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

James C. Taggart[1,4], Julia Dierksheide[1], Hannah LeBlanc[1], Jean-Benoît Lalanne[1,2,5], Sylvain Durand[3], Frédérique Braun[3], Ciarán Condon[3], and Gene-Wei Li[1,#]

[1]Department of Biology, Massachusetts Institute of Technology, Cambridge, MA 02142, USA

[2]Department of Physics, Massachusetts Institute of Technology, Cambridge, MA 02142, USA

[3]Expression Génétique Microbienne (EGM), CNRS, Université Paris Cité, Institut de Biologie Physico-Chimique, 13 rue Pierre et Marie Curie, 75005 Paris, France

[4]Current address: Department of Systems Biology, Harvard Medical School, Boston, MA, USA

[5]Current address: Department of Genome Sciences, University of Washington, Seattle, WA, USA.

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

ABSTRACT

RNA endonucleases are the rate-limiting initiator of decay for many bacterial mRNAs. However, the positions of cleavage and their sequence determinants remain elusive even for the well-studied *Bacillus subtilis*. Here we present two complementary approaches – transcriptome-wide mapping of endoribonucleolytic activity and deep mutational scanning of RNA cleavage sites – that reveal distinct rules governing the specificity among *B. subtilis* endoribonucleases. Detection of RNA terminal nucleotides in both 5′- and 3′-exonuclease-deficient cells revealed $>10^3$ putative endonucleolytic cleavage sites with single-nucleotide resolution. We found a surprisingly weak consensus for RNase Y targets, a contrastingly strong primary sequence motif for EndoA targets, and long-range intramolecular secondary structures for RNase III targets. Deep mutational analysis of RNase Y cleavage sites showed that the specificity is governed by many disjointed sequence features, each with mild contributions. Our results highlight the delocalized nature of mRNA stability determinants and provide a strategy for elucidating endoribonuclease specificity *in vivo*.

27 **INTRODUCTION**

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

37 In the Gram-positive bacterium *Bacillus subtilis*, cleavage by an endoribonuclease is thought to
38 initiate the decay of a large fraction of cellular mRNAs (Durand et al., 2012a; Lehnik-Habrink et al.,
39 2011a, 2012; Shahbabian et al., 2009). The cleaved RNA fragments typically have a 5′-phosphate and a
40 3′-hydroxyl group, which are subsequently degraded by a 5′-3′ exoribonuclease (RNase J1) and by a
41 combination of 3′-5′ exoribonucleases (PNPase, RNase R, RNase PH, and YhaM) (Trinquier et al., 2020),
42 respectively. The endoribonuclease RNase Y appears to drive the turnover of the majority of mRNAs in
43 *B. subtilis*, and its depletion results in the accumulation of hundreds of transcripts (Durand et al., 2012a;
44 Lehnik-Habrink et al., 2011a; Shahbabian et al., 2009). For several well-studied mRNAs, RNase Y
45 cleavage is specific to certain positions and can drive operon mRNA maturation (Braun et al., 2017;
46 Commichau et al., 2009; DeLoughery et al., 2018; Noone et al., 2014). Unfortunately, though RNase Y
47 has been identified as a key player shaping the transcriptomes of Gram-positive bacteria, our
48 understanding of which mRNA sequences *B. subtilis* RNase Y cleaves remains relatively limited. Studies
49 have suggested *B. subtilis* RNase Y recognizes single stranded RNA, is stimulated by downstream
50 secondary structure, and has a preference for adenosine and uridine-rich sequence (Braun et al., 2017;
51 Shahbabian et al., 2009). Orthologs of RNase Y in *Staphylococcus aureus* and *Streptococcus pyogenes*,
52 organisms for which RNase Y is not essential and appears to play a more limited role compared to *B.*

53 *subtilis*, additionally demonstrate a preference for cleavage downstream of a guanosine (Khemici et al.,
54 2015) (Benda et al., 2021). However, these limited sequence preferences cannot explain the highly
55 specific locations of cleavage that have been observed for the *B. subtilis* RNase Y (DeLoughery et al.,
56 2018).

57 In addition to RNase Y, several other endoribonucleases contribute to mRNA decay. RNase III
58 has been suggested to cleave and initiate decay for dozens of mRNAs in *B. subtilis* (DiChiara et al.,
59 2016). Although the *E. coli* RNase III strictly cleaves double-stranded RNAs, the *B. subtilis* RNase III has
60 been implicated to deviate from this rule in a way that remains unknown (DiChiara et al., 2016). Other
61 minor endoribonucleases, such as the RNase toxin EndoA (a homolog of *E. coli* MazF), may also initiate
62 mRNA decay and produce different 5′ and 3′ moieties (Cerullo et al., 2022; Ingle et al., 2022; Leroy et al.,
63 2017). Their positions of cleavage and subsequent decay processes remain poorly described (Pellegrini et
64 al., 2005).

65 A number of studies have mapped the target repertoire of major bacterial endoribonucleases
66 (Altuvia et al., 2018; Broglia et al., 2020; Chao et al., 2017; DiChiara et al., 2016; Gordon et al., 2017;
67 Khemici et al., 2015). These measurements detect the terminal nucleotides of RNA decay intermediates
68 and analyze their abundance upon endoribonuclease inactivation. The results are confounded by several
69 factors. First, because cleaved decay intermediates are typically unstable, the detected end signals are
70 often close to background levels and may be biased by their own stability. Second, the sequential actions
71 of several endoribonucleases and/or exoribonucleases on the same mRNA may result in RNA ends that
72 are shifted from the initial position that is cleaved by the endoribonuclease under investigation. As a
73 result, low-throughput approaches are often required to precisely determine exact positions of cleavage
74 for each target (Braun et al., 2017; Bruscella et al., 2011; DiChiara et al., 2016; Yao and Bechhofer,
75 2010).

76 In this work, we first present a global approach to map bacterial endoribonuclease cleavage sites
77 by combining high-resolution RNA end-mapping with genetic perturbation of all known exonucleases
78 participating in *B. subtilis* mRNA decay. By stabilizing mRNA decay intermediates in exonuclease

79 deficient *B. subtilis* cells, this "stabilized end-sequencing" approach revealed thousands of putative
80 positions of endoribonuclease cleavage and a previously unknown nuclease activity targeting RNA 5′
81 ends. Furthermore, using a panel of endoribonuclease knockouts, we dramatically expanded the known
82 target repertoire for RNase Y, RNase III, and EndoA. Our results captured a revised EndoA cleavage
83 motif, elucidated long-range RNA folds that dictate RNase III cleavage, and a surprisingly weak
84 consensus among RNase Y cleavage sites, despite their high specificity. To further investigate the
85 sequence determinants for RNase Y cleavage, we developed a deep mutational scanning approach to
86 interrogate individual cleavage sites. Interestingly, the residues that affect cleavage activity are scattered
87 across a region over 40 nucleotides long. These results suggest that the highly specific RNase Y cleavage
88 does not follow the conventional model of protein-nucleic acid interactions mediated by compact
89 sequence or structural motifs. Instead, the "cleavage code" is specified by a collection of weak
90 interactions embedded in an extended region.
91
92 **RESULTS**
93 **Stabilized end-sequencing reveals positions of endoribonucleolytic activity**
94 In the absence of additional nucleolytic activity, one can map the positions of cleavage by an
95 endoribonuclease by simply observing the termini of its cleavage products. *In vivo*, however,
96 exonucleases typically rapidly act on endonuclease cleavage products, rendering the original RNA
97 endonuclease cleavage sites difficult to detect. If all exoribonucleolytic activity targeting either 3′ or 5′
98 ends is ablated, endoribonuclease cleavage at a particular site should result in stably accumulating
99 products whose ends correspond to the position of cleavage (Figure 1A). Strains combining knockouts of
100 all known exonucleases in either the 5′ or 3′ direction are viable in *B. subtilis*, meaning such a
101 measurement is theoretically possible in this species (Figaro et al., 2013; Oussenko et al., 2005). If RNA
102 ends are separately mapped in both a knockout of RNase J1, the only known 5′-3′ exonuclease, and a
103 knockout of all four known 3′-5′ exonucleases (a knockout of *yhaM, rph, pnpA,* and *rnr*, hereafter referred

104 to as a 4-exo knockout), exact positions of endoribonucleolytic activity can be captured as pairs of
105 adjacent 3′ and 5′ ends that appear across the two strains.

106 Based on this principle, we implemented a "stabilized end-sequencing" approach to create a
107 transcriptome-wide map of endoribonuclease activity by capturing the 3′ hydroxyl and 5′ monophosphate
108 groups that are generated by endonucleases such as RNases Y and III (the primary endonucleases
109 targeting mRNAs in *B. subtilis*) (Figure 1A; (Bechhofer and Deutscher, 2019; Leroy et al., 2017)). To
110 quantitatively capture 3′ hydroxyl groups, we utilized a previously established method for RNA 3′
111 ligation and deep sequencing, hereafter referred to as 3′ end sequencing (Figure 1A, left; (Dar et al., 2016;
112 Herzel et al., 2022; Mandell et al., 2021)). To complement 3′ end mapping, we specifically captured 5′
113 monophosphates through T4 RNA Ligase 1 treatment, allowing quantification of cleavage-derived 5′ ends
114 by deep sequencing, hereafter referred to as 5′ end sequencing (Figure 1A, right). By performing 3′ end
115 sequencing on a 4-exo knockout (strain CCB396) and 5′ end sequencing on an RNase J1 knockout (strain
116 CCB434), we captured 3′/5′ end pairs derived from endoribonuclease activity transcriptome-wide.
117 Combining these exoribonuclease perturbations with endoribonuclease knockouts, we can attribute these
118 cleavage signatures to specific enzymes. End-enriched RNA sequencing (Rend-seq) can additionally be
119 performed to quantify expression of putative endoribonuclease targets and provide an orthogonal RNA
120 end-mapping approach (Lalanne et al., 2018).

121 This stabilized end-sequencing strategy robustly captures positions of previously documented
122 endoribonucleolytic cleavage sites. Among 13 previously characterized sites cleaved by either RNase Y,
123 RNase III, or Rae1 that were individually detected (Braun et al., 2017; Bruscella et al., 2011; DiChiara et
124 al., 2016; Durand et al., 2015; Leroy et al., 2017; Noone et al., 2014; Yao and Bechhofer, 2010), we
125 identified at least one clear 3′/5′ pair at or near the annotated cleavage position (Figure 1B, Figure S1). 5′
126 end sequencing of RNase Y+J1 or III+J1 double mutant strains (CCB760 and BG879) confirmed that we
127 can detect cleavage by both endoribonucleases and accurately assign the enzymes responsible for
128 cleavage (Figure 1B).

129 Importantly, the paired end-mapping strategy provides clear advantages over measurements of a
130 single end type. Because we capture both RNA ends generated by endoribonuclease activity, multiple
131 adjacent positions of cleavage can be resolved, as appears to be the case for the first characterized
132 substrate of Rae1, *S1025* (Figure 1B, bottom). Without these two paired channels of information, it is
133 difficult to determine whether signal at adjacent positions is derived from multiple positions of
134 endoribonucleolytic cleavage or residual exoribonuclease activity by unknown enzymes, the latter of
135 which may be occurring after endoribonucleolytic cleavage of the RoxS RNA (Figure 1B, middle).
136 Precisely resolving cleavage positions becomes particularly relevant in cases such as the
137 endoribonucleolytic cleavages observed in *hbs*, where many nearby positions are known to be subject to
138 RNase Y cleavage (Figure S1, (Braun et al., 2017)).
139 Using these data, we first identified putative positions of cleavage within transcripts whose
140 abundance is known to be regulated by RNA endonuclease activity. Recent work has shown that the RNA
141 decay machinery in *B. subtilis* is subject to extensive autoregulation and cross-regulation, suggesting that
142 cleavage sites should be detectable within the RNAs encoding decay-associated proteins (DeLoughery et
143 al., 2018; DiChiara et al., 2016; Korobeinikova et al., 2023). This expectation is borne out in our
144 stabilized end-sequencing, where multiple putative positions of endoribonucleolytic cleavage are detected
145 within the transcripts encoding RNases Y, III, and the Y-complex, a complex of three small proteins
146 (YlbF, YmcA, and YaaT) that modulates RNase Y activity (Deloughery et al., 2016) (Figure 2, Figure
147 S2). Further, a number of additional transcripts, including that encoding sporulation master regulator
148 SinR, have been experimentally characterized as stabilized upon deletion of RNase Y (Deloughery et al.,
149 2016; Lehnik-Habrink et al., 2011a). Paired peaks are clearly visible within each of these transcripts (Fig
150 2, Fig S2). Taken together, our ability to robustly identify known and previously unknown positions of
151 cleavage suggests that stabilized end-sequencing provide a high-resolution, genome-wide view of RNA
152 endonuclease activity in *B. subtilis*.
153
154 **Systematic identification of positions of endoribonuclease cleavage**

155 To systematically identify putative positions of endonuclease cleavage, we implemented a computational
156 approach to call abutting pairs of peaks in 3′ and 5′ end sequencing across our 4-exo and RNase J1
157 knockout data, respectively (Figure S3A). To do so, we assessed the signal at each position relative to the
158 average in a local window, calling positions exceeding a signal-to-background threshold as peaks.
159 Positions for which a peak is called in both 5′ and 3′ end sequencing with a 1-nucleotide offset (a 5′ peak
160 downstream of a 3′ peak) are called as peak pairs and represent putative sites of endoribonucleolytic
161 activity. This approach reveals 1798 putative positions of cleavage in mRNAs. It should be noted that we
162 are unable to identify cleavage events proximal to the ends of transcription units, because short RNA
163 decay fragments (<35 nt) are excluded during 5′ end sequencing library preparation (see Methods).
164 To assign putative cleavage positions to the activity of specific endoribonucleases, we identified
165 the positions that have diminished end-signals when an endoribonuclease is knocked out (Fig S3B). To
166 account for changes in expression levels in strains with and without the endoribonuclease, we first
167 normalized end-sequencing signals to local RNA abundance measured by Rend-seq in the same strain.
168 We then quantified the sensitivity of the expression-normalized signal to endoribonuclease knockout, by
169 computing the ratio of 5′-end signals (in the absence of RNase J1) between strains with and without the
170 endoribonuclease (Materials and Methods). This "sensitivity" score will be elevated for positions cleaved
171 by the corresponding endoribonuclease, as was previously shown by DiChiara et al. (DiChiara et al.,
172 2016).
173
174 ***B. subtilis* RNase III cleaves within long-range intramolecular structures**
175 We first characterized the cleavage profile of RNase III, motivated by previous reports of unexpected
176 cleavage patterns in *B. subtilis* (DiChiara et al., 2016). RNase III plays well-characterized roles in the
177 processing of noncoding RNAs (DiChiara et al., 2016; Herskovitz and Bechhofer, 2000; Oguro et al.,
178 1998) and the regulation of mRNAs (Durand et al., 2012b; Jahn and Brantl, 2013; Meißner et al., 2016).
179 Several potential mRNA targets of this enzyme have previously been proposed, but only a small number
180 of cleavage sites in these mRNAs have been identified (DiChiara et al., 2016; Gordon et al., 2017).

181 Intriguingly, such cleavage sites often did not appear as pairs at a defined 2-nt spacing within RNA
182 duplexes, contrasting with the hallmark of RNase III activity in *E. coli* (Altuvia et al., 2018). We reasoned
183 that a more systematic, high-resolution mapping of RNase III cleavage sites in *B. subtilis* will help resolve
184 this issue.

185 Our stabilized end-sequencing data revealed 68 RNase III cleavage sites located on 42 mRNAs
186 (Table S3). These sites constitute a small fraction of the putative endonuclease cleavage sites we
187 identified (5%) in the background strain (phage-cured (Gilet et al., 2015)), consistent with the notion that
188 RNase III targets a small subset of mRNAs (Fig. 3B). Despite the sparseness of these sites, they are often
189 present in the same mRNAs as other RNase III sites (45 of 68) (Figure 3A, Figure S1, and Table S3).
190 These results suggest that the *B. subtilis* RNase III acts in a manner similar to the *E. coli* RNase III in
191 cutting a pair of positions in RNA duplexes.

192 Indeed, of the 44 of these co-occurring cleavage sites that occur within 1 kb of each other, a large
193 majority (36 of 44) are predicted to reside in RNA secondary structures. The extent of secondary
194 structures is substantial, with up-to 505 nt between cleavage sites (Figure 3D). Despite the large contour
195 length, the cleavage positions are almost always (32 of 36) separated by 2 base pairs on the opposite sides
196 of RNA duplexes, which range in length from 10 to 36 basepairs (Figure 3C E, and F). 4 of the 36 co-
197 occurring sites are cleaved at 2 adjacent positions, which may still result from the 2-bp rule on individual
198 RNA molecules but are difficult to resolve in our ensemble measurement. It is possible that the minority
199 of sites that do not form a clear intra-molecular duplex (32 of 68) participate in inter-molecular pairing,
200 but the presence of such interactions remains to be determined.

201 We next sought to determine what sequence features within these specific secondary structures
202 might be recognized by RNase III. To increase the number of RNase III sites used in this analysis, we
203 included putative cleavage positions with a lower threshold on 5′ end sequencing depth (see Methods for
204 details, Tables S5,6 for a list of putative cleavage positions with this relaxed threshold, and Figures S3C,E
205 for endonuclease sensitivity distributions), restricting our analysis only to pairs of such putative cleavage
206 positions that were predicted to pair with 2-bp spacing. Using this approach, we identified an additional

207 10 pairs of RNase III cleavage sites for a total of 26 pairs. The genomic distance between these sites is
208 similar to those initially identified (Figure S3D).
209 From these pairs of RNase III cleavage sites, we calculated the position-wise sequence content
210 and along the predicted RNA duplex, as well as the fraction of sites for which each position is paired,
211 using the two cleavage positions as our reference points (Figure 3G). Through this, we generated a *B.*
212 *subtilis* RNase III sequence-structure motif as has previously been reported for *E. coli* (Altuvia et al.,
213 2018, Figure 3H). Like the *E. coli* homolog, the positions surrounding *B. subtilis* RNase III cleavage sites
214 appear to be consistently paired, with the probability of pairing increasing towards the cleavage positions
215 (reaching >90% at the -2/+2 positions) (Figure 3G). Similarly, a number of sequence features appear to be
216 recognized similarly by these enzymes, including position-specific preferences for A-U and G-C base-
217 pairing within the proximal box (PB) and an increased occurrence of U-G wobble pairing within the distal
218 box (DB) (Altuvia et al., 2018; Pertzev and Nicholson, 2006). At each position from -6 to 6, at least one
219 of the maximum-frequency base pairs is conserved between these species (Figure 3H, green text). In
220 contrast to *E. coli*, however, the base pairs between cleavage positions do not demonstrate a clear
221 preference for C-G/G-C pairing in our data (Altuvia et al., 2018) (Figure 3H, red box).
222 Together, these findings suggest that the *B. subtilis* RNase III displays similar substrate
223 preferences to its evolutionarily distant *E. coli* homolog (Altuvia et al., 2018; Pertzev and Nicholson,
224 2006). Furthermore, the presence of extensive long-range RNA secondary structure for RNase III targets
225 suggests that mRNA base pairing across hundreds of nucleotides is common in this organism and can
226 directly influence gene expression regulation (Braun et al., 2017; Ruiz de los Mozos et al., 2013).
227
228 **Induction of EndoA activity in exoribonuclease knockouts reveals additional 5′ end trimming**
229 **activity in *B. subtilis***
230 To our surprise, manual inspection of our Rend-seq data in all strains with an RNase J1 or 4-exo knockout
231 showed prominent peaks across the transcriptome within the motif U/ACAUA (/ denotes the cleavage
232 position, Figure 4A, B). This sequence resembles an extended version of the canonical U/ACAU cleavage

233 motif of the endonuclease EndoA, an RNA interferase of the MazF family associated with the EndoA-
234 EndoAI type II toxin-antitoxin system (Park et al., 2011; Pellegrini et al., 2005). The presence of adjacent
235 3′ and 5′ ends in Rend-seq data separating the first U/A dinucleotide within this motif strongly suggests
236 endoribonuclease activity (Figure 4A, B), possibly through activation of this toxin in the exoribonuclease
237 deficient strains used in this study. This U/ACAUA-specific cleavage is not observed in Rend-seq data
238 from wild-type *B. subtilis* (Figure S4A). Consistent with EndoA-mediated cleavage, which leaves a 5′
239 hydroxyl group (Bechhofer and Deutscher, 2019; Pellegrini et al., 2005), we did not observe
240 corresponding signals in our 5′ end sequencing data, which is specific to 5′ monophosphate groups,
241 shown for a representative site in the *pgk* gene (Figure 4E). Because Rend-seq relies on cDNA
242 circularization rather than RNA ligation to capture RNA 5′ ends, it is agnostic to the 5′ moiety generated
243 following cleavage (Lalanne et al., 2018). To further confirm that these cleavage sites are driven by
244 EndoA, which is encoded by the *ndoA* gene, we repeated these measurements upon deletion of *ndoA* in an
245 RNase J1 deficient strain and found complete ablation of the cleavage signature in Rend-seq (Figure 4C,
246 D). Furthermore, consistent with recent work indicating that EndoA recognizes a 6-mer motif and not a
247 previously described 5-mer motif (Guler et al., 2024), we found that only U/ACAU followed by an
248 adenosine leads to cleavage (Figure 4A, B).
249 Though our monophosphate-specific 5′ end sequencing approach did not detect the hydroxylated
250 5′ ends generated by EndoA, it unexpectedly captured peaks 2 nucleotides downstream of the canonical
251 EndoA cleavage positions (Figure 4E). These 5′ monophosphorylated ends only appear when EndoA is
252 present (Figure 4D, E), suggesting that two nucleotides are removed following cleavage by EndoA. To
253 test if this trimming activity is general against all 5′ hydroxyl groups, not simply EndoA sites, we looked
254 at the 5′ OH generated by the self-cleaving *glmS* ribozyme, whose downstream cleavage product is
255 subject to subsequent degradation by RNase J1 (Collins et al., 2007; Winkler et al., 2004). In both a
256 knockout of RNase J1 and a double knockout of RNase J1 and EndoA, 5′ end sequencing shows a clear 5′
257 monophosphorylated peak 2 nucleotides downstream of the site of self-cleavage in *glmS*, suggesting that
258 this novel trimming activity is independent of EndoA (Figure 4F). The trimming activity is also

259 independent of RNases Y, III, J1, the Y-complex constituent YlbF, or the RNA pyrophosphohydrolase
260 RppH, as downstream peaks are readily detectable in strains lacking these enzymes (Figure 4G). Finally,
261 this activity cannot be explained by RNase J2, as 5′-mapped Rend-seq measurements of a double
262 knockout of RNase J1 and J2 also show a clear peak two nucleotides downstream of the canonical *glmS*
263 cleavage position (Figure 4H).

