A high-resolution view of RNA endonuclease cleavage in *Bacillus subtilis*

James C. Taggart[1], Jean-Benoit Lalanne[2], Sylvain Durand[3], Frédérique Braun[3], Ciarán Condon[3] and Gene-Wei Li[4,#]

[1]Department of Systems Biology, Harvard Medical School, Boston, MA, USA
[2]Department of Genome Sciences, University of Washington, Seattle, WA, USA.
[3]Expression Génétique Microbienne (EGM), CNRS, Université Paris Cité 10 Institut de Biologie Physico-Chimique, 13 rue Pierre et Marie Curie, 75005 Paris, France
[4]Department of Biology, Massachusetts Institute of Technology, Cambridge, MA 02142, USA

[#]corresponding author (gwli@mit.edu)

ABSTRACT

RNA endonucleases are the rate-limiting initiator of decay for many bacterial mRNAs. Subsequent decay by exonucleases typically obscures these cleavage events, and even for the well-studied *Bacillus subtilis*, few precise positions of cleavage have been mapped. Here we present an approach for precisely mapping positions of endoribonucleolytic activity transcriptome-wide. Through the detection of RNA ends in exonuclease-deficient cells, we map $>10^3$ putative endonuclease cleavage sites, together with their dependence on RNase Y. Bioinformatic analysis reveals the targeting specificity of this enzyme, the mRNA interferase EndoA, as well as an unknown ribonuclease in *B. subtilis* that trims the 5′ end of RNAs with both 5′-P and 5′-OH. Our results provide a high-resolution view of mRNA decay in *B. subtilis* and a generalizable approach for elucidating endoribonuclease specificity *in vivo*.

INTRODUCTION

Predicting protein abundance from an arbitrary gene sequence remains an aspirational goal for quantitative biology. Decades of research has explored the encoding of rates of transcription and translation initiation (Taggart et al., 2021), resulting in predictive models for how RNA polymerase and the ribosome recognize promoters and ribosome binding sites. (Brewster et al., 2012; Salis, 2011; Salis et al., 2009; Urtecho et al., 2019). Comparatively poorly characterized, however, are RNA endonucleases, factors which exert widespread quantitative control on bacterial gene expression levels (Chen et al., 2015; Khemici et al., 2015; Moffitt et al., 2016). Understanding the encoding of endoribonuclease sites in bacterial mRNA sequences is therefore a critical question in quantitative biology.

In the Gram-positive bacterium *Bacillus subtilis*, cleavage by an endoribonuclease, primarily RNase Y, is thought to initiate the decay of a large fraction of cellular mRNAs (Lehnik-Habrink et al., 2011, 2012; Shahbabian et al., 2009). This cleavage is followed by exoribonucleolytic degradation of the two resulting products, mediated by RNase J1 in the 5-3′ direction and by exoribonucleases PNPase, RNase R, RNase PH, and YhaM in the 3-5′ direction (Trinquier et al., 2020). RNase Y appears to drive the turnover of the majority of mRNAs in *B. subtilis*, and its depletion results in the accumulation of hundreds of transcripts (Lehnik-Habrink et al., 2011; Shahbabian et al., 2009). Unfortunately, though RNase Y has been identified as a key player shaping the transcriptomes of Gram-positive bacteria, our understanding of which mRNA sequences *B. subtilis* RNase Y cleaves remains relatively limited. Studies have suggested *B. subtilis* RNase Y recognizes single stranded RNA, is stimulated by downstream secondary structure, and has a preference for adenosine and uridine-rich sequence (Braun et al., 2017; Shahbabian et al., 2009). Orthologs of RNase Y in *Staphylococcus aureus* and *Streptococcus*

pyogenes, organisms for which RNase Y is not essential and appears to play a more limited role, additionally demonstrate a preference for cleavage downstream of a guanosine. Whether this can be generalized to the quasi-essential (Benda et al., 2021) *B. subtilis* RNase Y remains to be determined. In addition to RNase Y, RNase III has been suggested to cleave in dozens of mRNAs, potentially initiating decay for a subset of the transcriptome (DiChiara et al., 2016). As new endoribonucleases are discovered, additional enzymes may also be implicated in decay-initiating events across the transcriptome (Cerullo et al., 2022; Ingle et al., 2022; Leroy et al., 2017).

A number of studies have mapped the target repertoire of major bacterial endoribonucleases, including *E. coli* RNase E, *B. subtilis* RNase III, and the RNase Y orthologs in *S. aureus* and *S. pyogenes* (Broglia et al., 2020; Chao et al., 2017; DiChiara et al., 2016; Khemici et al., 2015). These measurements typically rely on the mapping of RNA ends by high-throughput sequencing in strains defective in endoribonuclease activity and analyzing where signal is depleted. Although it can be effective at identifying endoribonuclease targets, such an approach requires caution as the subsequent activity of exoribonucleases on endoribonuclease-derived RNA ends can shift signal away from the point of cleavage. Such behavior can result in both false positives, with redundant calling of multiple positions derived from a single cleavage event, as well as false negatives if exoribonuclease activity is sufficiently efficient to suppress signal from accumulating at the cleavage site above the detection threshold. Low-throughput end mapping techniques are often therefore required to precisely determine exact positions of cleavage (Braun et al., 2017; Bruscella et al., 2011; DiChiara et al., 2016; Yao and Bechhofer, 2010).

In this work, we present a global approach to map bacterial endoribonuclease cleavage sites combining genetic perturbation of all known exonucleases participating in *B. subtilis* mRNA decay with high-resolution RNA end-mapping. By capturing both 3′ and 5′ ends in exonuclease deficient cells, we are able to determine exact positions of endoribonuclease activity transcriptome-wide. Through this, we have generated the first global map of endoribonuclease cleavage in *B. subtilis*, revealing thousands of putative positions of cleavage. By combining this approach with a panel of endoribonuclease knockouts, we have dramatically expanded the known target repertoire for RNase Y. Through this, we have directly measured the sequence and local secondary structural contexts preferred by RNase Y. We additionally present evidence for a more stringent *in vivo* specificity of the toxin EndoA and a previously unknown nuclease activity targeting RNA 5′ ends.

RESULTS

RNA end mapping in exonuclease deficient *B. subtilis* strains reveals positions of endoribonucleolytic activity

In the absence of additional nucleolytic activity, one can map the positions of cleavage by an endoribonuclease by simply observing the termini of its cleavage products. *In vivo*, exonucleases typically rapidly act on endonuclease cleavage products, rendering the original RNA endonuclease cleavage sites undetectable. If all exoribonucleolytic activity targeting either 3′ or 5′ ends is ablated, however, endoribonuclease cleavage at a particular site should result in a stably accumulating products with new ends at the position of cleavage (Figure 1A). Strains combining knockouts of all known exonucleases in each direction are viable in *B. subtilis*, meaning such a measurement is theoretically possible in this species (Figaro et al., 2013;

Oussenko et al., 2005). If RNA ends are mapped in both a knockout of RNase J1, the only known 5-3′ exonuclease, and a knockout of all four known 3-5′ exonucleases (a knockout of *yhaM, rph, pnpA,* and *rnr*, hereafter referred to as a 4-exo knockout), exact positions of endoribonucleolytic activity can be captured as pairs of adjacent 3′ and 5′ ends that appear across the two strains.

Based on this principle, we implemented an approach to create a transcriptome-wide map of endoribonuclease activity by capturing the 3′ hydroxyl and 5′ monophosphate groups that are generated by endonucleases such as RNases Y and III (the primary endonucleases targeting mRNAs in *B. subtilis)* (Figure 1A; (Bechhofer and Deutscher, 2019)). To quantitatively capture 3′ hydroxyl groups, we utilized a previously established method for RNA 3′ ligation and deep sequencing, hereafter referred to as 3′ end sequencing (Figure 1A, left; (Dar et al., 2016; Herzel et al., 2022; Mandell et al., 2021)). To complement this, we specifically captured 5′ monophosphates through T4 RNA Ligase 1 treatment, allowing quantification of cleavage-derived 5′ ends by deep sequencing (Figure 1A, right). By performing 3′ end sequencing on a 4-exo knockout (strain CCB396) and 5′ end sequencing on an RNase J1 knockout (strain CCB434), we captured 3′/5′ end pairs derived from endoribonuclease activity transcriptome-wide. Combining these exoribonuclease perturbations with endoribonuclease knockouts, we can attribute these cleavage signatures to specific enzymes. End-enriched RNA sequencing (Rend-seq) can additionally be performed to quantify expression of putative endoribonuclease targets and provide an orthogonal RNA end-mapping approach (Lalanne et al., 2018).

This paired end-mapping strategy robustly captures positions of previously documented endoribonucleolytic cleavage sites. Among 13 previously known sites cleaved by either RNase Y, RNase III, or Rae1 that were individually detected (Braun et al., 2017; Bruscella et al., 2011;

DiChiara et al., 2016; Durand et al., 2015; Leroy et al., 2017; Noone et al., 2014; Yao and Bechhofer, 2010), we identified at least one clear 3′/5′ pair at or near the annotated cleavage position (Figure 1B, Figure S1). End sequencing of knockouts of RNases Y or III coupled to RNase J1 perturbation (strains CCB760 and BG879) confirmed that we can detect cleavage by both enzymes, with accumulation of RNAs with these ends depended on the activity of the particular enzymes that were previously determined to drive cleavage (Figure 1B).

Importantly, the paired end-mapping strategy provides clear advantages over measurements of a single end type. Because we capture both RNA ends generated by endoribonuclease activity, adjacent positions of cleavage can be resolved, as appears to be the case for the first characterized substrate of Rae1, *S1025* (Figure 1B, bottom). Without these two paired channels of information, it is difficult to determine whether signal at adjacent positions may be derived from multiple positions of endoribonucleolytic cleavage or residual exoribonuclease activity by unknown enzymes, the latter of which may be occurring after endoribonucleolytic cleavage of the RoxS RNA (Figure 1B, middle). Precisely resolving cleavage positions becomes particularly relevant in cases such as the endoribonucleolytic cleavages observed in *hbs*, where many nearby positions are known to be subject to RNase Y cleavage (Figure S1, (Braun et al., 2017)).

Using these data, we sought to map putative positions of cleavage within transcripts whose abundance is known to be regulated by RNA endonuclease activity but for which precise positions of cleavage have yet to be mapped. Recent work has shown that the RNA decay machinery in *Bacillus subtilis* is subject to extensive autoregulation and cross-regulation, suggesting that cleavage sites should be detectable within these RNAs (DeLoughery et al., 2018; DiChiara et al., 2016). This is borne out in our end sequencing, where multiple putative positions

of endoribonucleolytic cleavage are detected within the transcripts encoding RNases Y, III, and the Y-complex (Figure 2, Figure S2). A number of additional transcripts, including that encoding sporulation master regulator SinR, have been experimentally characterized as stabilized upon deletion of RNase Y (Deloughery et al., 2016; Lehnik-Habrink et al., 2011). Paired peaks are clearly visible within each of these transcripts (Fig 2, Fig S2). Taken together, our ability to robustly identify known and previously unknown positions of cleavage suggests that these data provide a high-resolution, genome-wide view of RNA endonuclease activity in *B. subtilis*.

