

Title: Global analysis of transcriptionally engaged yeast RNA polymerase III reveals extended tRNA transcripts

Authors: Tomasz W. Turowski^{1,3}, Ewa Makala², Clementine Delan-Forino¹, Camille Sayou¹
Magdalena Boguta^{2,4}, David Tollervey^{1,4}

Affiliations:

¹Wellcome Trust Centre for Cell Biology, University of Edinburgh, Michael Swann Building, Kings Buildings, Mayfield Road, Edinburgh EH9 3JR, Scotland

²Institute of Biochemistry and Biophysics, Polish Academy of Sciences, ul. Pawinskiego 5a, 02-106 Warsaw, Poland

³Institute of Biotechnology, Faculty of Chemistry, Warsaw University of Technology, ul. Noakowskiego 3, 00-664 Warsaw, Poland

⁴ Correspondence to: d.tollervey@ed.ac.uk, magda@ibb.waw.pl

Running title: Mapping RNA polymerase III by CRAC

Key words: RNA polymerase III, tRNA, UV cross-linking, RNA-protein interaction, Maf1, Nab2, exosome, transcription termination, surveillance

ABSTRACT

RNA polymerase III (RNAPIII) synthesizes a range of highly abundant small stable RNAs, principally pre-tRNAs. Here we report the genome-wide analysis of nascent transcripts attached to RNAPIII under permissive and restrictive growth conditions. This revealed strikingly uneven polymerase distributions across transcription units, generally with a predominant 5' peak. This peak was higher for more heavily transcribed genes, suggesting that initiation site clearance is rate limiting during RNAPIII transcription. Down-regulation of RNAPIII transcription under stress conditions was found to be uneven; a subset of tRNA genes showed low response to nutrient shift or loss of the major transcription regulator Maf1, suggesting potential "housekeeping" roles. Many tRNA genes were found to generate long, 3'-extended forms due to read-through of the canonical poly(U) terminators. The degree of read-through was anti-correlated with the density of T-residues in the coding strand, and multiple, functional terminators can be located far downstream. The steady-state levels of 3'-extended pre-tRNA transcripts are low, apparently due to targeting by the nuclear surveillance machinery; especially the RNA-binding protein Nab2, cofactors for the nuclear exosome and the 5'-exonuclease Rat1.

INTRODUCTION

Transcription of nuclear DNA in eukaryotes is carried out by at least three different RNA polymerases. RNA polymerase III (RNAPIII) is specialized for high level synthesis of small noncoding RNAs. The most abundant products of RNAPIII-dependent transcription are the 5S rRNA and the many pre-tRNA species. In addition RNAPIII synthesizes numerous, less abundant small RNAs that are involved in diverse cellular processes, including protein translocation and the processing of pre-rRNA and pre-tRNAs (Dieci et al. 2013).

The nuclear genome of *Saccharomyces cerevisiae* contains 275 actively transcribed tRNA genes (including a tRNA of undetermined specificity (tX(XXX)D)). These are grouped into 20 isotypes each charged with a single amino acid, which are subdivided into 41 isoacceptors that each recognize the same anticodon sequence(s) (Hani and Feldmann 1998; Chan and Lowe 2009). The reported lengths of the primary transcripts vary between 72-133 nt, and around 25% of pre-tRNAs include introns. Primary pre-tRNA transcripts undergo 5' and 3' maturation and intron excision to generate the mature tRNAs.

In tRNA genes the transcription machinery recognizes conserved promoter elements, termed box A and box B, which are located within the transcribed region and form a bipartite binding site

for the six-subunit basal transcription factor TFIIC (Acker et al. 2013). Box A starts at position +8 of the mature tRNA, and the transcription start site is most frequently located 18-20 nt upstream (Dieci et al. 2013). Within yeast tRNA genes, boxes A and B are localized 31-93 nt apart and correspond to the universally conserved D- and T-loops in the tRNA structure. Internally located A and B boxes are also the main *cis*-acting control elements for most other RNAPIII transcription units.

Among nuclear RNA polymerases, RNAPIII has the most direct-acting termination signals, which consist of a tract of A residues on the template DNA strand; minimum lengths are reported to be A₄ for human and A₅ or A₆ for yeast (Braglia et al. 2005; Iben and Maraia 2012; Arimbasseri et al. 2013). The distinctive presence of a 3' poly(U) tract in the transcript apparently makes prediction of the 3' ends of RNAPIII transcription units relatively easy (Dieci et al. 2013). However, recent data indicate that the RNAPIII termination signals are present on both the template and non-template DNA strands. Interactions between oligo(dA) in the template strand and oligo(U) in the nascent transcript acts as a principal destabilizing signal, while the non-template strand oligo(dT) tract promotes polymerase pausing, formation of the pre-termination complex and transcript release (Arimbasseri and Maraia 2015).

RNAPIII transcription rates are regulated by the highly conserved repressor Maf1 and tightly coupled to environmental conditions. Maf1 binds the RNAPIII subcomplex C82/34/31, blocking recruitment of RNAPIII to a preinitiation complex formed by the initiation factor TFIIB and the promoter DNA (Cabart et al. 2008; Vannini et al. 2010). Several different signaling pathways have been reported to induce Maf1 repression (Pluta et al. 2001). Notably, Maf1 represses RNAPIII when cells are transferred from glucose containing medium to non-fermentable carbon sources, such as glycerol, particularly at elevated temperature (37°C) (Ciesla et al. 2007).

RNAPIII transcripts are targets for active surveillance pathways [reviewed in (Wichtowska et al. 2013)]. Pre-tRNAs with defects in modification or folding are degraded both by the 5'-exonuclease Rat1 and by the exosome nuclease complex (Chernyakov et al. 2008; Gudipati et al. 2012; Schneider et al. 2012). Targeting of pre-tRNAs to the exosome involves cofactors, including the Nrd1-Nab3 heterodimer and the Trf-Air-Mtr4 polyadenylation (TRAMP) complex (Kadaba et al. 2006; Jamonnak et al. 2011; Wlotzka et al. 2011). Within TRAMP, Trf4 oligoadenylates target RNAs, while the DEXH box RNA helicase Mtr4 loads the RNA into the exosome for degradation (Jia et al. 2011; Falk et al. 2014). Catalytic activity in the nuclear

exosome is provided by 2 subunits, Rrp44/Dis3, which shows both endonuclease and 3'-exonuclease activity, and the 3'-exonuclease Rrp6 [reviewed in (Chlebowski et al. 2013)]. Rrp6 interacts with the RNA-binding protein Nab2 (Schmid et al. 2012), which was recently reported to bind both RNAPIII transcripts and the polymerase (Reuter et al. 2015). Analyses of strains defective in both Rrp44 and Rrp6 activity indicate that a substantial fraction of pre-tRNA transcripts are normally degraded by the exosome (Gudipati et al. 2012; Schneider et al. 2012).

The putative total RNAPIII transcriptome was initially identified in *S. cerevisiae* by genome-wide analyses using chromatin immunoprecipitation (ChIP) (Harismendy et al. 2003; Roberts et al. 2003; Moqtaderi and Struhl 2004). However, ChIP analysis is not strand specific and generally have limited spatial resolution. Recently, the location of RNAPII has been more precisely mapped relative to the nascent transcript, rather than the DNA template, using nascent transcript sequencing (NETseq) or UV cross-linking and analysis of cDNA (CRAC) (Holmes et al. 2015; Mayer et al. 2015).

To better characterize overall RNAPIII transcription in yeast, we employed CRAC using a tagged form of the large, catalytic subunit of RNAPIII (Rpc160-HTP). High sequence coverage allowed accurate mapping of the location of RNAPIII along all nascent transcripts.

RESULTS

To permit the application of CRAC to RNAPIII, Rpc160 encoding the RNAPIII catalytic subunit was expressed as a fusion with a tripartite tag (His₆ - TEV protease cleavage site – protein A, HTP). The Rpc160-HTP fusion protein (Fig. 1A) was expressed under the control of the endogenous *RPC160* promoter from the chromosomal locus, and was the only source of Rpc160 in the cell. Strains expressing Rpc160-HTP showed wild type growth rates, demonstrating that the fusion protein is functional (data not shown).

To crosslink Rpc160-HTP to the RNA, cells actively growing in glucose medium at 30°C or glycerol medium at 37°C, were UV irradiated at 254 nm for approximately 100 s. As a negative control, the non-tagged, parental strain BY4741 (BY) was used. The CRAC analysis was performed as described (Granneman et al. 2009; Granneman et al. 2010) (see Fig. 1B for overview). Briefly, following cell lysis, protein-RNA complexes were isolated by immunopurification and affinity purification. RNA was partially digested with RNase A+T1, linkers

were ligated to the 5' and 3' termini, protein-RNA complexes were recovered by SDS-PAGE (Fig. S1A) and a cDNA library was prepared.

The sequencing generated 50 bp reads and, to ensure that *bona fide* 3' ends were mapped, we analyzed only reads that included both the 5' and 3' linkers. Sequences were mapped to the yeast genome using NovoAlign (Novocraft). Hits were counted using pyCRAC software (Webb et al. 2014) using a modified genome features file (GTF) as described in Methods. As expected, tRNA genes were highly represented in each of the Rrp160-HTP datasets, relative to the BY control strain, in both the wild type (wt) and *maf1* Δ backgrounds (Table S1A and Fig. 1C). Transfer to glycerol medium at 37°C reduced the relative recovery of tRNA reads and increased rRNA reads (predominantly 5S rRNA) (Table S1B; Fig. 1C) in the wt strain but not in *maf1* Δ , consistent with repression of tRNA transcription.

Analysis of hit distribution confirmed the association of RNAPIII with nuclear-encoded tRNA genes (Fig. 1D), since mitochondrial tRNA genes represented only 0.001% of all tRNA reads (Supplementary Table S1C). A low recovery of sequences annotated as protein-coding was observed, but further analysis revealed that these predominately represent regions that are in close proximity to tRNA genes. Known RNAPIII transcripts were also predominant among other small RNAs recovered (Fig. 1D). As an example, Table S1C shows that U6 snRNA was much more frequently recovered than U1 – U5 snRNAs, which are not RNAPIII transcripts. Similarly, the 5S rRNA represented 70% of all annotated rRNA sequences (Fig. 1D and Supplementary Table S1B).

Together these analyses confirm that the Rrp160 CRAC predominately recovered authentic RNAPIII transcripts.

Genome-wide analysis of tRNA transcription

The density of RNAPIII was compared across all individual tRNA genes. Figs. 2A and 2B show cumulative plots of primary transcripts aligned by the 5' end or 3' end of the mature tRNA, for yeast grown on glucose medium at 30°C. Apparent transcription read-through, indicated by RNAPIII density beyond the canonical termination site, just after 3' end of the mature tRNA (red line in Fig. 2B), was observed on many tRNA genes. Broadly similar distributions of RNAPIII density reads were observed during growth in glycerol medium at 37°C and in the absence of Maf1 (Fig. S1B).

Inspection of individual tRNA transcription units revealed notably uneven RNAPIII density across the mature tRNA and flanking pre-tRNA regions, as well as over 3' extensions when present. Representative, individual tRNAs are presented in Fig. 2C; other RNAPIII transcription units are shown in Supplementary Fig. S2. Graphs showing the distribution of RNAPIII across all individual tRNA genes are provided in Supplementary Dataset D1. Fig. 2D shows a heat map displaying the distribution of reads across each tRNA gene, relative to the maximum for the gene. This revealed a predominant pattern across most genes, with a high density of reads over the 5' end of the transcription unit and a weaker peak before 3' end of the mature tRNA. A minority of genes showed similar 5' and 3' peaks.

For each gene, the ratio between 5' and 3' peaks was determined. This was plotted relative to total number of reads mapped to that gene, as an indication of transcription rate (Fig. S3A). This revealed that the 5' "initiation" peak is relatively larger for highly transcribed genes. Measuring the areas under the 5' and 3' peaks confirmed that a higher transcription rate correlates with higher ratio of 5' to 3' peaks (Fig. S3B). Ordering genes according to their 5' : 3' peak ratio on global heat map clearly shows that the 5' peak is predominant in highly transcribed genes (Fig. 2E). However, under conditions of reduced RNAPIII transcription (glycerol medium at 37°C), the correlation between the 5' : 3' peak ratio and transcription rate was greatly reduced (Fig. S3C-E).

Clustering using 5' : 3' peak ratio revealed two major patterns: group 1 showed a very high 5' peak with much lower 3' peak and included 225/275 genes representing all 20 isotypes and 40 isoacceptors (Fig. 2F). Group 2 showed similar sized 5' and 3' peaks and was less common, including 50/275 genes representing 14 isotypes and 16 isoacceptors (Fig. 2G). Moreover, tRNAs genes falling into group 2 showed overall higher levels of RNAPIII transcription termination read-through (see below).

On tRNA genes, the spacing between the 5' and 3' peaks was ~50 nt. The approximate footprint expected for the RNAPIII complex is 38 nt (Hoffmann et al. 2015), suggesting that the observed peaks might reflect pileups caused by the high RNAPIII density. However, a consistent pattern of ~50 nt spacing across all tRNA genes was not observed (Fig S4A). Rather, the 5' and 3' peaks, indicating a high density of RNAPIII, appear to be localized with respect to the A and B boxes of the internal promoter. Two examples of tRNA genes are presented in supplementary Fig S4B, with A and B boxes marked. Genome-wide profiles of RNAPIII density aligned to A or B boxes revealed that the initiation 5' peak is located at the beginning of the A box (Fig. 2H) and the

second, 3' peak at the beginning of the B box (Fig. 2I). The same was true for the group 2 tRNAs (50 genes with 5':3' peaks ratio <1.5) (Fig. S4C, panels I and II) and for intron-containing genes, in which the relative position of B box to box A is variable since they are separated by the intronic sequence (Fig S4C, panels III and IV). We postulate that TFIIIC bound to the A and B boxes can interact with and impede transcribing RNAPIII, leading to transient pausing.

