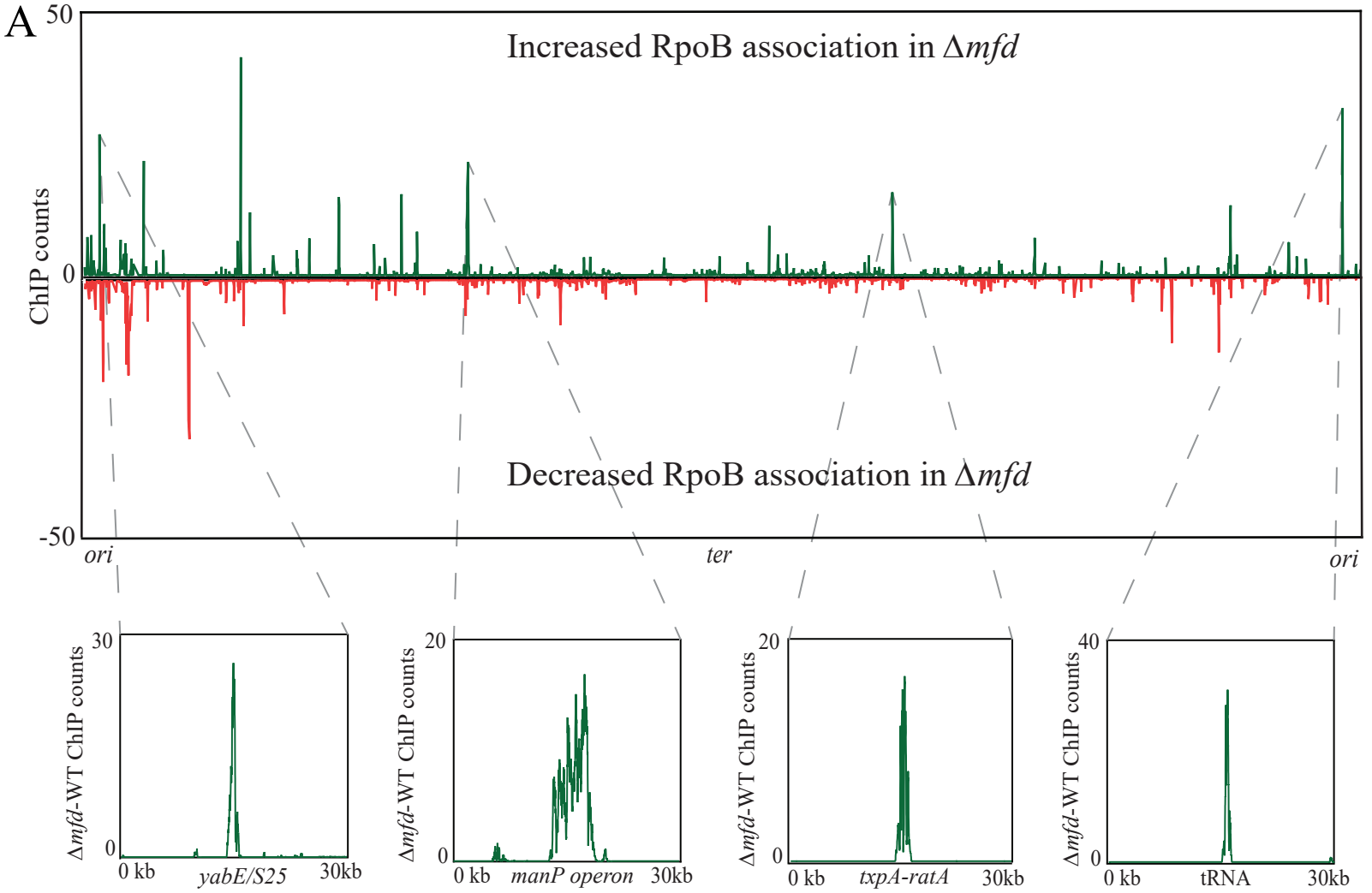


Fig. 2



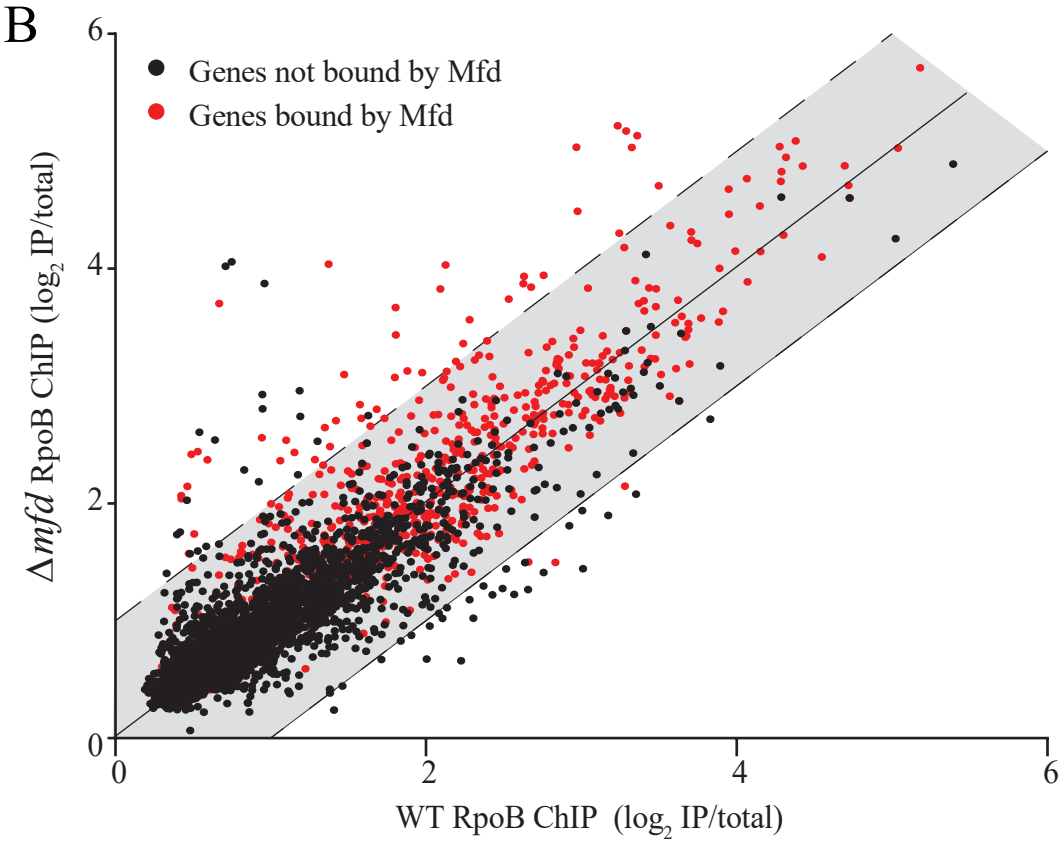