Systematic identification of positions of endoribonuclease cleavage

In order to systematically identify putative signatures of endonuclease cleavage transcriptome-wide, we implemented an approach to call abutting pairs peaks in 3′ and 5′ end sequencing across our 4-exo and RNase J1 knockout data (Figure S3, top). To do so, we assessed the signal at each position in both end sequencing datasets relative to the average in a local window, calling positions exceeding a signal-to-background threshold of 7.5 as peaks. Positions for which a peak is called in both 5′ and 3′ end sequencing with a 1 nucleotide offset are called as peak pairs and represent putative sites of endoribonucleolytic activity. This approach reveals 1427 putative positions of cleavage in our data. It should be noted, however, that due to constraints of our experimental approach, we are unable to identify cleavage events proximal to RNA ends, as short RNAs (<35 nts) are excluded during 5′ end sequencing library preparation (see Methods).

To assign putative endoribonuclease signatures to the activity of RNase Y, we sought to identify all putative cleavage sites that are absent in the context of each endoribonuclease knockout. To do so, we implemented a workflow to call endoribonuclease sites for which RNA end signals are diminished in a particular knockout relative to an unperturbed strain (Fig S3,

bottom). In this approach, end sequencing signals in an exoribonuclease deficient strain with and without a particular endoribonuclease knockout are normalized to local gene expression measured by Rend-seq. The ratio of these expression-normalized 5′ end sequencing signals between the wild-type and endoribonuclease knockout strains (all in an exoribonuclease deficient background) will be elevated for sites dependent on that endoribonuclease, as these sites will have low end sequencing signal in the knockout. This metric is hereafter referred to as the ratio of ratios or "RR" or a particular site. A similar strategy was previously used to map RNase III sites in *B. subtilis* (DiChiara et al., 2016), and details of our approach can be found in the Materials and Methods section.

Determinants of cleavage by RNase Y

To define the substrate features for RNase Y cleavage, we sought to identify a set of putative cleavage sites that are dependent on the activity of this enzyme. For the 1427 endonuclease cleavage sites, we calculated their RR values using RNase J1 deficient cells with or without an RNase Y knockout. This metric was bimodally distributed, and any cleavage signature with an RR value over 12.15 was called as RNase Y dependent based on a double-Gaussian fit of these data (Figure 3A). Using this approach, we mapped the location of 493 putative RNase Y cleavage sites across 322 genes. We were unable to detect any clear bias in localization towards RNA 3′ or 5′ ends, though this may be influenced by our inability to detect short cleavage products as previously discussed.

From this set of putative RNase Y sites, we can begin to understand the substrate features which guide RNase Y activity. Frequency and information content logos constructed from the RNA sequences surrounding putative RNase Y sites suggest a sequence preference for cleavage

downstream of a purine, with a particular preference for guanosine, and upstream of an adenosine (Figure 3B, top, middle). Similarly, considering k-mers of length ≤4, there is most significant enrichment for a GA at the -1 position, with additional enrichment for UGA and UUGA upstream, suggesting a preference for uridine in the -2 and -3 positions (Figure 3B, bottom). Given that our cleavage site calling depends on concordance between datasets generated through two distinct library preparation strategies involving distinct enzymatic reactions and adapter sequences, this observed specificity is unlikely to be a result of biases in our methodology. This is additionally consistent with the sequence context in several previously mapped cleavage sites (Braun et al., 2017; Shahbabian et al., 2009), as well as the measured sequence preferences of RNase Y in *S. aureus* and *S. pyogenes*, despite the more limited role this enzyme plays in those organisms (Broglia et al., 2020; Khemici et al., 2015). In a broader window around these sites of RNase Y cleavage, there is a clear depletion for GC-containing sequence, with the exception of the -1 position where guanosine is favored (Figure 3C).

Given this reduced GC content near positions of RNase Y cleavage, we suspected that RNase Y cleavage sites should have a lower propensity to form secondary structures. Computationally predicted secondary structures of 40 nucleotide sequences surrounding positions of cleavage are consistent with this idea, with weaker structures forming for 40-mers centered on RNase Y sites (Figure 3D). Randomly sampled positions corresponding to uncleaved GA dinucleotides do not recapitulate this structural signature. Interestingly, sequences centered just downstream of this lowly-structured region appear to fold more strongly than background, suggesting an enrichment for secondary structure proximal to the position of cleavage. A more modest enrichment for secondary structure may also be present upstream. The downstream signal is recapitulated in the GC content within this window (Figure 3C), and is consistent with

the influence of secondary structure on previously detected RNase Y cleavage events (Marincola and Wolz, 2017; Shahbabian et al., 2009). Taken together, these data indicate that *B. subtilis* RNase Y targets weakly-folding AU-rich sequences with a preference for GA dinucleotides and suggest a potential role for proximal secondary structure.

Induction of EndoA activity in exoribonuclease knockouts reveals unexpected *in vivo* specificity

To our surprise, manual inspection of our Rend-seq data in all strains with an RNase J1 or 4-exo knockout showed prominent peaks across the transcriptome within the motif UACAU (Figure 4A, B, 5A). This sequence is the canonical cleavage motif of the endonuclease EndoA, an RNA interferase of the MazF family associated with the EndoA-EndoAI type II toxin-antitoxin system (Park et al., 2011). This suggests activation of this toxin in the exoribonuclease deficient strains used in this study, though the mechanism of EndoA activation in these strains remains to be determined. Contrary to EndoA activity targeting a 5-mer, however, RNA 3′ ends were dramatically enriched in Rend-seq for UACAU motifs followed by an adenosine over any other nucleotide at hundreds of instances of this motif (Figure 4A). Rend-seq in an RNase J1 knockout revealed a similar distribution of 5′ ends 1 nucleotide downstream of these 3′ ends, corroborating this specificity and confirming that we are observing endonuclease activity (Figure 4B). This observed signature is ablated upon deletion of the *ndoA* gene (Figure 4C, D). Together, this suggests that EndoA may have a more stringent specificity *in vivo* than previously known, recognizing the 6-mer UACAUA. This extended motif has previously only been associated with a mutated variant of this protein engineered to incorporate canavanine instead of arginine (Ishida et al., 2013).

Evidence for additional 5′ end trimming activity in *B. subtilis*

Because the 5′ end sequencing approach presented in this work specifically captures 5′ monophosphates, we were surprised to find peaks 3 or 5 nucleotides downstream of the cleavage position within instances of the EndoA U/ACAUA motif (Figure 5A), given that EndoA cleavage leaves 5′ hydroxyl groups (Bechhofer and Deutscher, 2019). Presence of these peaks is dependent on the activity of EndoA and is detectable in Rend-seq, which is insensitive to RNA 5′ moiety, as minor peaks 3 and 5 nucleotides downstream of many EndoA sites (Figure 4, 5A). This suggests the generation of 5′ monophosphates through the removal of two or four nucleotides following cleavage by EndoA. To confirm that this activity is general against all 5′ hydroxyl groups, not simply EndoA sites, we looked at the *glmS* ribozyme, whose autocatalysis leaves a 5′ OH that is sensitive to subsequent RNase J1 activity. In both a knockout of *rnjA* and a double knockout of *rnjA* and *ndoA*, 5′ end sequencing shows a clear 5′ monophosphate peak +3 relative to the site of self-cleavage, suggesting that this currently unknown enzyme can act independently of EndoA (Figure 5B). This activity is additionally not performed by RNases Y, III, J1, or J2 as +3 position peaks are readily detectable in an *rny* or *rnjB* strain with RNase J1 depletion and in a complete RNase J1 knockout with or without *rnc* (Figure 5C, D).

Inspection of EndoA-derived cleavage events further reveals substrate preference for the detected nuclease activity. Clear patterns emerge in the densities of the 5′ end distribution downstream of the cleavage site depending on the identity of the two nucleotides following the UACAUA sequence (Figure 5E). An *rny* knockout with RNase J1 depletion was used in these analyses to amplify the detected EndoA signal. At EndoA sites that are followed by a guanosine, few 5′ monophosphates are seen in positions downstream of the +3 position, suggesting that 5′-AUAG blocks subsequent nucleotide removal. On the other hand, EndoA motifs followed by AA

result in a distributed 5′ end signal. In a 4-exo knockout, the 3′ end density remains centered on the -1 position relative to cleavage, suggesting that the shift in 5′ end density downstream is a product of successive rounds of removal, rather than a shift in EndoA cleavage position (Figure 5F). Because cleavage by this enzyme appears to leave a 5′ monophosphate, this implies an ability to cleave RNAs with both 5′ monophosphate and 5′ hydroxyl group (similar to RNase J1). Further work is required to reveal the identity and mode of action of this enzyme.

DISCUSSION

In this work we present the first genome-scale map of *B. subtilis* endoribonuclease cleavages at single-nucleotide resolution. Through RNA end mapping in exonuclease deficient strains, we identified over a thousand putative cleavage events, allowing us to learn substrate preferences of the central endonucleases involved in *B. subtilis* mRNA decay. This work revealed sequence and structural preferences of RNase Y and showed that the toxin EndoA recognizes a hexameric rather than pentameric motif *in vivo*. In these analyses, we additionally uncovered evidence for and potentially specificity of a new ribonucleolytic activity previously unknown in *B. subtilis*, motivating further work to understand RNA 5′ end processing and decay.

The transcriptome-wide atlas of endonuclease activity provided by this study represents a valuable resource in understanding the decay pathway of any given RNA species in the cell. By establishing a set of putative cleavage positions within any particular transcript, this information can help guide sequence design to avoid efficient endoribonuclease substrates. As high-throughput genome engineering technologies mature, this may allow for more controlled alterations to transcriptome and proteome composition, and may allow higher throughput probing of the importance of each putative endonuclease site for RNA stability.

It is worth noting that though the presented global map of endonuclease activity expands our list of known RNase Y cleavage positions by orders of magnitude, we do not expect it to be comprehensive. First, due to size-selection steps in the end-sequencing protocol, we are unable to detect cleavage sites within 15 nucleotides of a 5′ end or 40 nucleotides of a 3′ end. Secondary endoribonuclease cleavage followed by activity of the remaining exoribonucleases may additionally destabilize cleavage products, rendering them undetectable. Evidence of such behavior can be seen in the *cggR* gene, where a 5′ end associated with cleavage in the middle of the gene only appears when cleavage downstream is ablated through *ylbF* deletion (Figure S4). Finally, activity by unknown exoribonucleases may cause broadening or shifting of 3′ or 5′ ends, separating the peak pairs and obscuring potential cleavage positions (as seen in Figure 5). The absence of signal in these data therefore is not necessarily an indication that turnover of a particular RNA is independent of endoribonucleolytic cleavage.

In the process of mapping endonuclease cleavage events, we detected a number of unexpected signals, including EndoA induction and a previously uncharacterized 5′ end trimming activity in *B. subtilis.* We have confirmed that this activity is not dependent on the primary nucleases targeting mRNAs, including RNases Y, III, J1, J2, or EndoA. Whether this nuclease acts in a 5-3′ exonucleolytic manner or acts through endonucleolytic cleavage near 5′ ends remains to be determined. Further work is additionally necessary to understand whether this enzyme acts to facilitate RNase J1 cleavage at substrates with a 5′ hydroxyl moiety, as this would shape our understanding of how systems like the *glmS* ribozyme regulate gene expression. Given this, determining what gene encodes this activity is of great interest moving forward.

The approach described in this work can be generalized to study RNA decay in any bacterium for which exonuclease activity can be ablated or transiently depleted, representing a

new approach to discover factors influencing RNA decay. Additionally, application of this approach to conditions beyond exponential growth in rich medium will provide an insight into regulation of endonuclease activity and RNA decay more broadly. The high-resolution picture of RNA decay initiation provided by this approach, particularly when integrated with future measurements of RNA half-life, will inform future quantitative predictions of RNA stability in Gram-positive bacteria.