Novel RNAPIII transcripts

In addition to the protein coding genes and well-defined stable RNAs, yeast RNAPII transcribes very large numbers of non-coding RNAs, largely of unknown function, and also generates a low level of transcription throughout the majority of the genome [see (Porrua and Libri 2015) and references therein]. In contrast, yeast RNAPIII appears to be restricted to the expression of well-defined genes, with accurate recognition of transcription start sites. We were, however, able to identify a number of previously unrecognized, putative RNAPIII transcripts. Focusing on non-annotated regions of the genome (see Methods for details of regions selected), we found six potential new RNAPIII genes designated tRNA-like transcripts 1 to 6 (*TLT1-6*) (Table S2 and supplementary GTF file). In initial analyses, expression of *TLT2* and *TLT6* was confirmed by northern hybridization (Fig. 3A) and each is expressed at relatively high levels. Notably, *TLT6* is closely homologous to an apparent fragment of the 37S rDNA gene (*RDN37*) that is not located in the rDNA locus and has length of about 70 nt, characteristic of mature tRNAs and transcriptional read-through (see below). In addition, we identified distinct Rpc160-HTP-associated transcripts that were located within the RNAPI-transcribed *RDN37* rDNA, particularly at the 5' end of the 18S rRNA region (Fig. 3B, lower panel). These showed the characteristic double peak profile of RNAPIII, observed over tRNA genes, and were absent from the BY4741 control (Fig. 3B, upper panel). The prominent peaks visible in the 3' region of 25S rRNA in both datasets represent a common contaminant seen in many CRAC analyses (Granneman et al. 2010; Lebaron et al. 2013; Tuck and Tollervey 2013; Turowski et al. 2014). We speculate that RNAPIII transcription units may be active in the ~50% of rDNA repeats that are not transcribed by RNAPI.

RNAPIII transcription levels under different growth conditions

We compared relative transcription by RNAPIII over all nuclear tRNA genes under near optimal growth conditions (glucose medium at 30°C) and following transfer to stress conditions known to repress tRNA expression (glycerol medium at 37°C). Under stress conditions reduced

transcription was observed for nearly all tRNAs, as previously reported (Ciesla et al. 2007), but the degree of repression was highly variable between tRNA genes (Fig. 4A). Some tRNA genes appeared to have elevated RNAPIII occupancy (Gly : Glu ratios >1.0). This apparent increase is probably a consequence of normalization to total RNAPIII reads under conditions in which most tRNA genes are strongly repressed, but clearly shows that transcription of a subset of tRNA genes is less repressed following glucose withdrawal. This conclusion is broadly consistent with a previous microarray analysis, which revealed that the levels of mature tRNAs are reduced to variable extents on a non-fermentable carbon source (Ciesla et al. 2007).

Maf1 is an RNAPIII repressor that acts following nutritional downshift, particularly at elevated temperatures (Ciesla et al. 2007; Boguta 2013). To assess the role of Maf1 in repressing transcription of each tRNA gene, we plotted the ratio of RNAPIII occupancy following transfer between permissive (glucose at 30°C) and restrictive (glycerol at 37°C) conditions for wt (Fig. 4A) and *maf1Δ* datasets (Fig. 4B). Decreased repression in the absence of Maf1 is shown by the increased slope of the trend line ($y=1x-500$) compared to wt ($y=0.8x-23$).

In addition, it is evident that the heterogeneity in tRNA repression seen in the wild-type is substantially reduced in the absence of Maf1. This was confirmed by the increased correlation coefficient (R^2) in the absence of Maf1 ($R^2=0.85$) relative to the wild-type ($R^2=0.59$). This provides genome-wide evidence that Maf1 does not simply down regulate all tRNAs, but provides an additional layer of gene-specific RNAPIII regulation.

Figure 4C shows all tRNA genes ranked by the ratio of expression in glucose at 30°C versus glycerol at 37°C (line of black points in Fig. 4C), with the ratio of expression in *maf1Δ* versus *MAF1* shown in grey. In general, tRNA genes that show low repression under stress conditions also show low relative overexpression in *maf1Δ* strains (*maf1Δ* : wt ratio < 1.0). This indicates that a subset of tRNA genes shows low responsiveness to both environmental and cellular signals. Notably, this group contains at least one tRNA for each amino acid. In Figure 4C, the least affected tRNA for each amino acid is indicated in dark red. The only exception was *SUP53*, encoding tL(CAA)C, which was very insensitive to Maf1 relative to most tRNAs, but strongly repressed by transfer to glycerol at 37°C, suggesting a distinct mode of regulation. Together these findings suggest the existence of a basal subset of “housekeeping” tRNA genes.

Inspection of the RNAPIII profiles along all tRNA genes revealed very few differences between wt and *maf1Δ* strains (see Supplementary Fig. S1B and Dataset D1). This is in contrast to the

changes in total RNAPIII density and strongly indicates that Maf1, which acts during RNAPIII transcription initiation (Desai et al. 2005; Vannini et al. 2010), does not interfere with RNAPIII elongation.

A subset of other RNAPIII transcripts was strongly increased under stress conditions (Table S3). The most elevated was *RNA170*, transcription of which was upregulated nearly 100-fold in a Maf1-independent manner. Surprisingly, numerous ncRNAs that are generally transcribed by RNAPII, showed apparent increases in RNAPIII transcription of >2-fold under stress conditions. These included several small nucleolar RNAs (snoRNAs), notably snR189 and the U2 small nuclear RNA (Lsr1 in yeast). It appears that under stress conditions these RNAs are increasingly transcribed by RNAPIII, perhaps as a consequence of the loss of RNAPII transcription. Other known RNAPIII transcripts including *SNR6*, *SCR1* and *RPR1*, showed mild increases (1.4 - 1.9 fold) in apparent RNAPIII transcription following glucose withdrawal (Table S3). These modest apparent changes may simply reflect decreased tRNA transcription, since the reads are relative to total RNAPIII binding.

Termination read-through on RNAPIII transcription units

The major termination sites on yeast tRNA genes have long been identified by the presence of oligo(U) tracts in the transcript, located 5-15 nt downstream from the 3' end of the mature tRNA sequence (Orioli et al. 2011). Perhaps the most unexpected finding from our analyses was the identification of substantial read-through (RT) transcription on many tRNA genes, typically extending 50 – 200 nt beyond the expected terminator (Fig. 5). The RT distribution detected *in vivo* using CRAC was quite different from previous analyses of *in vitro* transcription (see Table 1 and Discussion) (Braglia et al. 2005).

To confirm the CRAC data and test whether RNAPIII generates the expected long, continuous transcripts, we performed northern hybridization for tRNAs showing high RT levels (Fig 5A and S5A-C). Probes specific for mature tRNA and extensions were used to assess the presence of 3' extended pre-tRNA transcripts and estimate ratios between mature and 3' extended tRNAs. Clear RT bands were identified using tRNA-specific probes for tH(GUG) (Fig 5A), tE(UUC), tK(UUU) and tY(GUA) (Figs. S5A-C). The RT bands were generally less strong for yeast grown on glycerol medium at 37°C, presumably due to downregulation of RNAPIII. RT products from the tH(GUG)G2 and tK(UUU)O genes were detected using a specific probes, against 54 nt 3' to the mature tH(GUG)G2 (Fig. 5A, left panel) and 40 nt 3' to the mature tK(UUU)O (Fig. S5B). The

abundance of the extended transcripts relative to mature tRNAs (Fig. 5A, right panel) was substantially lower than the degree of RT seen in the CRAC analyses (Fig. 5B), indicating they are rapidly processed or degraded (see below). The RT transcripts were not previously recognized as pre-tRNAs, probably due to the combination of their lengths and low abundance.

To identify features associated with read-through efficiency, RT levels were calculated for each individual gene. This was defined by the ratio between the number of reads mapping to extensions (regions >15 nt 3' to the mature tRNA) to total pre-tRNA reads (from -15 nt 5' to the mature tRNA gene to the 3' end of the RT product). Average RT levels for all tRNA genes were over 10% in strains growing under permissive conditions (Fig 5C). RT decreased following transfer to repressive conditions but was unaffected by the loss of Maf1 consistent with elongation being independent of Maf1. To visualize the overall frequency of transcriptional RT under permissive growth conditions, a global log2 heatmap is presented for all tRNA genes (Fig. 5D).

Even genes with relatively high RT levels showed a sharp drop in RNAPIII density following the 3' end of the mature tRNA (Fig. 2B and Fig. S5D), indicating that transcription termination generally occurs at canonical termination signals. However, the range of RT levels (Fig. 5E and Fig. S5E) shows that the efficiency of canonical termination is highly variable. To identify features associated with efficient termination, we compared genes showing different levels of RT (Fig. 6A and Fig. S5F). As expected, RT levels were negatively associated with the length of the oligo(U) tract at the canonical termination site. Genes showing higher read-through have clear tendency to have weaker canonical termination signals over the 50 nt 3' to the mature tRNA sequence (Fig. 6A). This also indicated that effective canonical termination of RNAPIII *in vivo* needs terminators stronger than a U6 tract. Moreover, for genes with >25% RT, over 60% have U6 tracts as the longest termination signal, whereas for genes with <5% RT 60% have U8 tracts. The most common poly(U) tract within 50 nt 3' to the mature sequence across all tRNA gene terminators are U7 and U8. Analysis of 3' extensions revealed that genes with high RT levels also lack a second canonical termination signal within the next 40 nt (Fig. 6B).

Within the 3' extensions, a clear negative correlation was seen between RNAPIII density and the frequency of U residues. Figure 6C and supplementary Figure S6A present hit distributions across RT regions with U residues indicated. A low frequency of U residues was correlated with accumulation of RNAPIII, even 100-150 nt downstream from the canonical termination site. We

confirmed these findings by statistical analysis of U enrichment within peaks and troughs in the RNAPIII density in the RT region. The canonical terminator was excluded from this analysis to avoid biasing the outcome. The results were highly significant (see legend to Fig. 6D). A box-and-whisker plot (Fig. 6D) clearly showed that U residues are more abundant within regions showing RNAPIII density minima (troughs). This was unexpected because oligo(U) tracts were reported to be associated with transcription pausing by RNAPIII [reviewed by (Arimbasseri et al. 2013)]. We speculate that on encountering a U-rich region the nascent transcript may be rapidly lost from RNAPIII, perhaps due to the low stability of oligo(dA:rU) regions. However, the polymerase may remain transiently bound to the DNA, perhaps stabilized by the interactions with the dT sequence of the non-template strand (Arimbasseri and Maraia 2015). Since CRAC relies on RNA crosslinking, these polymerases would not be detected, potentially giving rise to the observed RNAPIII density minima. Additionally, RNAPIII bound at these sites might transiently block elongation by subsequent polymerase molecules, potentially generating the upstream peaks in density.

The greatest RT was seen for a subset of tRNA isotypes (Fig. S6B); RT above 30% was observed for 20 genes representing 10 isotypes (Fig. S5D). However, RT above 15% was seen for 85 genes representing all isotypes, indicating only limited isotype-dependency. Formation of a RNA:DNA duplex (R-loop) facilitates termination by RNAPII (Komissarova et al. 2002; Skourti-Stathaki et al. 2014), suggesting that this might also be the case for RNAPIII. We therefore assessed the propensity for R-loop formation over the 20 nt downstream from each mature tRNA, by comparison of the predicted ΔG for DNA/DNA and RNA/DNA duplexes (Fig. S6C). Across all tRNA genes, no significant correlation between RT and ΔG was detected (Fig. S6C, left). However, a weak tendency towards R-loop formation was observed for isotypes with the highest RT (RT>25%; Fig S6C right) relative to other genes (Fig S6C, middle panel). We further sought to identify structural features that potentially facilitate termination. *In silico* analysis using RNA folding software failed to identify consistent correlations between the predicted presence or stability of structural elements in the nascent transcript and RT efficiency.

Altogether, oligo(U) tract length and uracil frequency were the only features in tRNA 3' flanking regions that were consistently correlated with transcription read-through levels *in vivo*.

Surveillance of 3' pre-tRNA extensions

The extended pre-tRNA species were not previously reported from RNA expression analyses in *S. cerevisiae*, and data in Fig. 5A and S5A-C also indicates their low steady-state levels, suggesting that they may be relatively unstable. Pre-tRNAs represent major targets for the nuclear surveillance machinery (Wlotzka et al. 2011; Gudipati et al. 2012; Schneider et al. 2012). To determine whether these 3'-extended RNAs are specifically targeted by the nuclear surveillance system, we analyzed CRAC data for known surveillance factors; Nab2, Mtr4, Rrp6, Rrp44 and Rat1 (Fig. 7). To compare binding profiles, full-length tRNA transcription units were annotated using the RNAPIII CRAC data (Fig. 7A) to limit background from RNAPII transcription, which is present at low-levels throughout the genome (Fig. 7B) (Holmes et al. 2015).

The RNA binding protein Nab2 is involved in surveillance (Schmid et al. 2012) and RNAPIII transcription (Reuter et al. 2015). Nab2 interacts directly with RNAPIII and TFIIIB at the 5'-end of tRNA genes (Reuter et al. 2015), and may couple transcription initiation and surveillance. In our analysis, Nab2 binding to tRNA 3'-extension was clearly elevated (Fig. 7C). Clear binding to 3' extended pre-tRNAs was also observed for Rat1 (Fig. 7D), the major 5'-exonuclease for tRNA surveillance (Chernyakov et al. 2008), and for Rrp44 (Fig. 7E), the major exosome-associated nuclease with 3'-exonuclease and endonuclease activities. Greatly reduced binding was seen for an Rrp44 mutant lacking exonuclease activity (Rrp44-exo) (Fig. 7F), and lower binding was also seen for the other exosome-associated nuclease Rrp6 (Fig. 7G). The major nuclear cofactor for the exosome is the RNA helicase Mtr4, which was also bound to tRNA 3'-extensions (Fig. 7H).

