1   Intronic non-coding RNAs within ribosomal protein coding genes can regulate biogenesis of

2   yeast ribosome

3   Akshara Pande[1,2], Rani Sharma[1], Bharat Ravi Iyengar[3,4], Vinod Scaria[1,2], Beena Pillai[1,2*], Samir K

4   Brahmachari[1,2*]

5   [1]CSIR-Institute of Genomics and Integrative Biology (CSIR-IGIB), New Delhi, India

6   [2]Academy of Scientific and Innovative Research (AcSIR)

7   [3]CSIR-National Chemical Laboratory (CSIR-NCL), Pune, India

8   [4]Department of Chemical Engineering, IIT Bombay, Mumbai, India

9   akshara.pande@igib.in, 07rani.sharma@gmail.com,   bharatravi.654@gmail.com,   vinods@igib.in,

10   beena@igib.in, skb@igib.in

11   *Corresponding authors: Beena Pillai, Samir K Brahmachari

12

13   **Abbreviations** ncRNA, non-coding RNA; isRNA, intronic small RNA; snoRNA, small nucleolar RNA;

14   TF, transcription factor; Pre-rRNA, precursor ribosomal RNA; rDNA, ribosomal DNA; rRNA, ribosomal

15   RNA; ITS, Internal Transcribed Spacer; ETS, External Transcribed Spacer; NTS, Non-Transcribed

16   Spacer; nt, Nucleotide

17

18   **Abstract**

19   The genome of the budding yeast (*Saccharomyces cerevisiae*) has selectively retained introns in

20   ribosomal protein coding genes. The function of these introns has remained elusive in spite of

21   experimental evidence that they are required for the fitness of yeast. Here, we computationally predict

22   novel small RNAs that arise from the intronic regions of ribosomal protein (RP) coding genes in

23   *Saccharomyces cerevisiae*. Further, we experimentally validated the presence of seven intronic small

24   RNAs (isRNAs). Computational predictions suggest that these isRNAs potentially bind to the ribosomal

25   DNA (rDNA) locus or the corresponding rRNAs. Several isRNA candidates can also interact with

26   transcripts of transcription factors and small nucleolar RNAs (snoRNAs) involved in the regulation of

27   rRNA expression. We propose that the isRNAs derived from intronic regions of ribosomal protein coding

28   genes may regulate the biogenesis of the ribosome through a feed-forward loop, ensuring the coordinated

29   regulation of the RNA and protein components of the ribosomal machinery. Ribosome biogenesis and

30   activity are fine-tuned to the conditions in the cell by integrating nutritional signals, stress response and

31   growth to ensure optimal fitness. The enigmatic introns of ribosomal proteins may prove to be a novel

32   and vital link in this regulatory balancing act.

33   **Keywords**: small RNA, rDNA, intron, ribosome, non-coding RNA, rRNA

34   **Introduction**

35   All the 401 introns present in its genome (Saccharomyces Genome Database (SGD)) [1] are found within

36   less than 370 (7%) of 6000 genes in the yeast, *Saccharomyces cerevisiae*. These include introns in

37   snoRNAs, tRNAs and the mitochondrial genome, pushing the number of spliceosomal introns to a shear

38   281. In contrast, the closely related *Schizosaccharomyces pombe*, has an abundance of 4698 introns in

39   2252 genes which constitutes about 45% of nearly 5000 genes in its genome (PomBase) [2].Compared to

40   other eukaryotic genomes of comparable size, the *Saccharomyces cerevisiae* genome is unusually intron

41   free. In spite of this loss of introns in the majority of yeast genes [3], 104 (out of 281) spliceosomal

42   introns are in ribosomal protein coding genes [4].The *S. cerevisiae* genome contains 137 coordinately

43   regulated ribosomal protein coding genes, 76% of which carry one or more introns. Evolutionary analysis

44   of the yeast genome has shown that the Saccharomyces family (exemplified by *S. cerevisiae*) has lost

45   introns extensively in spite of having intron rich ancestors [5]. This has led to the hypothesis that the

46   retained introns have a functional relevance, due to which these have been evolutionarily maintained in

47   the genome, despite the loss of introns in other gene classes[4,6].

48   In recent years, almost all organisms, prokaryotic or eukaryotic, have been shown to express non-coding

49   RNAs (ncRNAs) that carry out an ever growing set of diverse functions (reviewed in [7]). In prokaryotes,

50   some non-coding RNAs are part of innate anti-viral defense [8] while others regulate genes involved in

51   quorum sensing and motility (reviewed in [9]). In eukaryotes, non-coding RNAs have diverse functions:

52   microRNAs are mediators of post-transcriptional gene regulation (reviewed in [10]), while the slightly

53   longer piRNAs are involved in protecting genomes from transposable elements (reviewed in [11]) and

54   long-non-coding RNAs (lncRNAs) are involved in several functions ranging from chromatin organization

55   to structural scaffolding of proteins ([12], [13]). Intronic miRNAs and miRtrons are derived during

56   processing of messenger RNAs and participate in dynamic regulation of gene expression [14].

57   Introns are no longer considered as non-functional regions since it is now clear that introns may harbour

58   functional elements [6]. This, along with the fact that a few introns are selectively retained in yeast,

59   prompted us to explore the possibility that these intronic regions may have functional implications. We

60   used publicly available RNA-Seq data, and found sequence reads mapping to 12 introns in the

61   *Saccharomyces cerevisiae* genome. Seven of these show significant expression levels in at least one of

62   the reported studies. We find that these reads, four of which are derived from introns of ribosomal protein

63   coding genes, can arise from small RNAs that can potentially bind to the rDNA locus or rRNA

64   transcripts. These transcripts belong to the small and large subunits of the eukaryotic ribosome or

65   mitochondrial ribosome. Further, we could detect the expression of these small intronic RNAs

66   (henceforth referred to as isRNAs) in wild type yeast cells. In a strain where one of the introns

67   specifically from *IMD4* gene, was deleted, the expression of the corresponding isRNA was lost. Thus,

68   novel non-coding RNAs that could potentially target rRNA biosynthesis or ribosome assembly are

69   derived from introns of ribosomal protein coding genes. We propose that this constitutes a primitive

70   mechanism of feed-forward regulation that allows for coordinated expression of ribosomal proteins and

71   rRNA.