Fig. 3

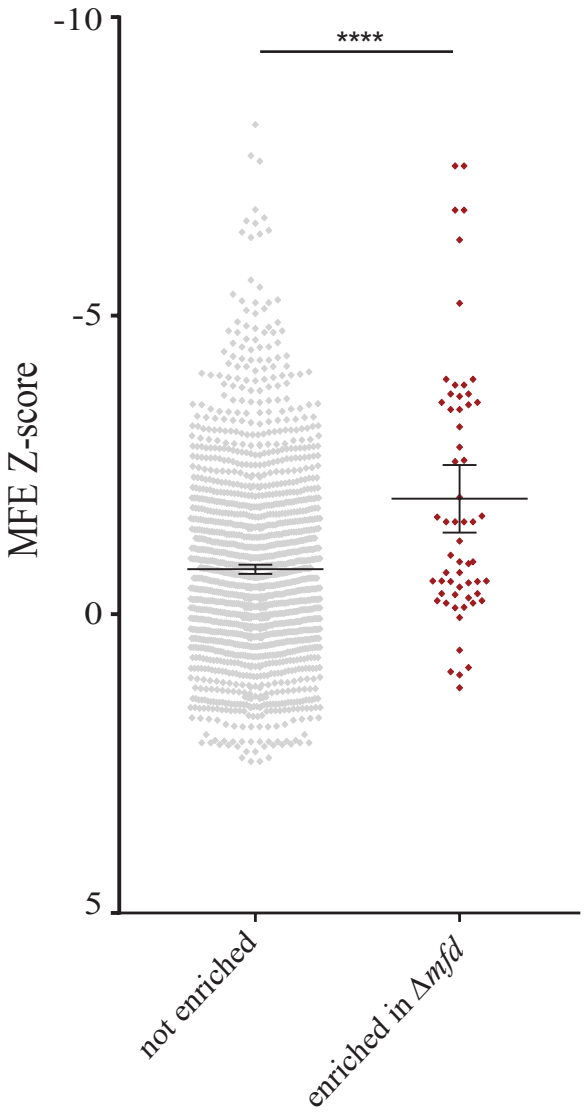


Fig. 4

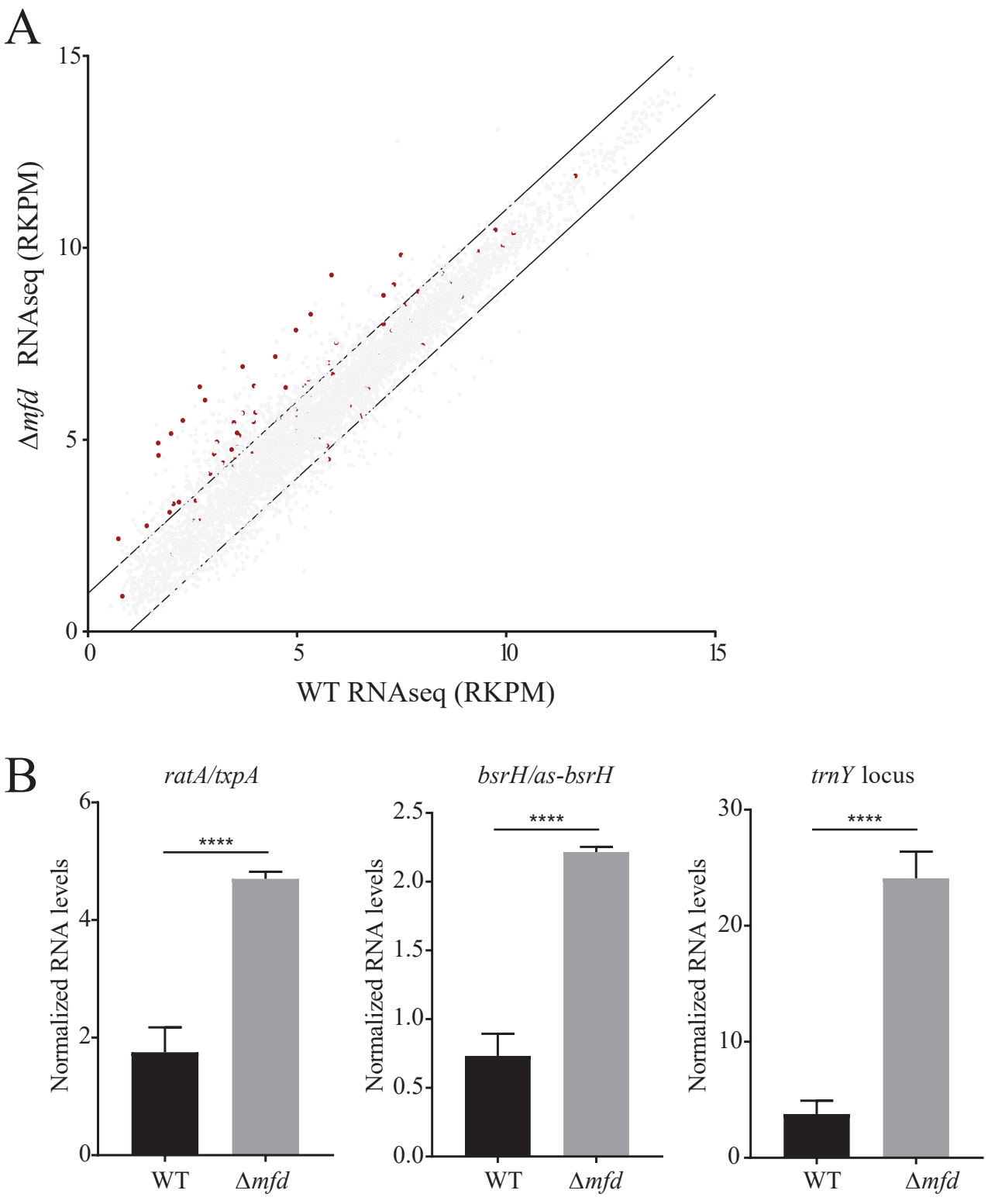


