Substrate Specificity of the TRAMP Nuclear Surveillance Complexes

Clémentine Delan-Forino[1], Christos Spanos[1], Juri Rappsilber[1], David Tollervey[1]

1: Wellcome Center for Cell Biology, University of Edinburgh, UK.

Running Title: Targeting RNAs for Nuclear Surveillance

ABSTRACT

During nuclear surveillance in yeast, the RNA exosome functions together with the TRAMP complexes. These include the DEAH-box RNA helicase Mtr4 together with an RNA-binding protein (Air1 or Air2) and a poly(A) polymerase (Trf4 or Trf5). To better determine how RNA substrates are targeted, we analyzed protein and RNA interactions for TRAMP components. Mass spectrometry identified three distinct TRAMP complexes formed *in vivo*. These complexes preferentially assemble on different classes of transcripts. Unexpectedly, on many substrates, including pre-rRNAs and pre-mRNAs, binding specificity was apparently conferred by Trf4 and Trf5. Clustering of mRNAs by TRAMP association showed co-enrichment for mRNAs with functionally related products, supporting the significance of surveillance in regulating gene expression. We compared binding sites of TRAMP components with multiple nuclear RNA binding proteins, revealing preferential colocalization of subsets of factors. *TRF5* deletion reduced Mtr4 recruitment and increased RNA abundance for mRNAs specifically showing high Trf5 binding.

INTRODUCTION

The transcription, processing and packaging of eukaryotic RNAs offers many opportunities for errors (reviewed in [1,2]). In consequence, quality control or "surveillance" of aberrant nuclear RNAs is an essential feature of eukaryotic gene expression. In addition, all stable RNA species undergo post-transcriptional maturation, including 3′ processing. A key component of both RNA surveillance and processing pathways is the exosome; an essential, multi-subunit complex that is highly conserved among eukaryotes.

The purified exosome shows weak activity *in vitro,* indicating that the rapid and processive activity inferred *in vivo* requires co-factors. In budding yeast, the Trf4/5-Air1/2-Mtr4 polyadenylation (TRAMP) complexes are major cofactors for the nuclear exosome and critical for nuclear RNA surveillance activity [3-6]. There are variants of the TRAMP complexes, but each has three components: Trf4 or Trf5, Air1 or Air2, and Mtr4, all of which are conserved to humans. Mtr4 is an ATP-dependent RNA helicase [7,8] and is essential for all known activities of the nuclear exosome. In contrast, the other TRAMP components are required for RNA surveillance in the nucleolus and nucleoplasm but are not known to participate in the accurate processing of stable RNA species. Air1 and Air2 are zinc-knuckle, putative RNA binding proteins [9], whereas Trf4 and Trf5 are noncanonical poly(A) polymerases. Together, they add a short oligo(A) tail, significantly shorter than poly(A) tail added by the canonical polymerase (Pap1) [10]. This is presumed to provide a single stranded "landing pad" that makes the RNA a better substrate for 3′ end degradation. It was previously suggested that TRAMP recruitment to substrates would be driven by Air1/2 through their RNA binding activities [4,11] (reviewed in [12]). In addition, Mtr4 includes a distinctive "Arch" or "KOW" domain, consisting of a beta-barrel stalk inserted into the typical DExH core [13-15]. The Arch domain is specifically bound by ribosomal biogenesis factors carrying an Arch Interacting Motif (AIM), enabling direct Mtr4 recruitment to pre-rRNAs [16]. The Arch was also proposed to function as a docking platform for Trf4 and Air2, acting independently of the helicase activity of the DExH core.

Human cells express at least three complexes that each contain MTR4 and a zinc finger protein. These include TRAMP, which is predominately nucleolar in humans, as well as NEXT (Nuclear EXosome Targeting) and PAXT (Poly(A) eXosome Targeting) complexes, which are nucleoplasmic [17-19]. Budding yeast appears to lack NEXT and PAXT homologues, but the TRAMP complexes are present in both the nucleolus and nucleoplasm. The Air1/2 and Trf4/5 pairs show some functional redundancy, since the single mutants are viable whereas double mutants are inviable or severely growth impaired, depending on strain background [4,9,20]. Structural and functional analyses have focused on the Trf4-Air2 TRAMP

complex [21-25]. However, both the Air1/2 and Trf4/5 pairs have very significantly diverged in sequence (Figs. 1A and 1B). The N-terminal regions of Air1 and Air2 are only ~44% identical (aa 1-72 for Air1, from 360 total; aa 1-59 for Air2 from 344 total), while the C termini are substantially different (15% identity between Air1 aa 204-360 and Air2 aa 193-344) [19]. Similarly, Trf4 and Trf5 show homology over their central regions (67% identical amino acids over 117 to 519 in Trf4, from 584 total), but diverge in both the N- and C-terminal regions. These sequence divergences are much greater than most duplicated gene pairs in yeast, suggesting that the functions of the different forms of the yeast TRAMP complex may also have diverged and developed distinct specificities *in vivo*.

The two C-terminal zinc knuckles of Air2 (ZK4 and ZK5, residues 119–199) mediate the interaction with the central domain of Trf4 [23,24,26], which requires the presence Air1 or Air2 to adenylate its substrate *in vivo* [6]. However, a catalytically inactive Trf4 mutant (Trf4-DADA) can support degradation of most Trf4 targets and rescue the lethality of a *Δtrf4 Δtrf5* double mutant, indicating that Trf4 can target RNAs to the exosome independently of adenylation [20]. In strains depleted of Mtr4, TRAMP substrates are both stabilized and hyper-adenylated [27]. This shows that Trf4/5 can be recruited to target RNAs and activated independent of Mtr4 and the exosome.

Here, we aimed to determine which TRAMP complexes form *in vivo*, how they bind to different substrate classes, and how they cooperate with the exosome. To this end, we characterized TRAMP protein-protein and protein-RNA interactions via mass spectrometry and CRAC (UV-crosslinking analyses), respectively. Unexpectedly, the results indicate major roles for Trf4 and Trf5 in TRAMP targeting and recruitment, and a specific role for Trf5 in mRNA stability. In contrast, Air1 and Air2 appear to be highly redundant.

RESULTS

Three distinct TRAMP complexes are detected *in vivo*

The TRAMP complexes potentially comprise four different combinations of Trf4/5 and Air1/2, together with Mtr4. We assessed the actual combinations formed *in vivo*, by tandem-affinity purification and mass spectrometry (MS). Air1, Air2, Trf4 and Trf5, were each tagged with His6-TEV-proteinA (HTP). Purifications were initially performed under high salt conditions (1M NaCl) to recover only stable interactions. Associated proteins were identified by mass spectrometry and subjected to label-free quantification. iBaq scores [28,29] were calculated for each protein recovered (Supplementary Table S2). These indicated that Air1 interacts with both Trf4 and Trf5 while Air2 interacts almost exclusively with Trf4 (Fig. 1C). Previous analyses reported that Mtr4 is not efficiently retained in TRAMP above 125 mM NaCl [4]. Consistent with this, Mtr4 was weakly recovered in the 1M NaCl preparations. These data demonstrate that three distinct TRAMP complexes co-exist *in vivo*, containing Trf4 + Air1 (TRAMP4-1), Trf4 + Air2 (TRAMP4-2) and Trf5 + Air1 (TRAMP5-1) (Fig. 1D). Each presumably also associated with Mtr4.