3

72 **Materials and Methods**

73 RNA-Seq data analysis: To identify small RNAs, that potentially arise from intronic regions, we collected

74 and downloaded publicly available RNA-Seq data from Short Read Archive (SRA) of *S. cerevisiae* in the

75 European Nucleotide Archive (http://www.ebi.ac.uk/ena/). Ten RNA-Seq experiments [ENA: SRP002218

76 [15], ENA:SRP017603 [16], ENA:SRX111550, ENA:SRX111551, ENA:SRX111552,

77 ENA:SRX111553, ENA:SRX391768 [17], ENA:SRX506368, ENA:SRX506456, ENA: SRX525284 [18]

78 available from the database were used for further analysis. FastQC [19] was applied initially to check the

79 quality of RNA-Seq data. Pre-built *S. cerevisiae* genome (version R64, 2011) package was downloaded

80 (http://bowtie-bio.sourceforge.net/tutorial.shtml). Subsequently, SolexaQA [20] was used to filter and

81 trim the reads of bad quality (phred score <20). Bowtie (version 0.12.9) [21] was used to align the RNA-

82 seq data with the reference genome (*S. cerevisiae*). We designed a computational pipeline to screen all the

83 intronic regions to identify regions, which have a read coverage of more than 10X. We downloaded the

84 intron sequences and their positions from yeast intron databases (http://www.yeastgenome.org/,

85 http://metarray.ucsc.edu/cgi-bin/intron/yirIntrondb and http://intron.ucsc.edu/yeast4.1/). We merged all

86 the BAM files with read coverage >10X. Subsequently we converted BAM to BED for the specific co-

87 ordinates and filtered out alignments that are opposite to the orientation of the gene. For instance, all +

88 strand alignments for *RPL18A* would be removed because the gene is on the negative strand. Further, read

89 count per nucleotide for each intron was calculated and plotted.

90 Potential targets of isRNA in yeast mRNA sequences were predicted using miRanda v3.3a algorithm and

91 RNAhybrid 2.2. A ΔG cutoff of -30kcal/mol for putative small RNAs-ncRNA interaction was used. We

92 found that putative small RNAs could bind to several regions of the RDN-locus. RiboVision software

93 (http://apollo.chemistry.gatech.edu/RiboVision/) was used to visualize the mapping of putative small

94 RNAs in the small subunit of rRNA and large subunit of rRNA [22]. To establish the ribosomal targets in

95 the RDN-locus, UCSC BLAT [23] was used.

4

96   We re-shuffled isRNA sequences to generate 1000 scrambled sequences to explore the occurrence of

97   target sites expected by chance. Potential targets of these thousand scrambled sequences were predicted

98   using miRanda v3.3a algorithm with ΔG cutoff of -30kcal/mol. The positions of potential target sites of

99   isRNAs checked with predicted target sites of thousand scrambled sequences with permitted overlap of

100 90%. The targets of isRNA from *COF1*, *IMD4*, *RPL28*, *RPL7B*, *RPL18A* and *RPS29A* (p-value < 0.1

101 with 90% overlap) were found to be significantly enriched compared to corresponding scrambled

102 sequences.

103 Experimental Validation: Total RNA was isolated using TRIzol, followed by cDNA synthesis with the

104 oligodT linked primer (5'-CTCAATCGTACATAGAAACAGGGATCTTTTTTTTTTTTTTTTTT-3') in

105 a two-step reaction using M-MuLV Reverse Transcriptase. PCR was carried out with the cDNA using

106 gene   specific   forward   primer   (*RPL7B*-5'-TTGTTAGTATCTATTACTTG-3';   *YRA1*-5'-

107 GACATGTTTCCCATAGCTAT-3';   *COF1*-5'TATGTCCTTTCTTTCCTCCC-3';   *RPS29A*-5'-

108 ATCATTGCGTATGCTGGCTG-3';   *RPL18A*-5'   CTGATGTGACAAATTATTGG-3';   *RPL28*-5'-

109 TTACGGCTATAAAAGGTAAC-3';   *IMD4*-5'-TTTCTTTTCTGGCTGGGCTG-3')   and   a   universal

110 reverse primer (5'-CTCAATCGTACATAGAAACAGGGATC-3').

111 **Results**

112 We used publicly available RNA-Seq data from 10 experiments (see methods for details) to map reads to

113 the 401 introns of *S. cerevisiae*. We selected seven putative small RNAs after filtering out candidates with

114 less than of 10X coverage. Read depths of the seven selected cases are shown in Figure 1. In the intron of

115 *COF1* and *RPL18A*, distinct peaks corresponding to the isRNA were seen. But in all other introns, no

116 such peak was evident. The *IMD4* intron includes a snRNA (snR54). The abundantly expressed reads

117 corresponding to the snRNA are clearly seen as a distinct peak within the first half of intron. At much

118 lower levels, reads corresponding to the predicted isRNA are also seen between 260-380 nt.

5

119 If the predicted non-coding RNAs are functionally important, the region is likely to be conserved in

120 related yeast. We used whole genome data for 11 related species (*Lachancea waltii*, *S. arboricola*, *S.*

121 *bayanus*, *S. boulardii*, *S. kudriavzevii*, *S. mikatae*, *S. paradoxus*, *S. pastorianus*) to assess the conservation

122 of these regions. All seven intronic non-coding RNA sequences were conserved in several Sacharomyces

123 species, with *S. boulardii* and *S. pastorianus* showing the maximum similarity (Figure 2). We also

124 examined the regions in the context of flanking regions to compare the degree of conservation in small

125 RNAs with the overall level of conservation in intronic regions.

126 We validated the presence of putative small RNAs in yeast cells, by artificially tailing it using polyA

127 polymerase and amplifying the specific product using an isRNA specific forward primer and oligodT

128 fused to the universal reverse primer. Using this assay (Figure 3A), we could detect the presence of seven

129 isRNAs (*RPL7B*, *YRA1*, *COF1, RPS29A*, *RPL18A, RPL28* and *IMD4)*. A strain with a deletion that spans

130 the intron of *IMD4* (originally created by J. Parenteu [4]) did not express the small RNA (Figure 3B). We

131 also verified the presence of the isRNA derived from the intron of *RPS29A* using an RT-PCR assay with

132 small RNA specific forward and reverse primers (Figure 3C). The observed size of the product was close

133 to the size predicted. For instance, in the *COF1* intron (179nt), the predicted isRNA spans 82-142nd

134 position in the intron, where an increase in reads is seen (Figure 1).