Fig. 5

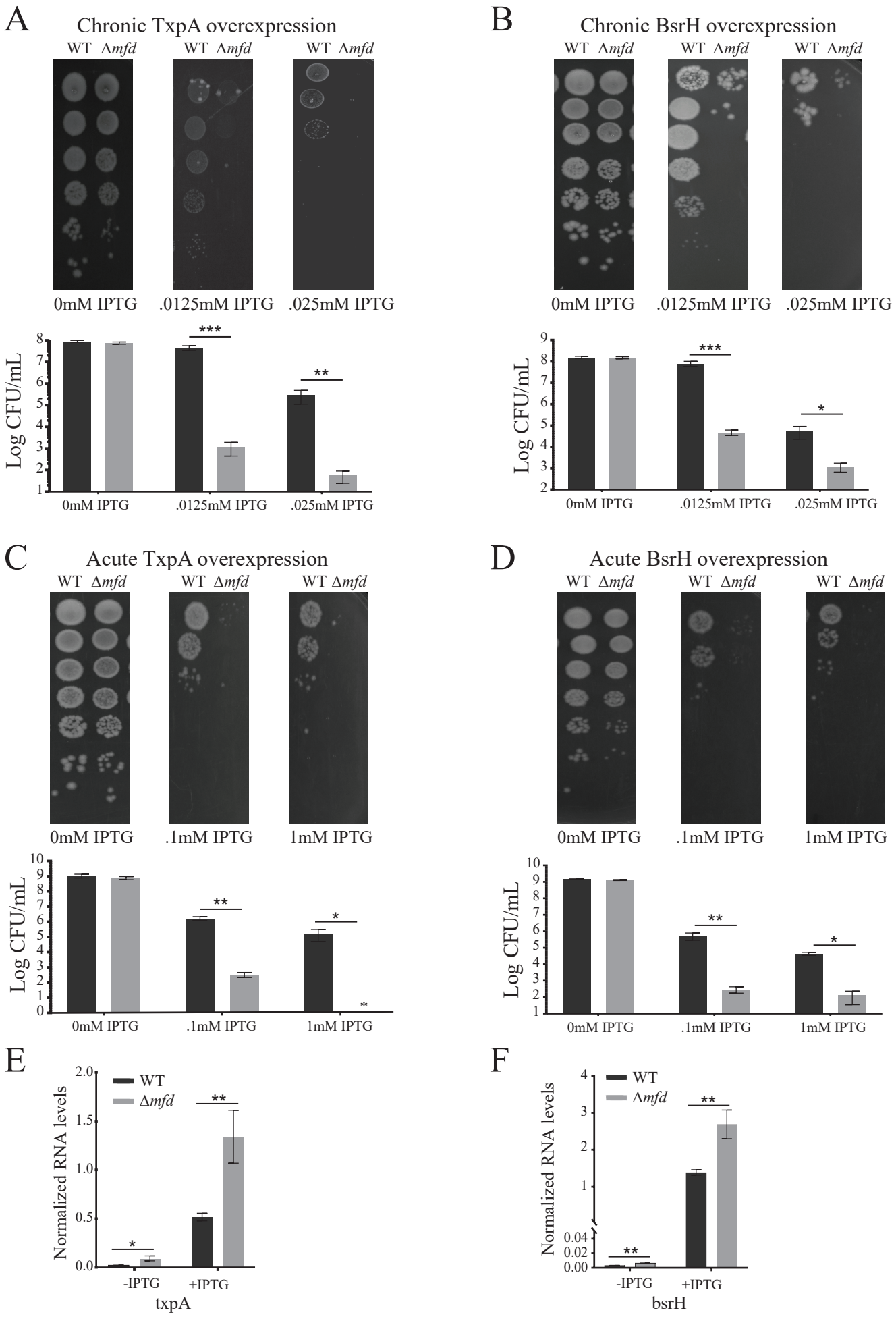
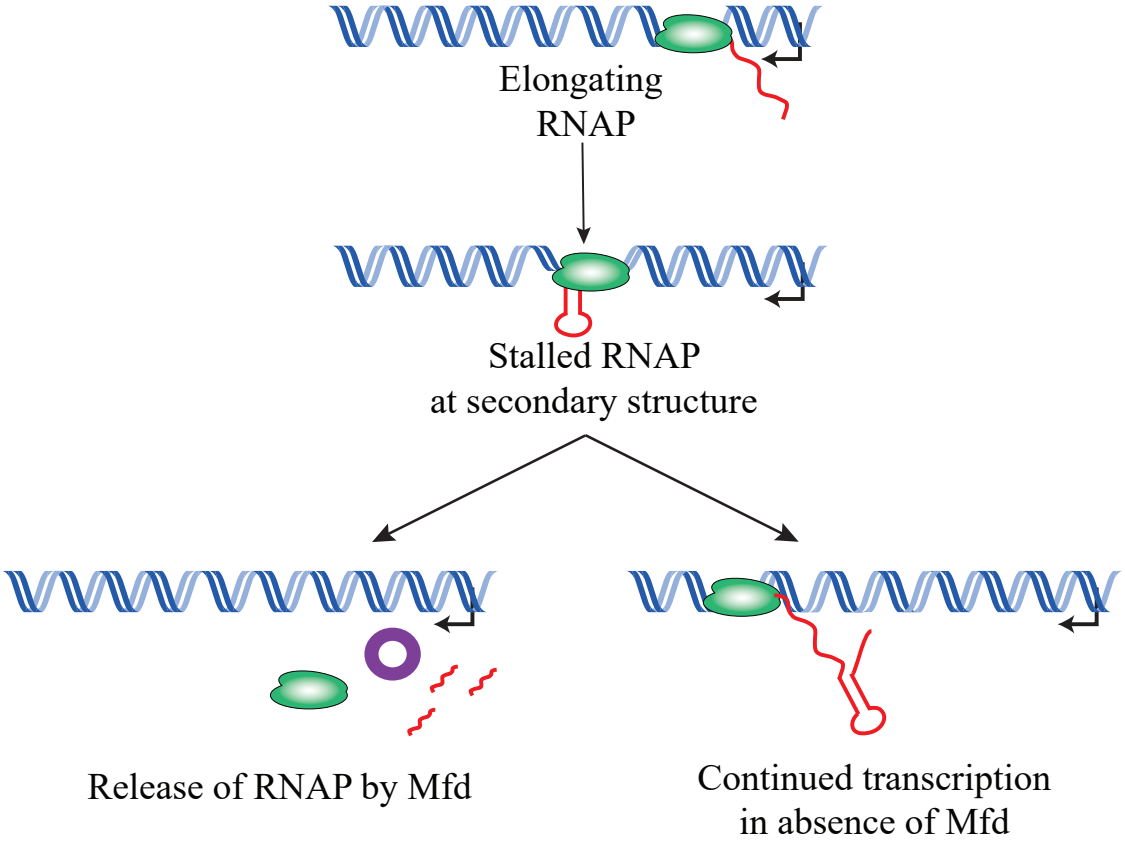