Trf5 and Air2 were not detected in association with either reciprocal precipitation, but it was unclear whether this reflected an inability to interact, or simply a higher affinity for Trf5-Air1 binding. To test this, *AIR1* was deleted in the strain expressing Trf5-HTP. In the *air1*Δ strain, Air2 was well recovered with Trf5-HTP, indicating competition for Trf5 association and redundancy between Air1 and Air2 (Fig. 1C).

To further characterize factors binding to Trf4 and Trf5, these purifications were repeated at 150mM NaCl, with and without the inclusion of RNase treatment (Supplementary Table S3 and Fig. 1E; colored boxes indicate fold enrichment in the precipitation indicated relative to the non-tagged control; <2 fold enrichment is grey). In this lower stringency purification, Mtr4 was well recovered with both Trf4 and Trf5, independent of RNA. However, in the *air1*Δ strain, recovery of Mtr4 with Trf5 was RNase sensitive (Trf5 *air1*Δ). This suggests that the Trf5-Air1 complex binds jointly to Mtr4, and that this interaction is not fully recapitulated by Trf5-Air2, making the complex less stable.

Nab3 and Nrd1, subunits of the Nrd1-Nab3-Sen1 complex, were recovered with Trf4 but not Trf5, consistent with the reported presence of a Nrd1-interacting motif in Trf4 but not Trf5 [25]. Both Nab3 and Nrd1 are implicated in RNA surveillance on RNAPII and RNAPIII transcripts but are not known to participate in pre-rRNA degradation. The 5′ exonuclease Xrn1 showed RNase sensitive recovery with Trf4 and was also identified with Air1 and Air2. Xrn1 is predominately cytoplasmic but several studies have reported nuclear roles [30-33], consistent

with this finding. Several early-binding, pre-mRNA packaging factors were also recovered with Trf4 and Trf5, likely reflecting the role of TRAMP in promoting rapid degradation of ncRNAs [6,34], and these interactions were largely RNase sensitive.

Multiple proteins associated with both the box C/D and box H/ACA classes of snoRNA were strongly recovered with Trf4, Air1, and Air2, consistent with the reported involvement of Trf4 in snoRNA maturation [35,36]. Curiously, recovery of the snoRNP proteins with Trf5 was enhanced in the *air1*Δ strain, which has increased Trf5-Air2 association, possibly indicating that Air2 contributes to snoRNP binding. A subset of ribosome synthesis factors was recovered, showing greater interaction with Trf5 than Trf4 and, particularly, with Air1 relative to Air2, for which none showed >2 fold enrichment (Fig. 1E). This suggests a preferential role for TRAMP5-1 in pre-rRNA degradation.

Notable omissions were Utp18 and Nop53, which were reported to bind Mtr4 directly in the context of pre-ribosomes and promote TRAMP-independent, pre-RNA processing [16]. The exosome was also very poorly recovered (Supplementary Table S3), consistent with the failure of the original exosome purifications to recover TRAMP components and *vice versa* [4,5,37]. TRAMP-exosome interactions may be too transient *in vivo* to be readily recovered.

The TRAMP complexes exhibit distinct substrate preferences

RNA targets of the different TRAMP subunits were identified by *in vivo* UV-crosslinking followed by protein purification and sequence analysis of cDNAs (CRAC). We compared strains in which the endogenous gene was HTP tagged for Air1, Air2, Trf4, Trf5, and Mtr4, as well as the exosome exonucleases Rrp44 and Rrp6. The Mtr4 arch domain is implicated in substrate recruitment [13,14,38,39], so we also constructed and analyzed a tagged Mtr4 mutant lacking this region (Mtr4Δarch) (Fig. S1A). The Mtr4Δarch construct was expressed from P_{MTR4} in a strain in which wild type Mtr4 was under P_{GAL} control, allowing its depletion on glucose medium (Fig. S1B).

Comparison of RNA target classes recovered with each protein showed clear differences (Fig. 2A). Notably, pre-rRNA spacer regions (ETS and ITS) were substantially more targeted by Air1 than Air2, whereas Trf4 and Trf5 exhibited similar recovery. They were also less recovered with Mtr4Δarch relative to Mtr4, consistent with pre-rRNA maturation defects reported for strains carrying arch mutations (Fig. S1C) [13,14,38,39]. In contrast, the major ncRNA classes, CUTs, SUTs and XUTs, were strongly recovered with Air2 and Trf4, relative to Air1 plus Trf5. Notably, protein-coding genes appeared to be major targets of Trf5, which showed the strongest recovery compared to the other factors.

TRAMP targets are expected to be subject to oligoadenylation, so we specifically analyzed cDNAs that carry additional, 3′ terminal non-templated A residues (Fig. 2B). These are very likely to represent the authentic 3′ ends of RNAs that have been recognized by one of the TRAMP complexes and targeted for degradation by oligo(A) addition. In these datasets, the preference for Air1 over Air2 in pre-rRNA binding was even more marked. Similarly, the preference for Air2 and Trf4 in recovery of ncRNAs was more striking in the oligo(A) population. This was seen for CUTs, SUTs and XUTs, as well as antisense and intergenic regions and RNAPIII transcripts ("other ncRNAs" in Fig. 2B).

Similarities and differences between TRAMP components were assessed by the degree of correlation between the recovered targets for RNAPII transcripts (Figs. 2C and S2, Supplementary Table S4) As expected, target RNA species ("Co-targeting") were more similar than the colocalization of the precise binding sites for different factors ("Colocalization"). Consistent with the proteomic analyses and target RNA classes, Air1 binding was similar to Trf4 and Trf5, while Trf4 showed similarity to both Air1 and Air2. In contrast, Trf5 binding was more similar to Air1 than to Air2, while Air 2 more resembled Trf4 that Trf5. Notably, targets for Mtr4Δarch were substantially less well correlated with all other TRAMP components than intact Mtr4 (Fig. 2C). This might not have been anticipated, since the arch domain was implicated in targeting Mtr4 to degrade specific pre-rRNA spacer regions that are independent of TRAMP [15,16]. Higher correlations between different TRAMP components were observed for oligo(A) RNAs than considering all reads (Figs. 2C and S2), confirming that filtering for adenylated species more reliably identifies *bona fide* surveillance targets.

Trf5-Air1 recruits Mtr4 to the A2-A3 region of the pre-rRNA

Mtr4 and the exosome are required for degradation of the excised 5′ ETS and the ITS2 region present in the 7S pre-rRNA. These are "default" activities, in that they are required during maturation of all pre-rRNAs, and neither is known to involve the TRAMP complexes. In both cases, Mtr4 is directly recruited to the pre-ribosome by binding to AIM-containing proteins, Utp18 and Nop53, respectively [16]. In contrast, the 23S RNA, an aberrant intermediate which extends from the transcription start to site A3 in ITS1, is also a characterized Mtr4 and exosome substrate, but is an subject to RNA oligo-adenylation and surveillance [27]. To assess the roles of other TRAMP components in Mtr4 recruitment, we tested strains lacking Air1, Air2, Tr4 or Trf5. Notably, loss of the other TRAMP components did not affect Mtr4 abundance (Fig. S1D).