135 We further extended the analysis by using the same program to identify potential targets of the seven

136 isRNAs in the non-protein coding genome. Several targets were found for each of the isRNAs (Table 2)

137 in genomic regions that give rise to rRNA, tRNA or snoRNA. We noted that the target sites in rRNAs had

138 a higher alignment score. The mature rRNAs are produced after post-transcriptional processing of long

139 precursor RNAs transcribed from a single rDNA locus on chromosome 12 [24]. The locus consists of

140 100-200 repeating units of 9.1kb, each consisting of a 5S and a 35S precursor which get processed to

141 form the 18S, 5.8S and 25S rRNA. Besides these rRNA coding regions, the locus consists of Internal

142 Transcribed Spacers (ITS1, 2); External Transcribed Spacers (ETS1, 2) and Non-Transcribed Spacer

143 regions (NTS1, 2). Therefore, we examined these 62 potential targets in the context of the rDNA locus.

144 As shown in figure 4 the small RNAs can potentially cover a substantial region of the rDNA locus, (up to

145 2.64kb of the 9.1kb locus; (Supplementary File 2). To ensure that these targets may not be predicted by

146 chance, we scrambled the isRNA sequence and predicted potential targets in the same loci. For all seven

147 isRNA targets were found to be significantly enriched compared to corresponding scrambled sequences

148 (p-value<0.1 with 90% overlap). The potential binding sites of the intronic small RNAs are enriched in

149 the internal (ITS), external (ETS) and non-transcribed (NTS) spacer sequences punctuating the rDNA

150 locus.

151 We visualized the putative binding sites in the context of the secondary structure of the rRNA scaffold by

152 mapping the seven non-coding RNAs on the rRNA using the RiboVision software [22]. In the large

153 subunit all except *RPS29A* derived ncRNA have potential binding sites unaffected by the secondary

154 structures. The 5S rRNA is potentially targeted by isRNA from *RPL7B* while isRNAs from *YRA1*, *RPL28*

155 and *RPL7B* find overlapping putative targets in the 5.8S region (Figure 4).

156 We searched for potential RNA-RNA hybrids between intronic small RNA and yeast mRNA based on

157 predicted ΔG of binding [25,26]. Speculating that isRNAs may also bind to protein coding transcripts, we

158 looked for potential targets in mRNAs of ribosome biogenesis genes. Potential targets of three isRNAs

159 (isCOF1, isIMD4, and isYRA1) were directly related to ribosomal protein or rRNA biogenesis (Table 3).

160 A particularly interesting target is in CURI complex involved in the transcription of ribosomal genes.

161 Rudra et al. [27] have earlier shown that IFH1 protein shuttles between the CURI complex, which is

162 involved in pre-rRNA processing and the RAP1 containing transcriptional complex at promoters of RP

163 genes. This shuttling has been proposed to interface between ribosomal protein production and pre-rRNA

164 processing. Our results suggest that the intronic RNA released during the splicing of *IMD4*, *YRA1* and

165 *COF1* mRNA can potentially bind to and regulate IFH1 mRNA (Table 3).

166 **Discussion**

167   The ribosomes are amongst the most ancient and largest multi-subunit ribonucleoprotein machinery of the

168   cell. A dedicated transcriptional machinery, run by the highly processive RNA polymerase I, produces a

169   6.6kb pre-rRNA transcript accounting for about 60% of the total cellular RNA transcription [28].The

170   processing machinery then cleaves and processes this large transcript into the mature rRNA transcripts

171   initially in the nucleus and subsequently in the cytoplasm. Folding of the rRNA, processing and assembly

172   of ribosomal proteins are highly regulated events involving several protein factors. The production of

173   ribosomes is closely tuned to the state of the cell by incorporating information like nutrient availability,

174   growth, cell division and stress [29]. It is widely held that ribosomal protein production is matched to

175   rRNA transcription and processing although the molecular basis of this crosstalk is not fully understood.

176   In prokaryotes, ribosomal protein biogenesis is coordinately regulated at the transcriptional level by a

177   common promoter that regulates RP genes from a polycistronic transcript and at the translational level

178   through a simple feedback by some of the ribosomal proteins that bind to the mRNA and prevent

179   translation [30,31] . In eukaryotes, it is known that the RP genes and rRNA are coordinately regulated but

180   they are neither clustered in the genome nor do they share obvious regulatory motifs. Here we show that

181   ncRNA arising from the introns of ribosomal protein coding genes could potentially form this missing

182   link in the coordinate regulation of ribosome biogenesis (Figure 5).

183   It was intuitively appealing to invoke the possibility that the small ncRNA could also modulate

184   transcription factors and RNA processing enzymes that are involved in ribosome biogenesis. Such ability

185   would allow feedback regulation of transcription and processing of the rRNA and ribosomal proteins.

186   Three isRNAs, namely, isCOF1, isIMD4 and isYRA1 can potentially bind to the mRNA of two proteins,

187   IFH1 and STB3 that are involved in regulation of rRNA or ribosomal protein genes. The *IFH1* mRNA

188   has potential binding sites for all three isRNAs: isCOF1, ncIMD4 and isYRA1. The IFH1 protein is a

189   critical component of two complexes that regulate ribosome biogenesis. Its presence in the transcriptional

190   complex with RAP1 and Fhl1 triggers the rapid transcription of Ribosomal protein genes. When the

191   transcription of RP genes is repressed, IFH1 leaves the promoter and localizes with the CURI complex

8

192   involved in the processing of pre-rRNA [32].*STB3* codes for a protein that binds to ribosomal RNA

193   processing element motifs and regulates gene expression [33] and its mRNA harbors potential binding

194   sites for ncCOF1.