We conclude that 3'-extended pre-tRNAs are bound by the nuclear RNA surveillance machinery, potentially explaining their low steady-state levels.

DISCUSSION

Here we report the analysis of localization of actively transcribing RNAPIII by UV crosslinking to the nascent transcripts. A similar approach was recently applied to characterize RNAPII transcription (Holmes et al. 2015). The data analysis applied here was similar to reported NETseq analyses for RNAPII (Mayer et al. 2015; Nojima et al. 2015), with selection for short reads containing native 3' ends. The RNAPIII binding sites identified were reproducible between experiments and included all previously reported RNAPIII transcription units. In addition, we identified novel RNAPIII transcription units, including transcripts generated from within the rDNA repeats, two of which were verified by Northern hybridization.

The CRAC analysis revealed notably uneven distribution of RNAPIII across transcription units. Pre-tRNA genes predominantly showed a strong 5' peak and a weaker peak close to the 3' end of the mature tRNA. We interpret these RNAPIII peaks as representing regions where crosslinking is favored because the polymerase shows a decreased elongation rate and/or transient pausing. This could reflect many features, but comparison of peaks to the locations of the internal promoter elements suggests that transcription factors bound at these sites may contribute to decreased elongation. The 5' peak of RNAPIII crosslinking was located at the beginning of the A box and the 3' peak at the beginning of the B box (Fig. 2H-I). The transcription factor TFIIIC binds to both the A and B boxes (Acker et al. 2013) and *in vitro* studies indicate that TFIIIC-DNA interactions must be disrupted during RNAPIII elongation (Bardeleben et al. 1994). The complex of TFIIIB and TFIIIC on tRNA genes occupies a DNA length similar to nucleosomes, which are absent from tRNA genes (Nagarajavel et al. 2013). TFIIIC association with DNA increases during repression and it has been proposed that, in addition to its activity as transcription factor, TFIIIC occupies tRNA and other genes during transitions in RNAPIII activity (Arimbasseri et al. 2014b). We speculate that TFIIIC binding to the A and B boxes helps form a physical barrier in the chromatin that transiently interferes with RNAPIII elongation.

Following its recruitment by TFIIIB, RNAPIII becomes part of a closed preinitiation complex at the transcription start site (Vannini and Cramer 2012). Subsequent promoter melting and the initiation-to-elongation transition involve major conformational changes of RNAPIII, which likely enhance its processivity (Fernandez-Tornero et al. 2010). On highly transcribed genes a prominent 5' peak was observed close to the transcription initiation site (Fig. 2E), suggesting

that initiation site clearance, i.e. dissociation from transcription factors, may be the rate limiting during RNAPIII transcription.

Read-through of the canonical termination sites on tRNA genes emerged as a surprisingly prevalent feature of RNAPIII transcription. The resulting 3' extensions were confirmed by northern hybridization. Previous studies showed that human RNAPIII occupies regions downstream from 3' ends of many tRNA genes (Canella et al. 2010; Orioli et al. 2011). Analyses of human genome also suggested the presence of terminators located ≥ 50 nt from the 3' end of tRNA coding sequence and terminator read-through (Orioli et al. 2011). Moreover, read-through tRNA transcripts detected in *Schizosaccharomyces pombe* lacking the mediator complex subunit Med20 are adenylated and targeted for degradation by the exosome (Carlsten et al. 2015). It seems likely that read-through of canonical termination sites is a feature of RNAPIII in many systems.

A run of six U residues in the nascent transcript was previously defined as the termination site in yeast RNAPIII genes [reviewed in (Arimbasseri et al. 2013)]. However, our analysis revealed that *in vivo* termination is significantly more effective with U₇-U₈ (Fig. 6A). We also sought to identify features that correlated with termination of the read-through transcripts, including sequence context, predicted secondary structure and spacing between short terminators. The most consistent correlation was between transcription termination and the frequency of U residues (Fig. 6D). Notably, uracil abundance was significantly higher for troughs in the RNAPIII density, suggesting a correlation with termination in the 3'-extended regions.

Recent work suggested that RNAPIII pauses and backtracks on termination signals, with termination facilitated by a structural element in the nascent transcript (Nielsen et al. 2013), whereas a subsequent study found termination to be independent of transcript terminator-proximal RNA structure (Arimbasseri et al. 2014a). In our analysis (Fig. S6C) we found no clear evidence for effects of structural elements on RNAPIII termination.

Extended tRNA transcripts likely escaped prior detection due to rapid turnover. This potentially involves recognition by the surveillance factor Nab2 and the TRAMP complex, followed by 5' degradation by the Rat1 and 3' degradation by the exosome (Fig. 8). Nab2 is required for full RNAPIII transcription (Reuter et al. 2015) suggesting dual roles in transcription and surveillance. Binding of Nab2 to tRNA RT transcripts is also evident in published data (Tuck and Tollervey

2013; Reuter et al. 2015) but was not discussed previously. The exosome component Rrp6 was previously implicated in the degradation of newly transcribed pre-tRNAs (Gudipati et al. 2012; Schneider et al. 2012). In contrast, stronger binding of the other exosome-associated nuclease Rrp44 was seen on RT transcripts. Pre-tRNAs can be 3' processed by the endonucleases RNase Z (Takaku et al. 2003) and cleavage might be followed by binding of Rat1. In northern analyses, extension fragments gave clear band(s) consistent with the lengths of RT transcripts.

Previous analyses have implicated tRNA spacers and cleavage fragments as potential regulators of mRNA expression (Thompson and Parker 2009; Haussecker et al. 2010; Pederson 2010; Gebetsberger et al. 2012). Transcription read-through appears to be a conserved feature and we speculate that the products may also function in the regulation of gene expression.

METHODS

Strains

All yeast analyses were performed in strains derived from (BY4741, *MATa*; *his3Δ1*; *leu2Δ0*; *met15Δ0*; *ura3Δ0*). Strains used are listed in Supplementary Table S4.

Crosslinking, preprocessing and aligning of Illumina Sequence Data (CRAC).

CRAC experiments using Rpc160-HTP, Rpc160-HTP *maf1Δ*, Rrp44-HTP, Rrp44-exo-HTP, Rrp6-HTP, Mtr4-HTP and Nab2-HTP were performed on cultures grown in SD medium with 2% glucose, lacking Trp to OD₆₀₀=0.5. For nutrient shift, Rpc160-HTP and Rpc160-HTP *maf1Δ* strains were grown to OD₆₀₀=0.5 in SD medium were harvested by filtration, transferred to medium containing 2% glycerol instead glucose and grown for additional 2h at 37°C. Actively growing cells were cross-linked in culture media (Granneman et al. 2011) and processed as previously described (Granneman et al. 2009; Granneman et al. 2010). To generate Nab2 datasets all enzymatic steps were performed at 20°C. Illumina sequencing data were collapsed to remove PCR duplicates. Rpc160-HTP and Rpc160-HTP *maf1Δ* data were preprocessed using flexbar with parameters -at 1 -ao 4, filtered to leave only reads containing 3' adaptor. Other data were preprocessed using flexbar with default parameters. Data for Rat1 (Granneman et al. 2011) and RNAPII (Holmes et al. 2015) were previously published. During data processing, the raw sequences read were "collapsed" by the removal of reads that are identical, including the random barcodes present in the linkers, and potentially representing PCR duplicates. Across the mature tRNA regions, some fragments will be the same for multiple different genes that

encode the same isoacceptor. It therefore seemed possible that collapsing the data might lead to the rejection of identical, but actually independent, reads with effects on global read distributions. To check this, we compared heatmaps for all tRNAs plotted using collapsed and uncollapsed data. Because the patterns were the same, further analyses were performed using the smaller, collapsed datasets.

All datasets were aligned to the yeast genome using Novoalign (<http://www.novocraft.com>) with both `-r random` and `-r None` parameter as indicated in the text.

Sequence and feature file

All Gene Transfer Format (GTF) annotation and genomic sequence files were obtained from ENSEMBL (*S. cerevisiae* genome version EF4.74). After initial analysis, genomic coordinates of tRNAs were arbitrary extended by 50 nt on 5' end and 250 nt on 3' end using bedtools (Quinlan and Hall 2010). Experimentally confirmed 3' tRNA extensions were annotated in a separate GTF file (Supplementary File 1). Potential new RNAPIII transcripts were annotated in separate GTF files (Supplementary File 2).

Downstream analysis

Downstream analysis were performed using pyCRAC software (Webb et al. 2014) and the gwide toolkit developed for this analysis (<https://github.com/tturowski/gwide>). PyReadCounters (pyCRAC) was used to calculate overlaps between aligned cDNAs and yeast genomic features, using a GTF file with extended tRNA genes. Single gene and genome-wide plots were generated using gwide toolkit. GENE-E was used to generate the heatmaps. Relative gene transcription was calculated as number of reads from 15 nt upstream of the mature tRNA to the 3' end of mature tRNA. 5':3' peak ratios were calculated dividing total read numbers from the 5' half of each gene by reads in the 3' half, where the genes were defined as extending from 15 nt upstream of mature tRNA to the 3' end of mature tRNA. A and B boxes were identified using fimo (Grant et al. 2011) from the MEME suite (Bailey et al. 2009) among all nuclear encoded tRNA genes.

tRNA read-through was calculated from 15 nt to 250 nt downstream from the mature tRNA 3' end. The longest U stretch in the 50 bp 3' to the end of the tRNA was identified from fasta sequences using simple unix command line tools awk, grep and wc. For identification of maxima and minima, the pypeaks package in the gwide toolkit was used (<https://github.com/gopalkoduri/pypeaks>). U-richness statistics using Wilcoxon and paired T-test

were calculated for various conditions: different average peak cut-offs, 10 nt, 20 nt and 50 nt regions around or before peaks and troughs. All p-values for 10 nt and 20 nt around peaks and troughs were significant (p-value < 0.0002). Calculations for 50 nt surrounding sequences served as negative controls and resulted in p-values above 0.1.

New genes were identified using an `sgr_peaks_identif.awk` script (Supplementary File 3) and `bedtools` (Quinlan and Hall 2010). Novofiles were converted into `sgrs` file (`sgr` format with added information about strand). All peaks at least 50 reads high and 10 nt long in `*.sgrs` files were identified using `sgr_peaks_identif.awk` and annotated. Using `bedtools`, annotated features were combined with `merge -d 5` and intersected with all annotated overlapping genome features. New features were combined with `merge -d 200` and annotated. All new features were manually analyzed; exclusion of overlaps with tRNA extensions, comparison with RNAPIII, RNAPII and exosome data. Six new transcripts were annotated and two were confirmed using northern blots. To compare datasets from different growth conditions we excluded all reads mapping to more than one site in the genome. This eliminates tRNA fragments that would be ambiguously mapped to the mature tRNA regions of more than one gene. However, gene-specific reads overlapping the region between the TSS and the mature tRNA 5' end, or between the 3' end of the tRNA and terminator, will reflect transcription rates for individual tRNA genes. Datasets were normalized to hits per million mapped reads.

Northern blotting

Yeast strains were grown in SC-trp medium containing 2% glucose at 30°C to exponential growth phase, then transferred to SC-trp medium containing 2% glycerol, incubated for 2 h at 37°C. Then RNA was isolated as described previously (Foretek et al. 2016). Quantity and purity of RNA were analyzed using a NanoDrop 1000 (Thermo). To identify read-through products for individual tRNAs, 5µg of the total RNA and 2µl low range ssRNA ladder (NEB) were resolved by 10% PAGE with 8 M urea. Gels were stained with ethidium bromide for visualization of RNA ladder. Then RNA was transferred onto a Hybond-N+ membrane (Amersham) by electroblotting in 0.5× TBE and crosslinked by UV radiation. For RNA detection a nonradioactive northern method was used (Wu et al. 2013). For some probes, the temperature of hybridization and the washing protocol were modified to increase specificity of the signal. Blots were developed using photographic films for higher resolution (CL-XPosure Film from Thermo Scientific Pierce). DIG-labeled oligonucleotides used for Northern hybridization are listed in supplementary table S5.

DATA ACCESS

All RNA sequencing data from this study have been submitted to the NCBI Gene Expression Omnibus (GEO; <http://www.ncbi.nlm.nih.gov/geo/>) under accession number GSE77863.

ACKNOWLEDGEMENTS

We thank Aleksandra Helwak for critical reading of the MS. We thank Grzegorz Kudła and Hywel Dunn-Davis for support and valuable advice during data analysis. This work was supported a grant from the Polish Ministry of Science and Higher Education Mobility Plus program to TWT [1069/MOB/2013/0], National Centre of Science in Cracow, Poland [Project Maestro; UMO-2012/04/A/NZ1/00052] and Foundation for Polish Science [Project Mistrz 7/2014] to EM and MB; EMBO Short Term Fellowship [ASTF 364-2014] to EM, EMBO Long Term Fellowship [ALTF 625-2014] to CS and Wellcome Trust Fellowship [077248] to DT Work in the Wellcome Trust Centre for Cell Biology is supported by Wellcome Trust core funding [092076].

DISCLOSURE DECLARATION

The authors declare that they have no competing interests.