Fig. 6



Supplementary Information for:

Genome-wide association and function studies identify Mfd as a critical RNA polymerase co-factor at hard-to-transcribe regions

Mark Ragheb, Houra Merrikh

Corresponding author: Houra Merrikh
Email: houra.merrikh@Vanderbilt.edu

This PDF file includes:

Expanded Materials and Methods

Detailed strain construction

In order to construct marker-less point mutations of Mfd in *B. subtilis*, the pminiMAD2 plasmid was used as previously described (1). Briefly, HM2916 was constructed by transforming pHM707 into HM1, respectively and grown at in LB broth containing MLS antibiotics at 22°C, the permissive temperature. *B. subtilis* strains were then subsequently incubated for 12 hours at 42°C while maintaining MLS selection. Cells were then serially diluted and passaged multiple times at 22°C. Individual colonies were then plated on LB plates with or without MLS to identify colonies which were MLS sensitive and had evicted the plasmid.

HM2769 was constructed by transforming pHM430 and pHM439 into HM2747. HM2771 was constructed by transforming pHM430 and pBRα into HM2747. HM2773 was constructed by transforming pHM439 and pACλCI into HM2747. HM2965 was constructed by transforming pHM441 and pHM439 into HM2747. HM2932 was constructed by transforming the HM2916 plasmid into HM1.

HM3157 was constructed using the transformation of SOE PCR product into HM1. First, Mfd-myc amplicon was generated using primers HM3759 and HM3760 and HM1 genomic DNA as a template in order to add a 1x myc sequence to the 3' end of the Mfd gene. Erm resistance cassette was amplified using pCAL215 plasmid DNA as a template and primers HM3854 and HM3969. These two respective amplicons were used as templated to generate a PCR SOE product using primers HM3759 and HM3854.

HM3808 was constructed by transformation of HM712 genomic DNA into HM1. HM3933 was constructed by transformation of HM1333 genomic DNA into HM1451. HM3945 and HM3947 were constructed by transforming HM3157 genomic DNA into HM2916 and HM2522, respectively. HM3986, HM3988, and HM3990 were constructed by transforming plasmid pHM676 into *E. coli* DH5α, HM1, and HM2521, respectively. HM4002, HM4003, and HM4004 were constructed by transforming plasmid pHM682 into *E. coli* DH5α, HM1, and HM2521, respectively.

Detailed plasmid construction

pHM430 was built using Gibson cloning from pACλCI-β-flap backbone (BamHI/NotI digested) and *B. subtilis* Mfd amplicon (AA492-AA625) amplicon with stop codon added (using primers HM3286 and HM3287). pHM431 was built using site-directed mutagenesis of pHM430 using primers HM3540 and H3541. pHM439 was built using Gibson cloning from pBRα-β-flap backbone (BamHI/NotI digested and *B. subtilis* rpoB amplicon (AA21-AA131) with stop codon added (using primers HM3292 and HM3293).

pHM676 was built using digestion of pDR111 with sphI and subsequent ligation of a PCR amplicon generated with primers HM5418 and HM5419 and HM1 genomic DNA as a template.

pHM682 was built using digestion of pDR111 with sphI and subsequent ligation of a PCR amplicon generated with primers HM5462 and HM5463 and HM1 as a template.

pHM707 was built using digestion of pminiMAD2 with kpnI and bamHI and subsequent ligation of a PCR amplicon with primers HM1004 and HM1005 with HM1 genomic DNA as a template. Mutations were subsequently introduced via site-directed mutagenesis using primers HM3540 and HM3541.

Western blot assay

Exponentially growing cultures were centrifuged, resuspended in Tris/Salt buffer (50 mM Tris-HCl pH 8, 300mM NaCl), and pelleted. Cell lysis buffer (10mM Tris-HCl pH7, 10mM EDTA, .1mM AEBSF, .1mg/ml lysozyme) was added and samples were incubated at 37° C for 15 minutes. SDS loading buffer was added to samples and 20µl was loaded onto Mini-PROTEAN TGX Precast Gels (BioRad) and run in Tris/SDS/Glycine running buffer in a Mini-PROTEAN Electrophoresis Cell (BioRad) at 200V for 40 minutes. Transfer was performed using the Trans-Blot Turbo Transfer System (BioRad). Anti-c-Myc antibody (1:5000 dilution) was added and blots were incubated overnight at 4° C. Anti-mouse antibodies (Li-Cor) (1:15000 dilution) was added and blot was imaged using the Odyssey CLx imaging system (Li-Cor).

Bacterial 2-hybrid assays

Bacterial 2-hybrid assays were performed as previously described(2). Briefly, RNAP interacting domains of B. subtilis Mfd (WT and L522A) and the Mfd interacting domain of RpoB were fused to Lambda repressor the N-terminal domain of E. coli RNAP alpha subunit. Fusions were subsequently transformed into a strain of E. coli containing the lambda operator sequence upstream of a luciferase reporter gene. In order to measure relative light units (RLUs), E. coli strains were grown overnight at in LB + 20mM IPTG at 30° C. The following day, cells were diluted 1:100 into LB+20mM IPTG and growing until OD600 ~2.0. Measurement of RLUs was performed using the Nano-glo substrate (Promega), according to the manufacturer's instructions. Luminescence was measured using the SpectraMax M3 96-well plate reader.

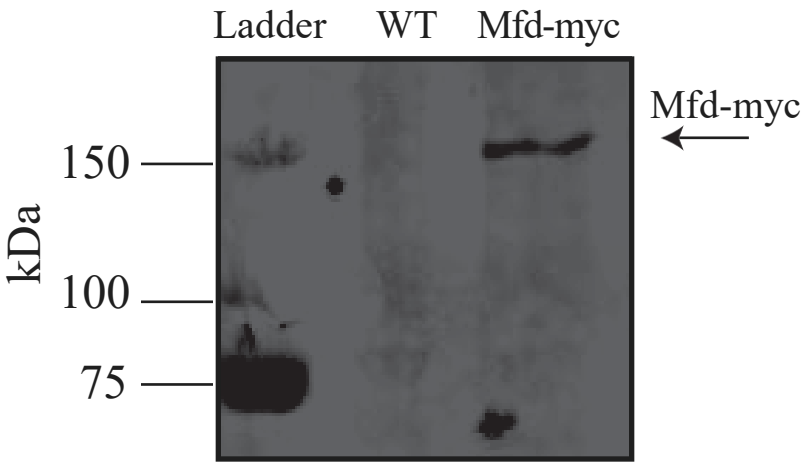


Fig. S1. Western blot of *B. subtilis* Mfd-myc

Western blot of *B. subtilis* WT and Mfd-myc. Anti-c-Myc antibody was used to probe blot.