264 We observed multiple rounds of 2-nt trimming at some, but not all, EndoA cleavage sites,
265 suggesting sequence-specific trimming activities (Figure S4B,C). To determine the sequence specificity,
266 we examined the downstream nucleotide content following the U/ACAUA EndoA motif (Figure 4I),
267 focusing on our *rny* knockout and RNase J1 depletion strain. This strain was chosen for its particularly
268 clear EndoA activity, which is likely driven by further stabilization of EndoA cleavage products
269 following inactivation of an additional downstream decay pathway, though EndoA activity may also be
270 increased. For sites with an EndoA motif followed by a U or C, monophosphorylated 5′ ends are found at
271 both 2 and 4 nucleotides downstream of the initial cleavage, indicating discrete trimming (Figure 4I). For
272 U/ACAUA motifs followed by an A (particularly by AA), additional 5′ ends are also generated five or
273 more nucleotides downstream of the initial cleavage. By contrast, motifs followed by a G only produce 5′
274 ends at 2 nucleotides downstream. These 5′ ends are not driven by changes to the EndoA cleavage
275 position, as the 3′ end density in the 4-exo knockout remains constant regardless of the downstream
276 sequence context (Figure 4J). These data are consistent with successive rounds of nucleotide removal
277 from the RNA 5′ end following endoribonucleolytic cleavage, with a preference for trimming sequences
278 rich in adenosine and against those rich in guanosine. We are unable to rule out that differential stability
279 following EndoA cleavage and 5′ end trimming may also influence the apparent sequence preferences we
280 observe, however, and further work is required to reveal the identity and mode of action of this enzyme.

281 Cleavage 2 or 4 nucleotides downstream of the canonical EndoA cleavage site coincides with the
282 U/A and C/A dinucleotides that are recognized by the common laboratory contaminant RNase A. Several
283 lines of evidence argue against such a contaminant generating the trimming behavior we characterize in
284 this work. First, RNase A cleavage generates RNA termini with 5′ hydroxyl or 3′ phosphate groups,

285 similar to EndoA (Cuchillo et al., 2011). These ends cannot be captured by our RNA 5′ and 3′ end
286 sequencing approaches (Figure 4E). Second, the 5′ trimming activity observed at *glmS* occurs within a
287 C/G dinucleotide, suggesting that this activity is not restricted to U/A and C/A dinucleotides. Finally, our
288 observation of 5′ end trimming is highly consistent across many replicate measurements of strains lacking
289 RNase J1, as is an observed lack of trimming when this enzyme is present. It is thus highly unlikely that
290 this trimming is driven by *in vitro* RNase activity downstream of cell lysis and RNA extraction.
291
292 **Determinants of cleavage by RNase Y**
293 Though RNase Y is thought to be the primary endonuclease driving RNA decay in *B. subtilis*, only a
294 small number of cleavage positions have been mapped in this bacterium, and what specifies RNase Y
295 cleavage is poorly understood (Trinquier et al., 2020). To define the substrate features driving RNase Y
296 cleavage, we sought to identify a set of putative cleavage sites that are dependent on the presence of this
297 enzyme. For the 1330 putative endonuclease cleavage positions identified in our W168 background, we
298 calculated their RNase Y sensitivity using RNase J1 deficient cells with or without an RNase Y knockout.
299 This metric was bimodally distributed, and any cleavage signature with a sensitivity value over 10.56 was
300 called as RNase Y dependent based on a double-Gaussian fit of these data (Figure 5A). Using this
301 approach, we identified 596 cleavage sites across 387 genes that can confidently assigned to RNase Y
302 sites (Table S4).
303 This set of putative RNase Y sites sheds light onto the substrate features that guide RNase Y
304 activity. Frequency and information content logos constructed from the RNA sequences surrounding
305 putative RNase Y sites suggest a sequence preference for cleavage downstream of a purine, with a
306 particular preference for guanosine, and upstream of an adenosine (Figure 5B, top, middle). Similarly,
307 considering k-mers of length ≤4, there is most significant enrichment for the dinucleotide sequence GA
308 spanning the cleavage site, with an additional preference for uridine in the -2 and -3 positions (Figure 5B,
309 bottom). This observed specificity is unlikely to be a result of biases in our methodology; identification of
310 these cleavage sites requires concordance between 3′ and 5′ end sequencing datasets that are generated

311 through distinct enzymatic reactions and adapter sequences. This sequence preference is additionally
312 consistent with the sequence context in several previously mapped cleavage sites (Braun et al., 2017;
313 Shahbabian et al., 2009), as well as the measured sequence preferences of RNase Y in *S. aureus* and *S.*
314 *pyogenes*, despite the more limited role this enzyme plays in those organisms (Broglia et al., 2020;
315 Khemici et al., 2015). However, the GA dinucleotide motif does not appear to be absolutely required, as
316 63% of the putative cleavage sites do not have this motif.

317 In a broader window around cleavage sites, there is a clear depletion for GC-containing sequence
318 (Figure 5C) and a lower propensity to form secondary structures (Figure 5D). Computationally predicted
319 folding of 40-nt RNAs surrounding positions of cleavage shows weaker structures for RNase Y sites
320 compared to randomly sampled positions at uncleaved GA dinucleotides (Figure 5D). Interestingly,
321 sequences centered just downstream of this lowly-structured region appear to fold more strongly than
322 background, suggesting an enrichment for secondary structure proximal to the position of cleavage
323 (Figure 5D). The downstream signal is recapitulated in the GC content within this window (Figure 5C),
324 and is consistent with the influence of secondary structure on previously detected RNase Y cleavage
325 events (Marincola and Wolz, 2017; Shahbabian et al., 2009). Taken together, these data indicate that *B.*
326 *subtilis* RNase Y targets share a weak motif comprising a GA dinucleotide surrounded by AU-rich
327 sequences that have reduced secondary structure.

328

329 **The complete Y-complex is not a strict requirement for RNase Y activity**

330 Recent work has shown that the Y-complex, comprising proteins YlbF, YaaT, and YmcA, can act as a
331 specificity factor for RNase Y for the maturation of nearly two-dozen operon mRNAs (DeLoughery et al.,
332 2018). To understand if this complex plays a role in specifying the observed RNase Y cleavage events,
333 we deleted the *ylbF* gene in both an RNase J1 knockout (CCB434) and 4-exo deficient background
334 (CCB396). We then repeated our Rend-seq and end-sequencing measurements in these backgrounds,
335 calculating the sensitivity of each putative cleavage site to YlbF (Figure 5E). Compared to RNase Y, few
336 sites appeared dependent on the presence of YlbF, suggesting that the Y-complex plays a limited role in

337 influencing RNA turnover outside of its known role in mRNA processing. Consistent with this
338 observation, knocking out *ylbF* has a limited effect on the *B. subtilis* transcriptome (Figure S4D), as has
339 been previously observed (DeLoughery et al., 2018).

340 Surprisingly, canonical maturation events, such as those within the *cggR-gapA*, *glnRA*, and *atp*
341 operons are still observed in the absence of YlbF, albeit strongly attenuated (Figure 5E-H). Previously,
342 the YlbF-dependence of these sites were determined using Rend-seq for cells that maintain the entire
343 repertoire of exonucleases, which is less sensitive at detecting cleavage events. Here, by ablating
344 exonucleases in cells lacking YlbF, we observed weak but detectable signals of cleavage products using
345 Rend-seq, as well as unambiguous peaks in both 5′ and 3′ end sequencing (Figure 5F-K). These results
346 underscore the sensitivity of our approach and suggest a revision of the role of the Y-complex. That is,
347 instead of being a strictly necessary specificity factor for RNase Y cleavage, the Y-complex plays an
348 enhancing role at these operon maturation events.

349

350 **Deep mutational scanning of RNase Y substrates**

351 Our census of RNase Y cleavage events showed a degenerate motif, which is insufficient to explain its
352 highly precise site selection for previously established mRNA targets. To identify additional sequence or
353 structural features required, we developed a massively parallel reporter assay (MPRA) that allows high-
354 throughput measurement of cleavage efficiency among a library of 10^4-10^5 mutated cleavage sites (Figure
355 6, Table S7). The reporter consists of a >100-nt region surrounding a cleavage site inserted into a stable
356 native mRNA encoding the gene *aprE*, truncated such that it lacks an active site (Figure 6A). These
357 reporter constructs are integrated at a neutral genomic locus (*ganA*) and contain a randomized 15-
358 nucleotide barcode associated with each mutant variant, allowing relative quantification of each variant's
359 expression by RNA and DNA sequencing.

360 To simultaneously quantify cleavage activity at each mutant cleavage site, a pooled library of *B.*
361 *subtilis* cells expressing our MPRA construct were grown and harvested in exponential phase for genomic
362 DNA (gDNA) and RNA extraction (Figure 6B). Purified RNA is then subjected to reverse transcription

363 and a size selection that separates cleaved from full-length RNA, allowing specific detection of each
364 isoform. A reduction in cleavage should result in greater accumulation of the full-length RNA, with
365 concomitant decrease in cleaved RNA. The barcode abundance in the uncleaved RNA fraction for each
366 variant, normalized by its frequency in the pool as measured by gDNA sequencing, thus provides a
367 readout of cleavage efficiency for that mutated sequence (Figure 6B).
368
369 **RNase Y cleavage is specified by distributed substrate features**
370 We initially focused our MPRA measurements on a cleavage site found within the *cggR* gene, which in its
371 native context drives differential expression within the glycolytic operon by producing a stable
372 downstream product following cleavage (Ludwig et al., 2001; Meinken et al., 2003) (Figure 5F, G). The
373 cleaved downstream product, which includes the *gapA* gene, is thought to be stabilized by an RNA
374 hairpin structure at the cleaved end (henceforth referred to as the downstream hairpin) (Meinken et al.,
375 2003), but whether proximal structure also plays a major role in specifying the cleavage site by *B. subtilis*
376 RNase Y remains unknown. To capture surrounding sequences that may contribute to this cleavage, we
377 inserted 148 bp of this sequence into our reporter, approximately centering the cleavage position within
378 this window (Figure 6C). The site of cleavage, which in its native context resides in the *cggR* open
379 reading frame, was similarly translated in the same reading frame of the reporter. Also adhering to the
380 native context, the site of cleavage was followed by a stop codon and a short open reading frame derived
381 from the start of the downstream *gapA* gene (Figure 6C). Using this approach, we characterized the effect
382 of mutations from positions -30 to +36 relative to the cleavage position. After sequencing of the gDNA
383 and RNA, 91,580 uniquely barcoded variants were detected across all experiments. These data cover
384 virtually all possible single-nucleotide changes (Figure 6D) and 32% and 44% of all double-nucleotide
385 changes across experiments detecting full-length and cleaved RNAs, respectively (Figure S5A, B). For
386 each mutant, we average the effects over different barcode variants (median 44 barcodes for single-
387 nucleotide changes) to reduce noise and barcode-driven variations (Figure S5C, D).

388 Mutations across a wide region surrounding the *cggR* cleavage site showed modest but detectable
389 impact on steady-state levels of the full-length mRNA (Figure 6D). Between -30 and +1 nucleotide
390 relative to the cleavage site, only a small number of mutations appear to reduce cleavage efficiency. The
391 strongest effects are seen at the cytosine at position -11 (denoted as C[-11]), and the effect sizes are at
392 most twofold. Interestingly, mutations of G at position -1 showed no statistically significant effect, even
393 though it is the strongest consensus among RNase Y cleavage sites across the transcriptome. The fact that
394 the G[-1] is not strictly required is consistent with our finding that 37% of the RNase Y cleavage sites
395 detected do not have a guanine at this position (Figure 5B). Therefore, mutations upstream of the *cggR*
396 cleavage site lend further support to our transcriptome-wide results that showed a weak consensus.
397 Downstream of the cleavage site, mutations in the hairpin region also produced limited effects on
398 full-length RNA levels. Single-nucleotide substitutions in this region did not have significant effects
399 (Figure 6D). Combinations of substitutions which further destabilize the hairpin also did not affect the
400 apparent cleavage efficiency (Figure 6H). Surprisingly, the only mutations in this structured region that
401 reduce cleavage are specific substitutions in a bulge of the downstream hairpin, including an A[+6]G
402 substitution (fourfold) and a C[+24]U substitution (twofold) (Figure 6D). These two single-substitution
403 mutants close the bulge and complete the hairpin, and other bulge-closing double substitutions at these
404 two locations also showed reduced cleavage. Together, these results suggest that the downstream hairpin
405 is not essential for cleavage, and that a stronger hairpin can in fact negatively impact cleavage.
406 Immediately downstream of the hairpin, an AU-rich region appears to play a role in cleavage.
407 Mutations to AUAAAUAU located at position +28 to +35 lead to increases in full-length RNA level by
408 up to 4-fold. In the folded RNA, this distal AU-rich region is brought to be adjacent to the cleavage site
409 by the intervening hairpin (Figure 6F). Although it is plausible that such spatial proximity enables the
410 distal region to recruit RNase Y or facilitate its activity, the fact that mutations in the hairpin region have
411 no detectable effects suggests that other mechanisms may be at play.
412 To further elucidate the role of the downstream hairpin, we tested the hypothesis that this
413 secondary structure mainly serves as a stabilizer for the cleaved product that encodes *gapA,* by preventing

414 exoribonucleolytic decay (Meinken et al., 2003). This role is distinct from the model for the
415 *Staphylococcus aureus cggR-gapA* operon, where the downstream hairpin has been suggested to specify
416 RNase Y cleavage sites (Marincola and Wolz, 2017; Le Scornet et al., 2024). To distinguish these roles,
417 we conducted deep mutational scanning on the accumulation of processed mRNA, measuring 349,266
418 mutant variants in total, and compared it to the results on full-length mRNA levels. Unlike full-length
419 mRNA, the level of the downstream processed mRNA should depend on both cleavage efficiency and its
420 own half-life post-cleavage.

421 Indeed, most mutations observed to increase full-length mRNA level resulted in reduced
422 accumulation of the cleavage product, but not vice versa (Figure 6E). The effects of C[-11] mutations,
423 bulge-closing mutation, and distal AU-rich region are recapitulated when we detected the downstream
424 cleavage product. However, in contrast to the full-length RNA, mutations in the downstream hairpin
425 region were among the most deleterious to cleavage product accumulation. Consistent with the secondary
426 structure playing a role, double-substitution variants that complement these mutations (Figure 6F) have
427 dramatically reduced effect sizes, and so are stem mutations that maintain base pairing (through wobble
428 interactions, Figure 6E, red arrows). Considering all variants, including those with two or more
429 substitutions, we found a clear trend that a stronger downstream hairpin is predictive of the cleaved RNA
430 level (Figure 6G), in stark contrast with a lack of correlation for full-length RNA (Figure 6H). Finally, we
431 repeated our measurements of cleaved RNA abundance in the context of an RNase J1 depletion, and
432 found that, with the exception of closing the bulge, substitutions within the downstream stem no longer
433 have a measurable impact when exoribonucleolytic activity against the processed RNA is attenuated
434 (Figure S6A). Together, these results suggest that the downstream hairpin plays a much more substantial
435 role dictating cleavage product abundance than the efficiency of RNase Y cleavage.

436 To assess whether the observed role of proximal secondary structure is specific to the *cggR*
437 operon, we repeated the described measurements and analysis for a second cleavage site, located within
438 the *glnR* gene of the *glnRA* operon (Figure S7A). As before, this cleavage occurs after the first gene in an
439 operon and a stem-loop structure is predicted to form at the 5′ end of a stabilized downstream cleavage

440 product (Figure S7B). Measuring accumulation of the cleaved RNA, we captured 54,431 variants with
441 mutations occurring from positions -5 to +31 relative to the position of cleavage. For nucleotides inside
442 the stem, both single mutations (Figure S7C) and higher-order substitutions that disrupt this RNA folding
443 (Figure S7D) attenuated the accumulation of the cleavage product (up to approximately 4-fold),
444 recapitulating the pattern seen for *cggR*. When we quantified the accumulation of full-length RNA
445 (Figure S7E) or cleaved RNA upon RNase J1 depletion (Figure S7F), we found that disruption of the
446 stem did not appear to stabilize the full-length RNA. Together, these results suggest that the behavior seen
447 for *cggR* may be shared with other similarly structured mRNA processing sites.
448 Finally, we placed the *cggR* cleavage site into a second reporter construct encoding the
449 exogenous gene *tetM* to assess whether the features driving cleavage product accumulation are context-
450 dependent. Although most key effects are observed in both reporter constructs, the *tetM* reporter showed
451 deleterious substitutions in C[-19]G and, to a lesser degree, A[-6]G (Figure S6B). The nucleotide C[-19]
452 is situated in an otherwise purine-rich region, AGA **C** GAAGGAG, inside the *cggR* open reading frame.
453 We initially hypothesized that the C[-19]G substitution may affect cleavage by slowing down elongating
454 ribosomes through its Shine-Dalgarno-like sequence. However, the deleterious effects were still observed
455 when we introduced an upstream stop codon, suggesting that these substitutions do not act through
456 translation but other context-specific means (Figure S6C). These results reinforce the notion that the
457 determinants of RNase Y cleavage are not localized to a narrow region on the mRNA.
458
459 **DISCUSSION**
460 In this work, we present the first genome-scale map of *B. subtilis* endoribonuclease cleavages at single-
461 nucleotide resolution. Using stabilized end-sequencing, we identified over a thousand putative cleavage
462 events, allowing us to learn substrate preferences of the central endonucleases involved in *B. subtilis*
463 mRNA decay. This work revealed the importance of long-range RNA base pairing for RNase III cleavage
464 and confirmed that the toxin EndoA recognizes a hexameric rather than pentameric motif *in vivo*. We
465 additionally uncovered evidence for and potential specificity of a new ribonucleolytic activity previously

466 unknown in *B. subtilis*, motivating further work to understand RNA 5′ end processing and decay. Finally,
467 we identified hundreds of new positions of cleavage by the core endonuclease driving mRNA decay,
468 RNase Y. Combining this global cleavage mapping with a new massively parallel reporter assay, we
469 demonstrate that cleavage by this enzyme is specified by features distributed over many positions across
470 regions tens of nucleotides in length. This stands in stark contrast to enzymes like EndoA, which
471 recognizes a strict consensus cleavage sequence, and RNase III, which has a strongly stereotyped
472 cleavage behavior within secondary structures.

473 The broad suite of features that contribute to RNase Y processing at the *cggR* site helps explain
474 the highly degenerate motif that we observed through global cleavage site mapping. Our data support a
475 model in which RNase Y cleavage is localized not just by the sequence proximal to the cleaved bond, but
476 rather by combined contributions of features distributed as many as tens of nucleotides away from the
477 cleavage site. The ability for features distal in primary sequence to strongly influence cleavage may be
478 mediated by RNA folding or by higher-order RNA-protein and protein-protein interactions between these
479 features and components of the bacterial RNA degradosome. As such, the particular positions recognized
480 to specify cleavage at each site may be variable and highly context-specific. Interestingly, a common
481 feature we identified between many RNase Y sites, the presence of a -1G, only have a modest effect on
482 the accumulation of processed *cggR* RNA in our reporter system, if any. This discrepancy may suggest a
483 similar model to what is observed in in *S. aureus,* where an upstream guanosine fine-tunes the position of
484 cleavage but is not required for RNase Y activity (Le Scornet et al., 2024).