REFERENCES

- Acker J, Conesa C, Lefebvre O. 2013. Yeast RNA polymerase III transcription factors and effectors. *Biochim Biophys Acta* **1829**(3-4): 283-295.
- Arimbasseri AG, Kassavetis GA, Maraia RJ. 2014a. Transcription. Comment on "Mechanism of eukaryotic RNA polymerase III transcription termination". *Science* **345**(6196): 524.
- Arimbasseri AG, Maraia RJ. 2015. Biochemical analysis of transcription termination by RNA polymerase III from yeast *Saccharomyces cerevisiae*. *Methods Mol Biol* **1276**: 185-198.
- Arimbasseri AG, Rijal K, Maraia RJ. 2013. Transcription termination by the eukaryotic RNA polymerase III. *Biochim Biophys Acta* **1829**(3-4): 318-330.
- . 2014b. Comparative overview of RNA polymerase II and III transcription cycles, with focus on RNA polymerase III termination and reinitiation. *Transcription* **5**(1): e27639.
- Bailey TL, Boden M, Buske FA, Frith M, Grant CE, Clementi L, Ren J, Li WW, Noble WS. 2009. MEME SUITE: tools for motif discovery and searching. *Nucleic Acids Res* **37**(Web Server issue): W202-208.
- Bardeleben C, Kassavetis GA, Geiduschek EP. 1994. Encounters of *Saccharomyces cerevisiae* RNA polymerase III with its transcription factors during RNA chain elongation. *J Mol Biol* **235**(4): 1193-1205.
- Boguta M. 2013. Maf1, a general negative regulator of RNA polymerase III in yeast. *Biochim Biophys Acta* **1829**(3-4): 376-384.
- Braglia P, Percudani R, Dieci G. 2005. Sequence context effects on oligo(dT) termination signal recognition by *Saccharomyces cerevisiae* RNA polymerase III. *J Biol Chem* **280**(20): 19551-19562.

- Cabart P, Lee J, Willis IM. 2008. Facilitated recycling protects human RNA polymerase III from repression by Maf1 in vitro. *J Biol Chem* **283**(52): 36108-36117.
- Canella D, Praz V, Reina JH, Cousin P, Hernandez N. 2010. Defining the RNA polymerase III transcriptome: Genome-wide localization of the RNA polymerase III transcription machinery in human cells. *Genome Res* **20**(6): 710-721.
- Carlsten JO, Zhu X, Davila Lopez M, Samuelsson T, Gustafsson CM. 2015. Loss of the Mediator subunit Med20 affects transcription of tRNA and other non-coding RNA genes in fission yeast. *Biochim Biophys Acta* **1859**(2): 339-347.
- Chan PP, Lowe TM. 2009. GtRNAdb: a database of transfer RNA genes detected in genomic sequence. *Nucleic Acids Res* **37**(Database issue): D93-97.
- Chernyakov I, Whipple JM, Kotelawala L, Grayhack EJ, Phizicky EM. 2008. Degradation of several hypomodified mature tRNA species in *Saccharomyces cerevisiae* is mediated by Met22 and the 5'-3' exonucleases Rat1 and Xrn1. *Genes Dev* **22**(10): 1369-1380.
- Chlebowski A, Lubas M, Jensen TH, Dziembowski A. 2013. RNA decay machines: the exosome. *Biochim Biophys Acta* **1829**(6-7): 552-560.
- Ciesla M, Towpik J, Graczyk D, Oficjalska-Pham D, Harismendy O, Suleau A, Balicki K, Conesa C, Lefebvre O, Boguta M. 2007. Maf1 is involved in coupling carbon metabolism to RNA polymerase III transcription. *Mol Cell Biol* **27**(21): 7693-7702.
- Desai N, Lee J, Upadhyay R, Chu Y, Moir RD, Willis IM. 2005. Two steps in Maf1-dependent repression of transcription by RNA polymerase III. *J Biol Chem* **280**(8): 6455-6462.
- Dieci G, Conti A, Pagano A, Carnevali D. 2013. Identification of RNA polymerase III-transcribed genes in eukaryotic genomes. *Biochim Biophys Acta* **1829**(3-4): 296-305.
- Falk S, Weir John R, Hentschel J, Reichelt P, Bonneau F, Conti E. 2014. The Molecular Architecture of the TRAMP Complex Reveals the Organization and Interplay of Its Two Catalytic Activities. *Mol Cell* **55**(6): 856-867.
- Fernandez-Tornero C, Bottcher B, Rashid UJ, Steuerwald U, Florchinger B, Devos DP, Lindner D, Muller CW. 2010. Conformational flexibility of RNA polymerase III during transcriptional elongation. *EMBO J* **29**(22): 3762-3772.
- Foretek D, Wu J, Hopper AK, Boguta M. 2016. Control of *Saccharomyces cerevisiae* pre-tRNA processing by environmental conditions. *RNA*.
- Gebetsberger J, Zywicki M, Kunzi A, Polacek N. 2012. tRNA-derived fragments target the ribosome and function as regulatory non-coding RNA in *Haloferax volcanii*. *Archaea* **2012**: 260909.
- Granneman S, Kudla G, Petfalski E, Tollervey D. 2009. Identification of protein binding sites on U3 snoRNA and pre-rRNA by UV cross-linking and high throughput analysis of cDNAs. *Proc Natl Acad Sci USA* **106**: 9613-9818.
- Granneman S, Petfalski E, Swiatkowska A, Tollervey D. 2010. Cracking pre-40S ribosomal subunit structure by systematic analyses of RNA-protein cross-linking. *Embo J* **29**: 2026-2036.
- Granneman S, Petfalski E, Tollervey D. 2011. A cluster of ribosome synthesis factors regulate pre-rRNA folding and 5.8S rRNA maturation by the Rat1 exonuclease. *The EMBO journal* **30**(19): 4006-4019.
- Grant CE, Bailey TL, Noble WS. 2011. FIMO: scanning for occurrences of a given motif. *Bioinformatics* **27**(7): 1017-1018.
- Gudipati RK, Xu Z, Lebreton A, Séraphin B, Steinmetz LM, Jacquier A, Libri D. 2012. Massive degradation of RNA precursors by the exosome in wild type cells. *Mol Cell* **48**: 409-421.
- Hani J, Feldmann H. 1998. tRNA genes and retroelements in the yeast genome. *Nucleic Acids Res* **26**(3): 689-696.
- Harismendy O, Gendrel CG, Soularue P, Gidrol X, Sentenac A, Werner M, Lefebvre O. 2003. Genome-wide location of yeast RNA polymerase III transcription machinery. *EMBO J* **22**(18): 4738-4747.

- Haussecker D, Huang Y, Lau A, Parameswaran P, Fire AZ, Kay MA. 2010. Human tRNA-derived small RNAs in the global regulation of RNA silencing. *RNA* **16**(4): 673-695.
- Hoffmann NA, Jakobi AJ, Moreno-Morcillo M, Glatt S, Kosinski J, Hagen WJ, Sachse C, Muller CW. 2015. Molecular structures of unbound and transcribing RNA polymerase III. *Nature* **528**(7581): 231-236.
- Holmes RK, Tuck AC, Zhu C, Dunn-Davies HR, Kudla G, Clauder-Munster S, Granneman S, Steinmetz LM, Guthrie C, Tollervey D. 2015. Loss of the Yeast SR Protein Npl3 Alters Gene Expression Due to Transcription Readthrough. *PLoS Genet* **11**(12): e1005735.
- Iben JR, Maraia RJ. 2012. tRNAomics: tRNA gene copy number variation and codon use provide bioinformatic evidence of a new anticodon:codon wobble pair in a eukaryote. *RNA* **18**(7): 1358-1372.
- Jamonnak N, Creamer TJ, Darby MM, Schaughency P, Wheelan SJ, Corden JL. 2011. Yeast Nrd1, Nab3, and Sen1 transcriptome-wide binding maps suggest multiple roles in post-transcriptional RNA processing. *RNA*.
- Jia H, Wang X, Liu F, Guenther UP, Srinivasan S, Anderson JT, Jankowsky E. 2011. The RNA helicase Mtr4p modulates polyadenylation in the TRAMP complex. *Cell* **145**(6): 890-901.
- Kadaba S, Wang X, Anderson JT. 2006. Nuclear RNA surveillance in *Saccharomyces cerevisiae*: Trf4p-dependent polyadenylation of nascent hypomethylated tRNA and an aberrant form of 5S rRNA. *RNA* **12**: 508-521.
- Komissarova N, Becker J, Solter S, Kireeva M, Kashlev M. 2002. Shortening of RNA:DNA hybrid in the elongation complex of RNA polymerase is a prerequisite for transcription termination. *Mol Cell* **10**(5): 1151-1162.
- Lebaron S, Segerstolpe Å, French Sarah L, Dudnakova T, de lima Alves F, Granneman S, Rappsilber J, Beyer Ann L, Wieslander L, Tollervey D. 2013. Rrp5 Binding at Multiple Sites Coordinates Pre-rRNA Processing and Assembly. *Mol Cell* **52**(5): 707-719.
- Mayer A, di Iulio J, Maleri S, Eser U, Vierstra J, Reynolds A, Sandstrom R, Stamatoyannopoulos JA, Churchman LS. 2015. Native elongating transcript sequencing reveals human transcriptional activity at nucleotide resolution. *Cell* **161**(3): 541-554.
- Moqtaderi Z, Struhl K. 2004. Genome-wide occupancy profile of the RNA polymerase III machinery in *Saccharomyces cerevisiae* reveals loci with incomplete transcription complexes. *Mol Cell Biol* **24**(10): 4118-4127.
- Nagarajavel V, Iben JR, Howard BH, Maraia RJ, Clark DJ. 2013. Global 'bootprinting' reveals the elastic architecture of the yeast TFIIB-TFIIC transcription complex in vivo. *Nucleic Acids Res* **41**(17): 8135-8143.
- Nielsen S, Yuzenkova Y, Zenkin N. 2013. Mechanism of eukaryotic RNA polymerase III transcription termination. *Science* **340**(6140): 1577-1580.
- Nojima T, Gomes T, Grosso AR, Kimura H, Dye MJ, Dhir S, Carmo-Fonseca M, Proudfoot NJ. 2015. Mammalian NET-Seq Reveals Genome-wide Nascent Transcription Coupled to RNA Processing. *Cell* **161**(3): 526-540.
- Orioli A, Pascali C, Quartararo J, Diebel KW, Praz V, Romascano D, Percudani R, van Dyk LF, Hernandez N, Teichmann M et al. 2011. Widespread occurrence of non-canonical transcription termination by human RNA polymerase III. *Nucleic Acids Res* **39**(13): 5499-5512.
- Pederson T. 2010. Regulatory RNAs derived from transfer RNA? *RNA* **16**(10): 1865-1869.
- Pluta K, Lefebvre O, Martin NC, Smagowicz WJ, Stanford DR, Ellis SR, Hopper AK, Sentenac A, Boguta M. 2001. Maf1p, a negative effector of RNA polymerase III in *Saccharomyces cerevisiae*. *Mol Cell Biol* **21**(15): 5031-5040.
- Porrua O, Libri D. 2015. Transcription termination and the control of the transcriptome: why, where and how to stop. *Nat Rev Mol Cell Biol* **16**(3): 190-202.

- Quinlan AR, Hall IM. 2010. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**(6): 841-842.
- Reuter LM, Meinel DM, Strasser K. 2015. The poly(A)-binding protein Nab2 functions in RNA polymerase III transcription. *Genes Dev* **29**(14): 1565-1575.
- Roberts DN, Stewart AJ, Huff JT, Cairns BR. 2003. The RNA polymerase III transcriptome revealed by genome-wide localization and activity-occupancy relationships. *Proc Natl Acad Sci U S A* **100**(25): 14695-14700.
- Schmid M, Poulsen MB, Olszewski P, Pelechano V, Saguez C, Gupta I, Steinmetz LM, Moore C, Jensen TH. 2012. Rrp6p controls mRNA poly(A) tail length and its decoration with poly(A) binding proteins. *Mol Cell* **47**(2): 267-280.
- Schneider S, Kudla G, Wlotzka W, Tuck A, Tollervey D. 2012. Transcriptome-wide analysis of exosome targets. *Mol Cell* **48**: 422-433.
- Skourti-Stathaki K, Kamieniarz-Gdula K, Proudfoot NJ. 2014. R-loops induce repressive chromatin marks over mammalian gene terminators. *Nature* **516**(7531): 436-439.
- Takaku H, Minagawa A, Takagi M, Nashimoto M. 2003. A candidate prostate cancer susceptibility gene encodes tRNA 3' processing endoribonuclease. *Nucleic Acids Res* **31**(9): 2272-2278.
- Thompson DM, Parker R. 2009. The RNase Rny1p cleaves tRNAs and promotes cell death during oxidative stress in *Saccharomyces cerevisiae*. *J Cell Biol* **185**(1): 43-50.
- Tuck AC, Tollervey D. 2013. A transcriptome-wide atlas of RNP composition reveals diverse classes of mRNAs and lncRNAs. *Cell* **154**: 996-1009.
- Turowski TW, Lebaron S, Zhang E, Peil L, Dudnakova T, Petfalski E, Granneman S, Rappsilber J, Tollervey D. 2014. Rio1 mediates ATP-dependent final maturation of 40S ribosomal subunits. *Nucleic Acids Res*.
- Vannini A, Cramer P. 2012. Conservation between the RNA polymerase I, II, and III transcription initiation machineries. *Mol Cell* **45**(4): 439-446.
- Vannini A, Ringel R, Kusser AG, Berninghausen O, Kassavetis GA, Cramer P. 2010. Molecular basis of RNA polymerase III transcription repression by Maf1. *Cell* **143**(1): 59-70.
- Webb S, Hector RD, Kudla G, Granneman S. 2014. PAR-CLIP data indicate that Nrd1-Nab3-dependent transcription termination regulates expression of hundreds of protein coding genes in yeast. *Genome Biol* **15**(1): R8.
- Wichtowska D, Turowski TW, Boguta M. 2013. An interplay between transcription, processing, and degradation determines tRNA levels in yeast. *Wiley Interdiscip Rev RNA* **4**(6): 709-722.
- Wlotzka W, Kudla G, Granneman S, Tollervey D. 2011. The nuclear RNA polymerase II surveillance system targets polymerase III transcripts. *EMBO J* **30**(9): 1790-1803.
- Wu J, Huang HY, Hopper AK. 2013. A rapid and sensitive non-radioactive method applicable for genome-wide analysis of *Saccharomyces cerevisiae* genes involved in small RNA biology. *Yeast* **30**(4): 119-128.