Analysis of protein binding in ITS1 (Fig. 3, S3) using oligo(A) reads (Fig. 3) showed high binding of Air1, Trf5, and Mtr4 at positions 5′ to site A3 (the 3′ end of 23S) relatively to total reads (Fig S3A, S3D-E), whereas little association was seen for Air2 or Trf4 (Fig. 3B, 3C, S3B, S3C). Mtr4 binding across ITS1 was also tested in strains deleted for TRAMP component (Fig. 3F-I, S3F-I). Loss of Trf5 or Air1 strongly reduced Mtr4 association (~5.8 fold and ~3.7 fold, respectively) (Fig. 3F, 3I, 3M). The effects of *air1*Δ were more modest, probably due to its replacement by Air2 (see Fig. 1C). Strains lacking Air2 also showed a reduction in Mtr4 association with ITS1 (~3.2 fold decrease) while loss of Trf4 had a milder effect on Mtr4 binding at this site (~ 2.0 fold).

These observations probably reflect both alterations in the balance of the remaining TRAMP complexes, when one component is absent, and the significant redundancy between Air1/Air2 and Trf4/Trf5 proteins. In the absence of Air2, increased levels of free Trf4 are expected to compete with Trf5 for Air1 binding, leading to increased formation of TRAMP4-1 and correspondingly reduced TRAMP5-1. This will partially mimic effects of *air1*Δ, consistent with ITS1 binding specificity arising from Trf5. Similarly, in absence of Trf4, TRAMP5-1 is expected to partially occupy previous TRAMP4-1 and 4-2 binding sites, reducing recovery at sites specifically occupied by Trf5 in the wild type background.

Deletion of the Arch domain from Mtr4 (Fig. 3J, 3M) also strongly reduced its association upstream of site A3. Similar profiles were observed for total, unfiltered reads (Figs. 3N, S3E and S3J) confirming that the loss of Mtr4 binding in *trf5*Δ was due to a recruitment defect and not to reduced oligo-adenylation. Neither Rrp44 nor Rrp6 showed clear binding across the A2-A3 region. We re-analyzed published data for exonuclease deficient Rrp44 (Rrp44-exo), which is stabilized in binding to many exosome substrates, but this also did not show clearly elevated binding to A2-A3. Northern blot analysis showed strong accumulation of cleaved A2-A3 fragment in *trf5*Δ compared to wild-type, confirming the role of Trf5 in degradation of this RNA (Fig. 3O). No comparable effect was observed for *trf4*Δ.

We conclude that recruitment of Mtr4 to the A3-cleaved pre-rRNA requires both Trf5 and the Arch domain (see Fig. 3P for a cartoon). The finding that the loss of Mtr4 was modest in the absence of Air1, indicated that this is largely redundant with Air2. This was unexpected, since it had long been envisaged that RNA binding specificity would largely be determined by Air1 and Air2, which are Zn-knuckle RNA binding proteins [4,5,26,40] .

In an attempt to identify regions of Trf4 or Trf5 responsible for target specificity, we generated Trf5 constructs with non-conserved domains deleted or exchanged with Trf4.

Unfortunately, neither deletion nor exchange of the N-terminal domain resulted in stable protein accumulation. Stable fusion constructs were expressed in which the C-terminal domain was deleted (Trf5ΔCTD) or replaced by Trf4 C-terminal domain (Trf5-CTD4), or in which a 5 amino acid, AIM-like sequence was replaced by 5 Ala (Trf5-5xA). In CRAC analyses, these constructs all showed inefficient cross-linking, resulting in poor signal to noise ratios and variable results (Fig. S5C-S5G, Fig. 8). However, the Trf5 C-terminal domain appeared to be important for recruitment of Trf5 to pre-rRNA (Fig. S5C), particularly across the A2-A3 region of ITS1 (Fig. S5E, S5F), whereas the AIM-like domain appeared dispensable for recruitment (Fig. S5G).

Trf4 and Trf5 show enrichment on different regions of pre-mRNAs

The distribution of TRAMP components on protein coding genes was initially assessed in metagene analyses of mRNAs aligned by the TSS or pA sites (Figs. 4A-4M; all protein coding genes >500 nt in length were included in metagene plots).

Strong, promoter proximal enrichment of the exosome has previously been reported [41,42], reflecting degradation of ncRNAs generated by early RNAPII termination (shown for the exosome catalytic subunits Rrp6 and Rrp44 in Figs. 4F and 4L). Similar 5′ enrichment was seen for Mtr4 (Fig. 4A), Air2 (Fig. 4C), Trf4 (Fig. 4D) as well as the exosome associated nucleases Rrp6 and Rrp44 (Figs. 4L-N). In marked contrast, pA site-proximal peaks were seen for Trf5, Rrp6 and to a lesser extent Air1, but not for Trf4, Air2 or Rrp44. In addition, a modest pA proximal peak was seen for Mtr4. All TRAMP components showed substantial numbers of hits within pre-mRNA introns (Fig. S4), with TRAMP4-1 apparently most strongly targeting introns. Notably, the Arch domain appears to be dispensable for Mtr4 recruitment to introns.

The 5′ peak for Mtr4 was reduced by either deletion of the Arch (Fig. 4G) or the absence of Trf4 (Fig. 4J), but was not clearly altered by loss of Trf5, Air1 or Air2. Recruitment of Trf4 to the TSS was not clearly affected by Mtr4Δarch (Fig. 4F, Fig. S1E). These results are consistent with Arch-dependent recruitment of Mtr4, with specificity dependent on Trf4. Surprisingly, recruitment of Rrp44 was unaffected in Mtr4Δarch (Fig. 4M-N). The same analysis was performed considering only the mRNAs for which Mtr4 binding was most dependent on the Arch domain; i.e. the 200 mRNAs that showed the greatest reduction in recovery with Mtr4Δarch compared to Mtr4 (Fig. 4O-P). On these mRNA, the TSS proximal peak of Rrp44 was substantially decreased, indicating that the Mtr4 Arch domain plays a role in TSS proximal recruitment of the exosome on specific targets.

Distribution of TRAMP components on individual mRNAs

The metagene plots indicated differential recruitment of Trf4 and Trf5, so we next assessed whether this reflected preferential association with different mRNA subsets. The top 1000 mRNAs in the CRAC datasets for each TRAMP component were combined (2,005 mRNAs in total). All mRNAs were divided into 5 bins of equal length and the distribution of each factor was determined for every bin. These data were used to cluster genes with related patterns of factor distribution (Figs. 5A and S6, Table S5). This analysis identified groups of mRNAs with distinctly different TRAMP occupancy.

The largest group of mRNAs was Cluster 3 (Figs. 5A, S6C and S6I), which showed high TSS-proximal occupancy of Trf4, Air2, and Mtr4, with somewhat lower binding for Air1 and Trf5. This is consistent with the metagene profiles in Fig. 4. In contrast, Cluster 2 showed high poly(A) proximal occupancy of Trf5, Air1 and Mtr4, with lower total binding of Air2 and Trf4 (Figs 5A, S6B, S6H). These features are also seen in metagene plots (Fig. 4). Other clusters showed distinct features: In Cluster 1 (Figs. 5A, S6A, S6G), Trf5 showed substantial 3' occupancy, along with a mild enrichment for Air1 (Fig. S7). Cluster1 mRNAs are very likely targeted by 2 distinct pathways: a minor one in which TRAMP4-1 plus TRAMP4-2 bind proximal to the TSS and the major pathway in which Trf5 targets the 3' end, probably partly with Air1. In consequence, Air1 hits within the mRNA are split between TSS and the pA site. Since the data are normalized across each mRNAs, this results in a reduced 3' peak for Air1 relative to Trf5 (Fig. S7). Rrp6 and Rrp44 are not clearly 3' enriched on these mRNAs, suggesting that their interaction with Trf5 may not be linked to surveillance. In Cluster 4, Trf4, Trf5, Air1, Air2, Rrp6 and Rrp44 all show moderate TSS proximal binding, but Mtr4 is largely absent. Surprisingly, TSS proximal binding was relatively increased for Mtr4 in *trf4*Δ or *trf5*Δ strains and for Mtr4Δarch (Fig. 5A, S6D). However, total association for Mtr4 was reduced in the absence of Trf4 (Fig. S6J).