MATERIALS AND METHODS

Strains and strain construction

For all Rend-seq, 5′ end sequencing, and 3′ end sequencing experiments excluding the Rend-seq measurement of an *rnjB* deletion with RNase J1 depletion, derivatives of W168 were used (Table S1). All transformations for strain construction were performed using standard protocols relying on natural *B. subtilis* competence and flanking homology mediated recombination using media and protocol described in (Parker et al., 2020). BJT070 was created through SPP1 phage transduction of the *ylbF::erm* allele from BJT034 into CCB396 (Yasbin and Young, 1974). To create BJT074, gDNA from BJT034 was transformed into SSB1002 to generate a *ylbF* knockout in a W168 background. From this, SPP1 phage transduction from CCB434 was used to construct an *rnjA ylbF* double knockout, BJT074. CCB390 was constructed by first transferring an inducible RNase J1 allele (first described in (Britton et al., 2007)) into W168 (strain SSB1002), followed by deletion of *rnjA* through transformation with gDNA from CCB434. To construct BJT231, genomic DNA from GLB230 was transformed into CCB390, supplementing with 2% xylose in the liquid and solid media throughout the transformation protocol.

Cell growth and harvesting

All experiments were performed in 2xYT, supplemented with 2% xylose or glucose when RNase J1 induction or depletion was necessary. Except in cultures for which we actively sought to deplete RNase J1, all solid and liquid media used to culture strains with an inducible RNase J1 allele was supplemented with 2% xylose. For all sequencing experiments, overnight cultures in 2xYT were back-diluted to an OD_{600} of 0.0002 in 20 mL prewarmed 2xYT and allowed to grow to an OD_{600} of 0.2, at which point they were harvested. For any strain with an RNase J1 knockout, cultures were vortexed prior to and following inoculation to minimize clumping. To achieve RNase J1 depletion, 10x the required volume of overnight culture was collected, washed twice with 1 mL pre-warmed 2xYT, and resuspended in 1 mL 2xYT. 100 µL of this resuspension was used to inoculate the final culture. Samples were harvested by adding 7 mL of culture to 7 mL of ice cold methanol, spinning, and flash freezing with liquid nitrogen. Cell pellets were stored at -80°C prior to RNA extraction.

End-enriched RNA sequencing and 3′ end sequencing

Prior to RNA extraction, cell pellets were washed with 10-15 mL ice cold 10 mM Tris pH 7 to remove any contaminating nucleic acids that precipitate from 2xYT in methanol. For Rend-seq, RNA was extracted from cell pellets using either an RNeasy Mini Kit (Qiagen) with on-column DNase treatment, an RNeasy Plus Mini Kit (Qiagen) or through RNAsnap extraction (Stead et al., 2012) followed by Turbo DNase treatment. All 3′ end sequencing samples were purified through RNAsnap. In the RNAsnap protocol, cells were resuspended in 500 µL Extraction solution (95% Formamide, 1% BME, 0.025% SDS, 18 mM EDTA) and transferred to a 2 mL tube containing 200 mL chilled 0.1 mm diameter zirconia beads. These tubes were

vortexed for 10 minutes using a horizontal vortex adapter and subsequently incubated at 95°C for 7 minutes. Samples were spun at 15,000 g, after which 400 µL was removed and diluted 1:5 with 1.6 mL DEPC-treated water. This mixture was precipitated with isopropanol and resuspended in 10 mM Tris, pH 7. At least 30 µg of RNA was then treated for 30 minutes at 37°C with 10 µL Turbo DNase in a 100 µL reaction. Following RNA extraction and gDNA depletion, 20 µg of purified RNA was treated with MICROBExpress Bacterial mRNA Enrichment Kit (ThermoFisher) per the manufacturer's instructions and precipitated with isopropanol.

Rend-seq libraries were prepared as described previously (Lalanne et al., 2018). In short, rRNA depleted RNA was fragmented for 25 seconds (ThermoFisher #AM8740) and fragments of sizes 15-45 were size selected on a 15% TBE-Urea polyacrylamide gel (ThermoFisher). RNA was dephosphorylated with T4 PNK (NEB) and 3 pmol of this dephosphorylated RNA was ligated to an adenylated adaptor using truncated T4 RNA ligase 2 K277Q (Jonathan Weissman). Ligated RNA was purified by denaturing polyacrylamide size selection and reverse transcribed using SuperScript III (ThermoFisher) with a custom primer oCJ485. This cDNA was then circularized using CircLigase (Lucigen) and barcodes and full Illumina adapters were added through PCR using Phusion (NEB).

3′ end sequencing libraries were prepared in a manner similar to Rend-seq, with deviation in the order of reactions (Herzel et al., 2022). Prior to fragmentation, 1.8 µg of rRNA depleted RNA was ligated to an adenylated adapter as described for Rend-seq. This ligation product was then fragmented and RNA species of sizes 33-63 nt were purified via denaturing polyacrylamide gel extraction. From this point onwards, the library preparation is identical to Rend-seq. Note that this protocol lacks a dephosphorylation step prior to 3′ end ligation, such that RNA ends

with a 3′ phosphate will not be captured by this protocol. Both Rend-seq and 5′ end sequencing libraries were sequenced with 50 or 75 bp single end reads on either a HiSeq 2000 or NextSeq500.

Multiplexed 5′ end sequencing

For all 5′ end sequencing samples, RNA was extracted using the RNAsnap protocol and Turbo DNase treated as described for the Rend-seq and 3′ end sequencing experiments. rRNA was depleted from 10 µg of purified RNA with the MICROBExpress Bacterial mRNA Enrichment Kit, per manufacturer's instructions, precipitated, and resuspended in 10 mM Tris pH 7. 600 ng rRNA depleted RNA was then ligated to 260 pmol oJT900 (ACGACGCUCUUCCGAUCUNNNNNNNN, IDT) for 18 hours at 22°C using T4 RNA Ligase 1 (based on (Subtelny et al., 2014)). oJT900 is an RNA adaptor with a 5′ hydroxyl group and 8 random nucleotides at its 3′ end, allowing this ligation to selectively capture 5′ monophosphates with reduced sequence bias.

Following ligation, RNA was fragmented for 2 minutes (ThermoFisher) and fragments of either size 41-71 (corresponding to 15-45 plus a 26 nucleotide adapter) or 60-90 were isolated through denaturing polyacrylamide gel extraction. The larger size window was chosen to increase effective sequencing depth by removing dimerized fragments of oJT900, which were observed in high abundance in sequencing libraries prepared with a 41-71 nt size selection. RNA fragments were dephosphorylated, polyadenylated, and reverse transcribed following the BaM-seq multiplexed RNA sequencing protocol (Johnson et al., 2023). In short, in a single tube RNA fragments are sequentially dephosphorylated with T4 PNK (NEB), polyadenylated with *E. coli* poly(A) polymerase (NEB) and reverse transcribed with SuperScript III (ThermoFisher) and a

barcoded reverse transcription primer. The cDNA from each sample generated by this reaction can then be pooled and purified by denaturing polyacrylamide gel electrophoresis. To add Illumina adapters, pooled cDNA was amplified between 5-10 cycles with primers oDP128 and oDP161 (Johnson et al., 2023) using Q5 DNA polymerase and purified by polyacrylamide gel electrophoresis. 5′ end sequencing libraries were sequenced with 75 bp single end reads on a NextSeq500.

Sequencing analysis

Adapter or poly(A) sequence was trimmed from end sequencing and Rend-seq libraries using CutAdapt (Martin, 2011) with parameters "-a 'CTGTAGGCACCATCAAT' -m 15" or "-a A{15} -m 15 -u 8 --rename='{id} UMI={cut_prefix}'". 5′ end sequencing reads were additionally UMI collapsed using the 8 bp random sequence at the start of each read using custom Python scripts, requiring identical UMI sequences and read sequences within a hamming distance of 1 to be considered duplicates. Trimmed reads were then mapped to the *S. cerevisiae* genome (SacCer3) using bowtie v1.1.2 with parameters "-v 1 -k 1 –best –un" with retained unmapped reads subsequently mapped to the *B. subtilis* genome (NC_000964.3) with the same parameters. Mapping against yeast removes contaminating 2xYT-derived sequences and allows estimation of the degree of contamination in a particular sample. Using custom Python scripts, we converted the bowtie output to wig files, attributing signal from reads with a mismatch at their first position to the adjacent nucleotide to address mismatches due to non-templated addition during reverse transcription.

Identification of putative endonuclease sites

To identify putative sites of endoribonucleolytic cleavage, we sought to identify positions for which an adjacent 5′ and 3′ end could be detected in our end sequencing approaches. To do this, we first normalized all datasets to the total number of millions of CDS-mapped reads in each dataset. At each positions, we identified peaks in read density by calculating a ratio of 5′ end signal at that position to the 90% winsorized average in a window 50 nucleotides upstream (for 3′ ends) or downstream (for 5′ ends) of that position, excluding a 2 nucleotide gap on either side of the peak. For any adjacent positions (allowing a gap of 1 position) with a ratio over background greater than 5, these positions were identified as a group for the purposes of downstream 5′ end signal quantification. At each position with >10 reads per million CDS-mapped reads (RPM) in 5′ end sequencing, we detected paired 3′ ends by calculating an equivalent ratio centered around the position one nucleotide upstream in the 3′ end sequencing of the 4-exo deletion. Additionally, in the 50 nucleotide window downstream of each position, we calculated the local density of mapped 3′ ends of the RNase J1 deletion Rend-seq data to establish that we were considering positions only in sufficiently expressed genes. Putative cleavage sites were called as positions for which the average local Rend-seq depth exceeded 0.1 RPM and both the 3′ and 5′ end sequencing signals had a ratio of 7.5 over background.

Because the sequencing methods used in this work have a lower bound on the size of RNA that can be captured (15 nt for Rend-seq and 3′ end sequencing and 15-34 nt for 5′ end sequencing), we did not consider positions which fell within 40 nucleotides of an RNA 3′ end. To identify these ends, we called peaks in the 3′ end signal from our RNase J1 Rend-seq data using a z-score based peak calling method. At each position, a z-score is calculated as the signal at that position minus the average signal in a window of 50 nucleotides on either side of the peak (excluding a gap of 2 nucleotides on either side of the peak), divided by the standard deviation

within this window. Any position for which the calculated score exceeds 8 is flagged as a peak and therefore a detected 3′ end in our Rend-seq data. Putative cleavage positions located up to 40 nucleotides upstream of one of these 3′ end positions are discarded.

To attribute putative cleavage positions to specific endonucleases, we calculated the ratio of 5′ end sequencing signal at each position in an endonuclease knockout with RNase J1 deletion (or depletion) to the average Rend-seq signal of the same strain in a corresponding window 50 nucleotides on downstream of that position, excluding a 2 nt gap. To account for differences in sequencing depth between samples, this 5′ end sequencing signal at each position was normalized by the total 5′ read counts mapping to putative cleavage sites and Rend-seq signal at each position was normalized to total CDS-mapped reads. Additionally, signal from positions which were previously flagged as grouped in the 5′ end sequencing were summed, with the cleavage assigned upstream of the 5′-most position. This 5′ end sequencing to average Rend-seq ratio is calculated for the corresponding RNase J1 deletion or depletion that expresses the considered endonuclease. The ratio in the endonuclease expressing strain is then divided by the ratio in the endonuclease knockout strain to produce a ratio of ratios (RR), similar to an approach used to map RNase III sites in the past (DiChiara et al., 2016). Putative cleavage positions for which the RR value exceeds 12.15 in a particular endonuclease knockout are called as cleavage sites for that enzyme. This threshold was empirically determined by fitting of the distribution of log-transformed RR values to the sum of two Gaussian distributions and finding the point at which 95% of this fit is contributed by the underlying Gaussian with the larger mean (see Figure 3A). For this endonuclease assignment analysis, any position for which the local Rend-seq density in the endonuclease knockout strain (as calculated above) falls below 0.1 RPM is excluded to avoid spuriously calling positions in genes that are downregulated in this condition.