485 The absence of a clear consensus motif specifying RNase Y activity echoes previous findings for
486 RNase E in *E. coli* (Chao et al., 2017; Lin-Chao et al., 1994; McDowall et al., 1994). These observations
487 suggest that the code for RNA decay fundamentally differs from transcription and translation, whose rates
488 are influenced by well-defined interactions between the gene expression machinery and localized
489 sequence features within either promoters or ribosome binding sites (Taggart et al., 2021). Instead, decay-
490 initiating RNases such as RNase Y appear to be strongly influenced by features distal to the position of
491 cleavage, possibly facilitated by RNA secondary structure and RNA-protein interactions. mRNA stability

492 may also be strongly influenced by features extrinsic to primary RNA sequence. Decay can occur
493 concurrently with and may be influenced by transcription and translation (Condon, 2003; Yarchuk et al.,
494 1992), and the nonuniform distribution of decay machinery in bacterial cells also means that protein
495 binding and RNA localization may have a strong influence on RNA stability (Hamouche et al., 2020;
496 Lehnik-Habrink et al., 2011b; Moffitt et al., 2016). Robust prediction of endoribonuclease cleavage from
497 primary sequence will thus likely require integration of more modes of measurement and interrogation of
498 more cleavage sites than our current end-mapping approach and MPRA allow.

499 Although the presented global map of endonuclease activity expands our list of known
500 endoribonuclease cleavage positions by orders of magnitude, we do not expect it to be comprehensive.
501 First, due to size-selection steps in the end-sequencing protocol, we are unable to detect cleavage sites
502 within 15 nucleotides of a 5′ end or 40 nucleotides of a 3′ end of a transcription unit. Additionally, while
503 all known exoribonuclease activity in one direction is ablated in each of our experiments, secondary
504 endoribonuclease cleavage followed by activity of the unperturbed exoribonucleases acting in the
505 direction opposite may destabilize cleavage products, rendering them undetectable. Finally, activity by
506 unknown exoribonucleases may cause broadening or shifting of 3′ or 5′ ends, separating the peak pairs
507 and obscuring potential cleavage positions (as seen in Figure 4). The absence of signal in these data
508 therefore is not necessarily an indication that turnover of a particular RNA is independent of
509 endoribonucleolytic cleavage.

510 Unexpectedly, our analysis of EndoA cleavage sites, as well as the *glmS* ribozyme, revealed a
511 previously uncharacterized 5′ end trimming activity in *B. subtilis.* We have confirmed that this activity is
512 not dependent on the primary enzymes targeting mRNAs for decay, including RNases Y, III, J1, J2,
513 RppH, or EndoA. Whether this nuclease acts in a 5′-3′ exonucleolytic manner or through
514 endoribonucleolytic cleavage near 5′ ends remains to be determined. Further work is additionally
515 necessary to understand whether this enzyme acts to facilitate RNase J1 cleavage at substrates with a 5′
516 hydroxyl moiety, as such behavior would shape our understanding of how systems like the *glmS*

517 ribozyme regulate gene expression. Determining what gene encodes this activity is thus of great interest
518 moving forward.

519 Our finding that the full Y-complex is not strictly required for processing of *B. subtilis* operons
520 suggests a refinement of our understanding of this complex in RNA maturation and decay. Although our
521 data support that loss of a component of the Y-complex strongly attenuates cleavage at the operons it
522 canonically regulates, persistence of weak cleavage signatures suggests that the presence of YlbF is not a
523 strict requirement for RNase Y activity at these sites. This residual cleavage may represent dispensability
524 of the Y-complex as a whole or dispensability of YlbF itself for Y-complex activity. Indeed, recent work
525 by Dubnau *et al.* has suggested that, of the Y-complex members, only YaaT (RicT) stably associates with
526 RNase Y to form the functional complex required for operon maturation (Dubnau et al., 2023). It remains
527 to be tested whether deletion of YaaT eliminates the low-level cleavage activity we detect in the absence
528 of YlbF.

529 Mutational scanning of the *cggR* mRNA processing site revealed that downstream secondary
530 structure primarily influences *cggR* operon stoichiometry through stabilization of the product of cleavage
531 rather than driving cleavage itself. This result, while consistent with early studies of this operon in *B.*
532 *subtilis* (Meinken et al., 2003), may appear to contradict studies in *S. aureus* that suggest presence of a
533 downstream hairpin is a key driver of RNase Y cleavage (Marincola and Wolz, 2017; Le Scornet et al.,
534 2024). It is worth noting, however, that *S. aureus* RNase Y targets a more limited set of RNAs than its *B.*
535 *subtilis* homolog and regulation of cleavage may differ between the two enzymes (Khemici et al., 2015;
536 Marincola et al., 2012). Additionally, the single base substitutions investigated here are a milder
537 perturbation than used in other studies of the *cggR* operon (Le Scornet et al., 2024), and when secondary
538 structure is strongly ablated in our experiments, a mild stabilization of the full-length construct is
539 observed (Figure 6H). It is thus possible that although the effect of secondary structure on cleavage
540 product stability is more pronounced, downstream RNA structure is playing a dual role in this system.

541 The approach described in this work has the potential to be generalized to study RNA decay in
542 other bacteria for which exonuclease activity can be ablated or transiently depleted, representing a new

543 approach to discover factors influencing RNA decay. Additionally, application of this approach to
544 conditions beyond exponential growth in rich medium will provide an insight into regulation of
545 endonuclease activity and RNA decay more broadly. The high-resolution picture of RNA decay initiation
546 provided by this approach, particularly when integrated with future measurements of RNA half-life, will
547 inform future quantitative predictions of RNA stability and gene expression levels from primary sequence
548 in Gram-positive bacteria.
549

550 **MATERIALS AND METHODS**

551 *Strains and strain construction*

552 For all Rend-seq, 5′ end sequencing, and 3′ end sequencing experiments, excluding the Rend-seq
553 measurement of an *rnjB* deletion with RNase J1 depletion, derivatives of W168 were used (Table S1). All
554 transformations for strain construction were performed using standard protocols relying on natural *B.*
555 *subtilis* competence and flanking homology mediated recombination using the media and protocol
556 described in (Parker et al., 2020). BJT070 was created through SPP1 phage transduction of the *ylbF::erm*
557 allele from BJT034 into CCB396 (Yasbin and Young, 1974). To create BJT074, gDNA from BJT034 was
558 first transformed into SSB1002 to generate a *ylbF* knockout in a W168 background, BJT069. From this,
559 SPP1 phage transduction from CCB434 was used to construct an *rnjA ylbF* double knockout, BJT074.
560 CCB390 was constructed by first transferring an inducible RNase J1 allele (first described in (Britton et
561 al., 2007)) into W168 (strain SSB1002), followed by deletion of *rnjA* through transformation with gDNA
562 from CCB434. To construct BJT231, genomic DNA from GLB230 was transformed into CCB390,
563 supplementing with 2% xylose in the liquid and solid media throughout the transformation protocol. To
564 construct BJT129, CCB390 was transformed with genomic DNA from an *rppH* knockout, strain
565 BKE30630 (Koo et al., 2017) .

566 A supercompetent *B. subtilis* strain BJT200, based on (Zhang and Zhang, 2011), was used for
567 experiments that do not require depletion or induction of RNase J1. This strain was constructed from
568 W168 *B. subtilis* strain SSB1002 through insertion of an anhydrotetracycline (aTc) inducible *comK* allele

569 at the *ganA* locus by transformation of ScaI-linearized plasmid pJT117. For experiments in which an
570 inducible RNase J1 was required, strain BJT201 was utilized. Similar to BJT200, this strain contains an
571 inducible *comK* gene. BJT201 is a derivative of strain CCB390, a W168 *B. subtilis* strain with xylose
572 inducible RNase J1, constructed through transformation with pJT117. These strains were used for all
573 MPRA experiments except the *tetM-cggR* data, which were collected in a *trpC* 168 strain. Plasmid
574 pJT117 and MPRA plasmids pJT028 and pJT124 were constructed by isothermal assembly.
575
576 *Cell growth and harvesting (Rend-seq, 5′-, and 3′-end sequencing)*
577 All Rend-seq, 5′-, and 3′-end sequencing experiments were performed using cells grown in 2xYT,
578 supplemented with 2% xylose or glucose for RNase J1 induction or depletion as necessary. Except when
579 we sought to deplete RNase J1, strains with an inducible RNase J1 allele were consistently grown in the
580 presence of 2% xylose. For all sequencing experiments, overnight cultures in 2xYT were back-diluted to
581 an OD_{600} of 0.0002 in 20 mL prewarmed 2xYT and grown to an OD_{600} of 0.2, at which point they were
582 harvested. For any strain with an RNase J1 knockout, cultures were vortexed prior to and following
583 inoculation to minimize clumping. To achieve RNase J1 depletion, 10 times the required volume of
584 overnight culture was collected, washed twice with 1 mL pre-warmed 2xYT, and resuspended in 1 mL
585 2xYT. 100 µL of this resuspension was used to inoculate the final culture. Samples were harvested for
586 sequencing by adding 7 mL of culture to 7 mL of ice cold methanol, pelleting, decanting, and flash
587 freezing with liquid nitrogen. Cell pellets were stored at -80°C prior to RNA extraction.
588
589 *End-enriched RNA sequencing and 3′ end sequencing*
590 Prior to RNA extraction, with the exception of our first Rend-seq experiment on the exonuclease-
591 deficient strains (profiling wild-type 168, 4-exo, and Δ*rnjA* strains with and without *ylbF* knockout), cell
592 pellets were washed with 10-15 mL ice cold 10 mM Tris pH 7.0 to remove contaminating nucleic acids
593 that precipitate from 2xYT in methanol. For Rend-seq, RNA was extracted from cell pellets using either
594 an RNeasy Mini Kit (Qiagen) with on-column DNase treatment, an RNeasy Plus Mini Kit (Qiagen) or

595 through RNAsnap extraction (Stead et al., 2012) followed by Turbo DNase treatment. All 3′ end
596 sequencing samples were purified through RNAsnap. In the RNAsnap protocol, cells were resuspended in
597 500 µL extraction solution (95% Formamide, 1% BME, 0.025% SDS, 18 mM EDTA) and transferred to a
598 2 mL tube containing 200 µL chilled 0.1 mm diameter zirconia beads. These tubes were vortexed for 10
599 minutes using a horizontal vortex adapter and subsequently incubated at 95 °C for 7 minutes. Samples
600 were spun at 15,000 g, after which 400 µL was removed and diluted 1:5 with 1.6 mL DEPC-treated
601 water. This mixture was precipitated with isopropanol and resuspended in 10 mM Tris, pH 7.0. At least
602 30 µg of RNA was then treated for 30 minutes at 37°C with 10 µL Turbo DNase in a 100 µL reaction.
603 Following RNA extraction and gDNA depletion, 20 µg of purified RNA was treated
604 with MICROBExpress Bacterial mRNA Enrichment Kit (ThermoFisher) per the manufacturer's
605 instructions and precipitated with isopropanol.
606 Rend-seq libraries were prepared as described previously (Lalanne et al., 2018). In short, rRNA
607 depleted RNA was fragmented for 25 seconds (ThermoFisher #AM8740) and fragments of sizes 15-45
608 were size selected on a 15% TBE-Urea polyacrylamide gel (ThermoFisher). RNA was dephosphorylated
609 with T4 PNK (NEB) and 3 pmol of this dephosphorylated RNA was ligated to an adenylated adaptor
610 using truncated T4 RNA ligase 2 K277Q (Jonathan Weissman). Ligated RNA was purified by denaturing
611 polyacrylamide size selection and reverse transcribed using SuperScript III (ThermoFisher) with a custom
612 primer oCJ485. This cDNA was then circularized using CircLigase (Lucigen) and barcodes and full
613 Illumina adapters were added through PCR using Phusion (NEB).
614 3′ end sequencing libraries were prepared in a manner similar to Rend-seq, with deviation in the
615 order of reactions (Herzel et al., 2022). Prior to fragmentation, 1.8 µg of rRNA depleted RNA was ligated
616 to an adenylated adapter as described for Rend-seq. This ligation product was then fragmented and RNA
617 species of sizes 33-63 nt were purified via denaturing polyacrylamide gel extraction. From this point
618 onwards, the library preparation is identical to Rend-seq. Note that this protocol lacks a
619 dephosphorylation step prior to 3′ end ligation, such that RNA ends with a 3′ phosphate will not be

620 captured by this protocol. Both Rend-seq and 5′ end sequencing libraries were sequenced with 50 or 75 bp
621 single end reads on either a HiSeq 2000 or NextSeq500.
622
623 *Multiplexed 5′ end sequencing*
624 For all 5′ end sequencing samples, pellets were washed, and RNA was extracted using the RNAsnap
625 protocol and Turbo DNase treated as described for the Rend-seq and 3′ end sequencing experiments.
626 rRNA was depleted from 10 µg of purified RNA with the MICROBExpress Bacterial mRNA Enrichment
627 Kit, per manufacturer's instructions, precipitated, and resuspended in 10 mM Tris pH 7.0. 600 ng rRNA
628 depleted RNA was then ligated to 260 pmol of the RNA adapter oJT900 (Table S2) for 18 hours at 22 °C
629 using T4 RNA Ligase 1 (based on (Subtelny et al., 2014)).
630 Following ligation, RNA was fragmented for 2 minutes (ThermoFisher) and fragments of either
631 size 41-71 (corresponding to 15-45 plus a 26-nucleotide adapter) or 60-90 were isolated through
632 denaturing polyacrylamide gel extraction. The larger size window was chosen to increase effective
633 sequencing depth by removing dimerized fragments of oJT900, which were observed in high abundance
634 in sequencing libraries prepared with a 41-71 nt size selection. RNA fragments were dephosphorylated,
635 polyadenylated, and reverse transcribed following the BaM-seq multiplexed RNA sequencing protocol
636 (Johnson et al., 2023). In short, in a single tube RNA fragments are sequentially dephosphorylated with
637 T4 PNK (NEB), polyadenylated with *E. coli* poly(A) polymerase (NEB) and reverse transcribed with
638 SuperScript III (ThermoFisher) and a barcoded reverse transcription primer. The cDNA from each sample
639 generated by this reaction can then be pooled and purified by denaturing polyacrylamide gel
640 electrophoresis. To add Illumina adapters, pooled cDNA was amplified between 5-10 cycles with primers
641 oDP128 and oDP161 (Johnson et al., 2023) using Q5 DNA polymerase and purified by polyacrylamide
642 gel electrophoresis. 5′ end sequencing libraries were sequenced with 75 bp single end reads on a
643 NextSeq500.
644
645 *Rend-seq, 5′-, and 3′-end sequencing data processing*

646 Adapter or poly(A) sequence was trimmed from end sequencing and Rend-seq libraries using CutAdapt
647 (Martin, 2011) with parameters "-a 'CTGTAGGCACCATCAAT' -m 15" or "-a A{15} -m 15 -u 8 --
648 rename='{id} UMI={cut_prefix}'". 5′ end sequencing reads were additionally collapsed using the 8 bp
649 unique molecular identifier (UMI) at the start of each read using custom Python scripts, requiring
650 identical UMI sequences and read sequences within a hamming distance of 1 to be considered duplicates.
651 Trimmed reads were then mapped to the *S. cerevisiae* genome (SacCer3) using bowtie v1.1.2 with
652 parameters "-v 1 -k 1 –best –un" with retained unmapped reads subsequently mapped to the *B. subtilis*
653 genome (NC_000964.3) with the same parameters (Langmead et al., 2009). Mapping against yeast
654 removes contaminating 2xYT-derived sequences and allows estimation of the degree of contamination in
655 a particular sample. Using custom Python scripts, we converted the bowtie output to wig files, attributing
656 signal from reads with a mismatch at their first position to the adjacent nucleotide to address mismatches
657 due to non-templated addition during reverse transcription.
658
659 *Identification of putative endonuclease sites*
660 To identify putative sites of endoribonucleolytic cleavage, we identified positions for which an adjacent 5′
661 and 3′ end could be detected in our end sequencing approaches. To do so, we first normalized each dataset
662 to the total number of millions of CDS-mapped reads. At each position, we identified peaks in read
663 density by calculating a ratio of 5′ end signal at that position to the 90% winsorized average in a window
664 50 nucleotides upstream (for 3′ ends) or downstream (for 5′ ends) of that position, excluding a 2-
665 nucleotide gap on either side of the peak. For any adjacent positions (allowing a gap of 1 position) with a
666 ratio over background greater than 7.5, these positions were grouped for the purposes of downstream 5′
667 end signal quantification. At each position with >10 reads per million CDS-mapped reads (RPM) in 5′
668 end sequencing, we detected paired 3′ ends by calculating an equivalent ratio centered around the position
669 one nucleotide upstream in the 3′ end sequencing of the 4-exo deletion. Additionally, in the 50-nucleotide
670 window downstream of each position, we calculated the local density of mapped 3′ ends of the RNase J1
671 deletion Rend-seq data to establish that we were considering positions only in genes expressed in this

strain. Putative cleavage sites were called as positions for which the average local Rend-seq depth exceeded 0.05 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 are unable to capture short RNA fragments (<15 nt for Rend-seq and 3′ end sequencing and <15-34 nt for 5′ end sequencing, depending on the exact size selection criteria used), we did not consider positions within 40 nucleotides of an RNA 3′ end. To identify these ends, we called peaks in 3′-mapped 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 (i.e. an RNA 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 count the number of putative cleavage positions across all our datasets, positions that were called in either the W168 (CCB434) or phage-cured W168 (BG880) background were considered. If two adjacent positions were called across these datasets, they were counted only a single position.

To attribute putative cleavage positions to specific endoribonucleases, 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 and Rend-seq signal at each position was normalized to the total CDS-mapped reads in each sample. 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 in the group. This ratio of 5′ end sequencing to Rend-seq signal is calculated for the corresponding RNase J1 deficient strain that expresses the considered endonuclease (e.g. a phage-cured RNase J1 deletion when considering RNase III). The ratio in the endoribonuclease expressing strain is then divided by the ratio in the endoribonuclease knockout strain to produce a “sensitivity” score. This approach is similar to that used to map RNase III sites in previous work (DiChiara et al., 2016).

698 Putative cleavage positions with a sensitivity score that exceeds 10.56 in a particular
699 endoribonuclease knockout are called as cleavage sites for that enzyme. This threshold was empirically
700 determined by fitting of the distribution of log-transformed sensitivity values for the RNase Y knockout
701 to the sum of two Gaussian distributions and finding the point at which 95% of this fit is contributed by
702 the underlying Gaussian with the larger mean (see Figure 5A). For this endonuclease assignment analysis,
703 any position for which the local Rend-seq density in the endonuclease knockout strain (as calculated
704 above) falls below 0.05 RPM is excluded to avoid spuriously calling positions in genes that are
705 downregulated in this condition. When calculating the fraction of putative cleavage sites that are
706 dependent on a particular protein, only those cleavage sites that were identified in an RNase J1 knockout
707 of the same background were considered (e.g. only sites appearing in the phage-cured RNase J1 knockout
708 strain BG880 were used when calculating the fraction of sites sensitive to RNase III knockout, which was
709 generated in the same phage-cured W168 background, see Table S1).
710
711 *Determination of RNase III site duplex length, genomic distance, and structural distance*
712 To identify pairs of RNase III cleavage positions that might represent intramolecular pairing, we first
713 identified all RNase III-dependent cleavage positions within 1 kb of one another. The distance between
714 sites in the primary sequence is the genomic distance as reported in Figure 3D. We then extracted the
715 intervening RNA sequence plus a 50 nucleotide buffer on either side for each pair of positions and
716 calculated its minimum free energy (MFE) structure using RNAfold from ViennaRNA v2.4.14 with
717 default parameters (Lorenz et al., 2011). We then used this MFE structure and a custom Python script to
718 calculate the length of the overhang generated (if any) by RNase III cleavage at this pair of sites. This
719 length is reported as structural distance in Figure 3E. The reported duplex length for each structure was
720 manually annotated, allowing for a maximum gap of two unpaired bases between base pairs.
721
722 *RNase III sequence-structure motif analysis*

723 For analyses of sequence content within RNA secondary structures recognized by RNase III, we
724 considered an expanded set of putative cleavage sites, using a threshold of >2 reads per million CDS-
725 mapping reads in 5′ end sequencing (rather than the previous >10 RPM). Results of putative cleavage
726 position identification and endonuclease sensitivity calculation using this threshold are summarized in
727 Tables S5,6. From this set of putative cleavage sites, identified pairs of sites within 1 kb of one another
728 and calculated a MFE secondary structure as described above. As before, we then used custom Python
729 scripts to quantify the overhang length generated by the two RNase III cleavages and retained any sites
730 which produced a 2-bp 3′ overhang.
731 Using the predicted MFE structures for these pairs of sites, we identified the base pair at each
732 position within the structure relative to the two positions of cleavage. The position-wise frequency of each
733 type of base-pairing interaction (A-U, U-A, G-C, C-G, U-G, or G-U) and fraction of sites for which that
734 position was paired was then determined. In cases where an asymmetric bulge was observed, those bases
735 were included as additional unpaired positions, rather than skipping over them in our indexing of base-
736 pairs. To compare our data to those previously-published for *E. coli* (Altuvia et al., 2018), we calculated
737 the difference between the frequency of each pairing type in our data and the equivalent observed at that
738 position by Altuvia et al.
739
740 *Local secondary structure prediction for RNase Y sites*
741 To determine folding propensity of sequences proximal to endoribonuclease cleavage sites, the *B. subtilis*
742 genome was tiled with 40-mers and folded with RNAfold from the ViennaRNA v2.4.14 package with
743 default parameters (Lorenz et al., 2011). The MFE values generated from this folding were assigned to
744 the center position within each k-mer and used to generate wig files containing positional folding
745 propensity. These values were extracted in 150 nucleotide windows centered around putative
746 endoribonuclease cleavage positions and a position-wise mean was calculated. If a cleavage site was
747 identified within 75 nucleotides of a 3′ or 5′ end present in wild-type W168 Rend-seq data (identified in
748 strain SSB1002 by the previously described z-score based peak calling method), that cleavage site was