195   Interestingly, in higher mammals the first intron in some ribosomal proteins has been shown to harbour a

196   motif [34] potentially involved in the coordinate regulation of ribosome biogenesis. Recently, Chak *et al*.

197   [35] have shown, from deep sequencing studies, that the highly repetitive rDNA arrays of Drosophila

198   harbor a highly conserved non-canonical miRNA. In this context, our study also implies that previously

199   unknown non-coding RNAs may be involved in dynamic regulation of ribosome activity in

200   *Saccharomyces cerevisiae*.

201   A yeast cell has about 150 tandem copies of the rDNA locus. Extrapolation of the RNA-Seq data [29]

202   suggests that there are about 50 mRNA transcripts of each ribosomal protein per cell (3000 RP gene

203   mRNA/60,000 mRNA transcripts: [36]) , suggesting that hundreds of intronic ncRNA may be present in

204   the cell. Furthermore, [3,6] have shown that introns are maintained in the cell even after splicing. As

205   shown in Figure 4 the ncRNAs derived from seven introns could potentially bind to 29% of the 9.1kb

206   rDNA locus. Thus the relative concentration of ribosomal proteins, mRNA, intronic ncRNA and rDNA

207   locus repeats in a cell suggest that the interactions are feasible.

208   Hooks *et al*. 2016 found evolutionary conserved structures in YRA1, RPL28, RPL7A/B (exactly the same

209   region as identified in this study), RPL18A/B and showed expression of the introns post splicing [6]. One

210   measure of the relevance of these introns for biology would be the loss in fitness due to mutations.

211   Parenteau *et al.* [4] through systematic deletion of introns in ribosomal proteins showed that the introns of

212   ribosomal proteins are involved in expression of other ribosomal protein genes. None of the intron

213   deletions resulted in lethality but several showed reduced fitness and altered expression of ribosomal

214   protein genes. The deletion of the intron of *RPS29A* resulted in a significant loss of fitness [4] and an

215   accumulation of the 20S precursor presumably due to reduced cleavage at D site between 18S and the

9

216   ITS1. In our analysis, we find that the *RPS29A* derived ncRNA has a potential binding site matching this

217   location.

218   The ribosome has proven to be a valuable drug target in prokaryotic systems. Many potent naturally

219   occurring antibiotics like streptomycin, penicillin and cephalosporin act by interfering the ribosome

220   biogenesis and function. Deletion of introns in *RPS29A*, which reduced fitness, also led to reduced

221   sensitivity of yeast to caffeine [4]. However, high conservation of ribosomal proteins and rRNA in

222   eukaryotes prevents its use as an anti-fungal drug target. Our findings imply that novel endogenous

223   isRNAs may bind to the rRNA and could potentially offer valuable sites for designing small molecule

224   inhibitors of translation.

Competing interests

226   The authors declare that they have no competing interests.

Authors' Contributions

228   SKB designed the project. AP and BRI analyzed the data. RS carried out experimental validation. BP and

229   SKB wrote the manuscript, VS improved the manuscript. All authors read and approved the manuscript.

Acknowledgments

References

235   [1] J.M. Cherry, E.L. Hong, C. Amundsen, R. Balakrishnan, G. Binkley, E.T. Chan, K.R. Christie, M.C.
236   Costanzo, S.S. Dwight, S.R. Engel, D.G. Fisk, J.E. Hirschman, B.C. Hitz, K. Karra, C.J. Krieger, S.R.
237   Miyasato, R.S. Nash, J. Park, M.S. Skrzypek, M. Simison, S. Weng, E.D. Wong, Saccharomyces
238   Genome Database: the genomics resource of budding yeast, Nucleic Acids Res 40 (2012) D700-
239   705.

240   [2] V. Wood, M.A. Harris, M.D. McDowall, K. Rutherford, B.W. Vaughan, D.M. Staines, M. Aslett, A. Lock,
241   J. Bahler, P.J. Kersey, S.G. Oliver, PomBase: a comprehensive online resource for fission yeast,
242   Nucleic Acids Res 40 (2012) D695-699.

243   [3] K.B. Hooks, D. Delneri, S. Griffiths-Jones, Intron evolution in Saccharomycetaceae, Genome Biol Evol
244   6 (2014) 2543-2556.

245   [4] J. Parenteau, M. Durand, G. Morin, J. Gagnon, J.F. Lucier, R.J. Wellinger, B. Chabot, S.A. Elela, Introns
246   within ribosomal protein genes regulate the production and function of yeast ribosomes, Cell
247   147 (2011) 320-331.

248   [5] J.E. Stajich, F.S. Dietrich, S.W. Roy, Comparative genomic analysis of fungal genomes reveals intron-
249   rich ancestors, Genome Biol 8 (2007) R223.

250   [6] N.S. Hooks KB, Parker S, Griffiths-Jones S, Delneri D., Novel Intronic RNA Structures Contribute to
251   Maintenance of Phenotype in Saccharomyces cerevisiae., Genetics. **203** (2016) 469-481.

252   [7] J.S. Mattick, I.V. Makunin, Non-coding RNA, Hum Mol Genet 15 Spec No 1 (2006) R17-29.

253   [8] P. Horvath, R. Barrangou, CRISPR/Cas, the immune system of bacteria and archaea, Science 327
254   (2010) 167-170.

255   [9] M.B. Miller, B.L. Bassler, Quorum sensing in bacteria, Annu Rev Microbiol 55 (2001) 165-199.

256   [10] R.J. Jackson, N. Standart, How do microRNAs regulate gene expression?, Sci STKE 2007 (2007) re1.

257   [11] T. Thomson, H. Lin, The biogenesis and function of PIWI proteins and piRNAs: progress and
258   prospect, Annu Rev Cell Dev Biol 25 (2009) 355-376.