The identified clusters in the heat map were analyzed for GO term enrichment of the protein products, as an indication of functional classes [43] (Fig. 5B). Each cluster showed highly significant enrichment for mRNAs encoding specific, functional subsets of proteins. This provides strong evidence that the TRAMP binding patterns observed have functional consequences for the regulation of protein production.

Correlations in pre-mRNA binding sites

Trf4 and Trf5 appear to recruit Mtr4 to at least some nuclear pre-mRNA sites, but the origin of their binding specificity was unclear. As an initial attempt to address this, we compared the distribution of TRAMP components to other nuclear RNA binding proteins that have been

similarly mapped using CRAC (Figs. 6 and S2, Tables S4, S6). For this we compared both the total relative recovery over each RNA species (designated as “co-targeting of RNAs” in Figs. 6 and S2) or the extent of binding at closely positioned RNA sites (closer than 50nt) (designated as “co-localization of sites” in Figs. 6A and S2). These analyses were performed either genome wide across all annotated genes (Fig. S2, Supplementary Table S4), or over all annotated RNAPII transcribed genes (Fig. 6A, Supplementary Table S6).

Some expected results were obtained; the nuclear poly(A) binding factors Pab1 and Nab2 were closely correlated and distinct from most other factors. However, Nab2, which has been implicated in RNA surveillance [41], showed closer correlation with other surveillance factors (Fig. 6A). Similarly, the snoRNP proteins Nop1, Nop56 and Nop58 were closely correlated over RNAPII transcripts (Fig. 6A) or all genes (Fig. S2). Ribosome synthesis factors (Utp proteins) were closely correlated on all genes, and showed correlations with surveillance factors, reflecting their interactions on pre-rRNAs.

For TRAMP components, Trf5 was better correlated with Air1 than with Trf4 or Air2, while Trf4 and Air2 were most highly correlated, consistent with the analyses of individual genes. The TRAMP components were also correlated with the exosome, as expected, although it was notable that Rrp6 was more correlated with Trf5, while Rrp44 correlated better with Trf4. Nab3 showed higher correlation with Air2 and Trf4 than Air1 and Trf5, consistent with Trf4 being recruited by Nrd1-Nab3 on some targets, including TSS-proximal ncRNA transcripts.

Hrp1, which was implicated in RNA surveillance [41], also showed correlations with TRAMP factors. A number of other proteins, showed notably close correlations, including the pre-mRNA binding protein Npl3 with Air1. It may be relevant that Air1 was initially identified through an interaction with Npl3 [9]. To follow up this observation, Air1 CRAC peaks across the genome were selected using a peak-calling algorithm and used as a reference point to align reads from pre-mRNA binding proteins and TRAMP components (Fig. 6B-6M, Fig S9). Binding sites were compared across the top 100 bound RNAPII transcripts, which includes the abundant snoRNA species (Fig. 6B-6G), or considering only mRNAs (Fig. 6H-M). Close colocalization was seen for Air1 and Air2 (Fig. 6E and 6K), supporting the conclusion that they are functionally redundant on many substrates. Consistent with the data in (Fig. 6A), the peak of Air1 binding sites closely correlated with a peak of Npl3 binding over the top RNAPII transcripts (Fig. 6B; these include many snoRNAs) or mRNAs (Fig. 6F). Colocalization with Air1 was also seen for the surveillance factor Hrp1 (Figs. 6F and 6L).

Similar analysis where performed on Trf5ΔCTD, Trf5-CTD4 and Trf5-5xA (Fig. S8). As for the pre-rRNA analyses (Fig. S5), these constructs gave poor crosslinking on mRNAs. However, mutation or deletion of the CTD increased Trf5 association with the TSS region (Fig. S8A). Notably, deletion of the CTD apparently lead to more delocalized binding across almost all mRNAs (Fig S8C), presumably reflecting a loss of specificity.

Loss of Trf5 alters Mtr4 association and abundance of targeted mRNAs

Previous functional and structural analyses of TRAMP have largely focused on Trf4 and Air2 [21-25]. However, the heat maps (Fig. 5) indicated that subsets of nuclear mRNAs show preferential enrichment for binding to Trf5 (clusters 1 and 2). We therefore determined whether Trf5 has a functional impact on Mtr4 recruitment and mRNA stability.

mRNAs were stratified by ranking in Trf5 association, based on reads per kilobase per million mapped reads (RPKM) across the genome. Bins of 200 mRNAs were compared between *TRF5* and *trf5*Δ backgrounds for changes in Mtr4 association by CRAC (Fig. 7A) or RNA abundance by RNA sequencing (Fig. 7B, Fig S10). Strikingly, mRNAs that were most strongly bound by Trf5 showed greatest reduction in Mtr4 binding when Trf5 is absent. Note that these are normalized data, so total recovery of Mtr4 across all genes is constant. The effects were quite substantial, with nearly 4-fold reduced Mtr4 association over the genes that were most strongly bound by Trf5. Conversely, mRNAs with the strongest Trf5 binding were modestly increased in the absence of Trf5, again using normalized sequencing data. These data suggest that loss of Trf5 reduces Mtr4 recruitment with a consequent increase in pre-mRNA or mRNA stability. Changes in abundance shown in the RNAseq data were confirmed by qRT-PCR on selected mRNAs (Fig. S10A, Table S7) relative to the ncRNA SCR1. The trend for increased mRNA abundance in *trf5*Δ was conserved between RNAseq and qPCR. However, greater increases were measured in qPCR, probably reflecting differences in normalization, which is complicated in surveillance mutants that potentially effect all RNA species.

A notable finding from the heat maps was the identification of group of mRNAs showing strong 3' association with Trf5 and Air2=1 in the absence of Mtr4 or the exosome (cluster 1 in Figs. 5 and S7) suggesting a distinct function for TRAMP5-1 here. To assess potential roles of Trf5, we assessed the distribution of RNAseq reads across mRNAs in WT and trf5Δ strains (Figs. 7C and 7D). For cluster 1, comparison of the WT and *trf5*Δ strains showed a clear deficit in reads close to the poly(A) site in the absence of Trf5 (Fig. 7C). Strikingly, comparable effects were not seen for any other cluster. This is presented for cluster 2 which also showed a 3' peak of Trf5, but accompanied by Trf4, Mtr4 and Rrp6, and cluster 3, which

showed only a 5' peak of Trf5 and other TRAMP components (Fig. 7C). When mRNAs were ranked by Trf5 binding in bin 5 (the 3' end), the same trend was seen (Fig. 7D). However, even for the most highly bound group, the 3' depletion in mRNAs reads was less marked than for cluster 1.

We conclude that mRNAs bound by Trf5 are selectively affected by its absence. Notably, the clustering in Fig. 5 was entirely orthogonal to the mRNA sequence data in Fig.7, indicating that the specific 3' binding by Trf5 identified in cluster 1 is robustly correlated with stabilization of this region of the transcripts.