749 not included in this analysis. To generate a background model that controls for sequence context in the -
750 1/+1 positions, this analysis was repeated 100 times sampling an equivalent number of random positions
751 within coding regions that are centered on GA dinucleotides (the most enriched dinucleotide flanking
752 putative RNase Y sites), excluding sites called as cleavage positions. From these predicted position-wise
753 average folding curves, the 25th and 75th percentile at each position across all curves was used to generate
754 the shaded region on Figure 5D.
755
756 *Information content and k-mer logo generation*
757 To generate sequence logos, the local sequence in a 30-nucleotide window surrounding all putative
758 cleavage sites for a given enzyme were identified and aggregated into a FASTA file. A background set of
759 30-mers evenly sampled every 100 nucleotides from *B. subtilis* coding regions (with sampling beginning
760 after the start codon) was generated as a background set. kpLogo v1.1 (Wu and Bartel, 2017) was then
761 used to generate information content and k-mer logos with parameters "-bgfile %s -startPos 16 -pc 0.01",
762 with %s pointing to our background set.
763
764 *MPRA plasmid library construction*
765 Plasmid libraries were constructed through restriction cloning of randomized mRNA processing site
766 sequences into suitable reporter plasmids (pJT028 or pJT124 for *tetM* or *aprE* experiments). Inserts for
767 plasmid libraries were constructed through a two-step protocol using mutated oligonucleotides generated
768 by IDT as a template. Oligonucleotides were designed with a 9% error rate (3% per non-native
769 nucleotide) over the specified positions (see Table S7 for enumeration of experiments and their variable
770 regions). In the first step, 10 fmol of oligo was PCR amplified for 12 cycles using Q5 DNA polymerase
771 (NEB). To produce a *cggR*-derived insert with mutations downstream of the cleavage site (experiments
772 2a, 2b, 4a, 4b, and 6), oJT837 was amplified with oJT851/852. To prepare the *glnR* insert (experiments
773 8a, 8b, 9), oJT839 was amplified with oJT857/858. For the *cggR* insert with upstream mutations
774 (experiments 1a, 1b, 3, 5, 7), instead of PCR amplification, oJT592 and oJT593 were mixed and extended

775 through one cycle of denaturing, annealing, and extension with Q5. These products were diluted 1:10 into
776 a second reaction in which they were amplified and barcoded through 4 cycles of PCR with Q5 using
777 oJT548/594 and oJT855/859 for all *cggR* and *glnR* inserts respectively. Amplified, barcoded inserts were
778 then cleaned up with a DNA Clean & Concentrator 5 column (Zymo Research).

779 To construct the plasmid pool, the amplified insert and 4 µg plasmid backbone were digested
780 with SalI-HF and NheI-HF for 60 minutes at 37°C and purified with a DNA Clean & Concentrator 5
781 column. 400 ng cut plasmid and insert were mixed at an approximately 1:10 ratio and ligated through a 5-
782 15 minute treatment with QuickLigase (NEB) at 25°C. This reaction was then cleaned up with a DNA
783 Clean & Concentrator 5 column and transformed into electrocompetent *E. coli* (NEB 10-beta or NEB 5-
784 alpha cells). Transformants were plated on LB with 100 µg/mL carbenicillin in 245 mm square bioassay
785 dishes (Corning) and grown at 37°C overnight. Cells were then scraped off the plates and pelleted for
786 ZymoPure II Maxiprep plasmid extraction (Zymo Research).

787

788 *Transformation and integration of MPRA library into B. subtilis genome*

789 To introduce our plasmid libraries into *B. subtilis*, we utilized an inducible competence system mediated
790 by an aTc-inducible *comK* allele introduced into BJT200 and BJT201, based on (Zhang and Zhang,
791 2011). A single colony was picked into LB (with 2% xylose for BJT201) and allowed to grow to an OD_{600}
792 of 1.0 at 37°C with vigorous shaking. At this point, 10 ng/mL aTc was added to the culture to induce
793 expression of ComK. Cells were shaken at 37°C for an additional 2 hours to allow induction of
794 competence, at which point 100 µL per mL culture of plasmid pool linearized with either ScaI-HF or AfeI
795 (NEB) was added to the culture. This restriction digest was prepared per the manufacturers protocol with
796 1 µg plasmid per 50 µL reaction. The cells were cultured with DNA for 1.5 hours, at which point the
797 cultures were pelleted at room temperature, resuspended in residual media, and plated on 245 mm
798 bioassay dishes containing 100 µg/mL spectinomycin or 5 µg/mL chloramphenicol. The *cggR-tetM*
799 library was transformed using standard protocols relying on natural competence. After an overnight
800 growth at 37°C, cells were scraped into LB with the appropriate antibiotic and approximately 500 million

801 cells were back-diluted into 200 mL LB with antibiotic. This culture was grown for approximately 4-8
802 hours, and several 1 mL aliquots of culture were mixed with equal volume 40% glycerol and frozen at -
803 80°C to prepare stocks for future experiments.
804
805 *Cell growth and harvesting (MPRA)*
806 For all MPRA experiments, cells were grown in LB medium with an inducer (20 ng/mL
807 anhydrotetracycline or 2% xylose) when noted. For all experiments in an inducible RNase J1 background,
808 2% xylose was consistently used to maintain RNase J1 expression, which was substituted for 2% glucose
809 when RNase J1 depletion was required. To collect samples, a single cryovial of a *B. subtilis* library was
810 thawed, washed twice with 10 mL LB to remove glycerol, and grown overnight in 200 mL of LB at 37°C
811 in a 2.8 L flask with vigorous shaking. These cultures were then back-diluted to an OD_{600} of 0.0003 in
812 200 mL LB with library inducer and grown at 37°C until an OD_{600} of 0.3, at which point samples were
813 harvested. Two cell pellets were harvested per sample, one for gDNA extraction and one for RNA
814 extraction. For RNA, 7 mL culture was added to 7 mL ice cold methanol, pelleted, decanted, and flash-
815 frozen in liquid nitrogen. For gDNA, 14 mL of culture was pelleted, decanted, and frozen in liquid
816 nitrogen. Cell pellets were stored at -80°C until nucleic acid extraction.
817
818
819 *MPRA RNA and DNA sequencing*
820 Three sequencing libraries were prepared for each sample: one to quantify the barcodes in cellular RNA,
821 one to quantify barcodes in the gDNA, and one to map barcodes to variant sequence. To prepare the RNA
822 barcode quantification library, RNA was first extracted from frozen cell pellets with gDNA depletion
823 using an RNeasy Plus Mini Kit (Qiagen). For experiments 1a, 1b, 2a, 2b, and 9 (see Table S7), 40 µg of
824 total RNA was reverse transcribed using primer oJT442 and SuperScript III. This cDNA was then run on
825 a 6% TBE-Urea gel, and the following size selection ranges were gel purified: 300-500 nt for full-length
826 *cggR* and *glnR* constructs (experiments 1a, 2a, 9) and 100-250 nt for cleaved *cggR* constructs
827 (experiments 1b and 2b). For the remaining experiments, cleaved RNA was isolated by size selection of

828 RNAs ranging from 200-400 nt for *tetM-cggR* sites, 150-300 nt for *aprE-cggR* sites, and 150-350 nt for
829 *aprE-glnR* sites on a 6% TBE-Urea gel (Novex). The size selected product was reverse transcribed using
830 primer oJT442 and SuperScript III (Thermo Fisher) and size selected on 6% TBE-Urea gel, extracting
831 approximately 120 to 200 nucleotides for *cggR* libraries and 100 to 200 for *glnR*. For all experiments,
832 size-selected cDNA was amplified using Phusion DNA polymerase (NEB), adding a barcode and
833 Illumina adapters with a Truseq indexing primer and an isoform-specific primer (see Table S2 for primer
834 sequences). These isoform-specific primers were oJT441/oJT1290 for cleaved/full-length products from
835 *cggR* constructs or oJT890/oJT1291 for cleaved/full-length products from *glnR* constructs. The resulting
836 libraries were sequenced with 50 or 75 bp single end reads on an Illumina NextSeq500.
837 To prepare both libraries for gDNA barcode quantification and barcode to variant sequence
838 mapping, a single gDNA sample was prepared from a frozen cell pellet using the Wizard Genomic DNA
839 Purification Kit (Promega). Both gDNA-derived libraries were generated using a two-step Phusion PCR
840 protocol. An initial amplicon was generated through 2 cycles of PCR using from 4 µg of purified nucleic
841 acid per 50 µL PCR reaction (reaction scaled up as needed to minimize UMI collapse in more complex
842 libraries) with primers oJT444/445 for *cggR* upstream mutations, oJT1125/445 for *cggR* downstream
843 mutations, and oJT891/445 for *glnR* downstream mutations. The untranslated *cggR* library required a
844 longer amplicon to capture the upstream stop codon and was instead amplified with oJT942 and oJT445.
845 This first PCR reaction was cleaned up with a double-sided Select-a-Size DNA Clean & Concentrator
846 column (Zymo) and a DNA Clean & Concentrator 5 column (Zymo). A quarter of this reaction was then
847 used for the second round of PCR, in which two different reactions were used to prepare barcode
848 quantification and barcode to variant sequence mapping libraries. For barcode quantification, the second-
849 round amplification was performed using the same primers as used for quantifying cleaved RNA
850 abundance (oJT441 or oJT890 with our Truseq indexing primer) and size selected on an 8% TBE gel.
851 These libraries were sequenced with 50 or 75 bp single end reads on an Illumina NextSeq500. Libraries
852 mapping barcodes to variant sequence were amplified using primers oJT446/447, size selected on an 8%
853 TBE gel, and sequenced with either 75 bp or 150 bp paired end reads on an Illumina MiSeq.

854

855 *Barcode to variant sequence mapping and barcode quantification*

856 Barcode quantification and variant sequence identification was handled using custom scripts written in
857 Python. To determine the variant sequence associated with each barcode, the paired end reads from the
858 gDNA-derived libraries were directly parsed as FASTQ files, with one read capturing the barcode and the
859 other the variable region of the construct. Variants with at least one mutation in the constant region
860 proceeding the barcode were discarded. To avoid rare mutations or sequencing errors confounding our
861 variant sequence mapping, variant sequence assignment required at least three reads per barcode. If,
862 among these reads, the second-most frequent sequence was at least 25% as frequent as the top sequence,
863 the barcode was discarded entirely.

864 Barcodes were counted by directly parsing the single-end sequencing reads derived from both
865 gDNA and RNA in a FASTQ format, with the barcode appearing at a fixed position within the read. As
866 before, any variant with mutations in the constant sequence proceeding the barcode was discarded.
867 Additionally, any two barcodes of identical sequence and unique molecular identifier were counted as a
868 single read. The number of unique reads mapping to each variant barcode was calculated for both the
869 RNA and gDNA-derived samples, and the ratio of RNA to gDNA-derived reads was calculated to extract
870 the relative accumulation of processed mRNA for each variant. Barcodes can then be grouped by
871 associated mutations, thresholded based on read count, and normalized to the median wild-type barcode's
872 ratio.

873
874
875 **AUTHOR CONTRIBUTIONS:**
876 J.T. and G.-W.L. conceptualized project, designed experiments and analysis, and wrote the manuscript
877 with input from all co-authors. J.T. collected data and performed analysis. H.L. and J.D. contributed to the
878 design and implementation of the MPRA experiments. J.-B.L. collected preliminary Rend-seq data for
879 exonuclease knockouts. C.C., S.D., and F. B. constructed strains, provided feedback on analysis, and
880 helped in conceptualization of project.
881
882 **ACKNOLWEDGEMENTS:**
883 The authors thank David Bechhofer for sharing the phage-cured RNase III knockout strains, Byoung-Mo
884 Koo and Carol Gross for sharing the RNase J2 knockout strain, the MIT BioMicro Center for advice and
885 high-throughput DNA sequencing, Sean McGeary for feedback on early versions of the manuscript, and

886 Michael Laub and David Bartel for feedback in developing this project. Thanks additionally to all
887 members of the Li Lab for valuable discussion. This research is supported by a National Science
888 Foundation Graduate Research Fellowship (to J.C.T.) and the NSF CAREER Award MCB1844668. The
889 C.C. lab is supported by funds from the CNRS (UMR 8261), Université Paris Cité, and the Agence
890 Nationale de la Recherche (CoNoCo and LabEx Dynamo).
891
892 **REFERENCES:**
893 Altuvia, Y., Bar, A., Reiss, N., Karavani, E., Argaman, L., and Margalit, H. (2018). In vivo cleavage rules
894 and target repertoire of RNase III in Escherichia coli. Nucleic Acids Res. *46*, 10380–10394.
895 Bechhofer, D.H., and Deutscher, M.P. (2019). Bacterial ribonucleases and their roles in RNA metabolism.
896 Crit. Rev. Biochem. Mol. Biol. *54*, 242–300.
897 Benda, M., Woelfel, S., Faßhauer, P., Gunka, K., Klumpp, S., Poehlein, A., Kálalová, D., Šanderová, H.,
898 Daniel, R., Krásný, L., et al. (2021). Quasi-essentiality of RNase Y in Bacillus subtilis is caused
899 by its critical role in the control of mRNA homeostasis. Nucleic Acids Res. *49*, 7088–7102.
900 Braun, F., Durand, S., and Condon, C. (2017). Initiating ribosomes and a 5- /3-UTR interaction control
901 ribonuclease action to tightly couple B. Subtilis hbs mRNA stability with translation. Nucleic
902 Acids Res. *45*, 11386–11400.
903 Brewster, R.C., Jones, D.L., and Phillips, R. (2012). Tuning promoter strength through RNA polymerase
904 binding site design in Escherichia coli. PLoS Comput. Biol. *8*, e1002811.
905 Britton, R.A., Wen, T., Schaefer, L., Pellegrini, O., Uicker, W.C., Mathy, N., Tobin, C., Daou, R., Szyk,
906 J., and Condon, C. (2007). Maturation of the 5′ end of Bacillus subtilis 16S rRNA by the essential
907 ribonuclease YkqC/RNase J1. Mol. Microbiol. *63*, 127–138.
908 Broglia, L., Lécrivain, A., Renault, T.T., Ahmed-begrich, R., Rhun, A. Le, and Charpentier, E. (2020). An
909 RNA-seq based comparative approach reveals the transcriptome-wide interplay between 3'-to-5'
910 exoRNases and RNase Y. Nat. Commun. *11*.
911 Bruscella, P., Shahbabian, K., Laalami, S., and Putzer, H. (2011). RNase Y is responsible for uncoupling
912 the expression of translation factor IF3 from that of the ribosomal proteins L35 and L20 in
913 Bacillus subtilis. Mol. Microbiol. *81*, 1526–1541.
914 Cerullo, F., Filbeck, S., Patil, P.R., Hung, H.C., Xu, H., Vornberger, J., Hofer, F.W., Schmitt, J., Kramer,
915 G., Bukau, B., et al. (2022). Bacterial ribosome collision sensing by a MutS DNA repair ATPase
916 paralogue. Nature *603*, 509–514.
917 Chao, Y., Li, L., Girodat, D., Förstner, K.U., Said, N., Corcoran, C., Śmiga, M., Papenfort, K., Reinhardt,
918 R., Wieden, H.J., et al. (2017). In Vivo Cleavage Map Illuminates the Central Role of RNase E in
919 Coding and Non-coding RNA Pathways. Mol. Cell *65*, 39–51.
920 Chen, H., Shiroguchi, K., Ge, H., and Xie, X.S. (2015). Genome-wide study of mRNA degradation and
921 transcript elongation in E scherichia coli. Mol. Syst. Biol. *11*, 781.
922 Collins, J.A., Irnov, I., Baker, S., and Winkler, W.C. (2007). Mechanism of mRNA destabilization by the
923 glmS ribozyme. Genes Dev. *21*, 3356–3368.
924 Commichau, F.M., Rothe, F.M., Herzberg, C., Wagner, E., Hellwig, D., Lehnik-Habrink, M., Hammer,
925 E., Völker, U., and Stülke, J. (2009). Novel activities of glycolytic enzymes in Bacillus subtilis:
926 interactions with essential proteins involved in mRNA processing. Mol. Cell. Proteomics *8*,
927 1350–1360.
928 Condon, C. (2003). RNA Processing and Degradation in Bacillus subtilis. Microbiol. Mol. Biol. Rev. *67*,
929 157–174.
930 Cuchillo, C.M., Nogués, M.V., and Raines, R.T. (2011). Bovine pancreatic ribonuclease: fifty years of the
931 first enzymatic reaction mechanism. Biochemistry *50*, 7835–7841.
932 Dar, D., Shamir, M., Mellin, J.R., Koutero, M., Stern-Ginossar, N., Cossart, P., and Sorek, R. (2016).
933 Term-seq reveals abundant ribo-regulation of antibiotics resistance in bacteria. Science *352*,
934 aad9822.
935 Deloughery, A., Dengler, V., Chai, Y., and Losick, R. (2016). Biofilm formation by Bacillus subtilis
936 requires an endoribonuclease-containing multisubunit complex that controls mRNA levels for the

937 matrix gene repressor SinR. Mol. Microbiol. *99*, 425–437.
938 DeLoughery, A., Lalanne, J.-B., Losick, R., and Li, G.-W. (2018). Maturation of polycistronic mRNAs
939 by the endoribonuclease RNase Y and its associated Y-complex in *Bacillus subtilis*. Proc. Natl.
940 Acad. Sci. *115*, E5585–E5594.
941 DiChiara, J.M., Liu, B., Figaro, S., Condon, C., and Bechhofer, D.H. (2016). Mapping of internal
942 monophosphate 5′ ends of Bacillus subtilis messenger RNAs and ribosomal RNAs in wild-type
943 and ribonuclease-mutant strains. Nucleic Acids Res. *44*, 3373–3389.
944 Dubnau, E., DeSantis, M., and Dubnau, D. (2023). Formation of a stable RNase Y-RicT (YaaT) complex
945 requires RicA (YmcA) and RicF (YlbF). MBio *14*, e0126923.
946 Durand, S., Gilet, L., Bessières, P., Nicolas, P., and Condon, C. (2012a). Three essential ribonucleases-
947 RNase Y, J1, and III-control the abundance of a majority of bacillus subtilis mRNAs. PLoS
948 Genet. *8*.
949 Durand, S., Gilet, L., and Condon, C. (2012b). The essential function of B. subtilis RNase III is to silence
950 foreign toxin genes. PLoS Genet. *8*, e1003181.
951 Durand, S., Braun, F., Lioliou, E., Romilly, C., Helfer, A.-C., Kuhn, L., Quittot, N., Nicolas, P., Romby,
952 P., and Condon, C. (2015). A Nitric Oxide Regulated Small RNA Controls Expression of Genes
953 Involved in Redox Homeostasis in Bacillus subtilis. PLOS Genet. *11*, e1004957.
954 Figaro, S., Durand, S., Gilet, L., Cayet, N., Sachse, M., and Condon, C. (2013). Bacillus subtilis mutants
955 with knockouts of the genes encoding ribonucleases RNase Y and RNase J1 are viable, with
956 major defects in cell morphology, sporulation, and competence. J. Bacteriol. *195*, 2340–2348.
957 Gan, J., Tropea, J.E., Austin, B.P., Court, D.L., Waugh, D.S., and Ji, X. (2006). Structural insight into the
958 mechanism of double-stranded RNA processing by ribonuclease III. Cell *124*, 355–366.
959 Gilet, L., Dichiara, J.M., Figaro, S., Bechhofer, D.H., and Condon, C. (2015). Small stable RNA
960 maturation and turnover in Bacillus subtilis. Mol. Microbiol. *95*, 270–282.
961 Gordon, G.C., Cameron, J.C., and Pfleger, B.F. (2017). RNA sequencing identifies new RNase III
962 cleavage sites in Escherichia coli and reveals increased regulation of mRNA. MBio *8*.
963 Guler, P., Bendori, S.O., Borenstein, T., Aframian, N., Kessel, A., and Eldar, A. (2024). Arbitrium
964 communication controls phage lysogeny through non-lethal modulation of a host toxin–antitoxin
965 defence system. Nat. Microbiol. *9*, 150–160.
966 Hamouche, L., Billaudeau, C., Rocca, A., Chastanet, A., Ngo, S., Laalami, S., and Putzer, H. (2020).
967 Dynamic membrane localization of RNase Y in bacillus subtilis. MBio *11*.
968 Herskovitz, M.A., and Bechhofer, D.H. (2000). Endoribonuclease RNase III is essential in Bacillus
969 subtilis. Mol. Microbiol. *38*, 1027–1033.
970 Herzel, L., Stanley, J.A., Yao, C.-C., and Li, G.-W. (2022). Ubiquitous mRNA decay fragments in E. coli
971 redefine the functional transcriptome . Nucleic Acids Res. *50*, 5029–5046.
972 Ingle, S., Chhabra, S., Chen, J., Lazarus, M.B., Luo, X., and Bechhofer, D.H. (2022). Discovery and
973 initial characterization of YloC, a novel endoribonuclease in Bacillus subtilis. RNA *28*, 227–238.
974 Jahn, N., and Brantl, S. (2013). One antitoxin-two functions: SR4 controls toxin mRNA decay and
975 translation. Nucleic Acids Res. *41*, 9870–9880.
976 Johnson, G.E., Parker, D.J., Lalanne, J., Parker, M.L., and Li, G. (2023). BaM-seq and TBaM-seq , highly
977 multiplexed and targeted RNA-seq protocols for rapid , low-cost library generation from bacterial
978 samples. *5*, 1–11.
979 Khemici, V., Prados, J., Linder, P., and Redder, P. (2015). Decay-Initiating Endoribonucleolytic Cleavage
980 by RNase Y Is Kept under Tight Control via Sequence Preference and Sub-cellular Localisation.
981 PLoS Genet. *11*, 1–27.
982 Koo, B.M., Kritikos, G., Farelli, J.D., Todor, H., Tong, K., Kimsey, H., Wapinski, I., Galardini, M.,
983 Cabal, A., Peters, J.M., et al. (2017). Construction and Analysis of Two Genome-Scale Deletion
984 Libraries for Bacillus subtilis. Cell Syst. *4*, 291-305.e7.
985 Korobeinikova, A., Laalami, S., Berthy, C., and Putzer, H. (2023). RNase Y Autoregulates Its Synthesis
986 in Bacillus subtilis. Microorganisms *11*, 1374.
987 Lalanne, J.B., Taggart, J.C., Guo, M.S., Herzel, L., Schieler, A., and Li, G.W. (2018). Evolutionary