259   [12] T.R. Mercer, J.S. Mattick, Structure and function of long noncoding RNAs in epigenetic regulation,
260   Nat Struct Mol Biol 20 (2013) 300-307.

261   [13] J.E. Wilusz, H. Sunwoo, D.L. Spector, Long noncoding RNAs: functional surprises from the RNA
262   world, Genes Dev 23 (2009) 1494-1504.

263   [14] S. Gokhale, M. Hariharan, S.K. Brahmachari, C. Gadgil, A simple method for incorporating dynamic
264   effects of intronic miRNA mediated regulation, Mol Biosyst 8 (2012) 2145-2152.

265   [15] R.P. Wu CY, Gifford DK, Fink GR, Control of transcription by cell size, PLoS Biol 8 (2010).

266   [16] U.B. Baker LA, Dewell S, Chait BT et al., The yeast Snt2 protein coordinates the transcriptional
267   response to hydrogen peroxide-mediated oxidative stress. , Mol Cell Biol 33 (2013) 3735-3748.

268   [17] E.S. Subtelny AO, Chen GR, Sive H et al. , Poly(A)-tail profiling reveals an embryonic switch in
269   translational control., Nature 508 (2014) 66-71.

270   [18] H.D. Jin M, Backues SK, Freeberg MA et al., Transcriptional regulation by Pho23 modulates the
271   frequency of autophagosome formation., Curr Biol 24 (2014) 1314-1322.

272   [19] A. S., FastQC: a quality control tool for high throughput sequence data. ,(2010).

273   [20] P.D. Cox MP, Biggs PJ. , SolexaQA: At-a-glance quality assessment of Illumina second-generation
274   sequencing data. , BMC Bioinformatics. (2010).

275   [21] T.C. Langmead B, Pop M, Salzberg SL., Ultrafast and memory-efficient alignment of short DNA
276   sequences to the human genome. Genome Biol., Genome Biol. **10** (2009) R25.

277   [22] C.R. Bernier, A.S. Petrov, C.C. Waterbury, J. Jett, F. Li, L.E. Freil, X. Xiong, L. Wang, B.L. Migliozzi, E.
278   Hershkovits, Y. Xue, C. Hsiao, J.C. Bowman, S.C. Harvey, M.A. Grover, Z.J. Wartell, L.D. Williams,
279   RiboVision suite for visualization and analysis of ribosomes, Faraday Discuss 169 (2014) 195-207.

280   [23] K.W. J, BLAT--the BLAST-like alignment tool, Genome Res. 12 (2002) 656-664.

281   [24] J. Venema, D. Tollervey, Ribosome synthesis in Saccharomyces cerevisiae, Annu Rev Genet 33
282   (1999) 261-311.

283   [25] A.J. Enright, B. John, U. Gaul, T. Tuschl, C. Sander, D.S. Marks, MicroRNA targets in Drosophila,
284   Genome Biol 5 (2003) R1.

285   [26] M. Rehmsmeier, P. Steffen, M. Hochsmann, R. Giegerich, Fast and effective prediction of
286   microRNA/target duplexes, RNA 10 (2004) 1507-1517.

287 [27] D. Rudra, J. Mallick, Y. Zhao, J.R. Warner, Potential interface between ribosomal protein production
288   and pre-rRNA processing, Mol Cell Biol 27 (2007) 4815-4824.
289 [28] J.L. Woolford, Jr., S.J. Baserga, Ribosome biogenesis in the yeast Saccharomyces cerevisiae, Genetics
290   195 (2013) 643-681.
291 [29] H. Adhikari, P.J. Cullen, Metabolic respiration induces AMPK- and Ire1p-dependent activation of the
292   p38-Type HOG MAPK pathway, PLoS Genet 10 (2014) e1004734.
293 [30] B. Li, J. Vilardell, J.R. Warner, An RNA structure involved in feedback regulation of splicing and of
294   translation is critical for biological fitness, Proc Natl Acad Sci U S A 93 (1996) 1596-1600.
295 [31] S.W. Fewell, J.L. Woolford, Jr., Ribosomal protein S14 of Saccharomyces cerevisiae regulates its
296   expression by binding to RPS14B pre-mRNA and to 18S rRNA, Mol Cell Biol 19 (1999) 826-834.
297 [32] Y. Zhao, K.B. McIntosh, D. Rudra, S. Schawalder, D. Shore, J.R. Warner, Fine-structure analysis of
298   ribosomal protein gene transcription, Mol Cell Biol 26 (2006) 4853-4862.
299 [33] D. Liko, M.K. Conway, D.S. Grunwald, W. Heideman, Stb3 plays a role in the glucose-induced
300   transition from quiescence to growth in Saccharomyces cerevisiae, Genetics 185 (2010) 797-810.
301 [34] H. Hu, X. Li, Transcriptional regulation in eukaryotic ribosomal protein genes, Genomics 90 (2007)
302   421-423.
303 [35] L.L. Chak, J. Mohammed, E.C. Lai, G. Tucker-Kellogg, K. Okamura, A deeply conserved, noncanonical
304   miRNA hosted by ribosomal DNA, RNA 21 (2015) 375-384.
305 [36] D. Zenklusen, D.R. Larson, R.H. Singer, Single-RNA counting reveals alternative modes of gene
306   expression in yeast, Nat Struct Mol Biol 15 (2008) 1263-1271.