DISCUSSION

To better understand the targeting of RNA degradation and surveillance targets *in vivo*, we characterized the protein composition of the TRAMP complexes and identified specific binding sites for the different TRAMP components in wild-type and mutant cells. These analyses identified three complexes containing Mtr4 together with Trf4 and Air1 (TRAMP4-1), Trf4 and Air2 (TRAMP4-2) or Trf5 and Air1 (TRAMP5-1). Substantial differences in RNA binding were observed, indicating that TRAMP4-1, TRAMP4-2 and TRAMP5-1 each exhibit substrate specificities.

We note that, based on 14 different publications, SDG (www.yeastgenome.org) lists median abundances (copies per cell) for TRAMP components as: Air1 (1851 +/- 466); Air2: (1750 +/- 420); Trf4 (2659 +/- 411); Trf5 (1329 +/- 608). If all components are in TRAMP complexes, this suggests very approximate abundances of: TRAMP4-1 (600); TRAMP4-2 (1700); TRAMP5-1 (1300). These values are consistent with the conclusion that Trf4 interacts with Air1 and Air2, whereas Trf5 binds only Air1.

TRAMP5-1 preferentially targeted the ITS1 spacer region of 35S pre-rRNA, a characterized exosome substrate for which no AIM-domain ribosome synthesis factor has been identified. Mtr4 binding to ITS1 in the pre-rRNA was strongly reduced by loss of Trf5, whereas loss of Trf4, Air1 or Air2 had only modest effects. In contrast, TRAMP4-2 was more strongly associated with RNAPII transcripts, particularly mRNA 5′ ends, close to the transcription start site and with the CUT, SUT and XUT ncRNAs. This implicated TRAMP4-2 in degradation of ncRNAs, including promoter-proximal ncRNAs generated by early termination of transcription, a model supported by colocalization with Rrp44 and Rrp6. Mtr4 exhibited lower binding to TSS proximal regions when Trf4 was absent, with comparable reduction on deletion of the Arch domain, but was little altered by loss of Air1 or Air2. These results

strongly indicate that substrate specificity can be provided by Trf4 and Trf5, while Air1 and Air2 appear to be largely interchangeable. This was unexpected, since it had previously been anticipated that RNA binding specificity would largely be determined by Air1 and Air2, which are Zn-knuckle RNA BPs [26,40], or by specific interactions of AIM-containing proteins with the Arch domain of Mtr4 [15,16].

When bound to the exosome, Mtr4 is positioned at the entrance to the central lumen, such that RNAs can pass through the helicase domain directly into the exosome [44,45]. The structure of the Trf4-Air2-Mtr4 TRAMP complex shows the Arch located opposite Trf4-Air2 [22]. N-terminal, low complexity regions of Air2 and Trf4 are jointly responsible for binding to the DExH core of Mtr4 *in vitro*. The residues in Trf4 and Air2 that bind the Mtr4 helicase core, are conserved in Trf5 and Air1, respectively. The N-terminus of Trf4 (aas 1-110) is absent from the TRAMP structure, but could conceivably extend to the Arch domain of Mtr4. However, a more likely interaction is between Air2 and the Arch. Indeed, in the crystal structure, the N-terminus from the adjacent Air2 molecule in the crystal lattice was shown to contact the Mtr4 Arch at the same sites as the AIM of Nop53 [15,22], through residues that are conserved in Air1. Association of Air2 and Air1 with the Arch domain, might require conformational changes in which the Arch moves toward the helicase core, switching Mtr4 (and the TRAMP) to a closed conformation. We therefore postulate that recruitment of Mtr4 to most nuclear RNA surveillance targets is based on the specificity of Trf4 or Trf5 interactions with the target, followed by (largely redundant) binding of Air1 or Air2 to the Arch domain.

Comparison of TRAMP binding with other factors expected to interact with nuclear transcripts showed a range of concordances. We speculate that combinations of proteins binding to nascent transcripts act together to promote or disfavor binding by the surveillance machinery and rapid nuclear RNA degradation. Particularly notable was the correlation between the hnRNP-like, pre-mRNA binding protein Npl3 and Air1, which was initially isolated via interaction with Npl3 [9]. Air1-Npl3 binding is bridged by the arginine-methyl transferase Hmt1 and blocks arginine methylation on Npl3, potentially changing its functional properties.

Clustering of the most recovered ~2,000 mRNAs, based on TRAMP factor occupancy, identified six clusters. Five of these showed statistically significant enrichment for functional classes of protein products, while the remaining class was enriched for types of transcript; introns with encoded features, notably snoRNAs, and unconfirmed protein products. Such

functional enrichment strongly supports the significance of TRAMP factor binding in regulating gene expression.

A notable finding was the apparently specific binding of Trf5 to a cluster of mRNAs. To better understand the significance of this association, we assessed the effects of deletion of *TRF5* on Mtr4 binding by CRAC and RNA abundance by RNAseq. When mRNAs were stratified by Trf5 association, mRNA species most highly bound by Trf5 in the WT, showed decreased binding by Mtr4 and increased total abundance in the *trf5*Δ strain. This is consistent with Trf5 promoting Mtr4 recruitment and degradation for many pre-mRNAs.

We also noted that clusters of mRNAs showed high p(A) proximal association of Trf5 (Clusters 1 and 2 in Fig. 5A). In particular, the 3' association of Trf5 and Air1 with mRNAs in cluster 1, was not accompanied by clear association with Mtr4 or the exosome, suggesting a possible function distinct from RNA surveillance and degradation. Comparing RNAseq data across mRNAs revealed a deficit in 3' reads, specifically for cluster 1. No such deficit was observed for cluster 2, in which 3' accumulation of Trf5 is accompanied by binding of Mtr4 and other TRAMP components. It remains unclear whether this apparent truncation reflects altered pre-mRNA synthesis or stability, but Trf5 clearly has significant effects on the 3' ends of highly bound pre-mRNAs.

We conclude that the functions of the different yeast TRAMP complexes have significantly diverged, particularly for the nucleotide transferases Trf4 and Trf5.

METHODS

Strains

All yeast strains are derived from (BY4741, *MATa; his3*Δ*1; leu2*Δ*0; met15*Δ*0; ura3*Δ*0*). Standard techniques were used to integrate C-terminal affinity tags and integration of galactose-driven genes. Strains used are listed in Supplementary Table S1.

For Mtr4 depletion, strains were grown in 2% Galactose minimal media, then shifted to 2% glucose media for 16 to 20 h until exponential phase (OD ~0.5) and used for CRAC or RNA extraction. Depletion was check by western blotting.

Western blotting

Yeast were grown to exponential phase (OD 0.5). Proteins extracted as described [46]. Yeast.] from 4OD of cells were loaded on NuPaGE Novex 4-12% Gel, liquid transferred and incubated

with anti-TAP (Thermo) and anti-PGK1 antibodies, followed by IRDye secondary antibodies (Licor) incubation. Membrane was visualized on Odyssey CLx scanner.

Northern blotting

RNA were extracted as described [47]. Oligoadenylated RNAs were purified from 75 μg of total RNAs using polyA+ kit (Ambion). All output along with 5μg total RNA were separated on a 10% acrylamide urea gel, stained with SybrSafe (Invitrogen) and liquid-transferred to a HybondN+ membrane. Hybridization with a radiolabeled oligonucleotide (GCGTTGTTCATCGATGC) was performed with ultraHyb (Ambion) and signal detected using a Fuji FLA-5100 PhosphoImager.