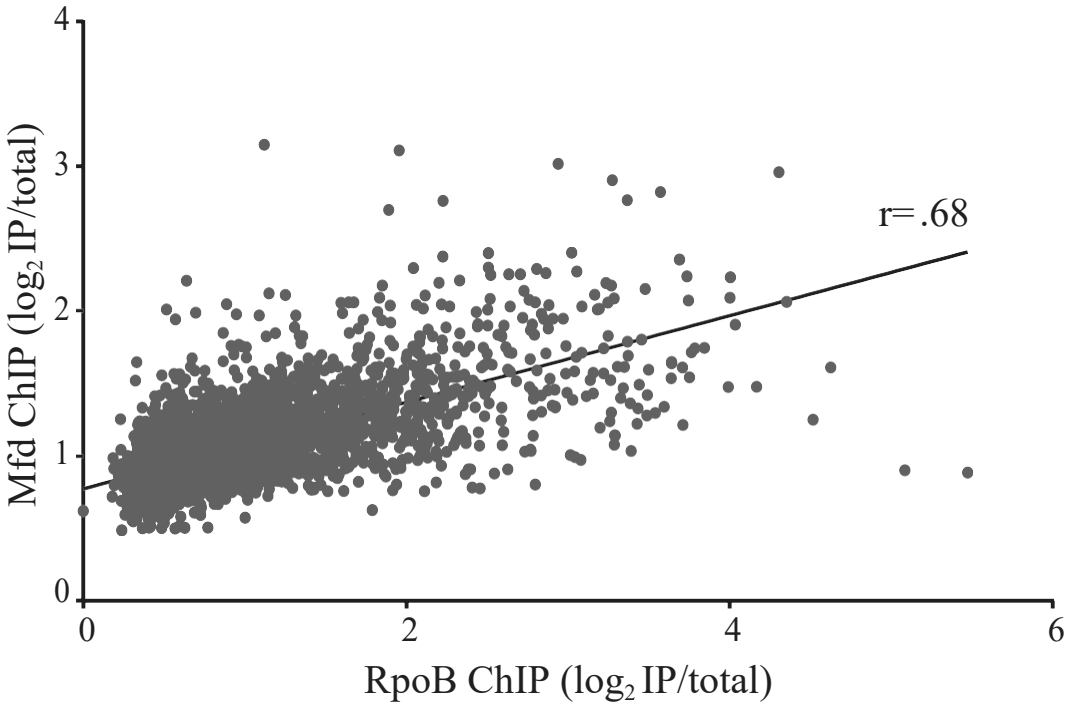


Fig. S2. *B. subtilis* Mfd and RpoB ChIP-seq are correlated

Linear regression analysis comparing binding of Mfd and RpoB at each gene in *B. subtilis*. Mfd-myc ChIP-seq (from Mfd-myc tagged *B. subtilis*) and RpoB ChIP-seq (from WT *B. subtilis*) read counts were determined for each gene in *B. subtilis* and normalized as described in Figure 2. Pearson's correlation coefficient for *B. subtilis* Mfd and RpoB = 0.68.

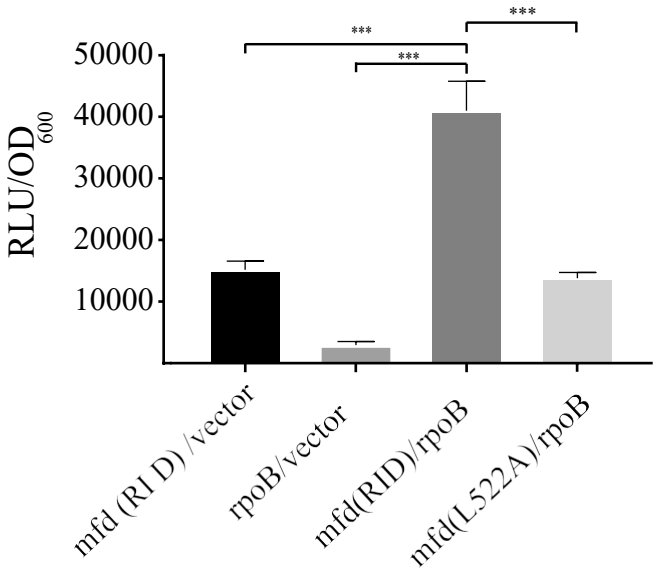
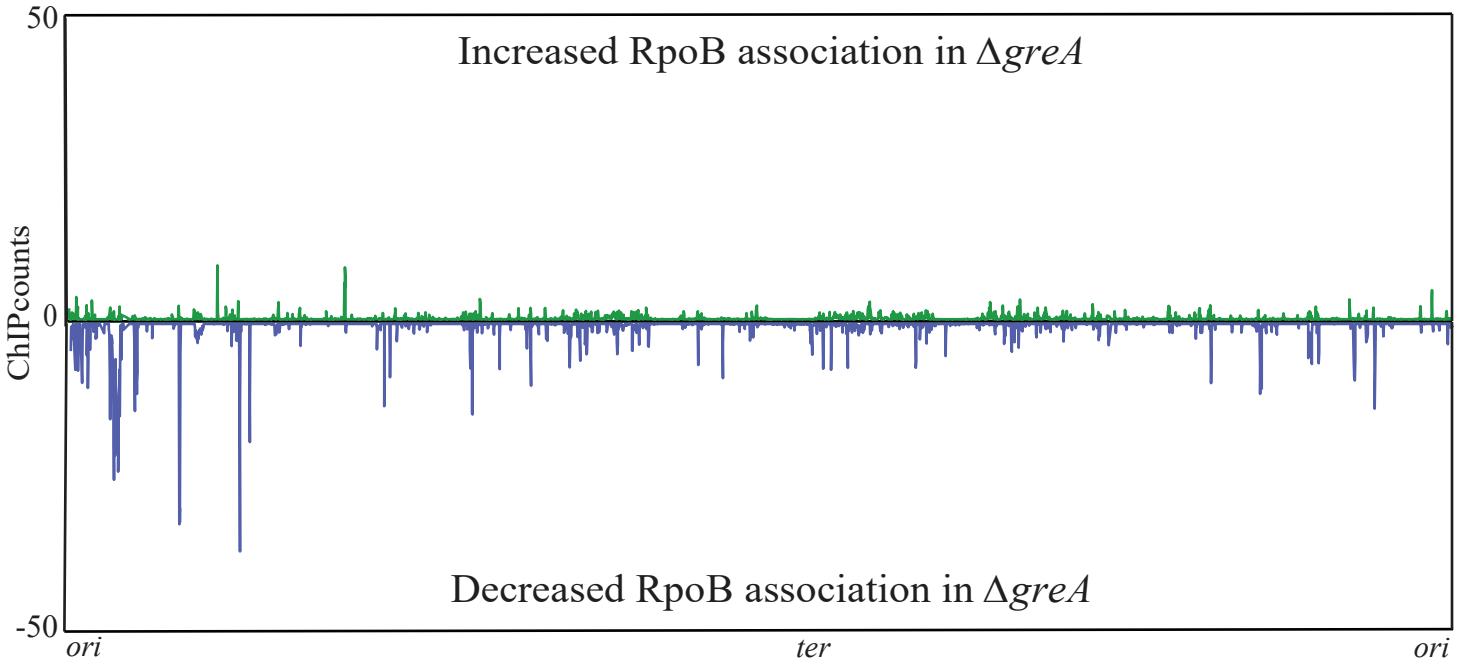