307
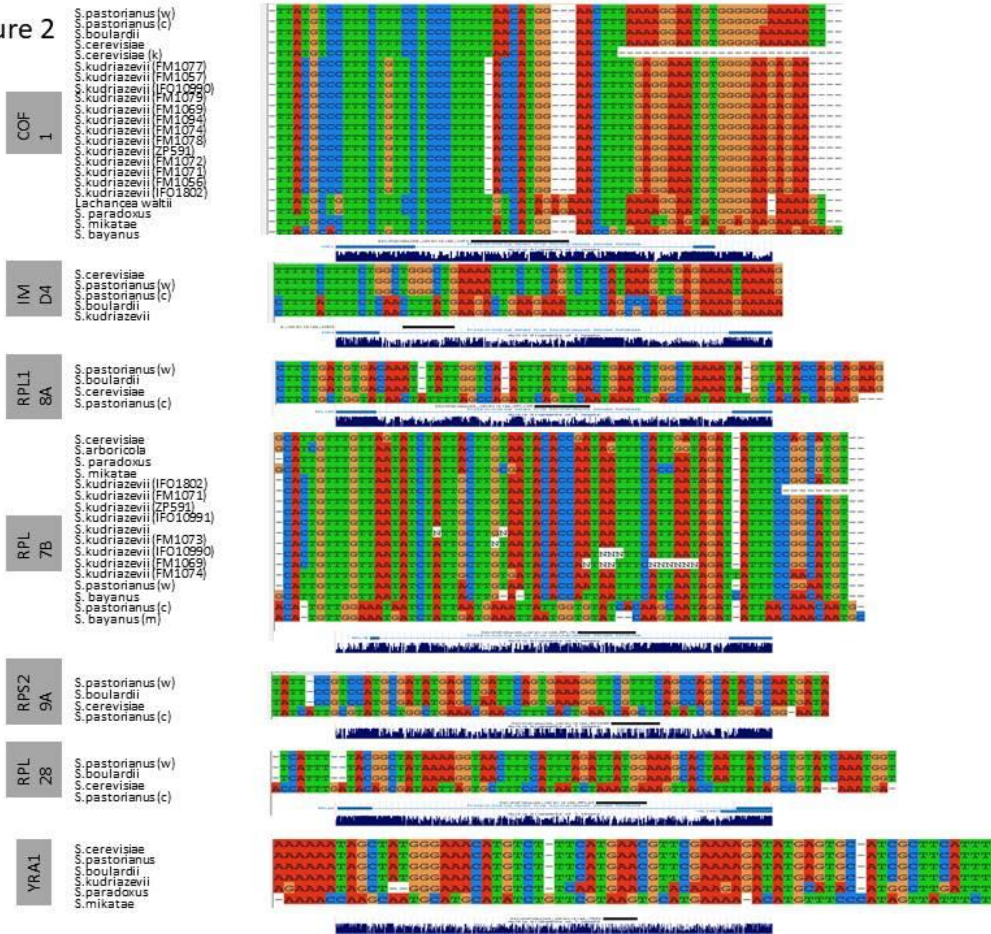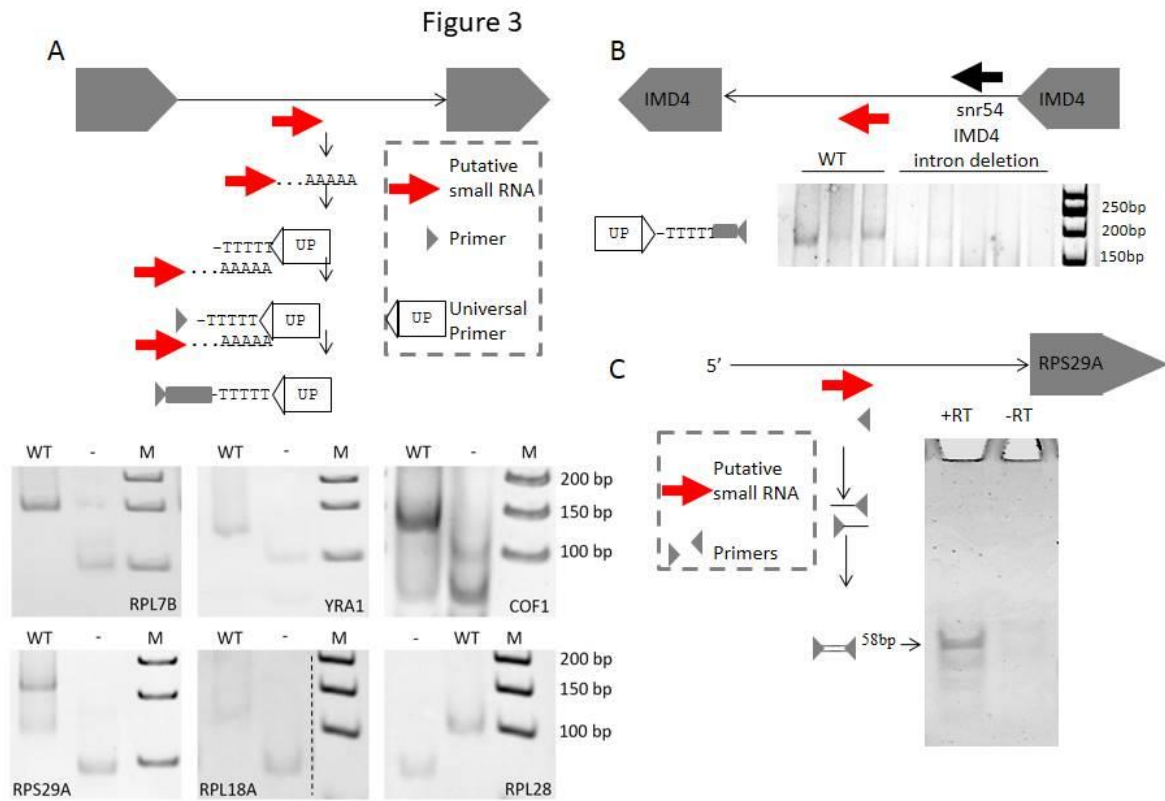
308   Figures