Protein Alignment

Alignment were generated with Clustal Omega and visualized using MView 1.63

Mass spectrometry analyses

Protein pull-downs were extracted from strains expressing the bait protein tagged with a C-terminal HisX6-Tev cleavage site-Protein A (HTP) tag. A non-tagged strain was used as a negative control. Protein purification has been made in 1 M NaCl conditions. For Trf4-HTP and Trf5-HTP, additional purification with lower stringency (150 mM NaCl) were performed in presence or absence of RNase A. Extracts were separated by SDS–polyacrylamide gel electrophoresis (PAGE), trypsin-digested, as previously described [48]. Following digestion, samples were diluted with equal volume of 0.1% TFA and spun onto StageTips as described[49]. Peptides were eluted and analyzed by LC-MS/MS on a Q Exactive Mass Spectrometer (Thermo Fisher Scientific) coupled on-line, to an Ultimate 3000 RSLCnano Systems (Dionex, Thermo Fisher Scientific). FTMS spectra were recorded at 70,000 resolution and the top 10 most abundant peaks with charge ≥ 2 and isolation window of 2.0 Thomson were selected and fragmented by higher-energy collisional dissociation [50] with normalized collision energy of 27. The maximum ion injection time for the MS and MS2 scans was set to 20 and 60 ms respectively and the AGC target was set to 1 E6 for the MS scan and to 5 E4 for the MS2 scan.

The MaxQuant software platform [51] version 1.6.1. 0 was used to process raw files and search was conducted against the *Saccharomyces cerevisiae* complete/reference proteome database (Uniprot, released in September, 2017), using the Andromeda search engine [52]. iBAQ (Intensity-based absolute quantification) values calculated by MaxQuant are the (raw)

intensities divided by the number of theoretical peptides (Figure 1C, Supplementary Table S2). In that way, iBAQ values are proportional to the molar quantities of the proteins [28]. Relative molar abundances, for each protein, was determined as its relative iBAQ (riBAQ), a normalized measure of molar abundance. We divided each yeast protein's iBAQ value by the sum of all non-contaminant iBAQ values [53]. Enrichment was calculated as the riBaq ratio between protein pull-down and negative control (Fig. 1C, Supplementary Table S3). All proteins showing a ratio inferior at 2 was considered as not enriched.

CRAC

CRAC was performed as described [54] on yeast strains expressing the protein of interest carrying a C-terminal HTP tag, grown in SD-medium to log phase and UV-crosslinked (254 nm, 100 sec) to covalently bind RNA to protein. RNA-protein complexes have been purified, RNAs are partially digested to leave only the “footprint” of the protein or protein complex. Mircat linkers and barcoded linkers (containing three random nucleotides) have been ligated on both 3′ and 5′ end respectively. Proteins were then digested by proteinase K, RNAs were reverse transcribed and PCR-amplified. cDNAs libraries were size fractionated on agarose gels then subjected to next-generation sequencing using Illumina HiSeq (Edinburgh Genomics) or Illumina Miniseq (our laboratory). Illumina sequence data from this publication have been submitted to the GEO database (http://www.ncbi.nlm.nih.gov/geo/) and assigned the identifier GSE135526.

RNAseq

Yeast were grown at 30°C to OD 0.5 in minimal media and RNAs were extracted using a standard acidic hot phenol method. RNA-seq libraries were prepared using a SENSE mRNA-seq Library Prep Kit V2 for Illumina (Lexogen), as recommended by the manufacturer. 2 µg of RNA were denatured and used as input for poly(A) tailed RNA purification through hybridization on oligodT magnetic beads. Purified Poly(A) RNA were subjected to reverse transcription and ligation generating short cDNA fragments with linker sequences at either end. The library was converted to dsDNA, purified and PCR amplified (13 cycles). Samples were checked on bioanalyser and then sequenced using standard Illumina protocol on NextSeq (75 cycles, High output).

Real-time PCR amplification

Yeast were grown at 30°C to OD 0.5 in minimal media and RNAs were extracted using a standard acidic hot phenol method. 5 µg RNA were submitted to TURBO DNase treatment (Thermo), checked for quality and reverse transcribed with Random decamers (Ambion) and

Superscript III Reverse Transcriptase (Invitrogen) following manufacturer's protocols. After RNAse H (thermo Scientific) treatment, cDNAs were diluted 20-fold. Real-time PCR amplifications were performed in a LightCycler480 system (Roche Diagnostics) in 384-well plates on 5 µL scale reactions using 2 µL diluted cDNAs and Brilliant III Ultra-fast SYBR Green qPCR Master Mix (Agilent) with primers listed in Table S7 at a final concentration of 0.4 µM. Technical triplicates for each independent biological replicates (2 for wild-type and 3 for *trf5*Δ) were carried out. For each sample, to ensure level of residual gDNA in the RNA prep were negligible, qPCR was performed without reverse transcription on DNase treated RNA ("no RT"). Primers linearity and efficiencies were tested by performing a standard amplification on serial dilution of wild-type genomic DNA in triplicates (100 ng/µL, 10 ng/µL, 1 ng/µL, 0.1 ng/µL, 0.01 ng/µL). Cycle threshold (Ct) values were averaged between triplicates for each RNA sample. Gene expression fold change were determined as reported [55]. For qPCR raw data, primer sequences and efficiencies, fold change calculation on wild-type and *trf5*Δ strains, see Table S7.

Bioinformatics analysis

CRAC data: pre-processing and alignment

CRAC sequencing data were quality filtered and adapter were trimmed using Flexbar 3.4.0 [56] with parameters -x 1 –ao 4 -g and only reads containing the 3′ adapter were kept. Then, the sequences were collapsed: reads having identical ends and identical random nucleotides in the 5′ barcode were counted as one, allowing removal of PCR duplicates. CRAC pre-processed reads containing non-encoded oligo-A tailed were identified using a pipeline developed by Grzegorz Kudla [10,41]. Reads were then aligned to the *Saccharomyces cerevisiae* genome (SGD v64) using Novoalign (Novocraft) with genome annotation from Ensembl (EF4.74) [57], supplemented with non-coding sequences as described in [41], with parameters –r Random.

RNAseq data: pre-processing and alignment

Low quality reads and 5′ extremities of reads were filtered out. Low complexity sequences (reads having more than 70% of their content corresponding to a single nucleotide stretch and that would be potentially misaligned) were filtered out before alignment. Then, RNAseq reads were aligned with STAR [58] using genome database from Ensembl (EF4.74).

Class distribution

Downstream analyses were performed using pyCRAC software [59]. pyReadCounters (pyCRAC) was used to count overlaps with genes and reads per millions (RPM) or RPM per kilobase (RPKM).

Correlation of binding

For correlation of binding analyses, low complexity sequences were filtered out before alignment. Overlaps either with transcripts (all or RNAPII transcripts when stated) or with a genome reference file divided into 50nt windows, independently of transcripts, were calculated, averaged between two biological repeats and used to calculate Pearson correlations between samples.

Plots, Binding profiles

Plot showing binding along single genes were generated using pyPileup (pyCRAC). Metagenomic plots were performed using homemade script using pyPileup on each individual transcript to count hits at each position. We obtain a table in which each row represents a transcript and each column represents the absolute position from 5′ end or 3′ end. The plot is summing up binding at each position allowing the display of a binding profile aligned either at 5′ end or 3′ end (Fig. 4). To calculate the enrichment or loss of mRNA signal between wild-type strain and *trf5*Δ strain, normalized coverage (reads per million) was calculated at each position along the genome for all datasets and averaged between replicate RNAseq datasets (4 and 3 independent replicates for wild-type and *trf5*Δ respectively). Log2 enrichment was calculated after addition of 5 pseudocounts to both the numerator and denominator as described [60].