Fig. S3. Bacterial two-hybrid assay exhibits abrogated binding between *B. subtilis* MfdL522A and RpoB.

Disruption of Mfd L522 in B. subtilis abrogates interaction with RpoB. The interacting domains of RpoB and Mfd were cloned into a luciferase based bacterial 2-hybrid assay. Interactions between RpoB and Mfd and an MfdL522A mutant were measured, along with appropriate empty vector controls. Interactions were measured using luminescence and normalized to OD600. Data is from at least two independent experiments and error bars indicate standard deviation. Two-tailed students T-test was used to determine statistical significance (***p-value <0.001).

A



B

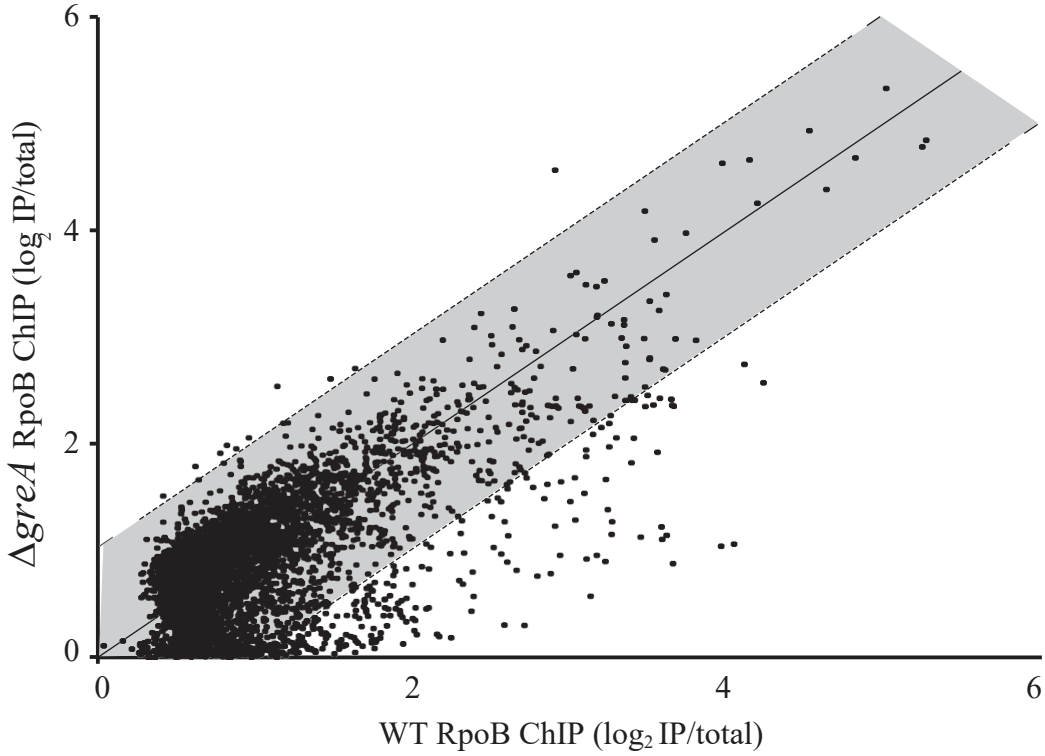


Fig. S4. RNAP termination at structured regulatory RNAs is specific to Mfd

RpoB ChIP-seq plots showing regions of RpoB enrichment in *ΔgreA* relative to WT. (A) Top half of graph (read counts in green) reflects normalized RpoB ChIP-seq counts where *B. subtilis ΔgreA* increased signal relative to WT. Bottom half of graph (read counts in blue) reflects RpoB ChIP-seq read counts where *ΔgreA* had decreased signal relative to WT. (B) Scatter plot of WT and *ΔgreA* RpoB ChIP-seq. Quantification of ChIP signal was performed as described in Figure 2B.S

TUs with Mfd binding and increased RpoB in *Δmfd*	regulatory RNA category
manP-manA-S439-yjdF	riboswitch, intergenic
S442-yjdH-S441-yjdG	5' UTR, intergenic
S81-ybeF-ybfA-ybfB	5' UTR
txpA (as ratA)	asRNA, ncRNA
bsrH (as bsrH)	asRNA, ncRNA
S782-S783-yopT	5' UTR
S345-S346-yhaX	independent transcript
S823-ilvD	5' UTR
yabE (S25 asRNA)	asRNA, ncRNA
yrrT-mtnN-S1033-mccA-mccB-yrhC	Intergenic
S1434-maeE	5' UTR
S1552-S1553-walR-walK-walH-walI-walJ-htrC	5' UTR
S655-S654	independent transcript
yhfO-yhfQ-S364-yhfP	intergenic
S460-mhqA	5' UTR
bsrG (SR4 asRNA)	asRNA, ncRNA
S438-yjdB-S437	5' UTR, 3' UTR
S27-rnmV-ksgA	5' UTR
S1123-nifZ-thiI-sspA	5' UTR
S1487-S1486-cydA-cydB-cydC-cydD	5' UTR
S492-clpE	5' UTR
manR	none
ndoAI-ndoA asRNA(S163-S164-S165)	asRNA
S811	independent transcript
trnY locus	tRNA, intergenic
alaR-alaT-S1201	3' UTR
S1175-S1174-mntA-mntB-mntC-mntD	5' UTR
S1427-S1426-atpI-atpB-atpE-atpF-atpH-atpA-atpG-atpD-atpC	5' UTR
S895-yqxK	5' UTR
S1513-bglP-bglH-yxiE	5' UTR, riboswitch
S321-glpF-glpK	5' UTR, riboswitch
mhqN-mhqO-mhqP	none
S1203-yugF	5' UTR
spoVM	none
S966-sdA (asRNA s965)	5' UTR, asRNA
polY2-yqjX-S898-S897-yqjY-yqjZ-yqkA-yqkB-yqkC	intergenic
S82-S83-glpT-glpQ	5' UTR, riboswitch

Table S1. TUs with Mfd binding and increased RpoB occupancy in *Δmfd*. Associated sRNA categories from previously defined work (3).