988 Convergence of Pathway-Specific Enzyme Expression Stoichiometry. Cell *173*, 749-761.e38.
989 Langmead, B., Trapnell, C., Pop, M., and Salzberg, S.L. (2009). Ultrafast and memory-efficient
990 alignment of short DNA sequences to the human genome. Genome Biol. *10*.
991 Lehnik-Habrink, M., Schaffer, M., Mäder, U., Diethmaier, C., Herzberg, C., and Stülke, J. (2011a). RNA
992 processing in Bacillus subtilis: Identification of targets of the essential RNase Y. Mol. Microbiol.
993 *81*, 1459–1473.
994 Lehnik-Habrink, M., Newman, J., Rothe, F.M., Solovyova, A.S., Rodrigues, C., Herzberg, C.,
995 Commichau, F.M., Lewis, R.J., and Stülke, J. (2011b). RNase Y in Bacillus subtilis: A natively
996 disordered protein that is the functional equivalent of RNase E from Escherichia coli. J. Bacteriol.
997 *193*, 5431–5441.
998 Lehnik-Habrink, M., Lewis, R.J., Mäder, U., and Stülke, J. (2012). RNA degradation in Bacillus subtilis:
999 An interplay of essential endo- and exoribonucleasesmmi. Mol. Microbiol. *84*, 1005–1017.
1000 Leroy, M., Piton, J., Gilet, L., Pellegrini, O., Proux, C., Coppée, J., Figaro, S., and Condon, C. (2017).
1001 Rae1/YacP, a new endoribonuclease involved in ribosome-dependent mRNA decay in
1002 Bacillus subtilis . EMBO J. *36*, 1167–1181.
1003 Lin-Chao, S., Wong, T.-T., Mcdowall, K.J., and Cohens, S.N. (1994). Effects of Nucleotide Sequence on
1004 the Specificity of me-dependent and RNase E-mediated Cleavages of RNA I Encoded by the
1005 pBR322 Plasmid. J. Biol. Chem. *269*, 10797–10803.
1006 Lorenz, R., Bernhart, S.H., Zu Siederdissen, C.H., Tafer, H., Flamm, C., Stadler, P.F., and Hofacker, I.L.
1007 (2011). ViennaRNA Package 2.0. Algorithms Mol. Biol. *6*, 1–14.
1008 Ludwig, H., Homuth, G., Schmalisch, M., Dyka, F.M., Hecker, M., and Stülke, J. (2001). Transcription of
1009 glycolytic genes and operons in Bacillus subtilis: Evidence for the presence of multiple levels of
1010 control of the gapA operon. Mol. Microbiol. *41*, 409–422.
1011 Mandell, Z.F., Oshiro, R.T., Yakhnin, A. V, Kashlev, M., Kearns, D.B., and Babitzke, P. (2021). NusG is
1012 an intrinsic transcription termination factor that stimulates motility and coordinates global gene
1013 expression with NusA. Elife *10:e61880*.
1014 Marincola, G., and Wolz, C. (2017). Downstream element determines RNase y cleavage of the saePQRS
1015 operon in Staphylococcus aureus. Nucleic Acids Res. *45*, 5980–5994.
1016 Marincola, G., Schäfer, T., Behler, J., Bernhardt, J., Ohlsen, K., Goerke, C., and Wolz, C. (2012). RNase
1017 Y of Staphylococcus aureus and its role in the activation of virulence genes. Mol. Microbiol. *85*,
1018 817–832.
1019 Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads.
1020 EMBnet.Journal *17*, 10.
1021 McDowall, K.J., Lin-Chaol, S., and Cohen, S.N. (1994). A+U content rather than a particular nucleotide
1022 order determines the specificity of RNase E cleavage. J. Biol. Chem. *269*, 10790–10796.
1023 Meinken, C., Blencke, H.M., Ludwig, H., and Stülke, J. (2003). Expression of the glycolytic gapA operon
1024 in Bacillus subtilis: Differential syntheses of proteins encoded by the operon. Microbiology *149*,
1025 751–761.
1026 Meißner, C., Jahn, N., and Brantl, S. (2016). In Vitro characterization of the type i toxin-antitoxin system
1027 bsrE/SR5 from Bacillus subtilis. J. Biol. Chem. *291*, 560–571.
1028 Moffitt, J.R., Pandey, S., Boettiger, A.N., Wang, S., and Zhuang, X. (2016). Spatial organization shapes
1029 the turnover of a bacterial transcriptome. Elife *5*, 1–22.
1030 Noone, D., Salzberg, L.I., Botella, E., Bäsell, K., Becher, D., Antelmann, H., and Devine, K.M. (2014). A
1031 highly unstable transcript makes CwlO D,L-endopeptidase expression responsive to growth
1032 conditions in bacillus subtilis. J. Bacteriol. *196*, 237–247.
1033 Oguro, A., Kakeshita, H., Nakamura, K., Yamane, K., Wang, W., and Bechhofer, D.H. (1998). Bacillus
1034 subtilis RNase III cleaves both 5'- and 3'-sites of the small cytoplasmic RNA precursor. J. Biol.
1035 Chem. *273*, 19542–19547.
1036 Oussenko, I.A., Abe, T., Ujiie, H., Muto, A., and Bechhofer, D.H. (2005). Participation of 3 ′ -to-5 ′
1037 Exoribonucleases in the Turnover of Bacillus subtilis mRNA Participation of 3 J -to-5 J
1038 Exoribonucleases in the Turnover of Bacillus subtilis mRNA. J. Bacteriol. *187*, 2758–2767.

1039 Park, J.-H., Yamaguchi, Y., and Inouye, M. (2011). Bacillus subtilis MazF-bs (EndoA) is a UACAU-
1040 specific mRNA interferase. FEBS Lett. *585*, 2526–2532.
1041 Parker, D.J., Lalanne, J.B., Kimura, S., Johnson, G.E., Waldor, M.K., and Li, G.W. (2020). Growth-
1042 Optimized Aminoacyl-tRNA Synthetase Levels Prevent Maximal tRNA Charging. Cell Syst. *11*,
1043 121-130.e6.
1044 Pellegrini, O., Mathy, N., Gogos, A., Shapiro, L., and Condon, C. (2005). The Bacillus subtilis ydcDE
1045 operon encodes an endoribonuclease of the MazF/PemK family and its inhibitor. Mol. Microbiol.
1046 *56*, 1139–1148.
1047 Pertzev, A. V, and Nicholson, A.W. (2006). Characterization of RNA sequence determinants and
1048 antideterminants of processing reactivity for a minimal substrate of Escherichia coli ribonuclease
1049 III. Nucleic Acids Res. *34*, 3708–3721.
1050 Ruiz de los Mozos, I., Vergara-Irigaray, M., Segura, V., Villanueva, M., Bitarte, N., Saramago, M.,
1051 Domingues, S., Arraiano, C.M., Fechter, P., Romby, P., et al. (2013). Base Pairing Interaction
1052 between 5′- and 3′-UTRs Controls icaR mRNA Translation in Staphylococcus aureus. PLoS
1053 Genet. *9*.
1054 Salis, H.M. (2011). The ribosome binding site calculator. Methods Enzym. *498*, 19–42.
1055 Salis, H.M., Mirsky, E.A., and Voigt, C.A. (2009). Automated design of synthetic ribosome binding sites
1056 to control protein expression. Nat. Biotechnol. *27*, 946–950.
1057 Le Scornet, A., Jousselin, A., Baumas, K., Kostova, G., Durand, S., Poljak, L., Barriot, R., Coutant, E.,
1058 Pigearias, R., Tejero, G., et al. (2024). Critical factors for precise and efficient RNA cleavage by
1059 RNase Y in Staphylococcus aureus. PLOS Genet. *20*, 1–30.
1060 Shahbabian, K., Jamalli, A., Zig, L., and Putzer, H. (2009). RNase Y, a novel endoribonuclease, initiates
1061 riboswitch turnover in Bacillus subtilis. EMBO J. *28*, 3523–3533.
1062 Stead, M.B., Agrawal, A., Bowden, K.E., Nasir, R., Mohanty, B.K., Meagher, R.B., and Kushner, S.R.
1063 (2012). RNAsnap^{TM}: a rapid, quantitative and inexpensive, method for isolating total RNA from
1064 bacteria. Nucleic Acids Res. *40*, e156.
1065 Subtelny, A.O., Eichhorn, S.W., Chen, G.R., Sive, H., and Bartel, D.P. (2014). Poly(A)-tail profiling
1066 reveals an embryonic switch in translational control. Nature *508*, 66–71.
1067 Taggart, J.C., Lalanne, J.B., and Li, G.W. (2021). Quantitative Control for Stoichiometric Protein
1068 Synthesis. Annu. Rev. Microbiol. *75*, 243–267.
1069 Trinquier, A., Durand, S., Braun, F., and Condon, C. (2020). Regulation of RNA processing and
1070 degradation in bacteria. Biochim. Biophys. Acta - Gene Regul. Mech. *1863*, 194505.
1071 Urtecho, G., Tripp, A.D., Insigne, K.D., Kim, H., and Kosuri, S. (2019). Systematic Dissection of
1072 Sequence Elements Controlling σ70 Promoters Using a Genomically Encoded Multiplexed
1073 Reporter Assay in Escherichia coli. Biochemistry *58*, 1539–1551.
1074 Winkler, W.C., Nahvi, A., Roth, A., Collins, J.A., and Breaker, R.R. (2004). Control of gene expression
1075 by a natural metabolite-responsive ribozyme. Nature *428*, 281–286.
1076 Wu, X., and Bartel, D.P. (2017). kpLogo: positional k-mer analysis reveals hidden specificity in
1077 biological sequences. Nucleic Acids Res. *45*, W534–W538.
1078 Yao, S., and Bechhofer, D.H. (2010). Initiation of decay of Bacillus subtilis rpsO mRNA by
1079 endoribonuclease RNase Y. J. Bacteriol. *192*, 3279–3286.
1080 Yarchuk, O., Jacques, N., Guillerez, J., and Dreyfus, M. (1992). Interdependence of translation,
1081 transcription and mRNA degradation in the lacZ gene. J. Mol. Biol. *226*, 581–596.
1082 Yasbin, R.E., and Young, F.E. (1974). Transduction in Bacillus subtilis by Bacteriophage SPP1 . J. Virol.
1083 *14*, 1343–1348.
1084 Zhang, K., and Nicholson, A.W. (1997). Regulation of ribonuclease III processing by double-helical
1085 sequence antideterminants. Proc. Natl. Acad. Sci. U. S. A. *94*, 13437–13441.
1086 Zhang, X.Z., and Zhang, Y.H.P. (2011). Simple, fast and high-efficiency transformation system for
1087 directed evolution of cellulase in Bacillus subtilis. Microb. Biotechnol. *4*, 98–105.
1088
1089

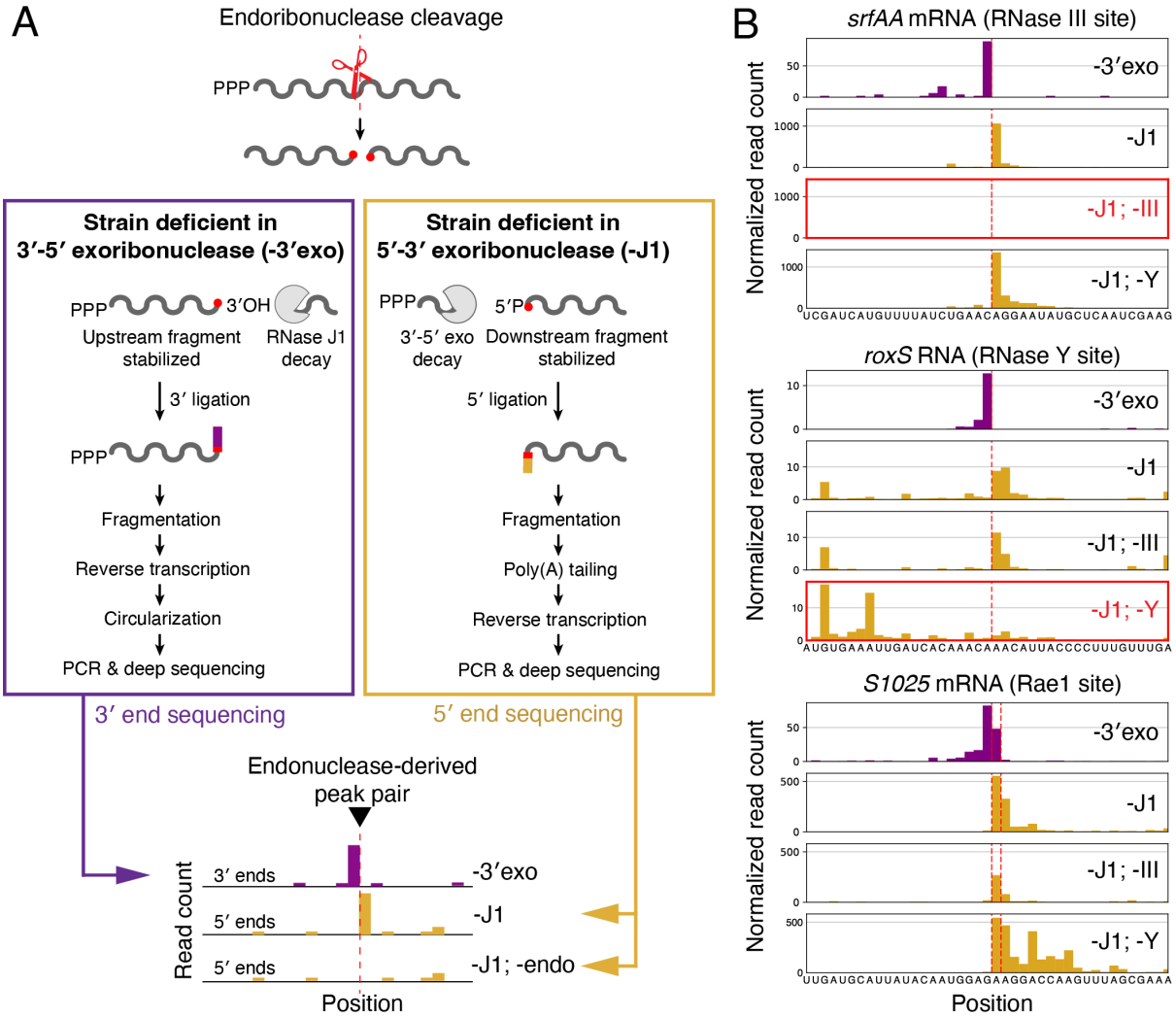

Figure 1. Workflow and validation of endoribonuclease cleavage mapping approach
(A) Deletion of exoribonucleases results in stable accumulation of RNA decay intermediates. A schematic mRNA is shown with endoribonuclease cleavage (red scissors) occurring, with remaining exoribonucleases indicated in gray. 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. Plotted are reads per million CDS-mapping reads, normalized to the average 3′-mapped Rend-seq RPM in this window.

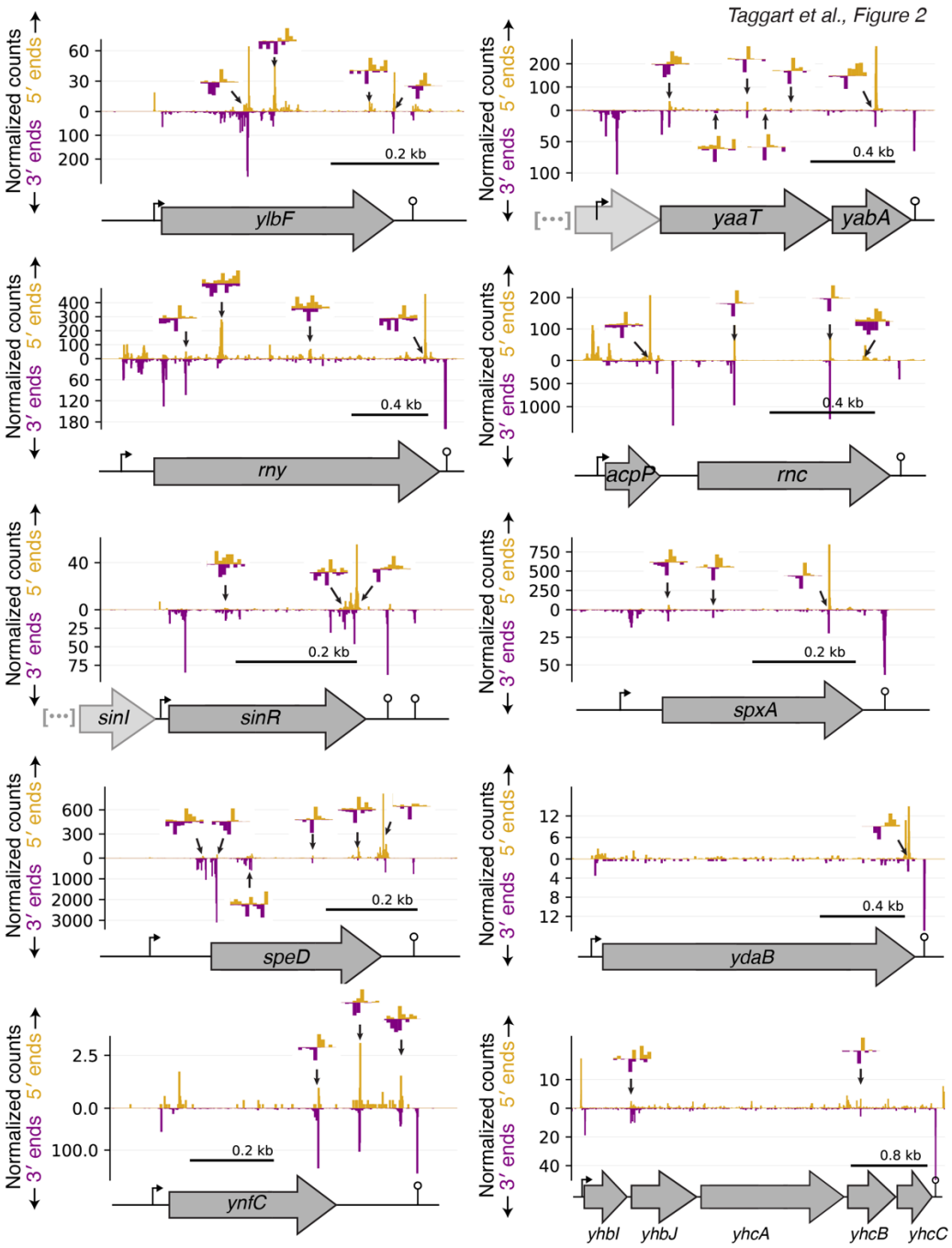

Figure 2. Cleavage within transcripts known to be destabilized by endoribonucleolytic cleavage. 5′ and 3′ end sequencing data are shown in yellow and purple, respectively. Plotted are reads per million CDS-mapping reads. Manually annotated putative cleavage sites (positions of adjacent 3′ and 5′ read 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. Annotated promoter and transcriptional terminators, indicated by bent arrows and lollipops, respectively, were identified using Rend-seq in SSB1002. Sequence context and RNase Y dependence of highlighted sites shown in Figure S2. Consistent with a recent report (Korobeinikova et al., 2023), multiple 5′ and 3′ end signals were observed in the 5′ UTR of *rny*.