Local secondary structure prediction

To determine folding propensity of sequences proximal to endonuclease cleavage sites, the *B. subtilis* genome was tiled with 40-mers and folded with RNAfold from the ViennaRNA v2.4.14 package with default parameters (Lorenz et al., 2011). The minimal free energy (MFE) values generated from this folding were assigned to the center position within each k-mer and used to generate wig files containing positional folding propensity. These values were extracted in 150 nucleotide windows centered around putative endonuclease cleavage positions and a position-wise mean was calculated. If a cleavage site was identified within 75 nucleotides of a 3′ or 5′ end present in wild-type W168 (identified in strain SSB1002 by z-score based peak calling with a threshold of z=8) Rend-seq data, that cleavage site was not included in this analysis. To generate a background model that controls for sequence context in the -1/+1 positions, this analysis was repeated 100 times sampling an equivalent number of random positions within coding regions that are centered on GA dinucleotides (the most enriched dinucleotide flanking putative RNase Y sites), excluding sites called as cleavage positions. From these predicted position-wise average folding curves, the 25th and 75th percentile at each position across all curves was used to generate the shaded region on relevant figures.

Information content and k-mer logo generation

To generate sequence logos, the local sequence in a 30 nucleotide window surrounding all putative cleavage sites for a given enzyme were identified and aggregated into a FASTA file. A background set of 30-mers evenly sampled every 100 nucleotides from *B. subtilis* coding regions (with sampling beginning after the start codon) was generated as a background set. kpLogo v1.1

(Wu and Bartel, 2017) was then used to generate information content and k-mer logos with parameters "-bgfile %s -startPos 16 -pc 0.01", with %s pointing to our background set.

Author contributions

J.T. and G.-W.L. conceptualized project, designed experiments and analysis, and wrote the manuscript. J.T. collected data and performed analysis. J.-B.L. collected preliminary Rend-seq data for exonuclease knockouts. C.C., S.D., and F. B. constructed strains, provided feedback on analysis, and helped in conceptualization of project.

Acknowledgements:

The authors thank David Bechhofer for sharing the phage-cured RNase III knockout strains, Byong-Mo Koo and Carol Gross for sharing the RNase J2 knockout strain, the MIT BioMicro Center for advice and high-throughput DNA sequencing, Sean E. McGeary for feedback on the manuscript, and Michael Laub and David Bartel for feedback in developing this project. Thanks additionally to all members of the Li Lab for valuable discussion. This research is supported by a National Science Foundation Graduate Research Fellowship (to J.C.T.) and the NSF CAREER Award MCB1844668. The C.C. lab is supported by funds from the CNRS (UMR 8261), Université Paris Cité, and the Agence Nationale de la Recherche (CoNoCo and LabEx Dynamo).

References

Bechhofer, D.H., and Deutscher, M.P. (2019). Bacterial ribonucleases and their roles in RNA metabolism. Crit. Rev. Biochem. Mol. Biol. *54*, 242–300.

Benda, M., Woelfel, S., Faßhauer, P., Gunka, K., Klumpp, S., Poehlein, A., Kálalová, D., Šanderová, H., Daniel, R., Krásný, L., et al. (2021). Quasi-essentiality of RNase Y in Bacillus subtilis is caused by its critical role in the control of mRNA homeostasis. Nucleic Acids Res. *49*, 7088–7102.

Braun, F., Durand, S., and Condon, C. (2017). Initiating ribosomes and a 5- /3-UTR interaction control ribonuclease action to tightly couple B. Subtilis hbs mRNA stability with translation. Nucleic Acids Res. *45*, 11386–11400.

Brewster, R.C., Jones, D.L., and Phillips, R. (2012). Tuning promoter strength through RNA polymerase binding site design in Escherichia coli. PLoS Comput. Biol. *8*, e1002811.

Britton, R.A., Wen, T., Schaefer, L., Pellegrini, O., Uicker, W.C., Mathy, N., Tobin, C., Daou, R., Szyk, J., and Condon, C. (2007). Maturation of the 5′ end of Bacillus subtilis 16S rRNA by the essential ribonuclease YkqC/RNase J1. Mol. Microbiol. *63*, 127–138.

Broglia, L., Lécrivain, A., Renault, T.T., Ahmed-begrich, R., Rhun, A. Le, and Charpentier, E. (2020). An RNA-seq based comparative approach reveals the transcriptome-wide interplay between 3'-to-5' exoRNases and RNase Y. Nat. Commun. *11*.

Bruscella, P., Shahbabian, K., Laalami, S., and Putzer, H. (2011). RNase Y is responsible for uncoupling the expression of translation factor IF3 from that of the ribosomal proteins L35 and L20 in Bacillus subtilis. Mol. Microbiol. *81*, 1526–1541.

Cerullo, F., Filbeck, S., Patil, P.R., Hung, H.C., Xu, H., Vornberger, J., Hofer, F.W., Schmitt, J., Kramer, G., Bukau, B., et al. (2022). Bacterial ribosome collision sensing by a MutS DNA repair ATPase paralogue. Nature *603*, 509–514.

Chao, Y., Li, L., Girodat, D., Förstner, K.U., Said, N., Corcoran, C., Śmiga, M., Papenfort, K., Reinhardt, R., Wieden, H.J., et al. (2017). In Vivo Cleavage Map Illuminates the Central Role of RNase E in Coding and Non-coding RNA Pathways. Mol. Cell *65*, 39–51.

Chen, H., Shiroguchi, K., Ge, H., and Xie, X.S. (2015). Genome-wide study of mRNA degradation and transcript elongation in E scherichia coli. Mol. Syst. Biol. *11*, 781.

Dar, D., Shamir, M., Mellin, J.R., Koutero, M., Stern-Ginossar, N., Cossart, P., and Sorek, R. (2016). Term-seq reveals abundant ribo-regulation of antibiotics resistance in bacteria. Science (80-.). *352*, aad9822.

Deloughery, A., Dengler, V., Chai, Y., and Losick, R. (2016). Biofilm formation by Bacillus subtilis requires an endoribonuclease-containing multisubunit complex that controls mRNA levels for the matrix gene repressor SinR. Mol. Microbiol. *99*, 425–437.

DeLoughery, A., Lalanne, J.-B., Losick, R., and Li, G.-W. (2018). Maturation of polycistronic mRNAs by the endoribonuclease RNase Y and its associated Y-complex in *Bacillus subtilis*. Proc. Natl. Acad. Sci. *115*, E5585–E5594.

DiChiara, J.M., Liu, B., Figaro, S., Condon, C., and Bechhofer, D.H. (2016). Mapping of internal monophosphate 5′ ends of Bacillus subtilis messenger RNAs and ribosomal RNAs in wild-type and ribonuclease-mutant strains. Nucleic Acids Res. *44*, 3373–3389.

Durand, S., Braun, F., Lioliou, E., Romilly, C., Helfer, A.-C., Kuhn, L., Quittot, N., Nicolas, P., Romby, P., and Condon, C. (2015). A Nitric Oxide Regulated Small RNA Controls Expression of Genes Involved in Redox Homeostasis in Bacillus subtilis. PLOS Genet. *11*, e1004957.

Figaro, S., Durand, S., Gilet, L., Cayet, N., Sachse, M., and Condon, C. (2013). Bacillus subtilis mutants with knockouts of the genes encoding ribonucleases RNase Y and RNase J1 are viable, with major defects in cell morphology, sporulation, and competence. J. Bacteriol. *195*, 2340–2348.

Gilet, L., Dichiara, J.M., Figaro, S., Bechhofer, D.H., and Condon, C. (2015). Small stable RNA maturation and turnover in Bacillus subtilis. Mol. Microbiol. *95*, 270–282.

Herzel, L., Stanley, J.A., Yao, C.-C., and Li, G.-W. (2022). Ubiquitous mRNA decay fragments in E. coli redefine the functional transcriptome . Nucleic Acids Res. *50*, 5029–5046.

Ingle, S., Chhabra, S., Chen, J., Lazarus, M.B., Luo, X., and Bechhofer, D.H. (2022). Discovery and initial characterization of YloC, a novel endoribonuclease in Bacillus subtilis. RNA *28*, 227–238.

Ishida, Y., Park, J.-H., Mao, L., Yamaguchi, Y., and Inouye, M. (2013). Replacement of all arginine residues with canavanine in MazF-bs mRNA interferase changes its specificity. J. Biol. Chem. *288*, 7564–7571.

Johnson, G.E., Parker, D.J., Lalanne, J., Parker, M.L., and Li, G. (2023). BaM-seq and TBaM-seq , highly multiplexed and targeted RNA-seq protocols for rapid , low-cost library generation from bacterial samples. *5*, 1–11.

Khemici, V., Prados, J., Linder, P., and Redder, P. (2015). Decay-Initiating Endoribonucleolytic Cleavage by RNase Y Is Kept under Tight Control via Sequence Preference and Sub-cellular Localisation. PLoS Genet. *11*, 1–27.

Koo, B.M., Kritikos, G., Farelli, J.D., Todor, H., Tong, K., Kimsey, H., Wapinski, I., Galardini, M., Cabal, A., Peters, J.M., et al. (2017). Construction and Analysis of Two Genome-Scale Deletion Libraries for Bacillus subtilis. Cell Syst. *4*, 291-305.e7.