TABLES

Table 1. Comparison between RNAPIII read-through for *in vivo* (this study) and *in vitro* (Braglia et al. 2005) studies.

	<i>in vivo</i>	<i>in vitro</i>
SUF2	30%	0%
tV(AAC)M2	16%	0%
tN(GUU)K	8%	2%
tN(GUU)O1	7%	4%
tS(AGA)E	6%	76%
IMT4	4%	13%
tL(CAA)L	3%	2%
tN(GUU)C	1%	83%

FIGURE LEGENDS

Figure 1. Outline and validation of RNAPIII CRAC.

- (A) Scheme of Rpc160-HTP fusion including a C-terminal His6-TEV protease cleavage site-protein A (HTP) tag for CRAC purification.
- (B) Outline of the CRAC crosslinking protocol.
- (C) Transcriptome-wide binding profiles for *MAF1* (wt) and *maf1* Δ strains encoding Rpc160-HTP and for the control BY strain producing untagged Rpc160. Bar diagrams illustrate the percentage of all sequences mapped to each RNA class indicated on the right of the figure.
- (D) Graphical representation of the numbers of reads corresponding to known RNAPIII transcripts found in CRAC and total reads found in RNAPIII CRAC divided into classes.

Figure 2. Unequal distribution of RNAPIII along tRNA genes indicated by genome-wide analysis.

- (A) Cumulative plot of all nuclear tRNA genes aligned to 5'-end of mature tRNA. Distance (nt) is shown on the x-axis.
- (B) As (A) but aligned to 3'-end of mature tRNA.
- (C) Hits distribution along particular genes (tA(AGC)K2, tF(GAA)G, tW(CCA)M) present spiky pattern of RNAPIII transcription. X-axis shows distance (nt) starting from the 5' end of mature tRNA. Green background indicates position of exon(s). Filled part of plot is recognized as RNAPIII read-through of termination signal. Plots for all tRNA genes are present in Supplementary Data.
- (D) Heatmap for all nuclear tRNA genes with 50 nt 5' flanks and 150 nt 3' flanks, clustered using 1-Pearson coefficient. Hits intensity scale is relative to the maximum value within each gene. The 5' peak is the predominant feature.
- (E) All nuclear tRNA genes were ordered according to 5' : 3' peak ratios. The color intensity compares to the number of reads to the global distribution of mapped reads.
- (F) Genome wide profile of 225 tRNA genes (group 1) showing predominant pattern with high 5' peak, aligned to 5' end of mature tRNA. Subpopulation of genes with 5' : 3' peak ratio >1.5
- (G) Genome wide profile of remaining 50 tRNA genes (group 2) showing less common pattern with two similar sized 5' and 3' peaks. Subpopulation of genes with 5' : 3' peak ratio <1.5
- (H) Genome wide profile of all nuclear tRNA genes aligned to A box. The start of the A box sequence is annotated by a vertical line
- (I) As (H) but aligned to B box. The start of the B box sequence is annotated by vertical line

Figure 3. RNAPIII CRAC reveals new RNAPIII transcripts.

- (A) Putative new RNAPIII transcripts were confirmed by Northern blotting using specific probes against *TLT2* (oTWT003) or *TLT6* (oTWT007) and the 5.8S rRNA loading control.

- (B) Inspection of *RDN37* gene encoding ribosomal rRNA reveals pattern characteristic for RNAPIII transcription units.

Figure 4. Comparison of genome-wide tRNA genes transcription under different growth conditions.

- (A) Correlation of relative expression of nuclear tRNA genes between permissive and repressive conditions for wild type strain. The correlation coefficient (R^2) reflects how well the data fit the trend line.
- (B) Correlation of relative expression of nuclear tRNA genes between permissive (glucose at 30°C) and repressive conditions (glycerol at 37°C) for *maf1Δ* strain.
- (C) Ratios of relative expression level of all nuclear tRNA genes under stress conditions reveals a subset of tRNA genes less responsive to regulation by Maf1. Genes are ordered according to gly/glu expression ratio (highest value for tD(GUC)N = 14.78 is not shown). In the same order *maf1Δ*/wt ratio is plotted. The least affected tRNA for each amino acid is colored in dark red.

Figure 5. Transcription termination read-through by RNAPIII

- (A) Northern blotting reveals extended forms of tH(GUG)G2 transcribed from gene tH(GUG)G2 (left panel). Total RNA extracts were analyzed by Northern blotting using probes specific to RT sequence 54 nt downstream of gene tH(GUG)G2 (left panel), overall tH(GUG) (right panel) and 5.8S rRNA (loading control). Extended forms are relatively low abundant in comparison to mature tH(GUG).
- (B) Distribution of reads density for tH(GUG)G2 gene and conditions presented on panel A shows consistency of Northern blotting with results of CRAC and indicates instability or processing of extended transcripts.
- (C) Average read-through (RT) level for all nuclear tRNA genes. RT was calculated for each gene under indicated conditions separately and average RT value for all nuclear tRNA genes is presented.
- (D) Heat map for all nuclear tRNA genes with 50 nt 5' flanks and 250 nt 3' flanks clustered using 1-Pearson coefficient. Hits intensity is log2 of global values. Regions of mature tRNA gene and RT are labeled.
- (E) Histogram showing distribution of RT among all tRNA genes.

Figure 6. Canonical and non-canonical RNAPIII termination signals

- (A) Read-through (RT) level depends on strength of canonical termination signal. Cumulative plot showing distribution of longest U stretch in 50 nt after 3' end of tRNA gene. All tRNA genes were divided according to RT level (< 5%, <15%, <25%, >15%, >25% and >35%) and number of genes is presented in parenthesis.

- (B) A weak first (canonical) terminator (T1) is often supported by second terminator (T2). Analysis of second termination signal in 40 nt range from first termination signal for weak terminators T1=5 and T1=6. The second terminator is classified as strong (more than six U), weak (less than six U), or absent (no T2).
- (C) Uracil abundance is a non-canonical termination signal within RT. Hits distribution along tH(GUG)G2 with marked all U nucleotides (grey lines) revealed correlation between U-richness and RNAPIII transcription/termination. Green background indicates position of exon. Filled part of plot is recognized as RNAPIII read-through.
- (D) Boxplot and line plot showing the distribution of uracil abundance in peaks and following troughs within RT. First, canonical, termination sites were excluded from this analysis. P-value = 2.43×10^{-13} was calculated using Wilcoxon test for peaks with average reads above 300 and considering 20 nt region – 10 nt before and 10 after each peak/valley.

Figure 7. Genome-wide analysis of surveillance factors and its role in targeting transcriptional read-through of tRNA genes.

Genome wide profile of nuclear tRNA transcripts aligned to 3' end of mature tRNA. Genome wide hits distribution for: (A) RNAPIII (Rpc160), (B) RNAPII (Rpb1; GSM1943529) (Holmes et al. 2015), (C) Nab2, (D) Rat1 (Granneman et al. 2011), (E) Rrp44, (F) Rrp44-exo, (G) Rrp6, (H) Mtr4.

Figure 8. Model for non-canonical termination of RNAPIII. Nab2 accompanies RNAPIII on tDNA chromatin. When canonical termination occurs transcript is released and RNAPIII starts another transcription cycle. Ever terminator read-through occurs the 3' extended pre-tRNAs are targeted by Nab2 and Mtr4/Rrp44 to degradation. Pre-tRNAs can be endonucleolytically cleaved by RNase Z. tRNA extensions could be degraded by Rat1 and/or the exosome, or potentially processed to functional ncRNAs.

SUPPLEMENTARY FIGURE LEGENDS

Figure S1. Genome wide plots.

- (A) SDS-PAGE purification of the RPC160-HTP protein, visualized by autoradiography of the cross-linked and radioactively labeled RNA. The red frame indicates the region excised for elution of RNA-protein complexes.
- (B) Genome wide profile of all nuclear tRNA genes under tested growth conditions with high 5' peak, aligned to 5' end of mature tRNA.

Figure S2. Hits density across non-tRNA RNAPIII transcripts

Mature RNA regions are colored:

- (A) *RDN5-1*
- (B) *RDN5-2*
- (C) *SCR1*
- (D) *SNR6*
- (E) *SNR52*
- (F) *RNA170*
- (G) *RPR1*
- (H) *ZOD1*

Figure S3. Predominant 5':3' peak ratio is correlated with highly transcribed genes.

- (A) Ratio of 5' : 3' peaks plotted against relative gene transcription (total number of reads mapped to a gene) is generally greater for highly expressed tRNA genes under permissive conditions
- (B) 5' : 3' peak ratio distribution among tRNA genes grouped by transcription rate.
- (C) Under restrictive growth conditions the 5' : 3' peak ratios are not well correlated with gene transcription rate.
- (D) As (B) but RNAPIII activity is repressed.
- (E) All nuclear tRNA genes were ordered according to 5' : 3' peak ratio. The color intensity shows number of reads in global scale of mapped reads, showing that under restrictive growth conditions 5' : 3' peak ratio is not well correlated with gene transcription rate.

Figure S4. Possible explanations of RNAPIII density patterns

- (A) Genome-wide distribution of peaks along tRNA genes. Each peak for each tRNA gene was counted as 1 and sum of peaks was plotted starting from the first peak. Raw plot (grey line) and trend line (black dotted line) are presented. Positions of the first four peaks are marked.

- (B) Hits distribution along particular genes (tF(GAA)H2 and tT(AGU)B) with annotated A box (red background) and B box (dark green background). Light green background indicates position of exon(s). Filled part of plot is recognized as RNAPIII read-through.
- (C) Genome wide profile for group 2 of 50 tRNA genes (with 5' : 3' peak ratio <1.5) (panel I and II) and for intron-containing and intron-less tRNA genes (panels III and IV). Vertical lines show start sites of A and B box sequences.

Figure S5. Analysis of transcriptional read-through of RNAPIII

- (A) Northern blotting reveals extended forms of pre-tRNA transcribed from tE(UUC) genes. Extended forms are relatively low abundant in comparison to mature tE(UUC). 5.8S rRNA is the loading control.
- (B) As (A) but probes specific for tK(UUU): 5'tK(UUU) and tK(UUU)O_RT (recognize sequence 40 nt downstream of gene tK(UUU)O) were used.
- (C) As (A) but tY(GUA) probe was used (named SRM15).
- (D) Cumulative plots of tRNA genes aligned to 3'-end of mature tRNA. Plots were generated for all tRNA genes and tRNA genes presenting indicated RT level.
- (E) Histogram showing distribution of RT under different growth conditions.
- (F) Table presenting relationship between RT level and 3'-extension sequence. T4 and longer T tracts in genomic sequence are marked in black.

Figure S6. Reasons of transcriptional read-through of RNAPIII

- (A) Hits distribution along some genes encoding glutamic acid tRNA (tE(UUC)E3, *SOE1*, tE(UUC)G2, tE(UUC)G3, tE(UUC)I, tE(UUC)J) with marked all U nucleotides (grey lines) revealed correlation between U-richness and RNAPIII transcription/termination. Green background indicates position of exon(s). Filled part of plot is recognized as RNAPIII read-through (RT).
- (B) High transcriptional RT is characteristic for group of tRNA isotypes. On the plot RT is ordered according to isotypes.
- (C) Correlation plot for transcriptional RT and ΔG between DNA/DNA and RNA/DNA duplexes for first 20 nt after 3' end of tRNA gene. Panels show correlation for all tRNA genes (left panel), isotypes genes with low RT (middle) and isotypes where RT of at least two genes exceeds 25% (right panel).

Fig.1 Turowski et al.