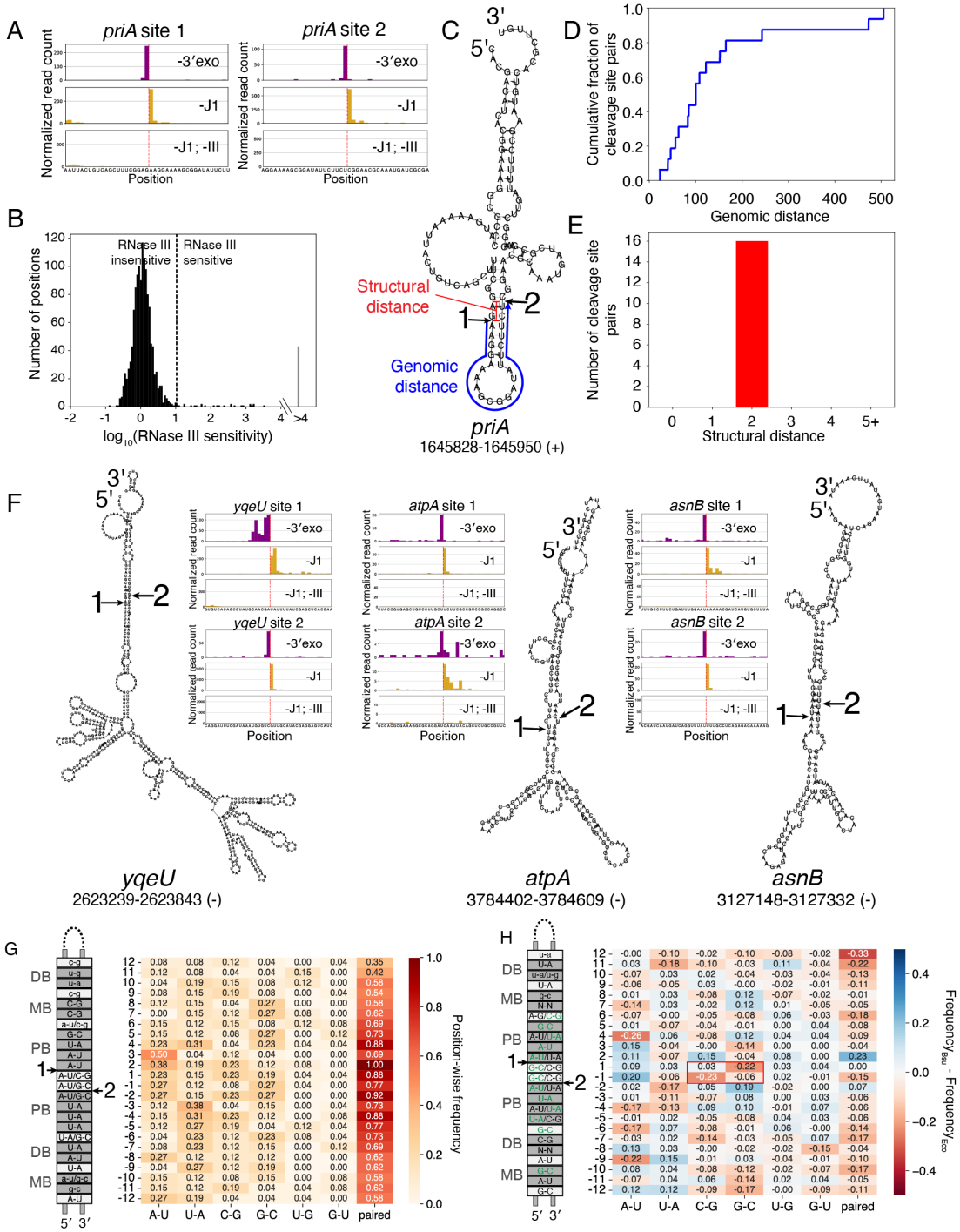

Figure 3. *B. subtilis* RNase III cleaves within long-range intramolecular secondary structures
(A) Identification of RNase III cleavage positions in the *priA* mRNA. 5′ and 3′ end sequencing data are shown in yellow and purple, respectively. Identified positions of cleavage are highlighted in red. Plotted are reads per million CDS-mapping reads, normalized to the average 3′-mapped Rend-seq RPM in this window.

1125 (B) Results of systematic identification of RNase III sites. Histogram shows distribution of endonuclease
1126 sensitivities for called peak pairs in 3′/5′ end sequencing of exoribonuclease knockouts (see Figure S3 and
1127 Methods), with black dashed line indicating threshold for calling dependence on RNase III. 68 sites
1128 exceeded this defined threshold. For 43 sites we were unable to calculate a sensitivity score due to an
1129 absence of 5′ end sequencing counts in our knockout. These sites are called as RNase III sensitive and are
1130 counted within the “>4” bin of the histogram.
1131 (C) Predicted secondary structure of the sequence surrounding the identified RNase III sites within the
1132 *priA* mRNA. Positions of cleavage are indicated with red arrows, with numbers corresponding to those in
1133 (A). Structural distance between these cleavage positions, defined as the defined as the length of the 3′
1134 overhang generated by the cleavages, is annotated in red. Genomic distance, defined as the distance
1135 between the cleavage sites in the primary sequence, is annotated in blue.
1136 (D) Genomic distance (as defined in C) between pairs of unambiguous RNase III cleavage positions
1137 identified within 1 kb of one another and predicted to fall on opposite sides of an RNA stem with a 2 bp
1138 spacing between cleavage positions (32 of 68 positions).
1139 (E) Structural distance (as defined in C) generated by RNase III cleavage at both positions within the
1140 pairs of cleavage positions plotted in (D).
1141 (F) Predicted secondary structures of the sequence surrounding the identified RNase III sites within the
1142 mRNAs encoding the *yqeU*, *atpA*, and *asnB* genes. Insets are plotted as in (A).
1143 (G) Consensus sequence-structure motif of RNase III targets in *B. subtilis*. The predicted structures for
1144 each pair of RNase III-sensitive cleavage sites that generate a 2-nt 3′ overhang were aligned to the
1145 positions of cleavage, “1” and “2”. The position-wise frequency of each type of base-pairing interaction is
1146 shown. Note that the base pairing frequencies sum to the fraction of sites paired at each position. The
1147 most frequent base pair at each position is shown in the schematic on the left, with lower-case characters
1148 used to designate positions for which the most frequent base pair occurs <20% of the time. When two
1149 base pair frequencies are within 0.02 of one another, both are separated by a slash (e.g., A-U/G-C). The
1150 gray regions labeled “PB”, “MB”, and “DB” correspond to the proximal, middle, and distal boxes (Gan et
1151 al., 2006; Zhang and Nicholson, 1997).
1152 (H) Comparison of base pairing frequences between *B. subtilis* and *E. coli* (Altuvia et al., 2018). The
1153 consensus *E. coli* sequence-structure motif is shown on the left with top base pairs at each position which
1154 are shared between *B. subtilis* and *E. coli* highlighted in green. Red box highlights the decrease in C-G or
1155 G-C pairing in *B. subtilis* relative to *E. coli*.
1156

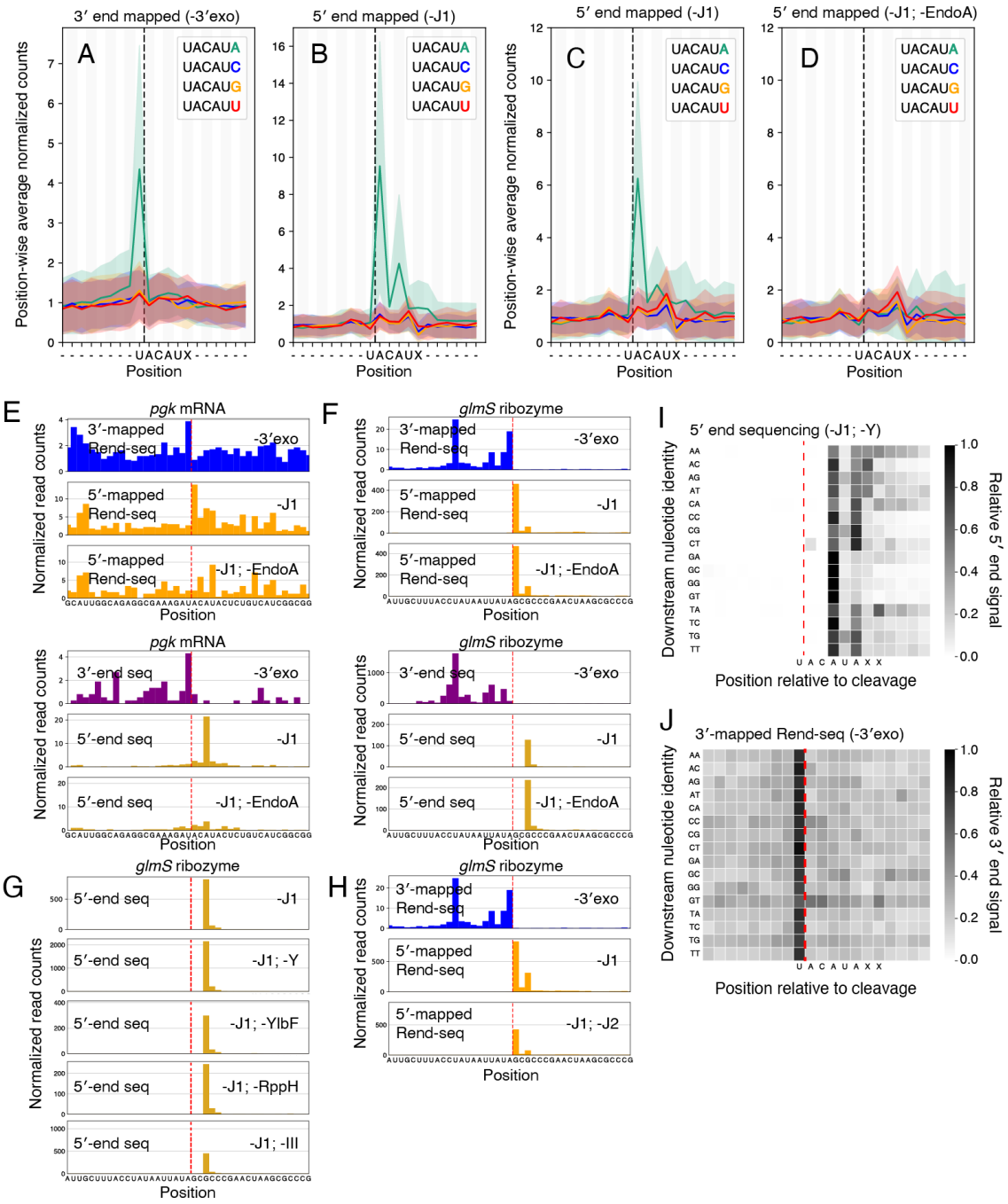

Figure 4. Revision of EndoA cleavage specificity and evidence for new ribonuclease activity in *B. subtilis*

(A, B) 3′ (A) and 5′ (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 (shaded interval) 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 435 (A) or 292 to 455 (B) per motif. The black 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 191 to 288 (C) or 86 to 117 (D) per motif.

(E) 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. Plotted are reads per million CDS-mapping reads. 5′/3′-end sequencing data are normalized to the average 3′-mapped Rend-seq signal in this window.

(F) 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 E. Red dotted line indicates position of cleavage. Data are plotted as in (E).

(G) 5′-end sequencing data at the position of cleavage by the *glmS* ribozyme for knockouts of additional RNA decay-associated proteins. All knockouts are coupled to either a depletion or deletion of RNase J1. Knocked out genes include *rnjA* (RNase J1, strain CCB434), *rny* (RNase Y, strain CCB760), *ylbF* (YlbF, strain BJT074), *rppH* (RppH, strain BJT129), and *rnc* (RNase III, strain BG879). Data are plotted as in (E), with the exception of *rppH*, which is normalized to an *rnjA* knockout alone, rather than the same genetic background, due to a lack of corresponding Rend-seq data.

(H) 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. Plotted are reads per million CDS-mapping reads.

(I, J) 5′ end sequencing (I) and Rend-seq 3′-mapped (J) 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 (I) or 10 to 62 per row (J). The red vertical line indicates the position of cleavage by EndoA.

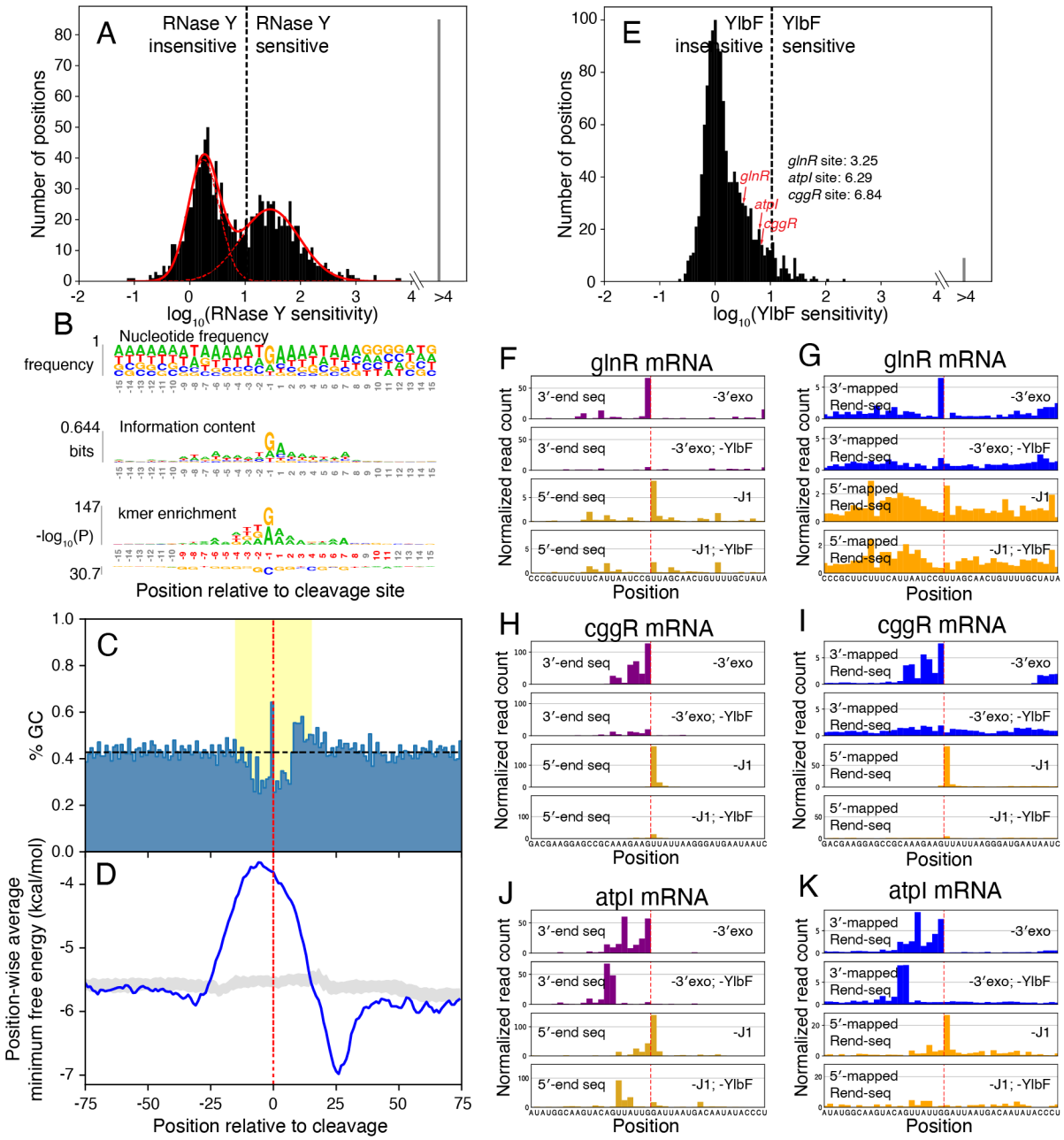

Figure 5. Sequence and structural features of putative RNase Y substrates
(A) Systematic identification of putative RNase Y cleavage sites. Histogram shows distribution of endonuclease sensitivities for called peak pairs in 3′/5′ end sequencing of exoribonuclease knockouts (see Figure S3 and Methods). 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 endonuclease sensitivity value (10.56) for which 95% of the Gaussian sum is derived from the right constituent distribution, a value above which sites are considered dependent on RNase Y. 596 of 1301 considered putative cleavage signatures exceed this threshold. For 85 sites we were unable to calculate a sensitivity score due to an

absence of 5′ end sequencing counts in the knockout. These sites are called as RNase Y sensitive and are counted within the “>4” bin of the histogram.

(B) Local sequence context around positions of RNase Y cleavage (N=596). 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. Significant positions (Bonferroni corrected P<0.01, one-sided binomial test), are highlighted in red.

(C) Local GC content around positions of RNase Y cleavage (N=484). 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. Yellow region indicates positions shown in (B).

(D) Predicted folding near RNase Y cleavage positions (N=484). 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.

(E) Systematic identification of YlbF-dependent endoribonuclease cleavage sites. Histogram shows distribution of endonuclease sensitivities for called peak pairs in 3′/5′ end sequencing of exoribonuclease knockouts (see Figure S3 and Methods). The dashed vertical line indicates the endonuclease sensitivity threshold used to call RNase Y sites. Red arrows indicate endonuclease sensitivity values of canonical processing sites within the *cggR*, *glnR*, and *atpI* transcripts. For 9 sites we were unable to calculate a sensitivity score due to an absence of 5′ end sequencing counts in the knockout. These sites are called as YlbF sensitive and are counted within the “>4” bin of the histogram.

(F-K) End mapping data showing evidence for inefficient but detectable endoribonucleolytic cleavage at canonical RNase Y processing sites within the transcripts encoding *glnR* (F, G), *cggR* (H, I), and *atpI* (J, K) in the absence of YlbF. (F, H, and J) show 5′/3′-end sequencing and (G, I, and K) show Rend-seq. 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 positions of endoribonucleolytic cleavage. Plotted are reads per million CDS-mapping reads. 5′/3′-end sequencing data are normalized to the average 3′-mapped Rend-seq signal in this window.

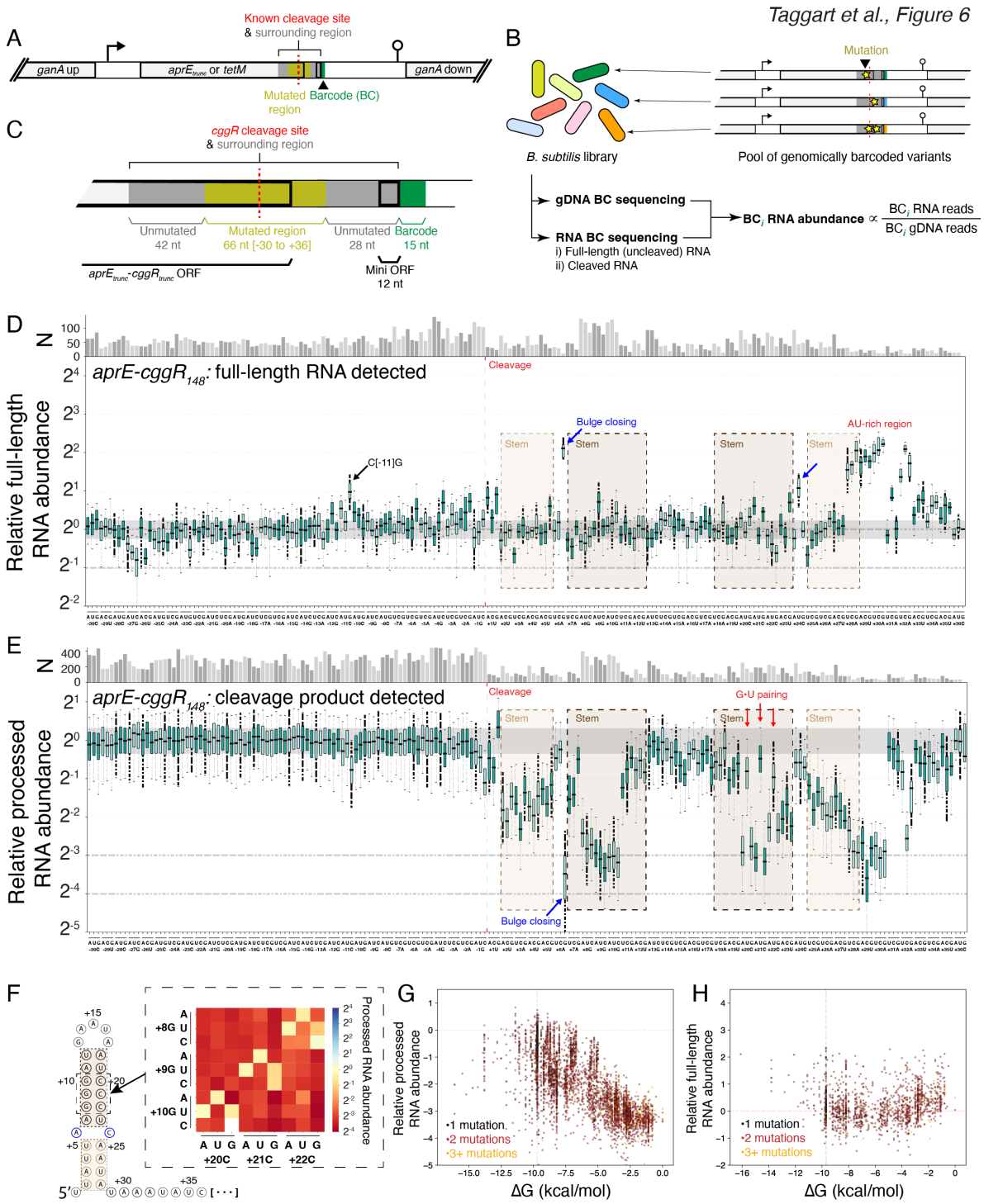

Figure 6. Mutational scanning of the *cggR-gapA* operon cleavage site

(A) Schematic design of MPRA construct. Cleavage region to mutate is indicated in dark gray with yellow used to indicate mutated region. Variant barcode indicated in green. Translated regions are indicated with a thick border. Promoters and transcriptional terminators are indicated by bent arrows and lollipops, respectively.