Clustering analysis of mRNAs

We selected and combined the top 1,000 protein-coding genes bound by each TRAMP component (2,005 mRNAs were included). pyBinCollector (pyCRAC) was used to calculate binding distribution along transcripts, each one being divided in 5 bins from TSS to pA site. Binding across each bin was calculated as a fraction of total binding across individual gene (set to 1). For each transcript, we averaged numbers of reads between two biological replicates to reduce the influence of experimental variation upon clustering analysis. We normalized the data for each gene. Air1, Air2, Trf4, Trf5, Mtr4 coverage along transcripts were then clustered using Cluster3.0 (k-medians, k = 6, Euclidian distance) and data displayed as heat maps. For data not included in the clustering analysis, pyBincollector output was sorted according to the clustered list and shown as heat map.

Peak calling

Low-complexity sequences trimmed dataset were used. Number of hits for each position of the genome were calculated excluding all non-RNA polymerase II transcripts and normalized to reads per millions of RNA polymerase II transcripts. Clusters consisting of 15 continuous

position with more than 50 hits across were detected. The highest position in each cluster was selected as "peak'. Air1 peaks were used as a reference file to which other protein peaks were aligned (Figs. 6B-6M and S9).

Availability of data and material

All sequence data from experiments made for this study and Rat1 datasets from Granneman et al, EMBO, 2011 are available from GEO under accession number GSE135526. Published data are available under GEO accession number GSE77683 (Mtr4), GSE69696 (Rrp44, Rrp6, Air2), GSE79950 (UTP proteins), GSE70191 (Nab3, Npl3, Sto1), GSE114680 (Nop1, Nop56, Nop58), GSE46742 (Trf4, Cbc1, Gbp2, Tho2, Hrp1, Nab2, Pab1, Mex67, Hek2, Xrn1, Ski2).

COMPETING INTERESTS

The authors declare no competing interests.

FUNDING

The Wellcome Trust generously funded this work through a Principle Research Fellowship to D.T. (077248), a Senior Research Fellowship to J.R. (103139) and an instrument grant (108504). CDF was supported by an FEBS Long Term Fellowship. Work in the Wellcome Centre for Cell Biology is supported by a Centre Core grant (203149).

ACKNOWLEDGEMENTS

We thank Stefan Bresson, Marie Winz and Aleksandra Helwak for critical reading of the MS.

REFERENCES

1 Bresson, S. & Tollervey, D. Surveillance-ready transcription: nuclear RNA decay as a default fate. *Open Biology* **8** (2018).

2 Schmid, M. & Jensen, T. H. Controlling nuclear RNA levels. *Nature Reviews Genetics* **19**, 518-529 (2018).

3 Carneiro, T. *et al*. Depletion of the yeast nuclear exosome subunit Rrp6 results in accumulation of polyadenylated RNAs in a discrete domain within the nucleolus. *Mol. Cell. Biol.* **27**, 4157-4165, doi:10.1128/MCB.00120-07 (2007).

4 LaCava, J. *et al*. RNA degradation by the exosome is promoted by a nuclear polyadenylation complex. *Cell* **21**, 713-724 (2005).

5 Vanacova, S. *et al*. A New Yeast Poly(A) Polymerase Complex Involved in RNA Quality Control. *PLoS Biol.* **3**, e189 (2005).

6 Wyers, F. *et al*. Cryptic Pol II transcripts are degraded by a nuclear quality control pathway involving a new poly(A) polymerase. *Cell* **121**, 725-737 (2005).

7 de la Cruz, J., Kressler, D., Tollervey, D. & Linder, P. Dob1p (Mtr4p) is a putative ATP-dependent RNA helicase required for the 3' end formation of 5.8S rRNA in Saccharomyces cerevisiae. *EMBO J* **17**, 1128-1140 (1998).

8 Liang, S., Hitomi, M., Hu, Y. H., Liu, Y. & Tartakoff, A. M. A DEAD-box-family protein is required for nucleocytoplasmic transport of yeast mRNA. *Mol. Cell. Biol.* **16**, 5139-5146 (1996).

9 Inoue, K., Mizuno, T., Wada, K. & Hagiwara, M. Novel RING finger proteins, Air1p and Air2p, interact with Hmt1p and inhibit the arginine methylation of Npl3p. *J. Biol. Chem.* **275**, 32793-32799 (2000).

10 Wlotzka, W., Kudla, G., Granneman, S. & Tollervey, D. The nuclear RNA polymerase II surveillance system targets polymerase III transcripts. *EMBO J* **30**, 1790-1803 (2011).

11 Schmidt, K. & Butler, J. S. Nuclear RNA Surveillance: Role of TRAMP in Controlling Exosome Specificity. *WIRES RNA* **4**, 217-231 (2013).

12 Houseley, J. & Tollervey, D. The many pathways of RNA degradation. *Cell* **136**, 763-776 (2009).

13 Jackson, R. N. *et al*. The crystal structure of Mtr4 reveals a novel arch domain required for rRNA processing. *Embo J* **29**, 2205-2216, doi:10.1038/emboj.2010.107 (2010).

14 Weir, J. R., Bonneau, F., Hentschel, J. & Conti, E. Structural analysis reveals the characteristic features of Mtr4, a DExH helicase involved in nuclear RNA processing and surveillance. *Proc. Natl. Acad. Sci. U.S.A.* **107**, 12139-12144, doi:10.1073/pnas.1004953107 (2010).

15 Falk, S. *et al*. Structural insights into the interaction of the nuclear exosome helicase Mtr4 with the pre-ribosomal protein Nop53. *RNA* (2017).

16 Thoms, M. *et al*. The Exosome Is Recruited to RNA Substrates through Specific Adaptor Proteins. *Cell* **162**, 1029-1038 (2015).

17 Silla, T., Karadoulama, E., Mąkosa, D., Lubas, M. & Jensen, T. H. The RNA Exosome Adaptor ZFC3H1 Functionally Competes with Nuclear Export Activity to Retain Target Transcripts. *Cell Rep.* **23**, 2199-2210 (2018).

18 Meola, N. *et al*. Identification of a Nuclear Exosome Decay Pathway for Processed Transcripts. *Mol. Cell* **64**, 520-533 (2016).

19 Lubas, M. *et al*. Interaction Profiling Identifies the Human Nuclear Exosome Targeting Complex. *Mol. Cell* **43**, 624-637 (2011).

20 San Paolo, S. *et al*. Distinct roles of non-canonical poly(A) polymerases in RNA metabolism. *PLoS Genet* **5**, e1000555 (2009).

21 Jia, H., Wang, X., Anderson, J. T. & Jankowsky, E. RNA unwinding by the Trf4/Air2/Mtr4 polyadenylation (TRAMP) complex. *Proc. Natl. Acad. Sci. U.S.A.* **109**, 7292-7297 (2012).

22 Falk, S. *et al*. The Molecular Architecture of the TRAMP Complex Reveals the Organization and Interplay of Its Two Catalytic Activities. *Mol. Cell* **55**, 856-867 (2014).

23 Holub, P. *et al*. Air2p is critical for the assembly and RNA-binding of the TRAMP complex and the KOW domain of Mtr4p is crucial for exosome activation. *Nucleic Acids Research* **40**, 5679-5693 (2012).

24 Hamill, S., Wolin, S. L. & Reinisch, K. M. Structure and function of the polymerase core of TRAMP, a RNA surveillance complex. *Proc. Natl. Acad. Sci. U.S.A.* **107**, 15045-15050 (2010).