Lalanne, J.B., Taggart, J.C., Guo, M.S., Herzel, L., Schieler, A., and Li, G.W. (2018). Evolutionary Convergence of Pathway-Specific Enzyme Expression Stoichiometry. Cell

173, 749-761.e38.
Lehnik-Habrink, M., Schaffer, M., Mäder, U., Diethmaier, C., Herzberg, C., and Stülke, J. (2011). RNA processing in Bacillus subtilis: Identification of targets of the essential RNase Y. Mol. Microbiol. *81*, 1459–1473.
Lehnik-Habrink, M., Lewis, R.J., Mäder, U., and Stülke, J. (2012). RNA degradation in Bacillus subtilis: An interplay of essential endo- and exoribonucleasesmmi. Mol. Microbiol. *84*, 1005–1017.
Leroy, M., Piton, J., Gilet, L., Pellegrini, O., Proux, C., Coppée, J., Figaro, S., and Condon, C. (2017). Rae1/YacP, a new endoribonuclease involved in ribosome-dependent mRNA decay in Bacillus subtilis . EMBO J. *36*, 1167–1181.
Lorenz, R., Bernhart, S.H., Zu Siederdissen, C.H., Tafer, H., Flamm, C., Stadler, P.F., and Hofacker, I.L. (2011). ViennaRNA Package 2.0. Algorithms Mol. Biol. *6*, 1–14.
Mandell, Z.F., Oshiro, R.T., Yakhnin, A. V, Kashlev, M., Kearns, D.B., and Babitzke, P. (2021). NusG is an intrinsic transcription termination factor that stimulates motility and coordinates global gene expression with NusA. Elife *10:e61880*.
Marincola, G., and Wolz, C. (2017). Downstream element determines RNase y cleavage of the saePQRS operon in Staphylococcus aureus. Nucleic Acids Res. *45*, 5980–5994.
Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet.Journal *17*, 10.
Moffitt, J.R., Pandey, S., Boettiger, A.N., Wang, S., and Zhuang, X. (2016). Spatial organization shapes the turnover of a bacterial transcriptome. Elife *5*, 1–22.
Noone, D., Salzberg, L.I., Botella, E., Bäsell, K., Becher, D., Antelmann, H., and Devine, K.M. (2014). A highly unstable transcript makes CwlO D,L-endopeptidase expression responsive to growth conditions in bacillus subtilis. J. Bacteriol. *196*, 237–247.
Oussenko, I.A., Abe, T., Ujiie, H., Muto, A., and Bechhofer, D.H. (2005). Participation of 3 ′ -to-5 ′ Exoribonucleases in the Turnover of Bacillus subtilis mRNA Participation of 3 J -to-5 J Exoribonucleases in the Turnover of Bacillus subtilis mRNA. J. Bacteriol. *187*, 2758–2767.
Park, J.-H., Yamaguchi, Y., and Inouye, M. (2011). Bacillus subtilis MazF-bs (EndoA) is a UACAU-specific mRNA interferase. FEBS Lett. *585*, 2526–2532.
Parker, D.J., Lalanne, J.B., Kimura, S., Johnson, G.E., Waldor, M.K., and Li, G.W. (2020). Growth-Optimized Aminoacyl-tRNA Synthetase Levels Prevent Maximal tRNA Charging. Cell Syst. *11*, 121-130.e6.
Salis, H.M. (2011). The ribosome binding site calculator. Methods Enzym. *498*, 19–42.
Salis, H.M., Mirsky, E.A., and Voigt, C.A. (2009). Automated design of synthetic ribosome binding sites to control protein expression. Nat. Biotechnol. *27*, 946–950.
Shahbabian, K., Jamalli, A., Zig, L., and Putzer, H. (2009). RNase Y, a novel endoribonuclease, initiates riboswitch turnover in Bacillus subtilis. EMBO J. *28*, 3523–3533.
Stead, M.B., Agrawal, A., Bowden, K.E., Nasir, R., Mohanty, B.K., Meagher, R.B., and Kushner, S.R. (2012). RNAsnap^{TM}: a rapid, quantitative and inexpensive, method for isolating total RNA from bacteria. Nucleic Acids Res. *40*, e156.
Subtelny, A.O., Eichhorn, S.W., Chen, G.R., Sive, H., and Bartel, D.P. (2014). Poly(A)-tail profiling reveals an embryonic switch in translational control. Nature *508*, 66–71.
Taggart, J.C., Lalanne, J.B., and Li, G.W. (2021). Quantitative Control for Stoichiometric Protein Synthesis. Annu. Rev. Microbiol. *75*, 243–267.
Trinquier, A., Durand, S., Braun, F., and Condon, C. (2020). Regulation of RNA processing and

degradation in bacteria. Biochim. Biophys. Acta - Gene Regul. Mech. *1863*, 194505.
Urtecho, G., Tripp, A.D., Insigne, K.D., Kim, H., and Kosuri, S. (2019). Systematic Dissection of Sequence Elements Controlling σ70 Promoters Using a Genomically Encoded Multiplexed Reporter Assay in Escherichia coli. Biochemistry *58*, 1539–1551.
Wu, X., and Bartel, D.P. (2017). kpLogo: positional k-mer analysis reveals hidden specificity in biological sequences. Nucleic Acids Res. *45*, W534–W538.
Yao, S., and Bechhofer, D.H. (2010). Initiation of decay of Bacillus subtilis rpsO mRNA by endoribonuclease RNase Y. J. Bacteriol. *192*, 3279–3286.
Yasbin, R.E., and Young, F.E. (1974). Transduction in Bacillus subtilis by Bacteriophage SPP1 . J. Virol. *14*, 1343–1348.

Figure legends

Figure 1. Workflow and validation of endoribonuclease cleavage mapping approach
(A) Deletion of exoribonucleases results in stable accumulation of RNA decay intermediates. A schematic mRNA is shown with endoribonuclease cleavage (red scissors) occurring, with remaining exoribonucleases indicated in green (RNase J1) or red (3-5′ exoribonucleases). RNAs are shown with 5′-P and 3′-OH characteristic of the most common endoribonucleases in *B. subtilis*. Newly generated RNA ends can be captured through RNA ligation allowing mapping of cleavage positions by 3′ and 5′ end sequencing. Adapters used to capture RNA ends indicated in purple for 3′ end sequencing and yellow for 5′ end sequencing, with matching color scheme in schematic data below.
(B) Validation of endonuclease cleavage site detection using known positions of endoribonucleolytic cleavage by RNase III, RNase Y, and Rae1 (DiChiara et al., 2016; Durand et al., 2015; Leroy et al., 2017). Yellow indicates 5′ end sequencing data and purple indicates 3′ end sequencing data. Red dotted line represents manually annotated cleavage positions and gray line indicates published positions of cleavage if different than positions highlighted in red.

Figure 2. Cleavage within transcripts known to be destabilized by endoribonucleolytic cleavage. 5′ and 3′ end sequencing data are shown in yellow and purple, respectively, with axes increasing in opposite directions. Manually annotated putative cleavage sites (positions of adjacent 3′ and 5′ read density) are highlighted with arrows. Insets show highlighted sites at single-nucleotide resolution, with the Y-axis in each direction rescaled to the maximal value within the inset region. Data are aligned to a schematic representation of the relevant genomic locus, with promoter and transcriptional terminator positions annotated using Rend-seq in SSB1002. Sequence context and RNase Y dependence of these sites shown in Figure S2.

Figure 3. Sequence and structural features of putative RNase Y substrates
(A) Identification of putative RNase Y cleavage sites. Histogram shows distribution of RR (ratio of ratios) values for called peak pairs in 3′/5′ end sequencing of exoribonuclease knockouts. Solid red line shows results of fitting the sum of two Gaussian distributions to this underlying log-transformed data, with the dashed lines showing the two constituent distributions produced by this fitting. The dashed vertical line indicates the lowest RR value at which 95% of the Gaussian sum is derived from the right constituent distribution (RR=12.15), a value above which sites are considered dependent on RNase Y. 493 of 1427 putative cleavage signatures exceed this threshold.
(B) Local sequence context around positions of RNase Y cleavage (N=493). Nucleotide frequency (top), information content (middle) and k-mer (bottom) logos in a 30 nucleotide window around called RNase Y sites. In k-mer logo, the most enriched k-mers are shown above the sequence and the most depleted below, with significant positions highlighted in red.
(C) Local GC content around positions of RNase Y cleavage (N=403). Sites near transcript boundaries are not considered (see Methods). Dotted gray line indicates average %GC across *B. subtilis* genome. Vertical red line indicates position of cleavage.
(D) Predicted folding near RNase Y cleavage positions (N=403). Sites near transcript boundaries are not considered (see Methods). Blue line shows position-wise average minimal free energy (MFE) predicted by RNAfold for 40-mers tiling cleavage sites. Gray band indicates per-position interquartile range of curves resulting from folding randomly sampled GA-dinucleotide centered

windows within coding regions, excluding putative cleavage positions (see Methods). Red vertical line indicates putative position of cleavage.

Figure 4. Stringent *in vivo* specificity of EndoA
(A, B) 5′ (A) and 3′ (B) mapped Rend-seq signal across all UACAU motifs in the genome, separated by downstream nucleotide. 5′ end sequencing data derived from an RNase J1 deletion (CCB434) and 3′ end sequencing derived from a 4-exo knockout (CCB396). Rend-seq data at each site is normalized to first 8 (for 3′-mapped) or last 8 (for 5′-mapped) positions within the window and a position-wise mean and standard deviation are calculated with 90% winsorization. Motif instances with fewer than 1 read per position or fewer than 10 reads within the normalization window are not considered. Following this filtering, the number of considered sites ranges from 264 to 455 per motif. The red vertical line indicates the position of cleavage by EndoA.
(C, D) 5′-mapped Rend-seq signal in an RNase J1 depletion (CCB390) (C) or and RNase J1 depletion with deletion of *ndoA* (BT231) (D). Analysis performed identically to (A, B). Following filtering, the number of considered sites ranges from 86 to 288 per motif.

Figure 5. Evidence for previously unknown 5′ end trimming activity
(A) Rend-seq (top) and end sequencing (bottom) data at a representative EndoA cleavage site located in the gene *pgk* for a 4-exo knockout (CCB396), RNase J1 depletion (CCB390), and RNase J1 depletion with an *ndoA* knockout (BJT231). 5′-mapped data shown in yellow (5′ end sequencing) or orange (Rend-seq) and 3′-mapped data shown in purple (3′ end sequencing) or blue (Rend-seq). Red dotted line indicates position of cleavage. To account for changes in gene expression, end sequencing data are normalized to the average Rend-seq 3′-mapped read counts in the same strain in the considered window.
(B) Rend-seq (top) and end sequencing (bottom) data at the position of cleavage by the *glmS* ribozyme. 5′-mapped data shown in yellow (5′ end sequencing) or orange (Rend-seq) and 3′-mapped data shown in purple (3′ end sequencing) or blue (Rend-seq). Strains included are identical to panel A. Red dotted line indicates position of cleavage. End sequencing data are normalized to the average Rend-seq 3′-mapped read counts in the same strain in the considered window.
(C) End sequencing data at the position of cleavage by the *glmS* ribozyme in a 4-exo knockout (CCB396) and panel of all decay-associated proteins for which we have collected 5′ end sequencing data. All endonuclease knockouts are coupled to either a depletion or deletion of RNase J1. Included knockouts include *rnjA* (CCB434), *ndoA* (BJT231), *ylbF* (BJT074), *rnc* (BG879), and *rny* (CCB760). End sequencing data are normalized to the average Rend-seq 3′-mapped read counts in the same strain in the considered window. 3′ end sequencing data are shown in purple and 5′ end sequencing are shown in yellow.
(D) Rend-seq data at the position of cleavage by the *glmS* ribozyme in a 4-exo knockout (CCB396), RNase J1 knockout (CCB434), and RNase J1 depletion with a *rnjB* knockout (GLB186). Red dotted line indicates position of cleavage. 3′-mapped data are shown in blue and 5′-mapped data are shown in orange.
(E, F) 5′ end sequencing (E) and Rend-seq 3′-mapped (F) signal across all UACAUA EndoA cleavage motifs, separated by downstream sequence. 5′ end sequencing data derived from an RNase J1 depletion with *rny* knockout (CCB760) and 3′ end sequencing derived from a 4-exo knockout (CCB396). The signal in a 20 nt window around each cleavage site was normalized to

its maximal value and a position-wise average across all normalized windows was calculated with 90% winsorization. UACAUA instances for which the local density in corresponding Rend-seq data is less than 1 read per position were discarded. Following this filtering, the number of considered sites ranges from 10 to 54 per row (top) and 10 to 62 per row (bottom). The red vertical line indicates the position of cleavage by EndoA.