Figure 1

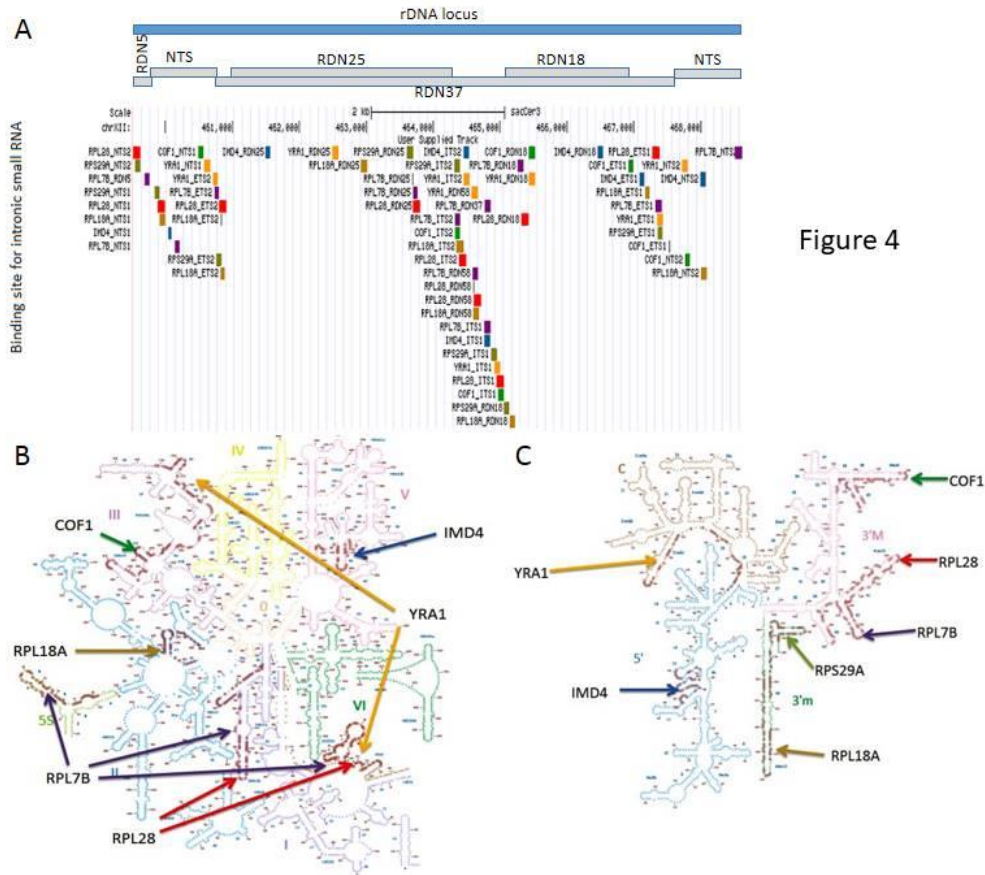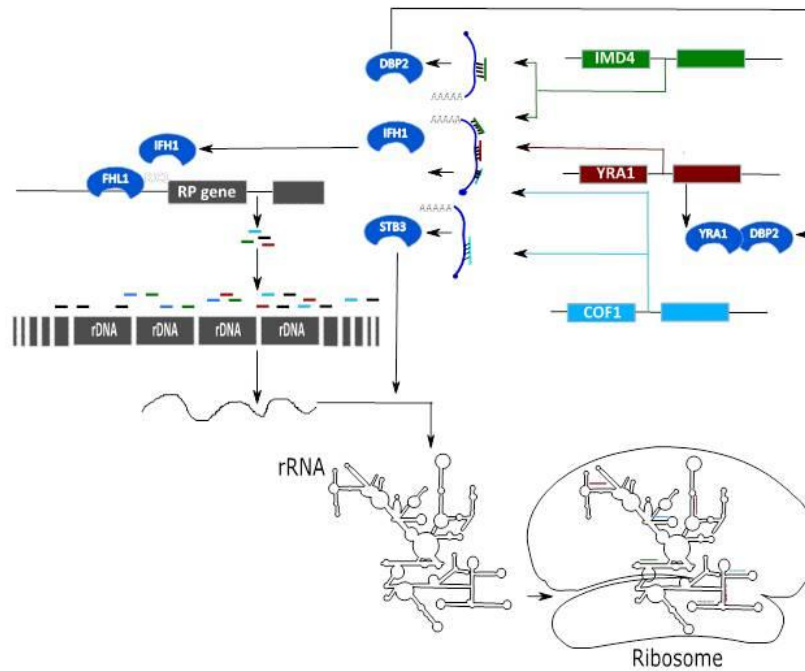

309

Figure 2

310

14
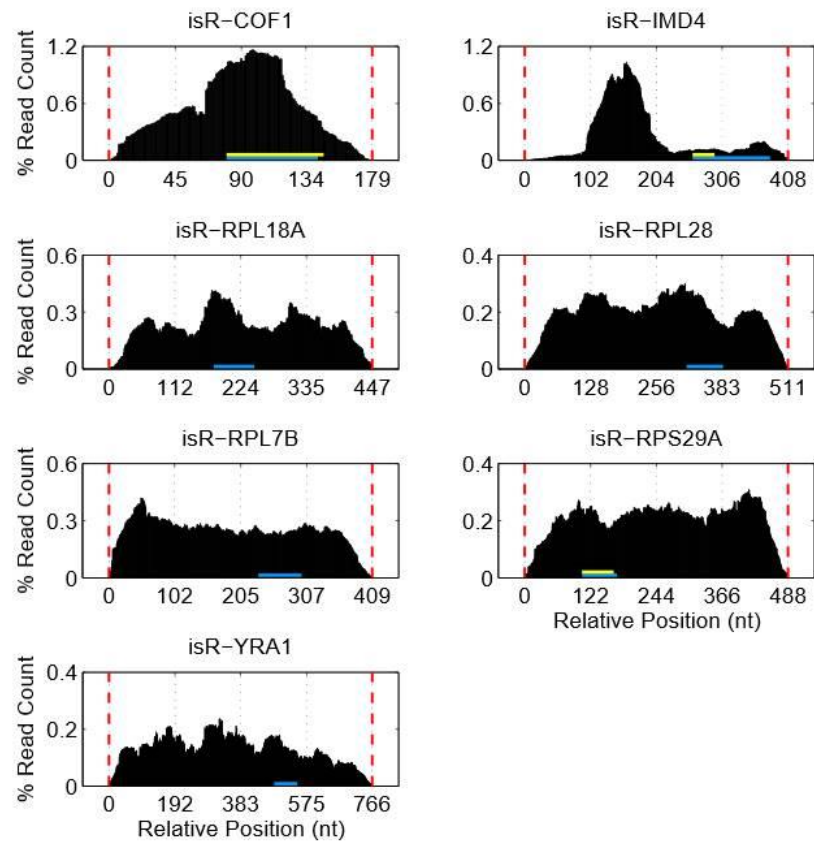
Figure 3

311

Figure 4

312

Figure 5

313

314 **Figure legends**

315 **Figure 1: Read depth within intronic regions.** X axis represents the relative position of the nucleotide
316 from start of intron and Y axis represents the percentage of read counts. Reads from several experiments
317 are mapped to the region between (isR-COF1) 80 to 142 nt of *COF1* gene, (isR-IMD4) 260 to 380 nt of
318 *IMD4* gene, (isR-RPL18A) 178 to 247 nt of *RPL18A* gene, (isR-RPL28) 314 to 385 nt of *RPL28* gene,
319 (isR-RPL7B) 232 to 299 nt of *RPL7B* gene, (isR-RPS29A) 106 to 171 nt of *RPS29A* gene and (isR-
320 YRA1) 481 to 548 nt of *YRA1* gene. Yellow horizontal arrow shows the inferred 3' end of isRNA and
321 blue horizontal line shows the end of bioinformatically predicted stem loop.

322 **Figure 2: Conservation of intronic region that give rise to non-coding transcripts in *S. cerevisiae*.**
323 The sequence of the non-coding transcripts in *S. cerevisiae* was used as a query to compare to genomes of
324 several Saccharomyces species. Multiple Sequence alignment of the results was done using ClustalX.
325 First letter of strain name is included in parentheses wherever applicable.

326 **Figure 3: Detection of small RNA from introns in yeast.** Line diagrams depict the position of the
327 predicted small RNA (in red), and the strategy used to detect it. **(A)** In the left panel, RT PCR was carried
328 out for the small RNA predicted from *RPL7B*, *YRA1*, *COF1, RPS29A*, RPL18A*, RPL28* and *IMD4*
329 (+RT).DNA contamination was ruled out by repeating the experiment in the absence of reverse
330 transcriptase (-RT). In the left panel, RNA from wild type strain was polyadenylated in vitro, and a
331 corresponding cDNA was synthesized using a universal primer.A small RNA specific primer and the

17

332 universal primer were used in combination to detect the small RNA. Control reactions were also carried
333 out in the absence of reverse-transcriptase (-) to rule out the possibility of trace DNA contamination
334 giving rise to the product. **(B)** In the right panel, RNA from wild type strain and a strain wherein the
335 intron of *IMD4* was deleted were compared. The expected product was seen in three biological replicates
336 (black line) of the wild type strain but not in the intron deleted *IMD4*.The PCR products were
337 electrophoresed on 8% polyacrylamide gels and visualized by Ethidium bromide staining. **(C)** Intronic
338 RNA from *RPS29A* was also detected at the predicted size using gene specific primers.