(B) Workflow for measurement of mRNA processing. Protocol begins from *B. subtilis* culture containing a pool of genetically encoded cleavage site variants, and splits into gDNA and RNA barcode

1252 quantification protocols. The relative barcoded RNA abundance is calculated as the ratio of RNA-derived
1253 to gDNA-derived reads.
1254 (C) Schematic of the *cggR*$_{148}$ construct inserted into the *aprE* MPRA transcript. *cggR*-derived sequence is
1255 colored gray, with the mutated positions highlighted in yellow. Translated regions are indicated with a
1256 thick border, and the variant barcode is indicated in green.
1257 (D) Impact of all single-nucleotide mutations on the accumulation of barcoded full-length RNA. Boxplots
1258 show variation between barcodes of identical variant sequence. Number of barcodes captured for each
1259 mutation is indicated above plot. Whiskers indicate 5th and 95th percentile. Gray shaded region indicates
1260 interquartile range for variants of wild-type sequence. Red line indicates position of cleavage. Blue
1261 arrows highlight mutations which close the bulge in the *cggR* downstream hairpin. Four variants (no more
1262 than one per mutation) have a value of zero are thus not visualized.
1263 (E) Impact of all single-nucleotide mutations on the accumulation of barcoded cleavage product. Boxplots
1264 show variation between barcodes of identical variant sequence. Number of barcodes captured for each
1265 mutation (N) is indicated by the bar height (top). Whiskers indicate 5th and 95th percentile. Gray shaded
1266 region indicates interquartile range for variants of wild-type sequence. Red line indicates position of
1267 cleavage and brown bars indicate regions predicted to pair in the formation of a downstream stem-loop
1268 structure (as illustrated in F). Red arrows indicate mutations which are predicted to generate G•U wobble
1269 pairing, and the blue arrow highlights a mutation which corrects the bulge in the downstream stem-loop
1270 structure (B). Four variants (no more than one per mutation) have a value of zero are thus not visualized.
1271 (F) Predicted stem-loop structure at 5′ end of processed RNA. Inset shows impact of double mutants on
1272 accumulation of processed RNA. Double mutant averages calculated over any variants with the two
1273 indicated mutations. Any variants which contain additional mutations that on their own result in greater
1274 than 10% change in processed RNA accumulation are discarded.
1275 (G) Relationship between predicted strength of downstream secondary structure and accumulation of
1276 barcoded cleavage product. The experiment each data point is derived from is denoted with a triangle
1277 (experiment 2b, Table S3), circle (4a), or square (6). The vertical dashed line indicates the ΔG of the
1278 unmutated sequence. Five variants fall outside of the bounds of this plot.
1279 (H) Relationship between predicted strength of downstream secondary structure and accumulation of
1280 barcoded full-length RNA. All data derived from experiment 2a (Table S7). The vertical dashed line
1281 indicates the ΔG of the unmutated sequence. Five variants fall outside of the bounds of this plot.
1282

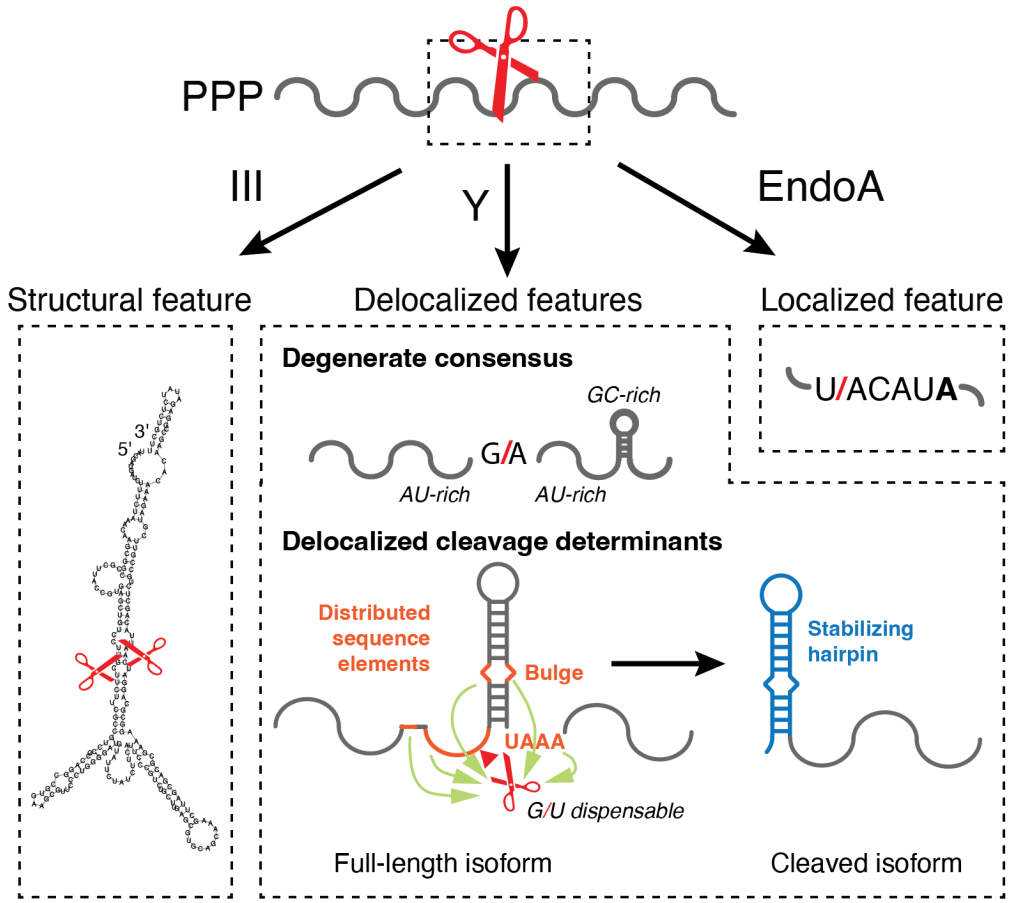

Figure 7. Determinants of endoribonuclease cleavage identified in this work
(A) Summary of the identified cleavage determinants of the three endoribonucleases profiled in this work. RNase III appears to cleave within double-stranded RNA sequences, including many duplexes formed through long range (>100 nt) intramolecular interactions, consistently leaving a 2 nucleotide 3′ overhang (left, "III"). EndoA appears to cleave specifically at a UACAUA primary sequence motif (right, "EndoA"). RNase Y cleavage appears to be targeted towards AU-rich regions of low secondary structure, appearing to prefer cleavage within a G/A dinucleotide (center, "Y"). Interrogation of the cleavage sites within *cggR* and *glnR* suggest that a -1G is dispensable for RNase Y cleavage, and for a given site the sequence and structural elements that drive cleavage (orange) may be distributed over many tens of nucleotides. Further, for some mRNA processing sites such as those in *cggR* or *gapA*, a downstream stem-loop structure (blue) appears to primarily drive stabilization of the cleaved isoform.

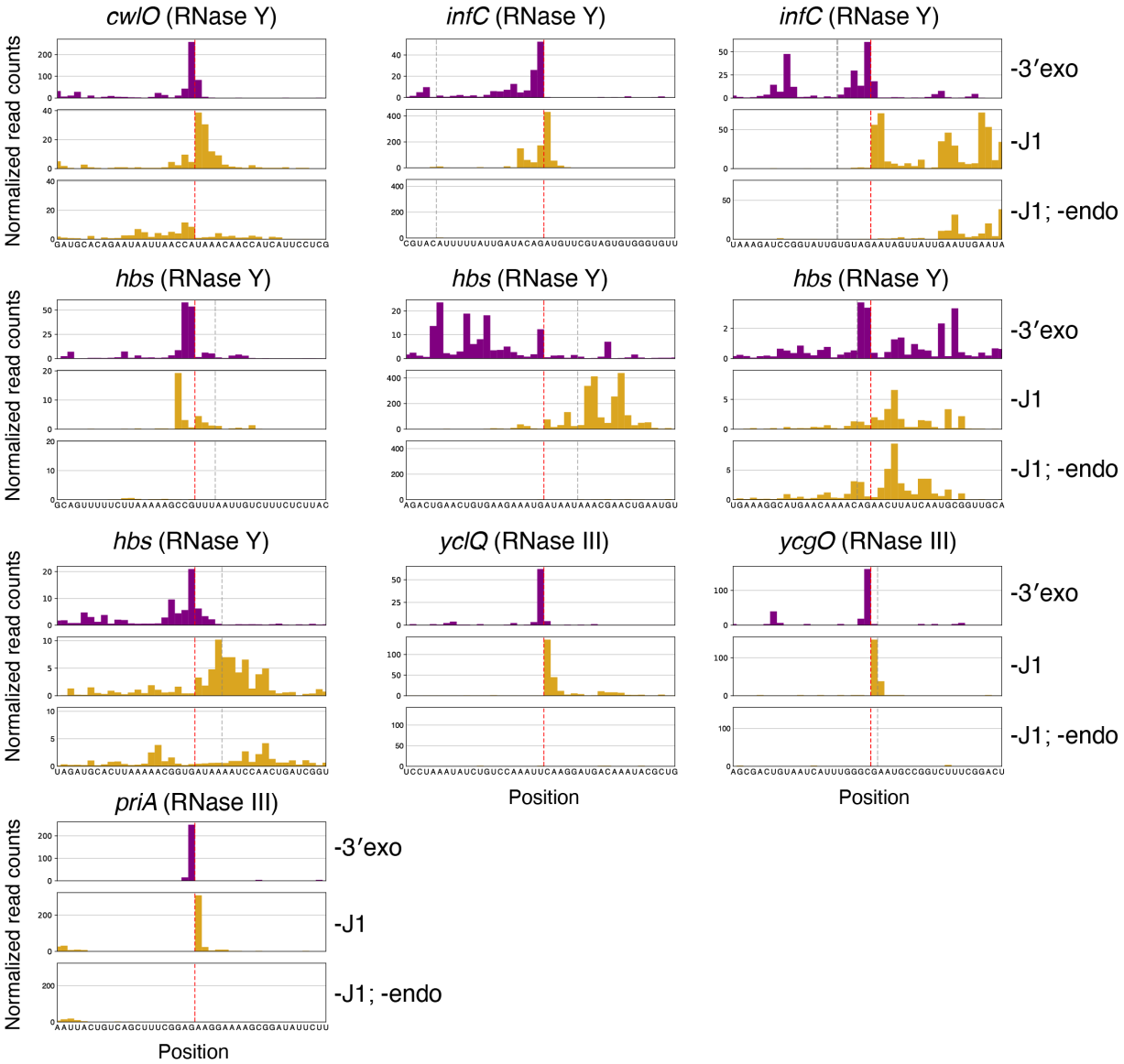

Figure S1. Validation of approach through detection of known positions of endonuclease activity. 5′ and 3′ end sequencing data at known positions of cleavage by RNases Y and III. Datasets considered are a 4-exo knockout (CCB396), *rnjA* knockout (CCB434), and either deletion or depletion of RNase J1 with a knockout of RNase Y (CCB760) or III (BG879). 5′-mapped data shown in yellow and 3′-mapped data shown in purple. Plotted are reads per million CDS-mapping reads, normalized to the average 3′-mapped Rend-seq RPM in this 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.

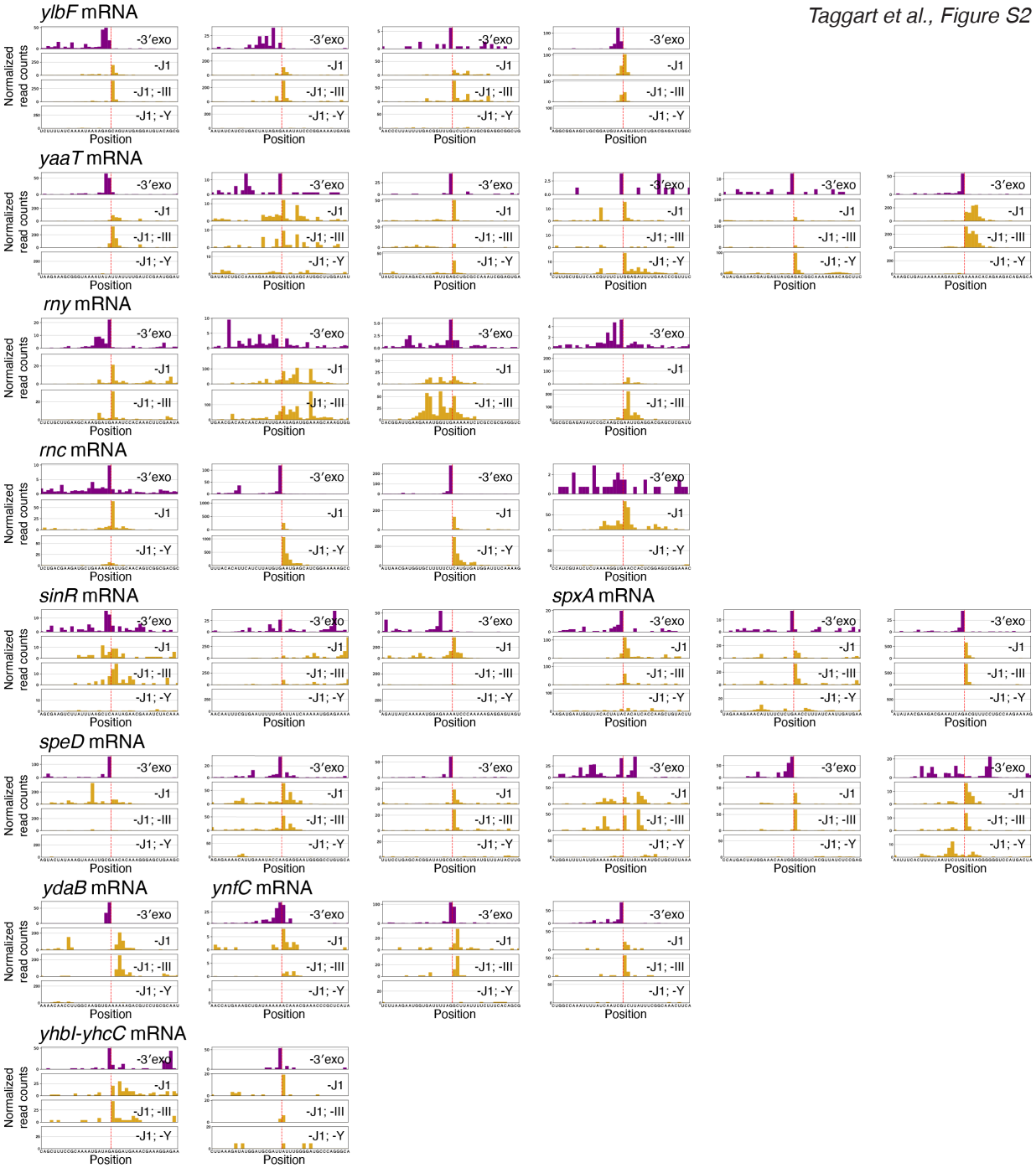

Figure S2. **Sequence context and RNase Y/III dependence of cleavage positions within mRNAs known to be destabilized by endoribonucleases**. Cleavage positions match order of those shown in Figure 2. Datasets considered are a 4-exo knockout (CCB396), *rnjA* knockout (CCB434), rnjA rnc double knockout in an SPβ and Skin-cured background (BG879), 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. Plotted are reads per million CDS-mapping reads, normalized to the average 3′-mapped Rend-seq RPM in this window. A manually annotated position of cleavage based on sequencing data is shown with a red dotted line. Data from the *rnc*- or *rny*-deficient strain are not shown for sites within the transcripts encoding these genes.

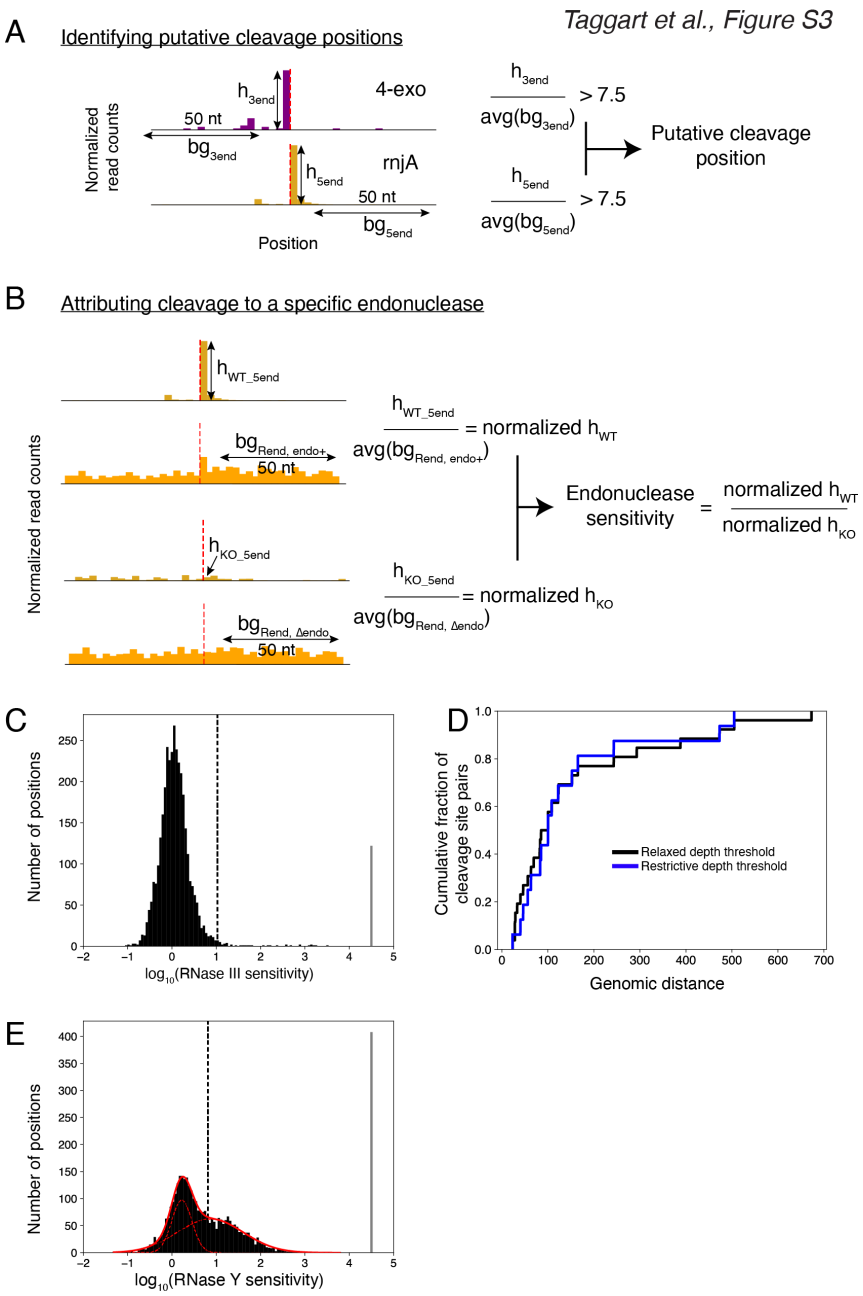

Figure S3. Schematic representation of strategy for calling putative cleavage positions and assigning them to the activity of endoribonucleases

(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 (termed "endonuclease sensitivity") is calculated as a metric for dependence on this endoribonuclease. Note that prior to illustrated calculation, 5′ end sequencing data (h_{WT_5end} and h_{KO_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.