Figure S1. Validation of approach through detection of known positions of endonuclease activity. 5′ and 3′ end sequencing data at known positions of cleavage by RNases Y and III. Datasets considered are a 4-exo knockout (CCB396), *rnjA* knockout (CCB434), and either deletion or depletion of RNase J1 with a knockout of RNase Y (CCB760) or III (BG879). 5′-mapped data shown in yellow and 3′-mapped data shown in purple. To account for changes in gene expression, end sequencing data are normalized to the average Rend-seq 3′-mapped read counts in the same strain in the considered window. A manually annotated position of cleavage based on sequencing data is shown with a red dotted line, and the published position is marked in gray of cleavage if this position disagrees with our annotation. In transcripts with many nearby peak pairs such as *hbs*, additional cleavage positions may be visible beyond that which is highlighted.

Figure S2. **Sequence context and RNase Y dependence of cleavage positions within mRNAs known to be destabilized by endonucleases**. Cleavage positions match order of those shown in Figure 2. Datasets considered are a 4-exo knockout (CCB396), *rnjA* knockout (CCB434), and depletion of RNase J1 with a knockout of *rny* (CCB760). 5′ end sequencing data shown in yellow and 3′ end sequencing data shown in purple. To account for changes in gene expression, end sequencing data are normalized to the average Rend-seq 3′-mapped read counts in the same strain in the considered window. A manually annotated position of cleavage based on sequencing data is shown with a red dotted line.

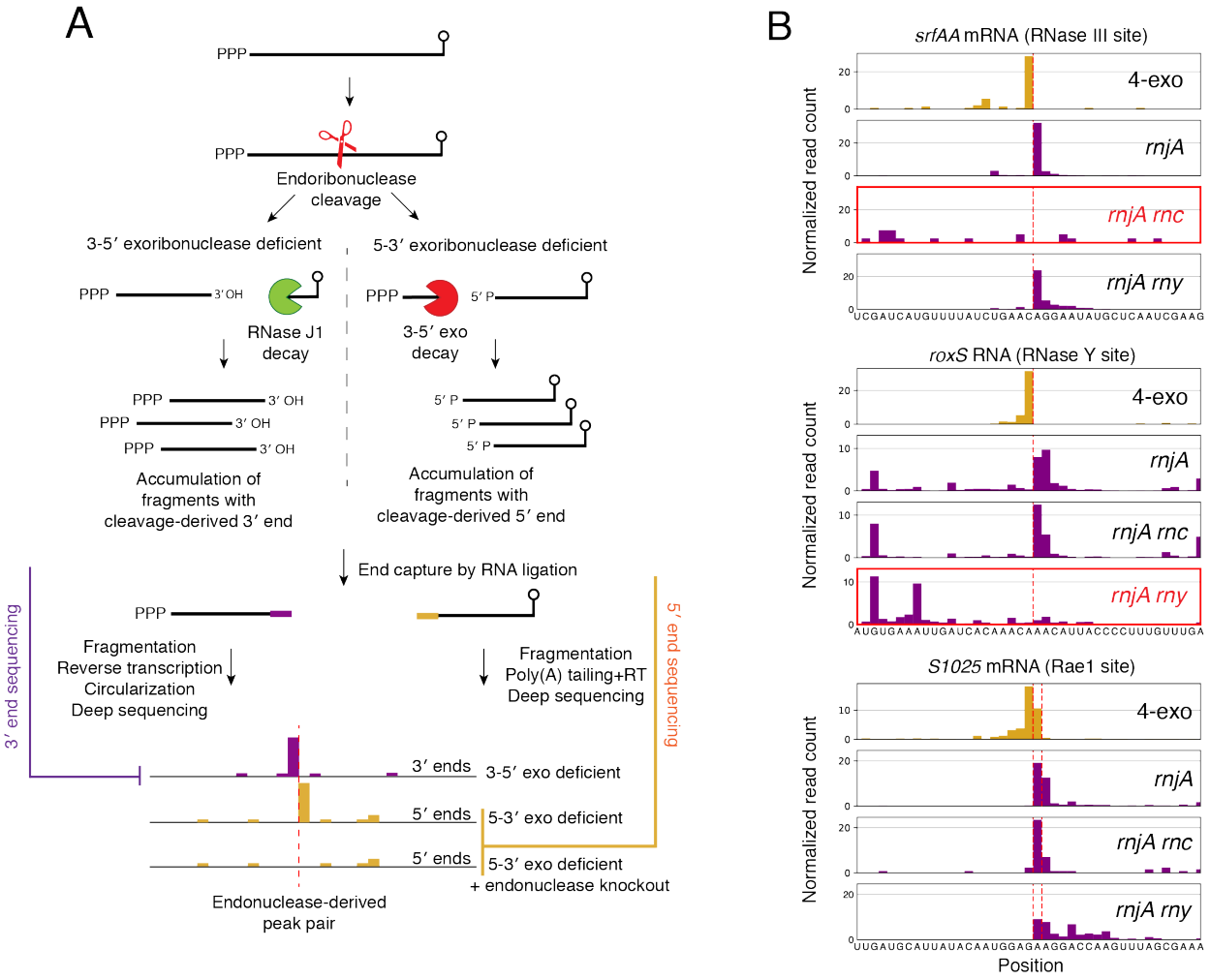
Figure S3. Schematic representation of strategy for calling putative cleavage positions and assigning them to the activity of endoribonucleases
(A) Strategy for calling paired 3′ and 5′ peaks in end sequencing data, with sites exceeding a peak-to-background ratio of 7.5 in the 3′/5′ at adjacent positions called as putative cleavage positions.
(B) Strategy for attributing sites to specific endonucleases. Peak heights in 5′ end sequencing are normalized to local density in a corresponding 5′-mapped Rend-seq dataset. A ratio of normalized peak heights between exoribonuclease strains with and without an endoribonuclease knockout (ratio of ratios, or RR) is calculated as a metric for dependence on this endoribonuclease. Note that prior to illustrated calculation, 5′ end sequencing data (h_{WT_5end} and $_{hKO_5end}$) are normalized by the sum of 5′ end sequencing reads mapped to putative cleavage positions, and the Rend-seq background (bg_{Rend}) is normalized to the total CDS-mapped reads to account for differences in sequencing depth between samples.

Figure S4. **Suppression of *cggR* cleavage site by downstream cleavage**. Rend-seq data at an RNase Y cleavage site located in the gene *cggR* for a 4-exo knockout (CCB396), *rnjA* knockout (CCB434), and *rnjA ylbF* double knockout (BJT074). 5′-mapped data shown in orange and 3′-mapped data shown in blue. Red dotted line indicates position of cleavage.

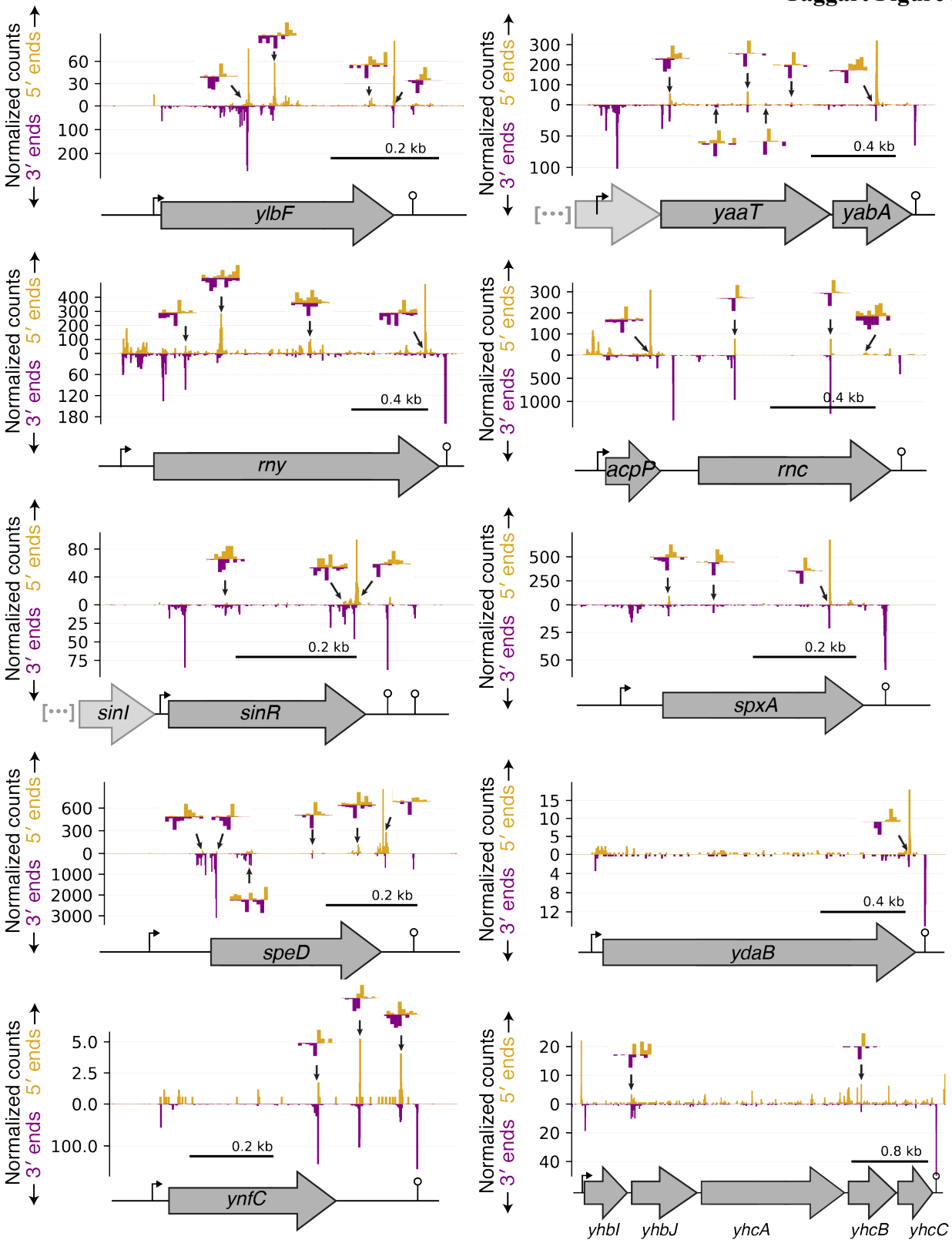
Table S1. List of strains used in this study.