TUs with decreased RpoB in *Δmfd*	regulatory RNA category
srfAA-srfAB-comS-srfAC-srfAD	none
rsbR-rsbS-rsbT-rsbU-rsbV-rsbW-sigB-rsbX	none
S1343-csbA-S1342	5' UTR, 3' UTR
Hpf	none
ytxG-ytxH-ytxJ	none
ywjC-S1446	3' UTR
mtlA-mtlF-mtlD	none
S408-yjbC-S409-spx	5' UTR, intergenic
S928-mgsR	5' UTR
S426-yjcD (asS427-yjzE)	asRNA
Ctc	none
ywiE-ywjA-ywjB	none
S294-csbB	5' UTR
ypiA-ypiB	none
ahpF-ahpC	none
ybeC	none
sunA-sunT-bdbA-yolJ-bdbB	none
Ybyb	none
ptsG-ptsH-ptsI	none
pdaC	none
gtaB-S1363	3' UTR
S476-ykoM	5' UTR
serA	none
rsbRD (as927)	asRNA
S1366	independent transcript
yjbB	none
Icd	none
S1171-ytkA-S1172-dps	5' UTR, intergenic
yjcH-yjcG-yjcF	none
S1301	independent transcript
ykzB-ykoL	none

Table S2. TUs with decreased RpoB association in *Δmfd*. Associated sRNA categories from previously defined work (3).

B. subtilis TA gene	Mfd ChIP-seq association	*Δmfd* RpoB ChIP-seq fold enrichment
txpA (toxin)	**18.1091**	**3.6902**
RatA (antitoxin)	**13.3760**	**3.4526**
bsrG (toxin)	**3.1852**	**2.7718**
SR4 (antitoxin)	**4.4111**	**2.8398**
bsrE (toxin)	1.9520	1.4331
SR5 (antitoxin)	2.1053	1.2556
yonT (toxin)	1.8159	**3.6664**
as-yonT (antitoxin)	1.7007	**3.7755**
bsrH (toxin)	**10.4646**	**3.5415**
as-bsrH (antitoxin)	**11.4629**	**3.2391**

Table S3. Mfd association and RpoB occupancy of *Δmfd* strains at toxin- antitoxin genes. Genes with bolded values fulfill criteria for significant differences in Mfd occupancy (defined as genes with an Mfd ChIP association one standard deviation greater than the mean) and/or significant increase in RpoB occupancy in *Δmfd* (criteria defined in detail in dataset S2).

Strain	Genotype and Features	Reference
HM1	WT *B. subtilis* JH642	Brehm et al. J Bacteriol.1973
HM712	*B. subtilis* 168 Δ*greA*::mls	Koo et al Cell Syst. 2017 (Bacillus Genetic Stock Center) (4)
HM2295	*E. coli* F' (Kan) placOL2–62-lacZ	Dove et al Nature. 1997
HM2521	*B. subtilis* JH642 Δ*mfd*::mls	Ragheb et al Mol Cell. 2019
HM2602	*E. coli* F' (Kan) placOL2–62-lacZ pSIM27(tet)	Ragheb et al Mol Cell. 2019
HM2747	*E. coli* F' (Kan) placOL2–62-Nanoluc(hyg)	Ragheb et al Mol Cell. 2019
HM2769	*E. coli* F' (Kan) placOL2–62-Nanoluc(hyg) pSIM27(tet) pHM430(cm)pHM439(amp)	This study
HM2771	*E. coli* F' (Kan) placOL2–62-Nanoluc(hyg) pSIM27(tet) pHM430(cm) pBRα (amp)	This study
HM2773	*E. coli* F' (Kan) placOL2–62-Nanoluc(hyg) pSIM27(tet) pACλCI (cm) pHM439(amp)	This study
HM2916	*E. coli* AG1111 pminiMAD2-mfdL522A	This study
HM2932	*B. subtilis* JH642 MfdL522A	This study
HM2965	*E. coli* F' (Kan) placOL2–62-Nanoluc(hyg) pSIM27(tet) pHM431(cm) pHM439(amp)	This study
HM3157	*B. subtilis* JH642 Mfd-1xmyc	This study
HM3808	*B. subtilis* JH642 Δ*greA*::mls	This study
HM3947	*B. subtilis* JH642 MfdL522A-1xmyc	This study

HM3986	*E. coli* DH5α pHM676	This study
HM3988	*B. subtilis* JH642 amyE::P_{spank}-txpA	This study
HM3990	*B. subtilis* JH642 amyE::P_{spank}-txpA *Δmfd::mls*	This study
HM4002	*E. coli* DH5α pHM682	This study
HM4003	*B. subtilis* JH642 amyE::P_{spank}-bsrH	This study
HM4004	*B. subtilis* JH642 amyE::P_{spank}-bsrH *Δmfd::mls*	This study
Plasmids	**Description**	**Reference**
pBRα	Used as a negative control in bacterial 2-hybrid assays	Dove et al Nature. 1997 (Addgene 53731)
pBRα-β-flap	Used to clone and express RNA polymerase α-subunit fusions in *E. coli*	Dove et al Nature. 1997 (Addgene 53734)
pACλCI	Used as a negative control in bacterial 2-hybrid assays	Dove et al Nature. 1997 (Addgene 53730)
pACλCI-β-flap	Used to clone and express λCI fusions in *E. coli*	Dove et al Nature. 1997 (Addgene 53733)
pHM430	Plac-CI-Bsubmfd(494-625)	This study
pHM431	Plac-CI-Bsubmfd(494-625)L522A	This study
pHM439	Plac-a-BsubrpoB(21-131)	This study
pHM676	amp^R, amyE::$P_{spank(hy)}$-txpA, lacI, $spec^R$	This study
pHM682	amp^R, amyE::$P_{spank(hy)}$-bsrH, lacI, $spec^R$	This study
pHM707	pminiMAD2-BsubMfdL522A	This study
pDR111	amp^R, amyE::$P_{spank(hy)}$, lacI, $spec^R$	Guérout-Fleury et al Gene. 1996 (5)
pminiMAD2	Scarless integration plasmid for *B. subtilis*	Patrick and Kearns Mol Micro. 2008 (1)
pNL1.1	NanoLuc expression vector	Promega (GenBank Accession #JQ513379)

Table S4. Bacterial strains and plasmids used in this study.