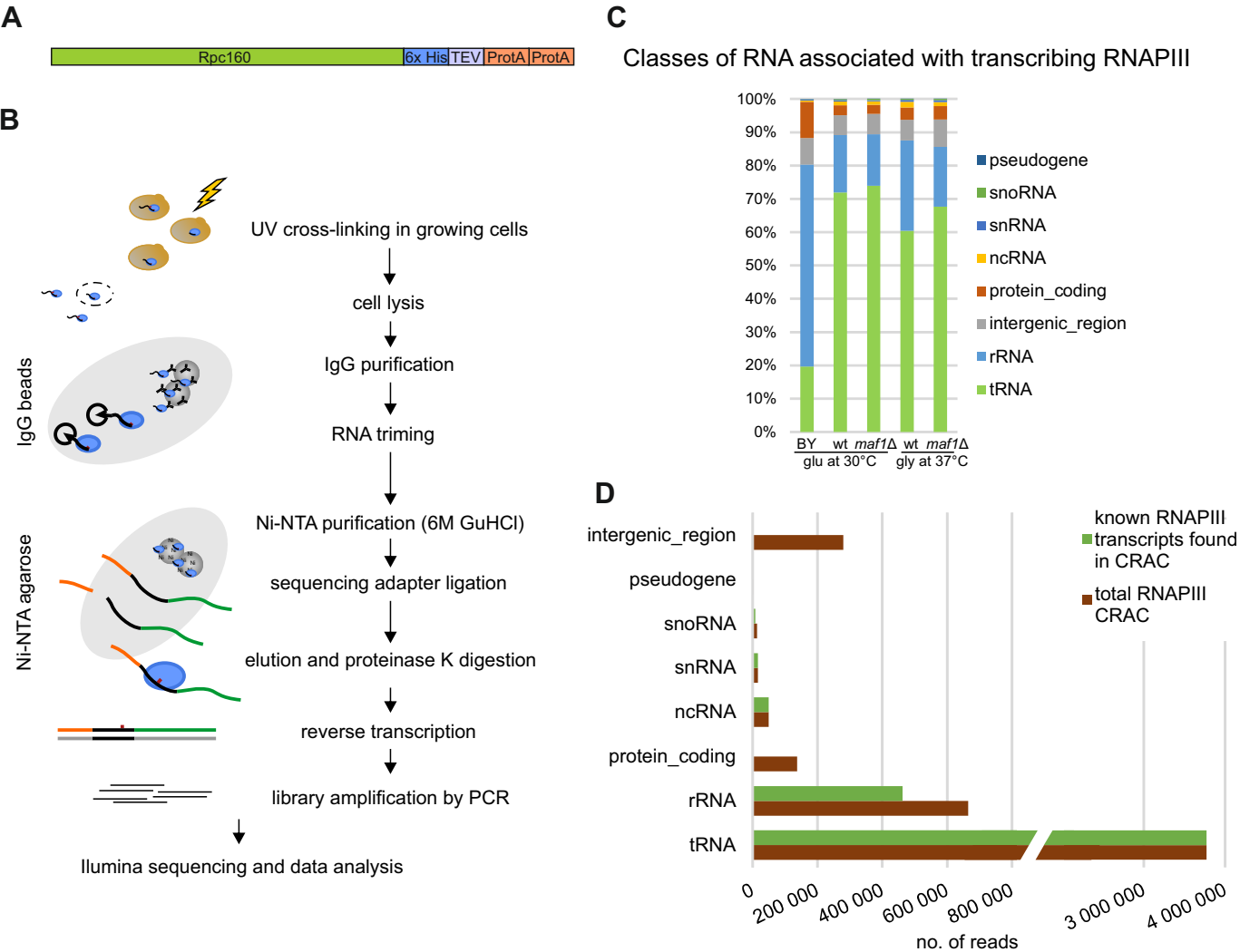


Fig.2 Turowski et al.

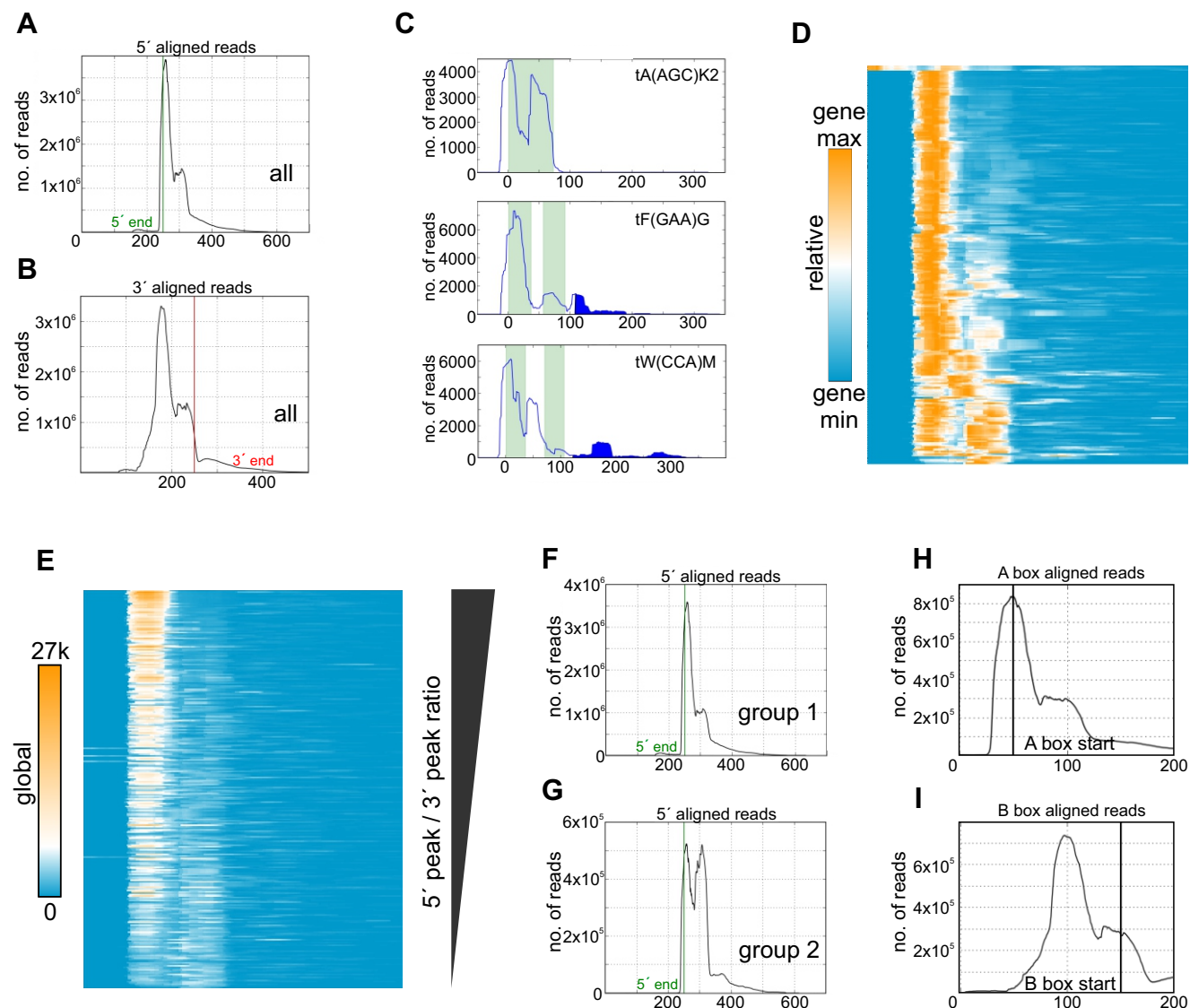


Fig.3 Turowski et al.

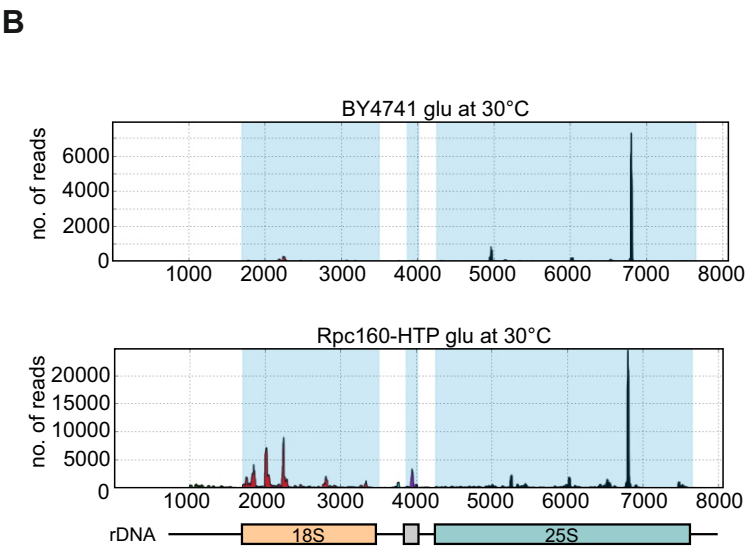
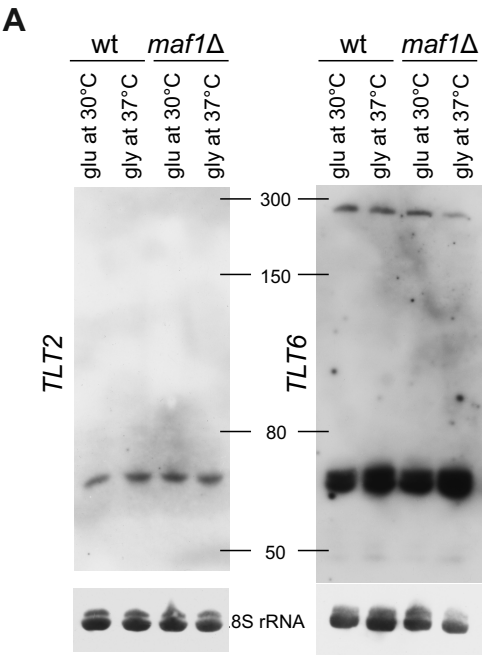


Fig.4 Turowski et al.

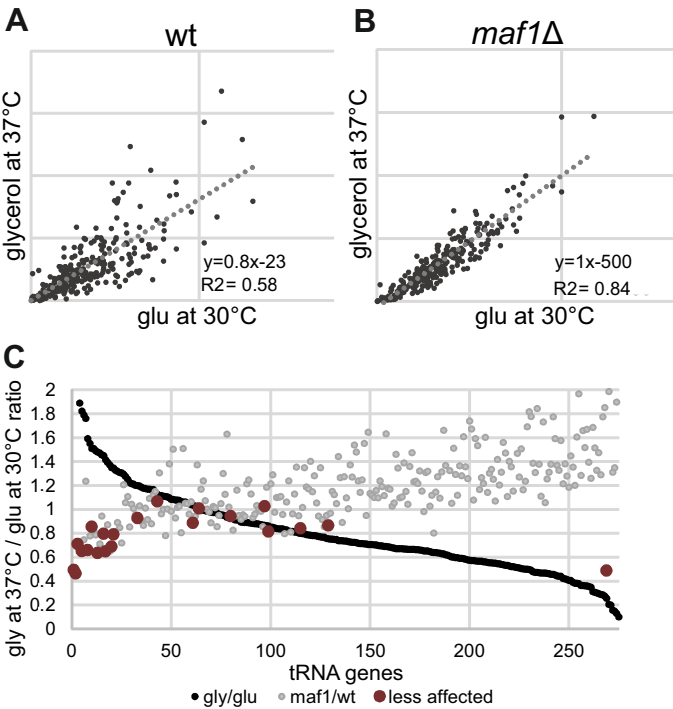


Fig.5 Turowski et al.

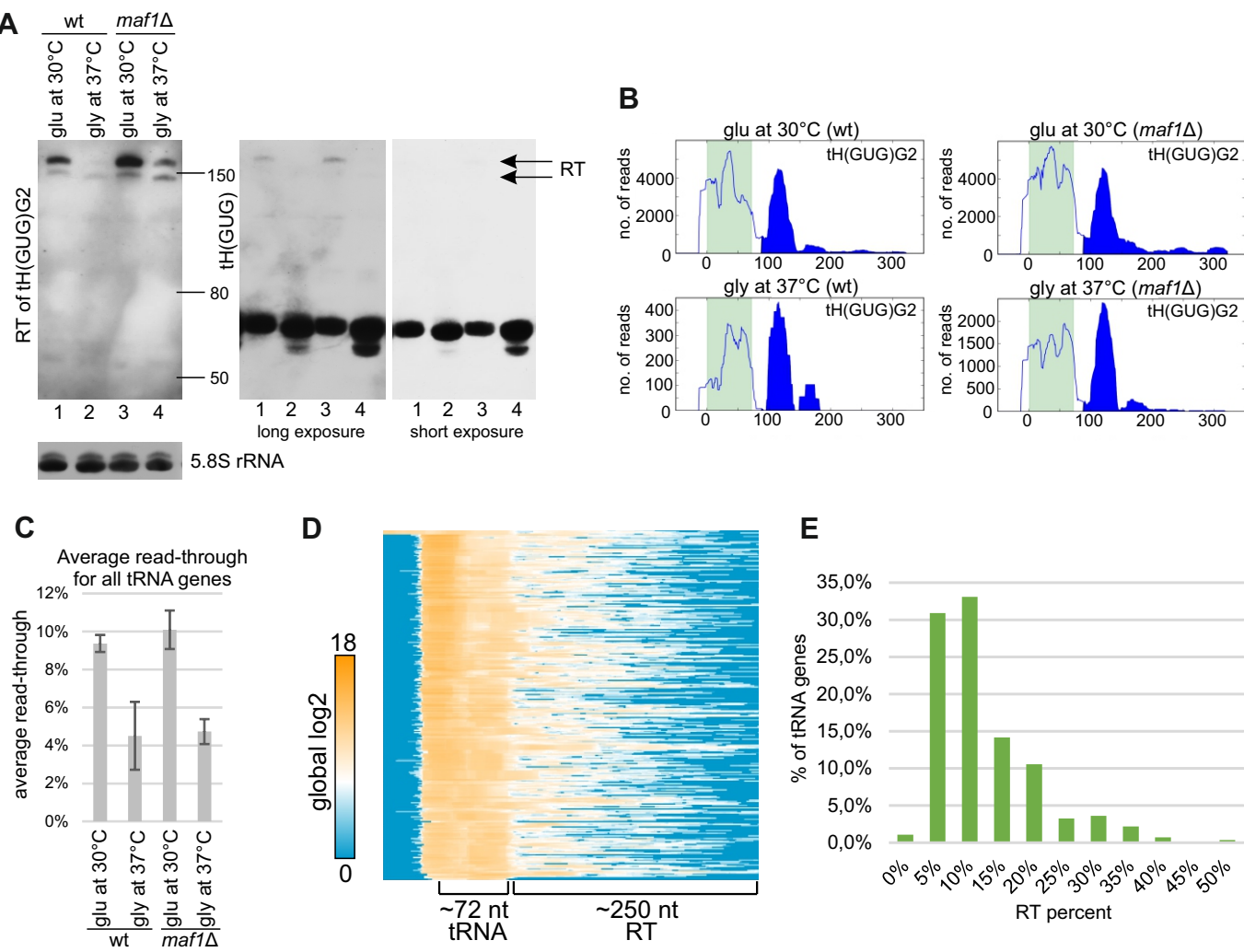


Fig.6 Turowski et al.