Figures

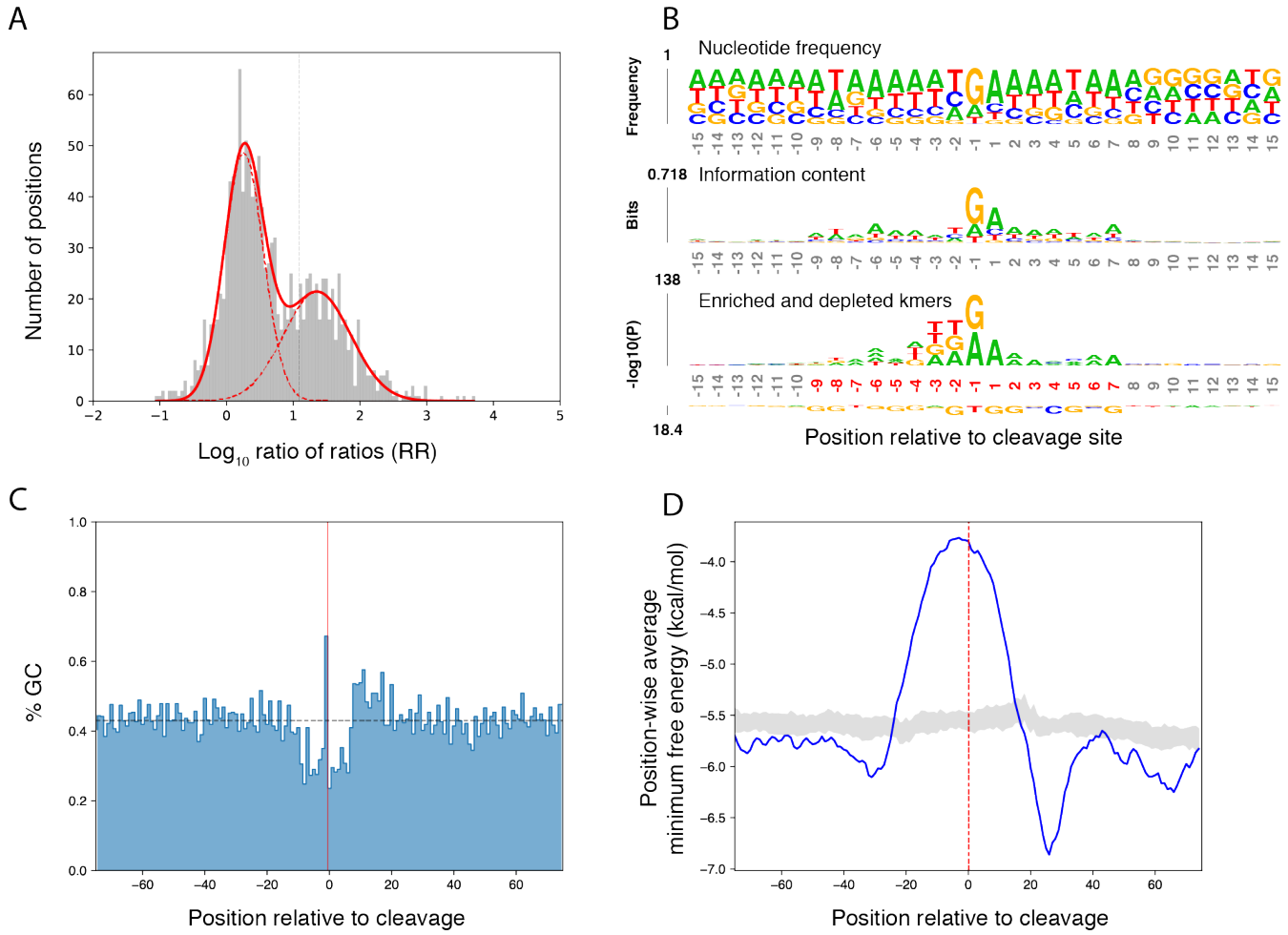
Taggart Figure 1



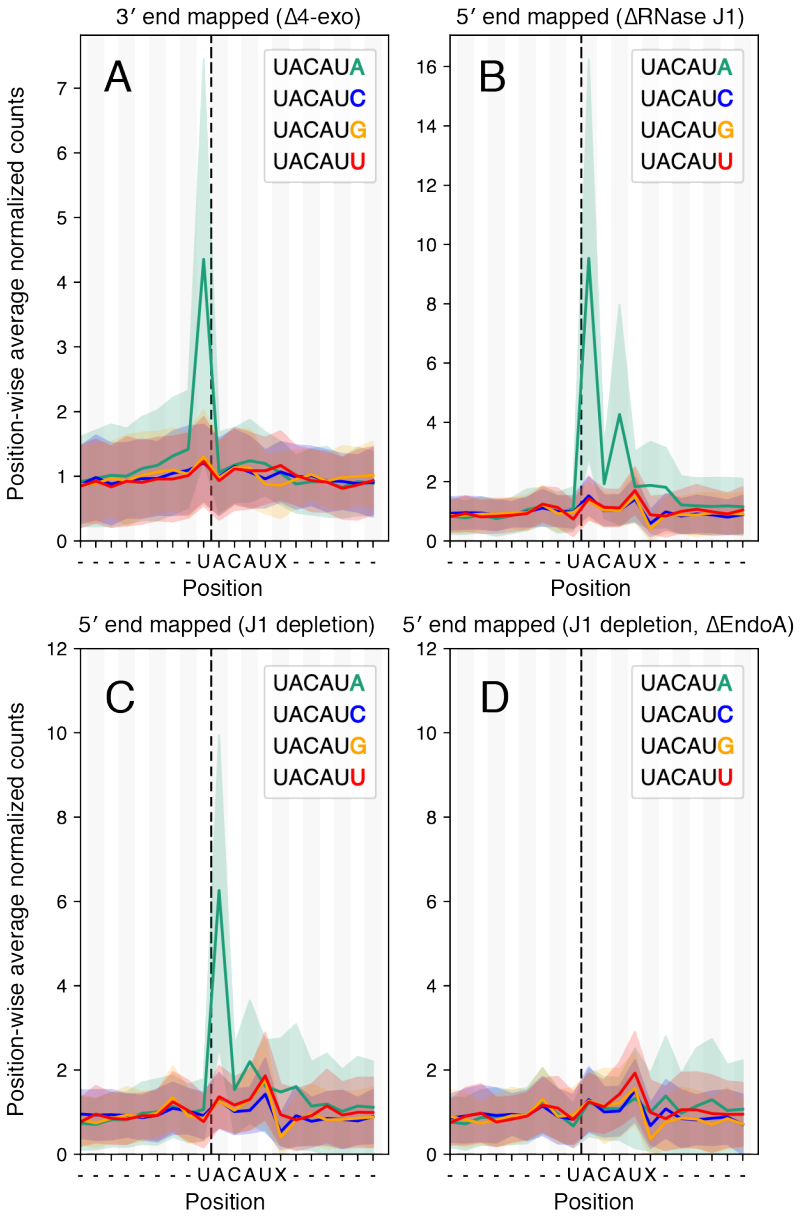
Taggart Figure 2



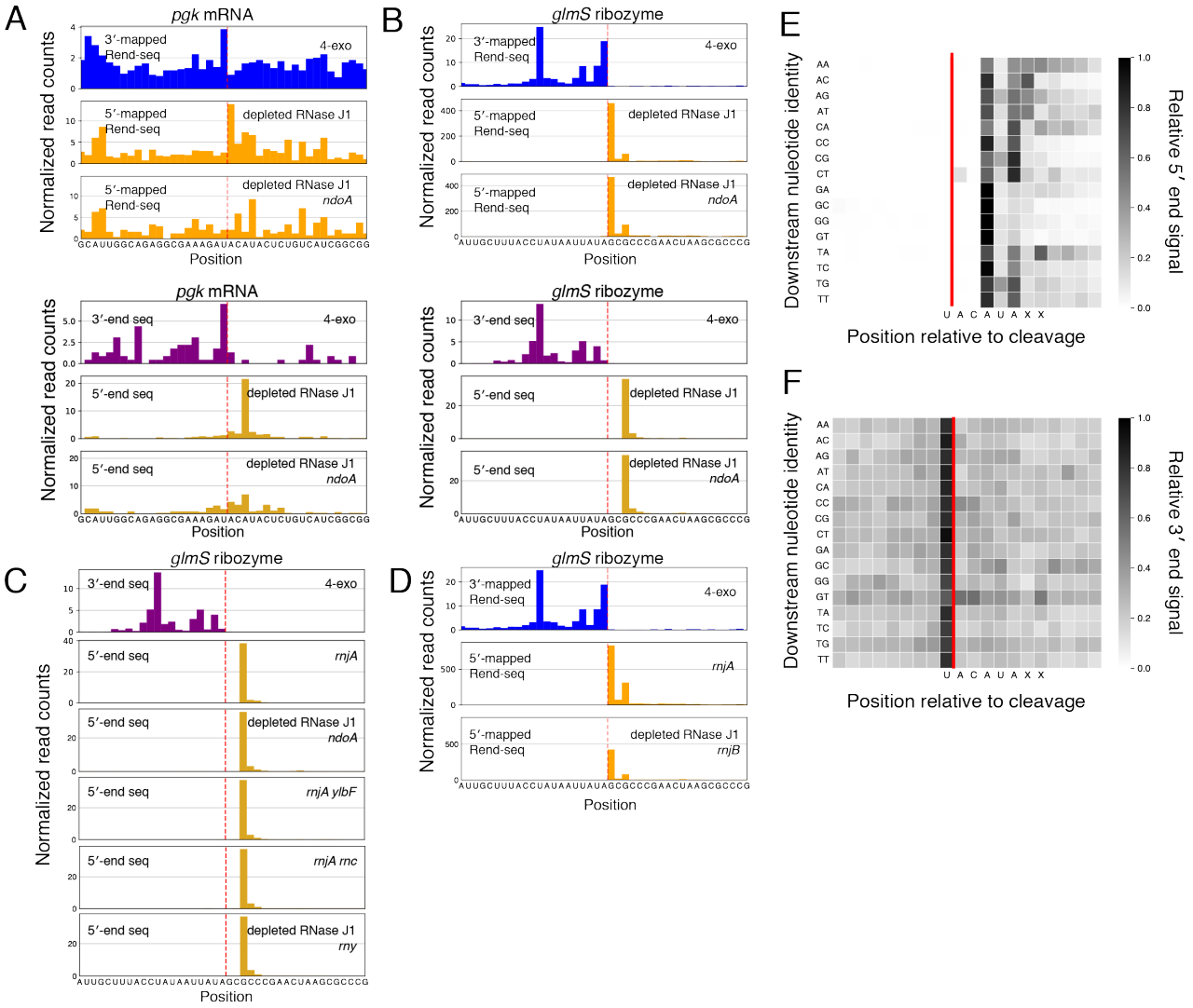
Taggart Figure 3



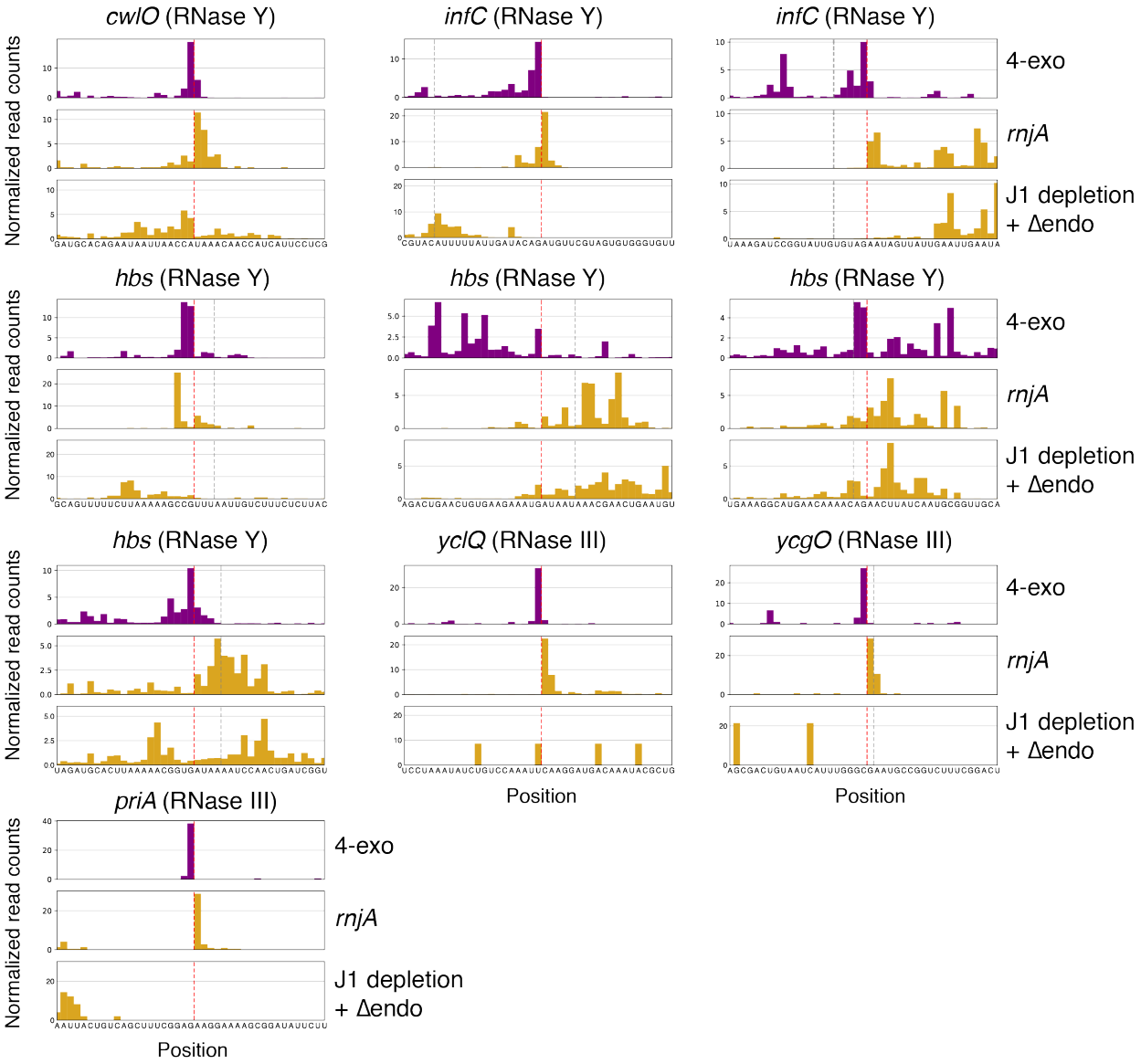
Taggart Figure 4



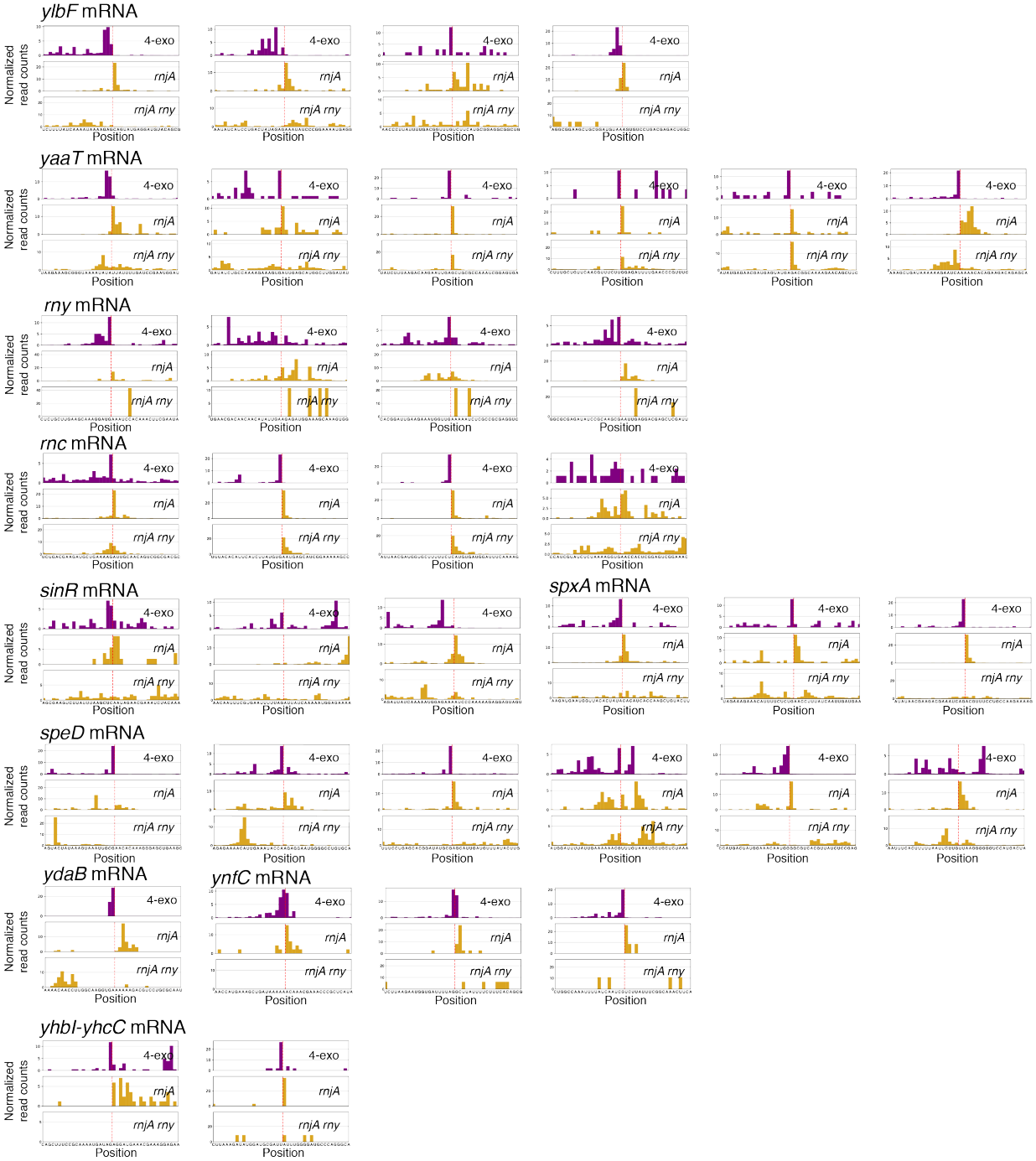
Taggart Figure 5



Taggart Figure S1

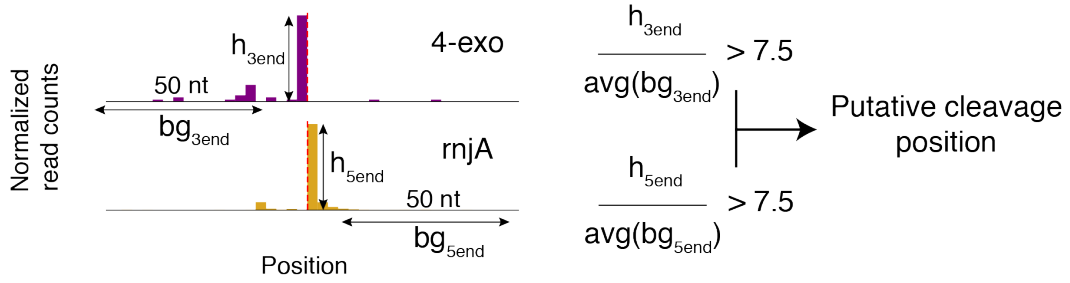


Taggart Figure S2

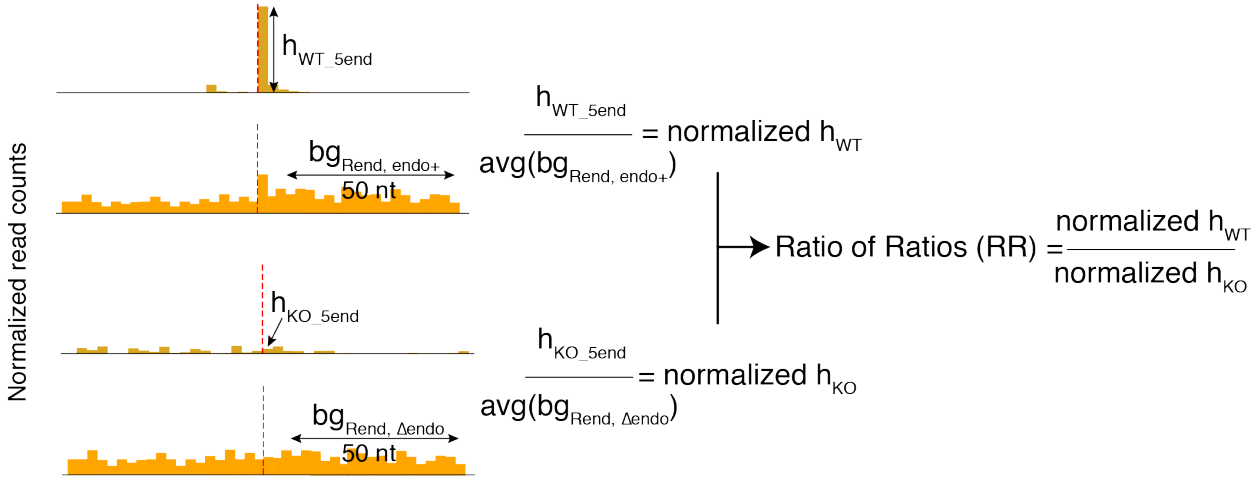


Taggart Figure S3

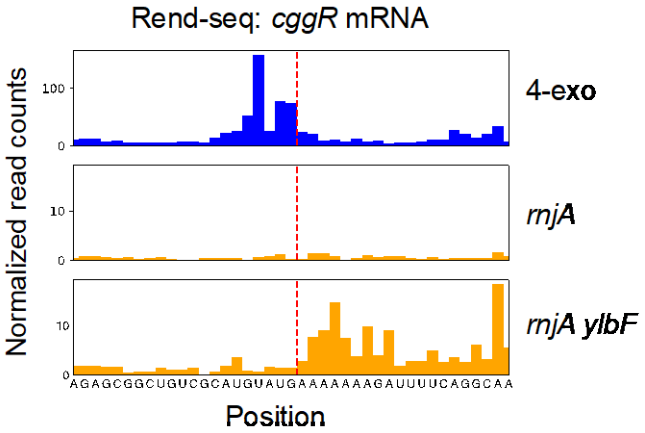
A Identifying putative cleavage positions



B Attributing cleavage to a specific endonuclease



Taggart Figure S4



Taggart Table S1

Table S1: Strains used in this study

Strain name	Genotype	Source
CCB434	*W168 rnjA::spc*	(Figaro et al., 2013)