Primer #	Sequence	Description
HM1004	CATGAGGGTACCGATGATCAGCGGT CAATTGA	For amplifying *B. subtilis mfd* to insert into pminiMAD2
HM1005	CATGAG GGATCCCATAGTGCTGCT GTGCCAA	For amplifying *B. subtilis mfd* to insert into pminiMAD2
HM3286	AGTGGCCTGAAGAGACGTTTGGCGCA AAAAGCTATTCTGAGCTTCAAATTG	For amplifying *B. subtilis mfd* (bp 1483-1875) with homology to pACCI for Gibson and extra base to maintain frame.
HM3287	CTGCGATGCAGATCTGTAAGGTAAGTT AAGTCTCTTGATAAGGGAAAGCC	For amplifying *B. subtilis mfd* (bp 1483-1875) with homology to pACCI for Gibson with stop codon added
HM3292	AAGTGAAAGAAGAGAAACCAGAGGCA GAAGTGTTAGAATTACCAAATCTCATT G	For amplifying *B. subtilis rpoB* (bp 64-393) with homology to pBRa for Gibson and extra base to maintain frame.
HM3293	CGGCCACGATGCGTCCGGCGTAGAGT TATTCCGCACCGTTAATGATAAAAG	For amplifying *B. subtilis rpoB* (bp 64-393) R with homology to pBRa for Gibson and stop codon added.
HM3540	GAATGCCGTTGATTTCAGCAGTTTCAA TCCCCAGGTATTTTCCG	Quickchange primer to make L522A *mfd* mutation
HM3541	CGGAAAATACCTGGGGATTGAAACTG CTGAAATCAACGGCATTC	Reverse quickchange primer to make L522A *mfd* mutation
HM3759	CAAGTCCTCTTCACTGATTAACTTCTG CTCCGTTGATGAAATGGTTTGCT	For amplifying C-terminally myc tagged *mfd* by SOE PCR
HM3760	GAGCAGAAGTTAATCAGTGAAGAGGA CTTGTAAATTTTGTTACTCTCTGGTGTA TATTAC	For amplifying C-terminally myc tagged *mfd* by SOE PCR
HM3854	CGAGGCTCCTGTCACTGCT	For amplifying erm-HI cassette
HM3969	GAGCAGAAGTTAATCAGTGAAGAGGA CTTGATTTTGTTACGCAGGCGAGAAAG GAGAGAG	For amplifying erm-HI cassette with myc tag at 5' end

HM5162	ACACTCCTCATGTTTGCCTT	For *B. subtilis tnrY* qPCR
HM5163	GTGTCGGCGGTTCGATT	For *B. subtilis tnrY* qPCR
HM5418	CATGATGCTAGCTGAAAGGAGGTGAA ATTATGTCGAC	For making *txpA* overexpression construct cloning
HM5419	CATGATGCATGCCTACCCTTTAATAGG AGGGT	For making *txpA* overexpression construct cloning
HM5437	CAAGCAAAAGTATTGCAACT	For *B. subtilis ratA* qPCR
HM5438	GGTAATGTGGTAATGTGGTA	For *B. subtilis ratA* qPCR
HM5441	ATGTCGACCT ATGAATCTCT	For *B. subtilis txpA* qPCR
HM5442	CCCATGTCATAATCCCGCCT	For *B. subtilis txpA* qPCR
HM5462	CATGATGCTAGCATGGTTTAGTATAAA TGAAT	For making *bsrH* overexpression construct cloning
HM5463	CATGATGCATGCAAGAGACCCGGTTG CCGCCGGG	For making *bsrH* overexpression construct cloning

Table S5. Oligonucleotides used in this study.

Dataset S1 (separate file) Quantification of Mfd association of genes in the *B. subtilis* 168 genome. Mfd-myc binding was calculated by taking the average read count across a given gene and normalizing internally to overall read counts as well as to WT *B. subtilis* (lacking a myc tag). Values were subsequently $\log_2$ normalized. Genes are sorted from highest to lowest Mfd binding values. Those genes with greater than one standard deviation from the mean Mfd-myc binding value were defined as Mfd associated.

Dataset S2 (separate file) Genes with increased RpoB ChIP association in Δ*mfd*, sorted by increasing p-value. (logFC= log-fold change, logCPM= log counts per million, FDR= false discovery rate). The following criteria were used to define increased RpoB association= logFC>1, logCPM>4, p-value< $1x10^{-4}$, FDR< .001.

Dataset S3 (separate file) Genes with decreased RpoB ChIP association in Δ*mfd*, sorted by increasing p-value. Criteria used to define decreased RpoB association is the same as described in Table S1.

Dataset S4 (separate file) Genes with altered RpoB ChIP association in Δ*greA*. Genes are sorted by increasing p-value, with the first 12 genes exhibiting increased RpoB occupancy in Δ*greA* and the remaining genes exhibiting decreased RpoB occupancy in Δ*greA*. To define significant differences, the same criteria were used as described in Tables S1 and S2.

Dataset S5 (separate file) Upregulated and Downregulated genes in *B. subtilis Δmfd* strain. Genes are sorted by increasing p-value. The following criteria was used to define transcriptional differences = logFC> 1, logCPM> 2, FDR< .05

Supplementary References

1. J. E. Patrick, D. B. Kearns, MinJ (YvjD) is a topological determinant of cell division in Bacillus subtilis. *Mol. Microbiol.* **70**, 1166–79 (2008).

2. Dove, S. L., Joung, J. K. & Hochschild, A. Activation of prokaryotic transcription through arbitrary protein- protein contacts. *Nature* **386,** 627–630 (1997).

3. B. Zhu, J. Stülke, SubtiWiki in 2018: from genes and proteins to functional network annotation of the model organism Bacillus subtilis. *Nucleic Acids Res.* **46**, 743–748 (2017).

4. B. M. Koo, *et al.*, Construction and Analysis of Two Genome-Scale Deletion Libraries for Bacillus subtilis. *Cell Syst.* **4**, 291-305.e7 (2017).

5. A. M. Guérout-Fleury, N. Frandsen, P. Stragier, Plasmids for ectopic integration in Bacillus subtilis. *Gene* **180**, 57–61 (1996).