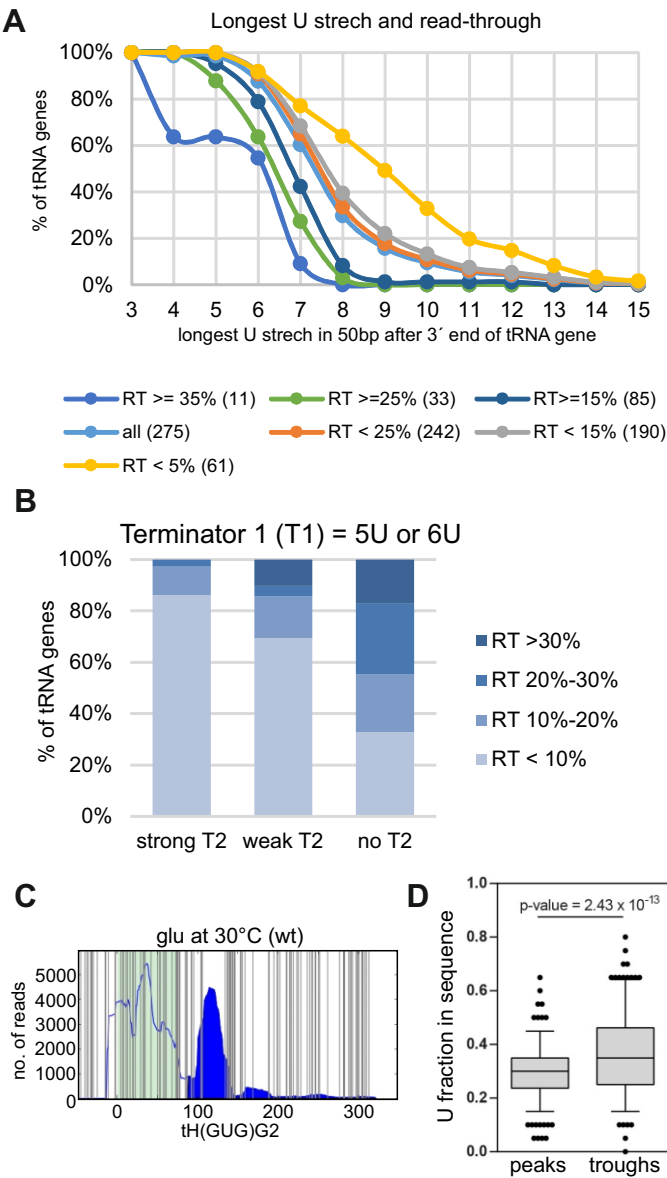


Fig.7 Turowski et al.

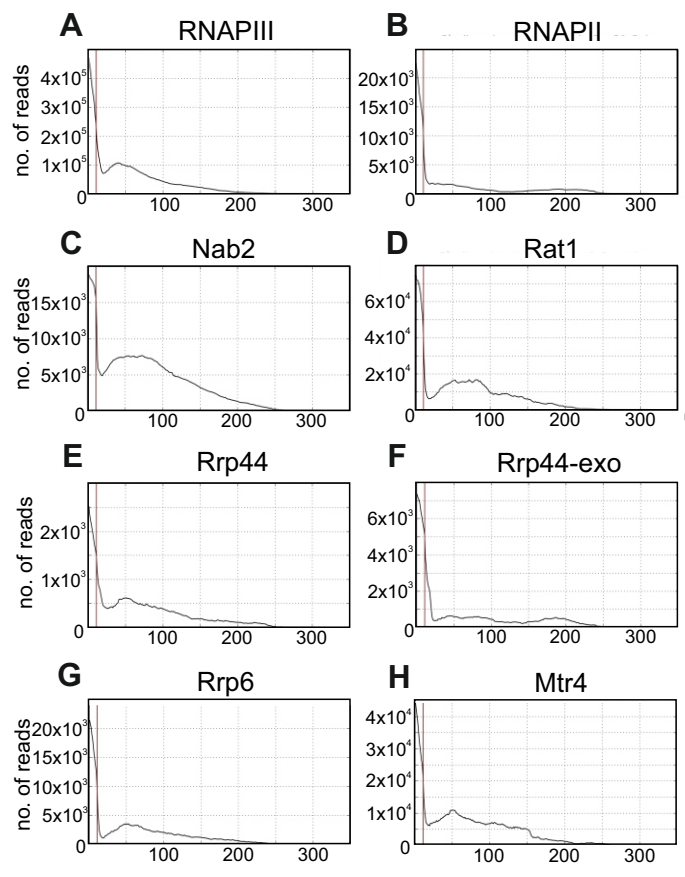
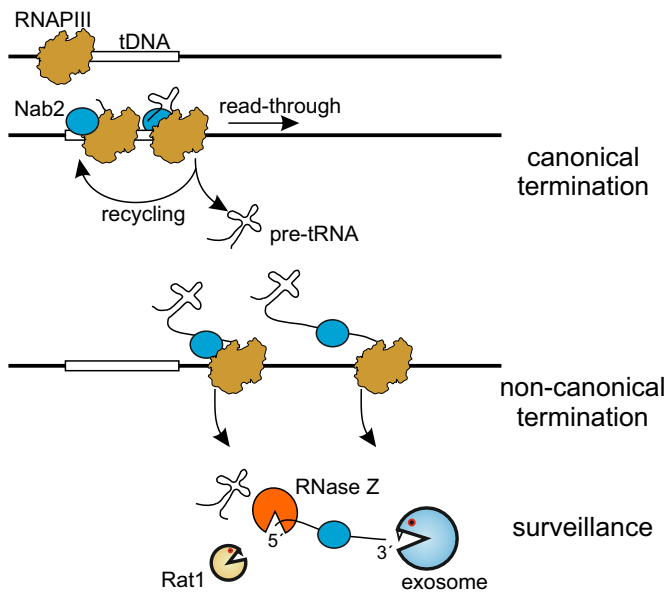
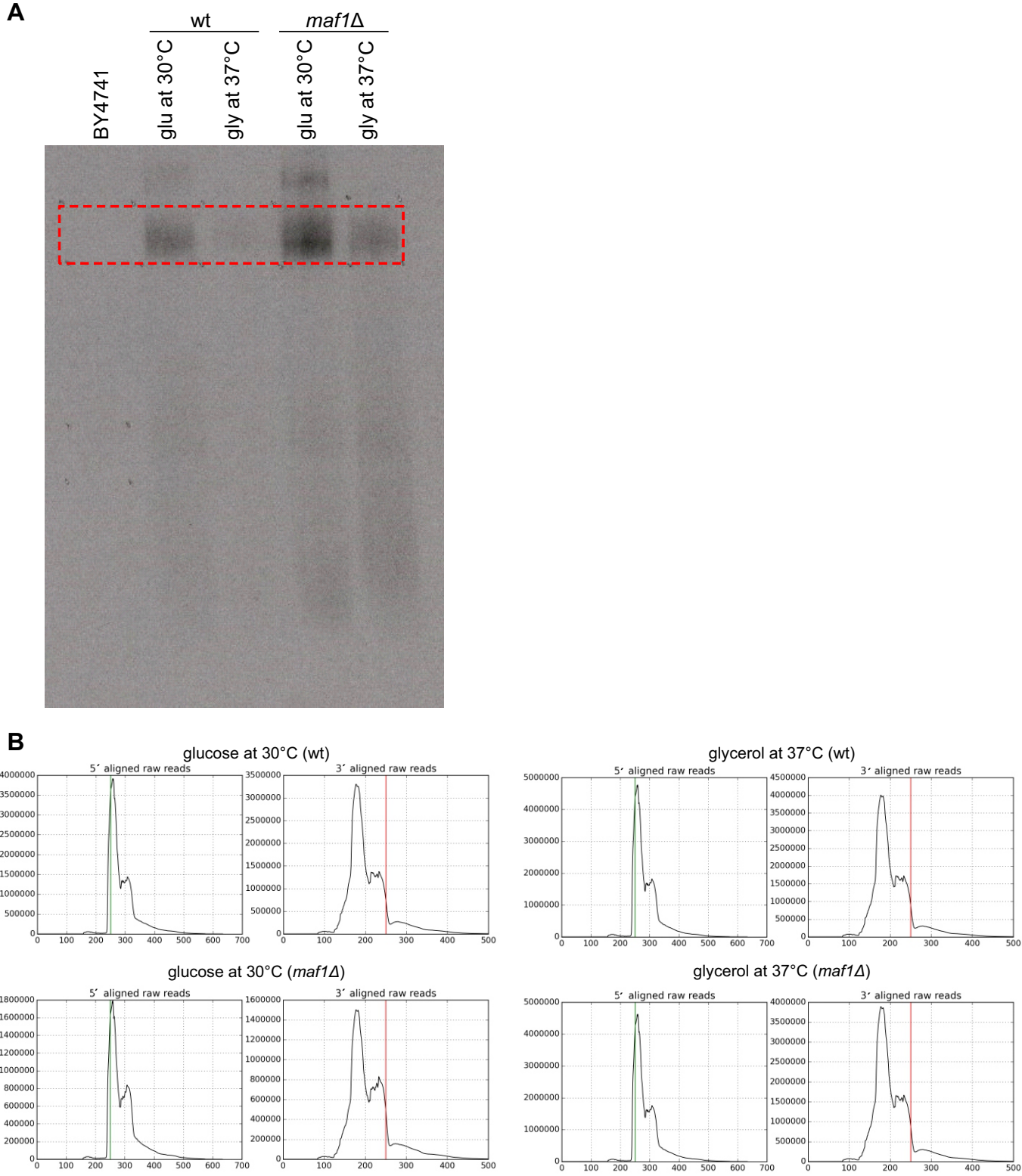


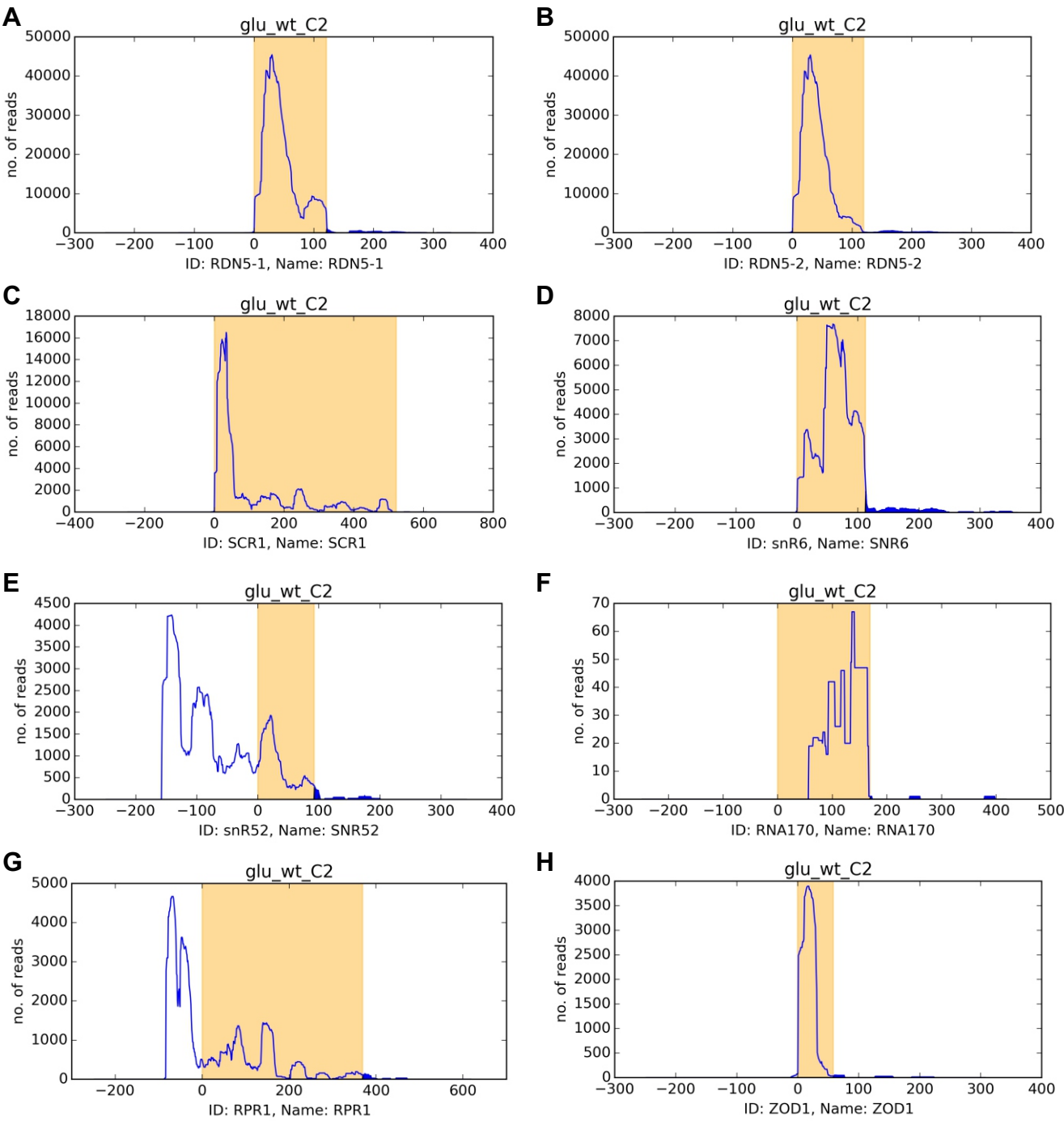
Fig.8 Turowski et al.