339 **Figure 4: A) Potential binding sites of small RNAs in rDNA locus.** Blue bars; visualized using UCSC
340 browser. Small RNAs arising from intronic regions of different genes have been assigned specific colors
341 (*RPL28*-red; *RPS29A*-olive; *RPL7B*-mauve, *RPL18A*-brown; *IMD4*-blue; *COF1*-green; *YRA1*-orange).For
342 instance, the putative binding sites of the small RNA derived from *RPL28* gene (shown in red) are at
343 NTS, ETS, ITS, RDN58, RDN18, and RDN25 regions of rDNA locus. The spacer sequences, both
344 internal (ITS) external (ETS) and non-transcribed (NTS) punctuating the rDNA locus, are rich in putative
345 binding sites for small RNAs. **B) Potential binding sites of small RNAs in the ribosomal RNA scaffold
346 of the Large Subunit**: Six of seven small RNAs, labeled as per their source gene, have potential binding
347 sites within the 5S, 5.8S and 25S subunits. **C) Potential binding sites of small RNAs in the ribosomal
348 RNA scaffold of the Small Subunit**: All seven small RNAs, labeled as per their source gene, have
349 potential binding sites within the 18S rRNA of the small subunit. One of the sites is potentially targeted
350 by two overlapping small RNAs.

351 **Figure 5: A brief overview of the hypothesized regulatory network.** Small RNAs arising from intronic
352 regions of ribosomal protein coding genes (RP genes) and *YRA1*, *COF1* and *IMD4* genes make a network.
353 The small RNAs arising from introns of *YRA1*, *COF1* and *IMD4* can potentially target the transcription
354 Factor IFH1, which is involved in expression of ribosomal protein genes and STB3 involved in rRNA
355 processing. The small RNAs from RP genes have several potential binding sites in the rDNA locus and
356 rRNA scaffold.

357 **Table 1: Details of prediction and validation of isRNAs**

isRNA (arising from intron of)	Intron size (nt)	Predicted RNA size (nt)
COF1	179	63
YRA1	766	68
IMD4	408	61
RPS29A	487	66
RPL28	511	72
RPL7B	409	68

358

18

359 **Table 2: Non-coding RNAs (ncRNAs) arising from intronic region interacts with other ncRNAs**

ncRNA (arising from intron of)	Number of non-coding target sites (ΔG <-30kcal/mol)	Percent of RDN locus covered (ChrXII; 9.1Kb)	rRNA	tRNA	snoRNA	miscRNA (snRNA, lincRNA etc)
RPL28	129	5.7%	21	68	25	15
RPL7B	117	6.8%	27	31	42	17
RPS29A	99	4.7%	19	25	38	17
YRA1	107	6.0%	21	8	57	21
COF1	58	3.5%	17	-	25	15
RPL18A	243	6.1%	20	140	62	20
IMD4	104	3.5%	19	37	30	18

360

361 **Table 3: Putative interactions between non-coding RNAs and transcription factors (TF)**

isRNA	Target Gene	mfe (Kcal/mol)	Binding pattern
COF1	IFH1	-39.8	
YRA1	IFH1	-42.8	
IMD4	IFH1	-40.0	

COF1	STB3	-38.5	

362

363

364

365

366