1331 (C) Results of systematic identification of RNase III sites with a relaxed 5′ end sequencing depth
1332 threshold (as used in Figure 3G, H). Histogram shows distribution of endonuclease sensitivities for called
1333 peak pairs in 3′/5′ end sequencing of exoribonuclease knockouts (see Figure S3A and Methods), with
1334 black dashed line indicating threshold for calling dependence on RNase III. 162 sites exceeded this
1335 defined threshold. For 122 sites we were unable to calculate a sensitivity score due to an absence of 5′ end
1336 sequencing counts in our knockout. These sites are called as RNase III sensitive and are counted within
1337 the “>4” bin of the histogram.
1338 (D) Genomic distance between pairs of unambiguous RNase III cleavage positions identified within 1 kb
1339 of one another and predicted to fall on opposite sides of an RNA stem, using both restrictive and relaxed
1340 5′ end sequencing depth thresholds.
1341 (E) Results of systematic identification of RNase Y sites with a relaxed 5′ end sequencing depth
1342 threshold. Histogram shows distribution of RNase Y sensitivities, plotted as described in Figure 5A. 1351
1343 sites exceeded this defined threshold. For 408 sites we were unable to calculate a sensitivity score due to
1344 an absence of 5′ end sequencing counts in our knockout. These sites are called as RNase Y sensitive and
1345 are counted within the “>4” bin of the histogram.
1346

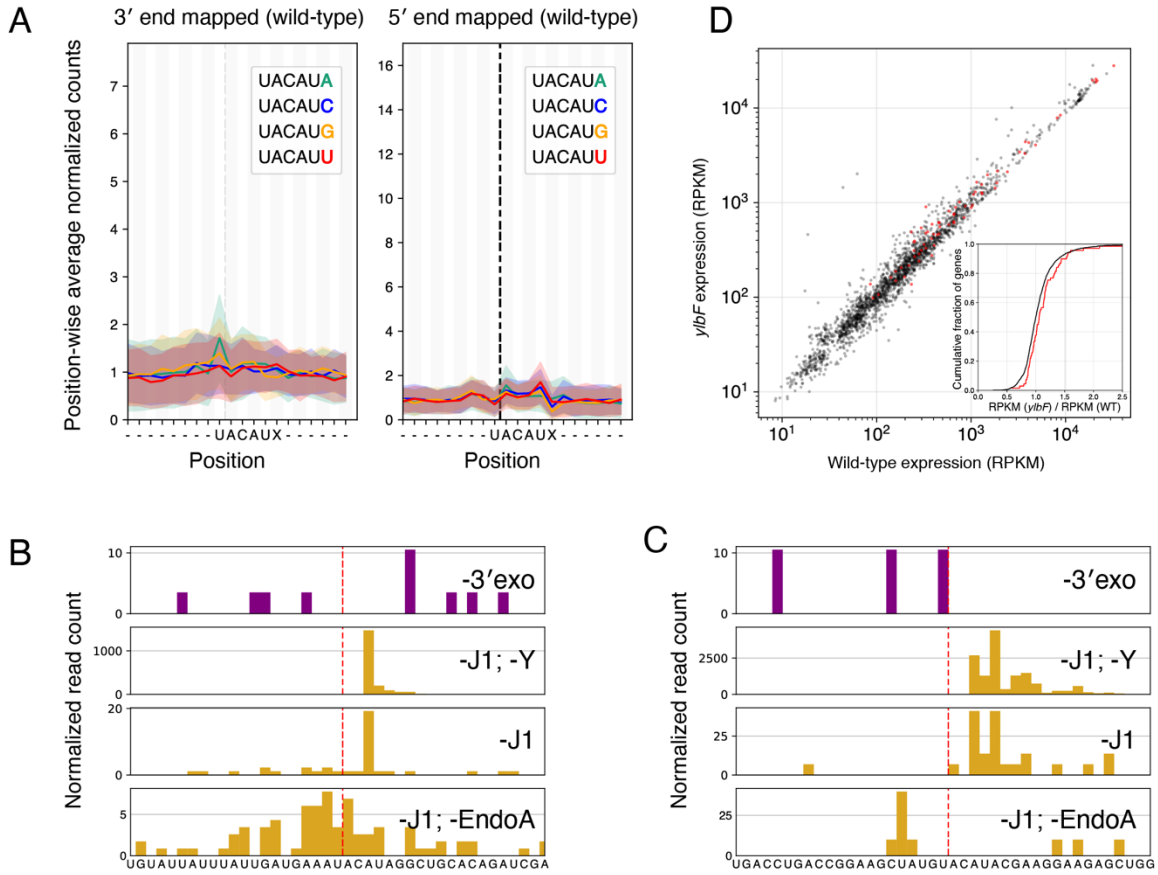

Figure S4. EndoA and YlbF have a limited influence on wild-type *B. subtilis* mRNA decay

(A) UACAUA-specific EndoA cleavage is not detectable in wild-type *B. subtilis.* 3′ (left) and 5′ (right) mapped Rend-seq signal across all UACAU motifs in the genome, separated by downstream nucleotide. Data are derived from wild-type W168. 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 193 to 313 (for 3′-mapped) and 203 to 311 (for 5′-mapped) per motif. The dashed vertical line indicates the position of cleavage by EndoA.

(B, C) Representative EndoA sites showing presence (B) or absence (C) of subsequent trimming after the first two nucleotides. Yellow indicates 5′ end sequencing data and purple indicates 3′ end sequencing data. Plotted are reads per million CDS-mapping reads, normalized to the average 3′-mapped Rend-seq RPM in this window. Red dotted line represents manually annotated EndoA cleavage positions. Note that cleavage product accumulation is dependent on EndoA (bottom two panels).

(D) mRNA abundances for all annotated *B. subtilis* genes in a wild-type and *ylbF* W168 background. Lowly expressed genes (<128 total reads) are excluded. RPKM calculated as reads per thousand bases of the gene per million total reads that map to *B. subtilis* coding regions. Inset shows differential expression distribution (Δ*ylbF*/WT RPKM) for genes with genes with (red) or without (black) a YlbF-dependent putative cleavage site.

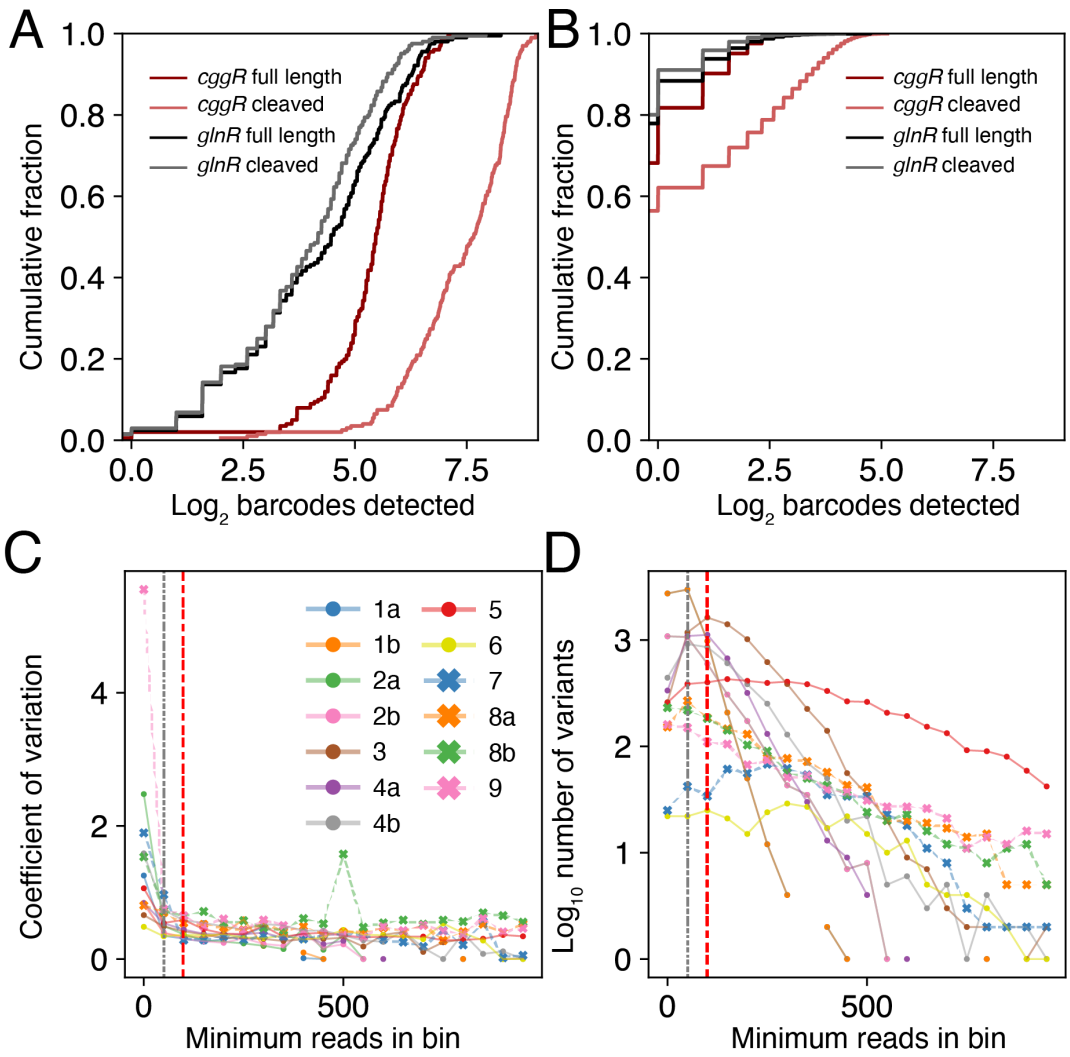

Figure S5. Variability and sequence coverage within our MPRA datasets

(A) Cumulative distribution of the number of unique barcodes associated with each possible sequence containing a single substitution relative to the wild-type *cggR* or *glnR* operon sequence. Undetected sequences are included with a pseudocount of 0.1 prior to $\log_2$ transformation.

(B) Cumulative distribution of the number of unique barcodes associated with each possible sequence containing two substitutions relative to the wild-type *cggR* or *glnR* operon sequence. Undetected sequences are included with a pseudocount of 0.1 prior to $\log_2$ transformation.

(C) Variation in MRPA readout as a function of genomic DNA sequencing depth. The normalized MPRA RNA abundance for each variant with a wild-type sequence was calculated as illustrated in Figure 6A. These values were grouped by the number of unique reads mapping to that variant in the genomic DNA barcode sequencing with a bin width of 50 reads. The coefficient of variation within each group is shown. The gray vertical dashed line indicates the 50 read cutoff used for experiments 4b, 8a, 8b, and 9, red vertical dashed line indicates the 100 read cutoff used in analysis of all remaining experiments. See Table S7 for more information.

(D) The number of variants within each bin for each dataset show in (A). Dashed lines show thresholds as described in (A).

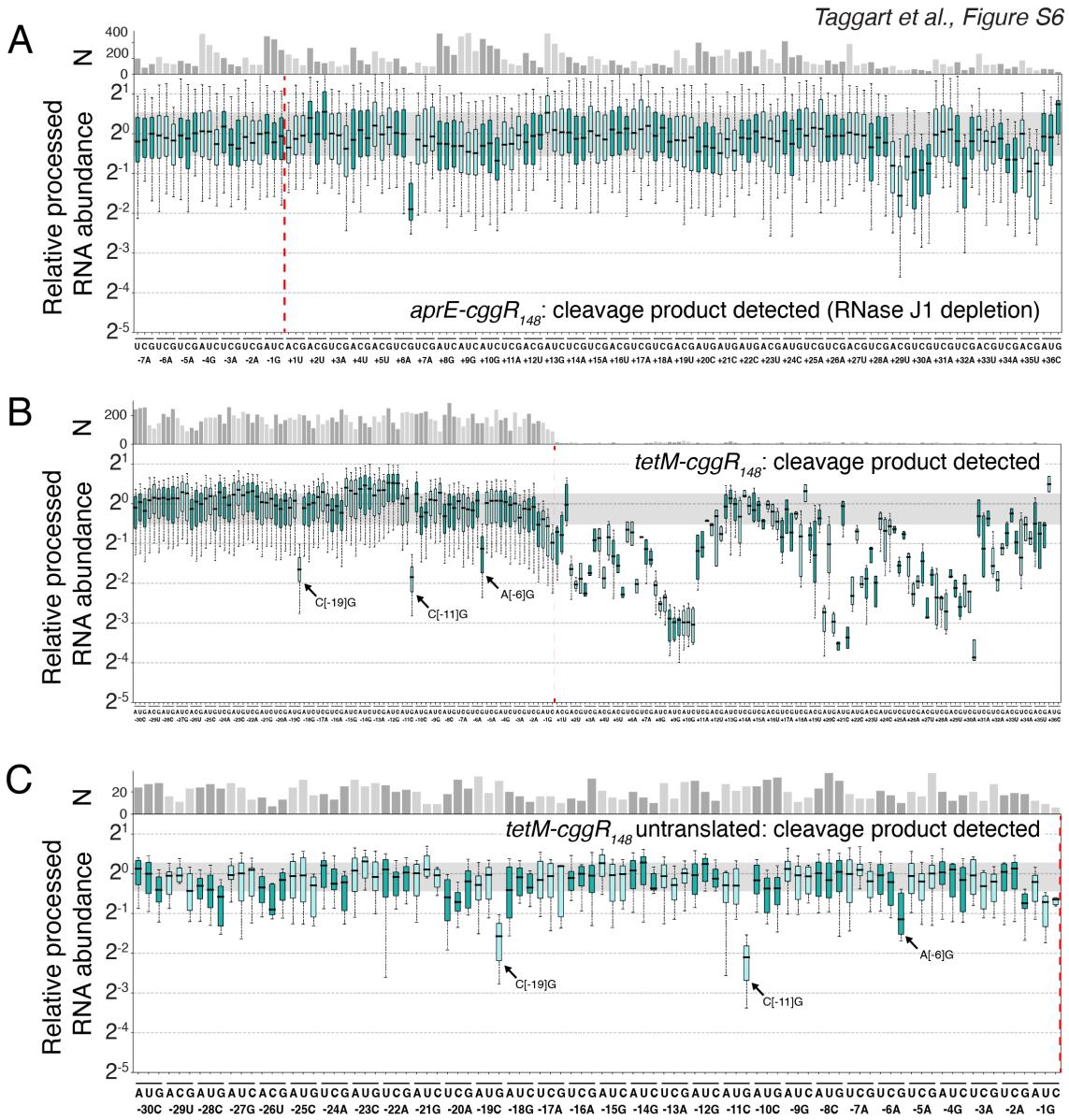

Figure S6. Features driving RNase Y cleavage within *cggR* are consistent across multiple MPRA contexts

(A) Impact of all single-nucleotide mutations on the accumulation of barcoded cleaved *aprE-cggR*$_{148}$ RNA in the context of RNase J1 depletion. Boxplots show variation between barcodes of identical variant sequence. Number of barcodes captured for each mutation is indicated above plot. Whiskers indicate 5th and 95th percentile. Gray shaded region indicates interquartile range for variants of wild-type sequence. Red line indicates position of cleavage. Ten variants have a value of zero (no more than one per mutation) and are thus not visualized.

(B) Impact of all single-nucleotide mutations on the accumulation of barcoded cleavage product when the *cggR*$_{148}$ construct is appended to a new scaffold RNA, *tetM*. Plotted as in (A). Two variants (no more than one per mutation) have a value of zero are thus not visualized.

(C) Impact of all single-nucleotide mutations in the 30 nucleotides upstream of the cggR$_{148}$ cleavage site on the accumulation of barcoded cleavage product when a stop codon is inserted upstream of the inserted sequence. *tetM* scaffold was used in this experiment, as in (C). Plotted as in (A).

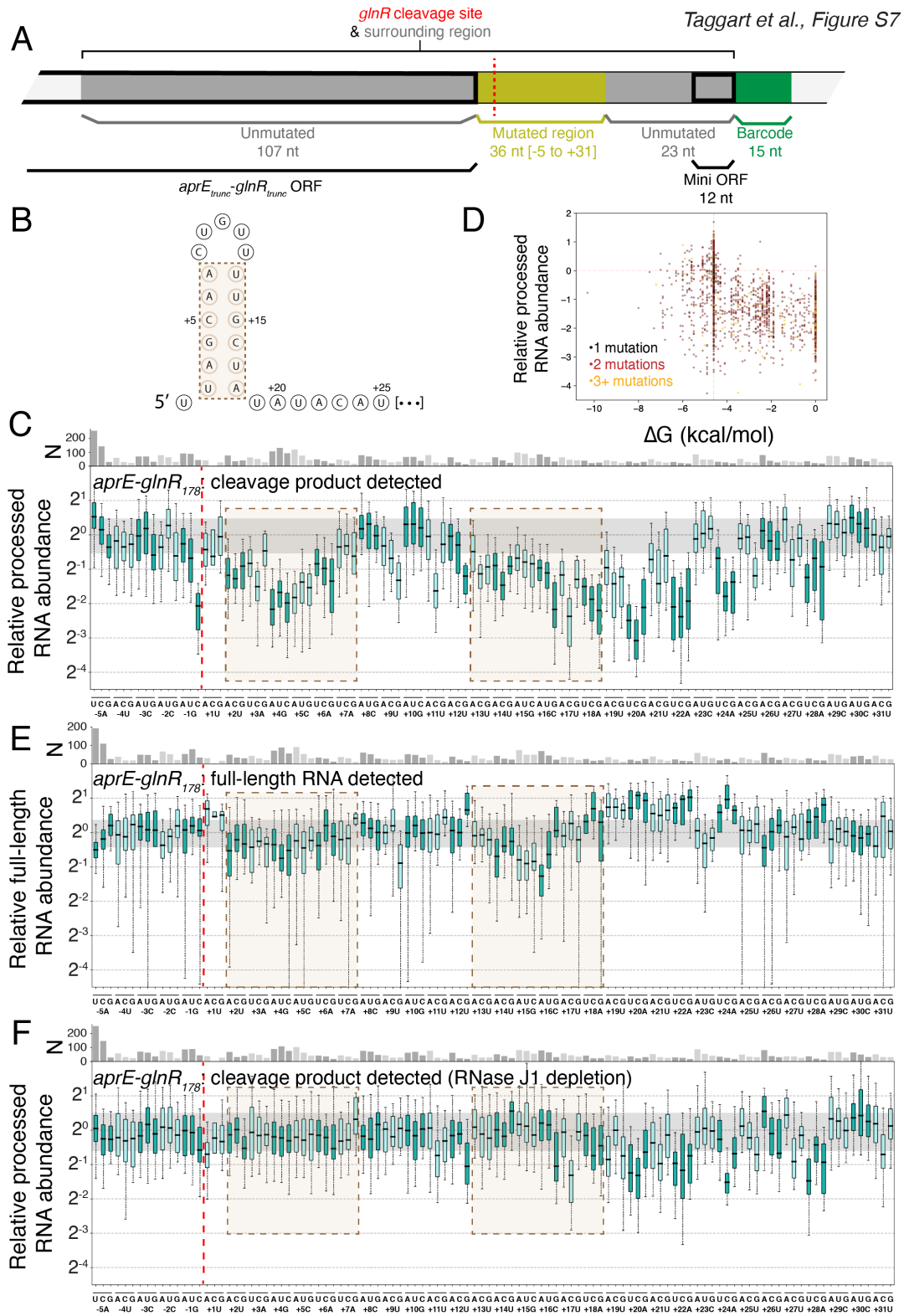

Figure S7. Determinants of *glnR-glnA* mRNA processing by RNase Y.
(A) Schematic of the $glnR_{178}$ construct inserted into the *aprE* MPRA transcript. *glnR*-derived sequence is colored dark gray, with the mutated positions highlighted in yellow. Translated regions are indicated with a thick border, and the variant barcode is indicated in green.

1407 (B) Predicted stem-loop structure at 5′ end of processed RNA. Brown bar corresponds to regions
1408 highlighted in panels C, E, and F.
1409 (C) Impact of all single-nucleotide mutations on the accumulation of barcoded cleaved RNA. Boxplots
1410 show variation between barcodes of identical variant sequence. Number of barcodes captured for each
1411 mutation is indicated above plot. Whiskers indicate 5th and 95th percentile. Gray shaded region indicates
1412 interquartile range for variants of wild-type sequence. Red line indicates position of cleavage. Brown bar
1413 corresponds to positions predicted to form the stem of the structure showed in (B). Nine variants (no more
1414 than two per mutation) have a value of zero are thus not visualized.
1415 (D) Relationship between predicted strength of downstream secondary structure and accumulation of
1416 barcoded cleavage product. All data are derived from experiment 8a (Table S7). The vertical dashed line
1417 indicates the ΔG of the unmutated sequence. 17 variants fall outside of the bounds of this plot.
1418 (E) Impact of all single-nucleotide mutations on the accumulation of barcoded full-length RNA, plotted as
1419 described in (C). 30 variants (no more than two per mutation) have a value of zero are thus not visualized.
1420 (F) Impact of all single-nucleotide mutations on the accumulation of barcoded cleaved RNA in the
1421 context of RNase J1 depletion, plotted as described in (C). Three variants (no more than one per
1422 mutation) have a value of zero are thus not visualized.
1423

1424 **Table S1. List of strains used in this study**
1425
1426 **Table S2. List of oligonucleotide sequences used in this study**
1427
1428 **Table S3. List of identified RNase III cleavage sites**
1429
1430 **Table S4. List of identified RNase Y cleavage sites**
1431
1432 **Table S5. List of putative RNase III cleavage sites with relaxed 5' end sequencing depth threshold**
1433
1434 **Table S6. List of putative RNase Y cleavage sites with relaxed 5' end sequencing depth threshold**
1435
1436 **Table S7. Summary of MPRA experiments conducted in this study**