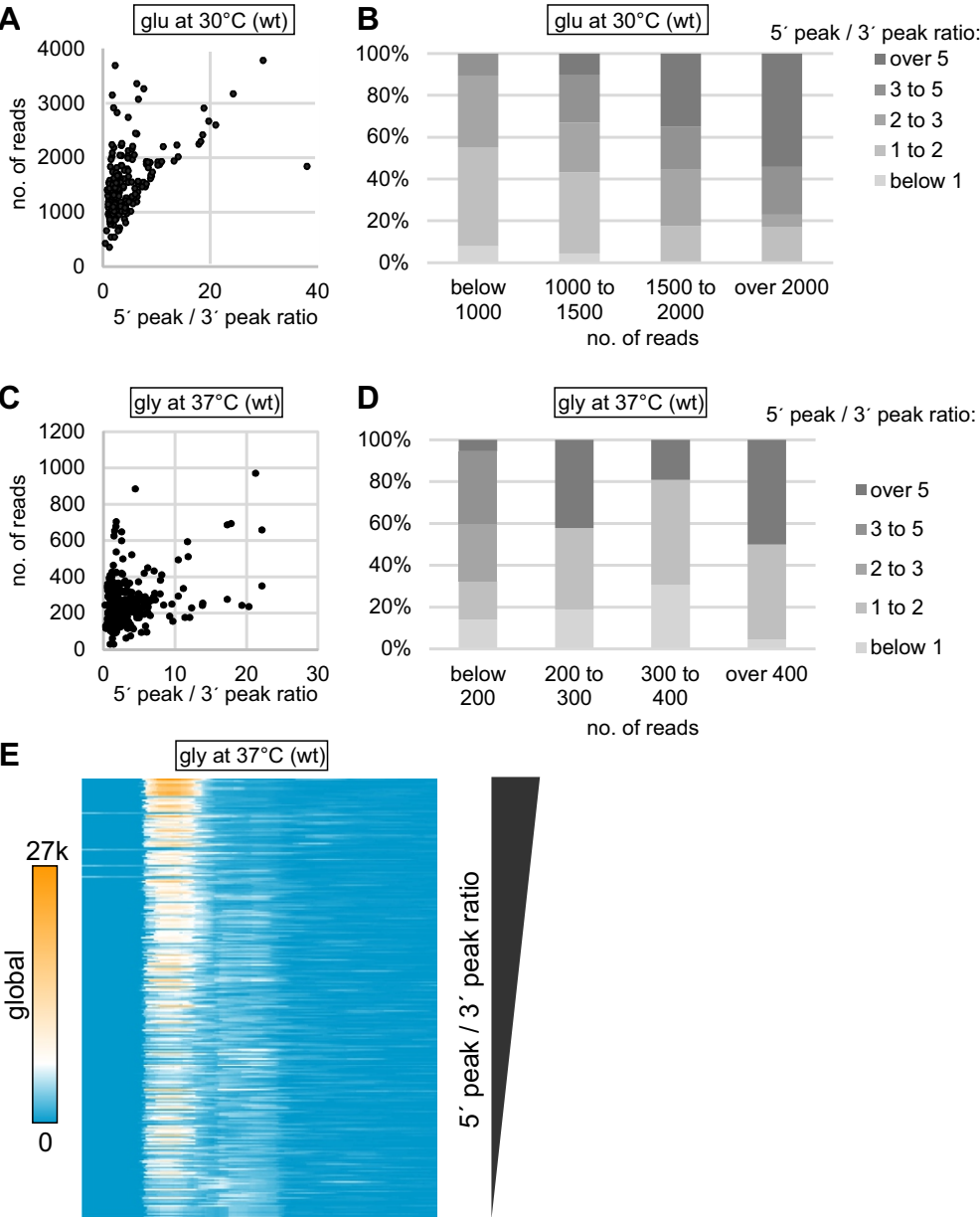
Supplemental Fig.S1 Turowski et al.



Supplemental Fig.S2 Turowski et al.

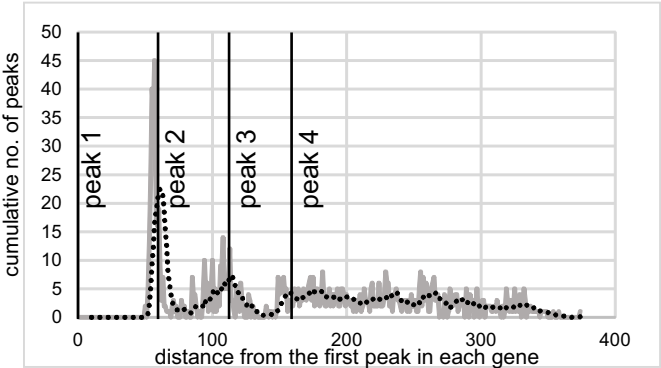


Supplemental Fig.S3 Turowski et al.

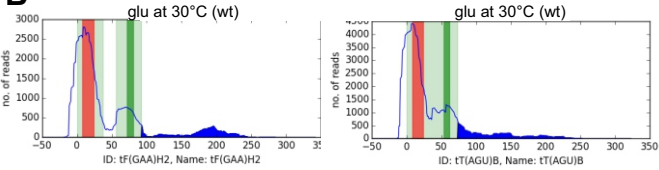


Supplemental Fig.S4 Turowski et al.

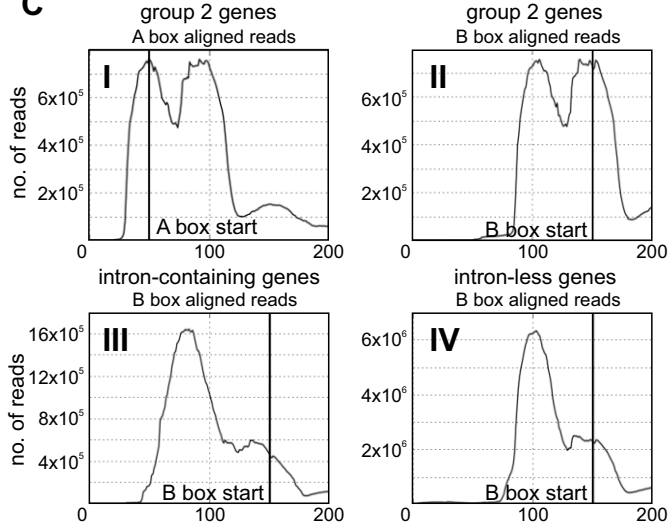
A



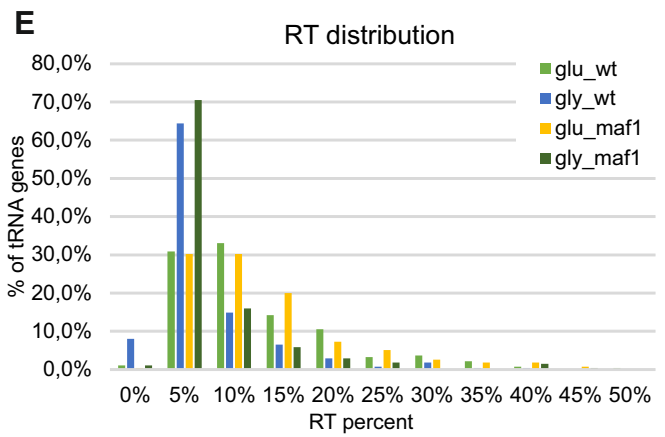
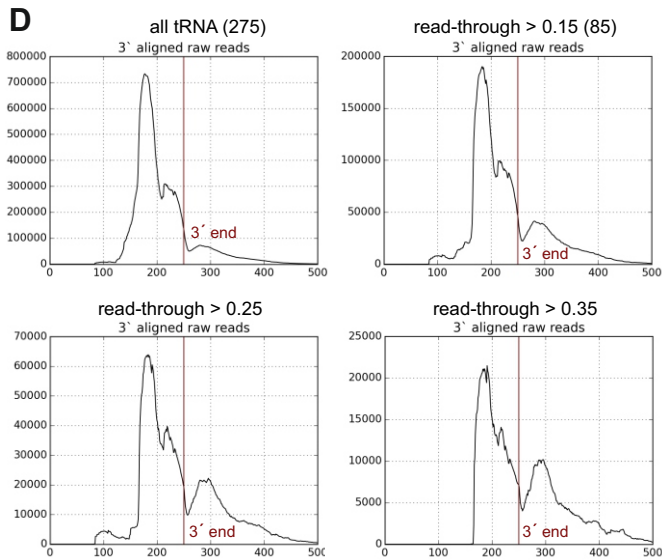
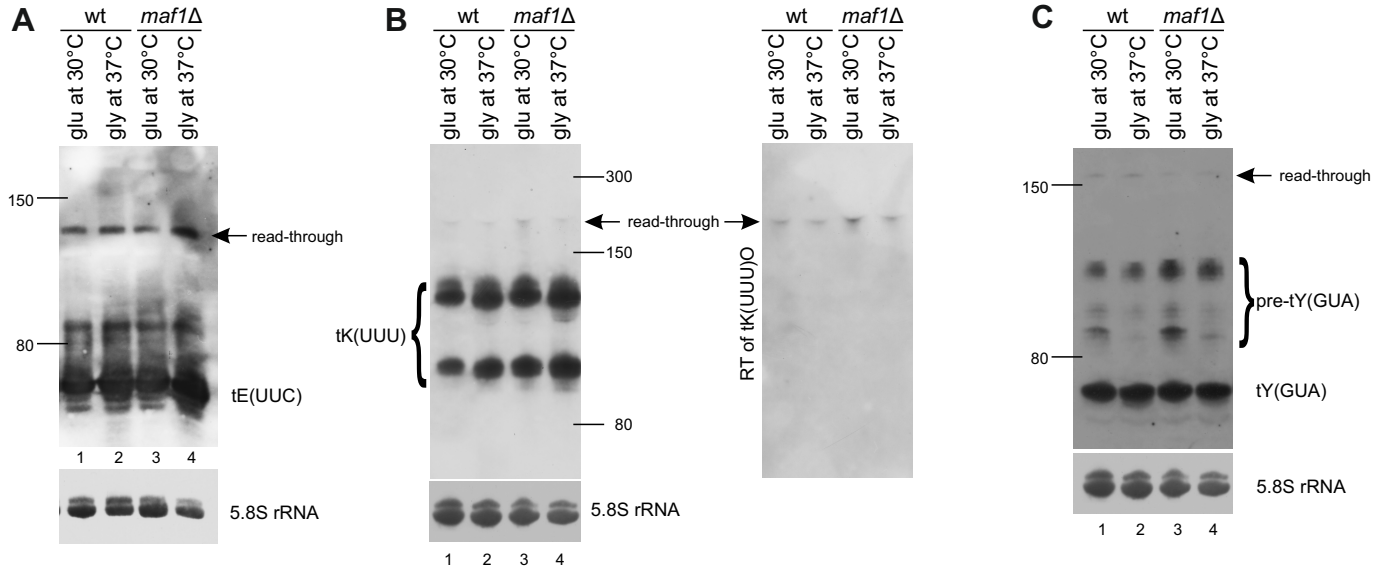
B



C



Supplemental Fig.S5 Turowski et al.



F

gene name	gene ID	RT	last 10 nt of tRNA gene + 50 nt after
tE (UUC) E2	tE (UUC) E2	56, 1%	CGTATCGGAGTACTTTTTGACATCATACAGTGTATGTATGGGGTTAGGAGCCACCTTCC
SOE1	tE (UUC) G1	47, 4%	CGTATCGGAGTTTTTTGTCGTAAGTTGAGAAAGTTTTTCCAGTTAAGTTATTATGACGC
tE (UUC) E1	tE (UUC) E1	39, 3%	CGTATCGGAGTACTTTTTTGATGGCTACAATCATAATTTGACATACCTCATATGAATATG
tE (UUC) G3	tE (UUC) G3	32, 5%	CGTATCGGAGTGTTCCTTTTTCTTACCTCTTGAAAAAGAGGTAGTAAAGTCAATATAAATT
tE (UUC) P	tE (UUC) P	29, 3%	CGTATCGGAGTTTTCTTTTTGCTTCCTTCCGCCCTTCCAAGTTGTTGCATAATCTTATCT
tE (UUC) I	tE (UUC) I	17, 7%	CGTATCGGAGTATGTTTATTTTATTTTTTTCATATCCGCTCGAATATTGCCTTCGTTTAA
tE (UUC) E3	tE (UUC) E3	15, 6%	CGTATCGGAGTATTCGTTATATTTTTTCAATCGCTTACGATCAAAAGCTTTGAAAACATT
tE (UUC) K	tE (UUC) K	13, 5%	CGTATCGGAGTTTTTCATTATTATTTTTTAATCTAGCAGTACTAGCGTGATGTTAATGTTAC
tE (UUC) L	tE (UUC) L	10, 8%	CGTATCGGAGATATATTTAATATTTTTTTGATTTGACTTCAAATGAGAACTATTTTTTT
tE (UUC) M	tE (UUC) M	10, 7%	CGTAGCGGAGTTTTCTTTTTGAAGAAATTTGAAAAATTTTTTCTTCACGTCGACCAAGGT
tE (UUC) B	tE (UUC) B	10, 7%	CGTATCGGAGTTATTTTTTGTTTCTATACCATTATTTTTCTTGAATGTTGCATTATGATA
tE (UUC) G2	tE (UUC) G2	9, 5%	CGTATCGGAGTATTTTTTGAAAGTTTTTTTTTTCACGTTAGTTCGTTGAAGAAAATTTTTT
tE (UUC) J	tE (UUC) J	9, 5%	CGTATCGGAGTATTTTTTATGTGATTAGCTAAATAGTTATTTTTTGAAATAGTTTGTAGT
tE (UUC) C	tE (UUC) C	8, 1%	CGTATCGGAGTTTTTTTTTCTGCTATTACAATTAAGATTTTTTCAACTATATCTGAGA

Supplemental Fig.S6 Turowski et al.