CCB396 (4-exo)	*W168 rph::spc rnr::tc yhaM::Pm pnp::kan*	(Gilet et al., 2015; Oussenko et al., 2005)
BJT074	*W168 rnjA::spc ylbF::erm*	This study
BJT070	*W168 rph::spc rnr::tc yhaM::Pm pnp::kan ylbF::erm*	This study
CCB390	*W168 amyE::{Pxyl-rnjA Cm} rnjA::spc*	This study
BJT231	*W168 amyE::{Pxyl-rnjA Cm} rnjA::spc ndoA::kan*	This study
BG880	*W168 ΔSkin ΔSPβ::PIID-sspB kan rnjA::spc*	(Gilet et al., 2015)
BG879	*W168 ΔSkin ΔSPβ::PIID-sspB kan rnc::ery rnjA::spc*	(Gilet et al., 2015)
CCB760	*W168 Pspac-rnjA ery rny::spc*	(Braun et al., 2017)
BJT034 (BKE14990)	*trpC2 ylbF::erm*	(Koo et al., 2017)
SSB1002	*W168 (trp+)*	Condon lab wild-type strain
GLB230 (BKK04660)	*trpC2 ndoA::kan*	(Koo et al., 2017)
GLB186	*trpC2 rnjB::lox72, amyE::{Pspank-rnjA spc}, rnjA::kan*	Provided by Byoung-Mo Koo and Carol Gross