25 Tudek, A. *et al*. Molecular Basis for Coordinating Transcription Termination with Noncoding RNA Degradation. *Mol. Cell* **55**, 467-481 (2014).

26 Fasken, M. B. *et al*. Air1 zinc knuckles 4 and 5 and a conserved IWRXY motif are critical for the function and integrity of the Trf4/5-Air1/2-Mtr4 polyadenylation (TRAMP) RNA quality control complex. *J. Biol. Chem.* **286**, 37429-37445 (2011).

27 Houseley, J. & Tollervey, D. Yeast Trf5p is a nuclear poly(A) polymerase. *EMBO Rep.* **7**, 205-211 (2006).

28 Schwanhäusser, B. *et al*. Global quantification of mammalian gene expression control. *Nature* **473**, 337 (2011).

29 Chatterjee, B. & Thakur, S. S. Single-Run Mass Spectrometry Analysis Provides Deep Insight into E. coli Proteome. *J. American Soc. Mass Spectrometry* **29**, 2394-2401 (2018).

30 Petfalski, E., Dandekar, T., Henry, Y. & Tollervey, D. Processing of the precursors to small nucleolar RNAs and rRNAs requires common components. *Mol. Cell. Biol.* **18**, 1181-1189 (1998).

31 Haimovich, G. *et al*. Gene Expression Is Circular: Factors for mRNA Degradation Also Foster mRNA Synthesis. *Cell* **153**, 1000-1011 (2013).

32 Blasco-Moreno, B. *et al*. The exonuclease Xrn1 activates transcription and translation of mRNAs encoding membrane proteins. *Nature Comm.* **10**, 1298-1298 (2019).

33 Stevens, A., Hsu, C. L., Isham, K. R. & Larimer, F. W. Fragments of the internal transcribed spacer 1 of pre-rRNA accumulate in *Saccharomyces cerevisiae* lacking 5'-3' exoribonuclease 1. *J. Bacteriol.* **173**, 7024-7028 (1991).

34 Houseley, J., Kotovic, K., El Hage, A. & Tollervey, D. Trf4 targets ncRNAs from telomeric and rDNA spacer regions and functions in rDNA copy number control. *EMBO J* **26**, 4996-5006 (2007).

35 Grzechnik, P. & Kufel, J. Polyadenylation linked to transcription termination directs the processing of snoRNA precursors in yeast. *Mol Cell* **32**, 247-258 (2008).

36 Egecioglu, D. E., Henras, A. K. & Chanfreau, G. F. Contributions of Trf4p- and Trf5p-dependent polyadenylation to the processing and degradative functions of the yeast nuclear exosome. *RNA* **12**, 26-32 (2006).

37 Mitchell, P., Petfalski, E., Shevchenko, A., Mann, M. & Tollervey, D. The exosome; a conserved eukaryotic RNA processing complex containing multiple 3'->5' exoribonuclease activities. *Cell* **91**, 457-466 (1997).

38 Klauer, A. A. & van Hoof, A. Genetic interactions suggest multiple distinct roles of the arch and core helicase domains of Mtr4 in Rrp6 and exosome function. *Nucleic Acids Res.* **41**, 533-541 (2013).

39 Losh, J. S. *et al*. Interaction between the RNA-dependent ATPase and poly(A) polymerase subunits of the TRAMP complex is mediated by short peptides and important for snoRNA processing. *Nucleic Acids Research* **43**, 1848-1858 (2015).

40 Schmidt, K., Xu, Z., Mathews, D. H. & Butler, J. S. Air proteins control differential TRAMP substrate specificity for nuclear RNA surveillance. *RNA* **18**, 1934-1945 (2012).

41 Tuck, A. C. & Tollervey, D. A transcriptome-wide atlas of RNP composition reveals diverse classes of mRNAs and lncRNAs. *Cell* **154**, 996-1009 (2013).

42 Delan-Forino, C., Schneider, C. & Tollervey, D. Transcriptome-wide analysis of alternative routes for RNA substrates into the exosome complex. *PLOS Genetics* **13**, e1006699 (2017).

43 Eden, E., Navon, R., Steinfeld, I., Lipson, D. & Yakhini, Z. GOrilla: a tool for discovery and visualization of enriched GO terms in ranked gene lists. *BMC Bioinformatics* **10**, 48-48, doi:10.1186/1471-2105-10-48 (2009).

44 Weick, E.-M. *et al*. Helicase-Dependent RNA Decay Illuminated by a Cryo-EM Structure of a Human Nuclear RNA Exosome-MTR4 Complex. *Cell* **173**, 1663-1677.e1621 (2018).

45 Schuller, J. M., Falk, S., Fromm, L., Hurt, E. & Conti, E. Structure of the nuclear exosome captured on a maturing preribosome. *Science* **360**, 219, doi:10.1126/science.aar5428 (2018).

46 Kushnirov, V. V. Rapid and reliable protein extraction from yeast. *Yeast* **16**, 857-860 (2000).

47 Tollervey, D. & Mattaj, I. W. Fungal small nuclear ribonucleoproteins share properties with plant and vertebrate U-snRNPs. *Embo J* **6**, 469-476 (1987).

48 Shevchenko, A., Wilm, M., Vorm, O. & Mann, M. Mass spectrometric sequencing of proteins from silver-stained polyacrylamide gels. *Anal. Chem.* **68**, 850-858 (1996).

49 Rappsilber, J., Ishihama, Y. & Mann, M. Stop and go extraction tips for matrix-assisted laser desorption/ionization, nanoelectrospray, and LC/MS sample pretreatment in proteomics. *Anal Chem* **75**, 663-670 (2003).

50 Olsen, J. V. *et al*. Higher-energy C-trap dissociation for peptide modification analysis. *Nature Methods* **4**, 709 (2007).

51 Cox, J. & Mann, M. MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat. Biotechnol.* **26**, 1367-1372 (2008).

52 Cox, J. *et al*. Andromeda: A Peptide Search Engine Integrated into the MaxQuant Environment. *J. Proteome Res.* **10**, 1794-1805 (2011).

53 Shin, J.-B. *et al*. Molecular architecture of the chick vestibular hair bundle. *Nat Neurosci* **16**, 365-374 (2013).

54 Granneman, S., Kudla, G., Petfalski, E. & Tollervey, D. 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. U.S.A.* **106**, 9613-9818 (2009).

55 Pfaffl, M. W. A new mathematical model for relative quantification in real-time RT-PCR. *Nucleic Acids Res.* **29**, e45-e45 (2001).

56 Dodt, M., Roehr, J. T., Ahmed, R. & Dieterich, C. FLEXBAR—Flexible Barcode and Adapter Processing for Next-Generation Sequencing Platforms. *Biology* **1**, 895-905 (2012).

57 Flicek, P. *et al*. Ensembl 2014. *Nucleic Acids Res.* **42**, D749-D755 (2014).

58 Dobin, A. *et al*. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15-21 (2013).

59 Webb, S., Hector, R. D., Kudla, G. & Granneman, S. PAR-CLIP data indicate that Nrd1-Nab3-dependent transcription termination regulates expression of hundreds of protein coding genes in yeast. *Genome Biol*. **15**, R8 (2014).

60 Milligan, L. *et al*. Strand-specific, high-resolution mapping of modified RNA polymerase II. *Mol*. *Sys*. *Biol*. **12**, 874 (2016).

FIGURE LEGENDS:

Figure 1: Protein interactions involving TRAMP components

A-B. Alignment of paralogs Air1 and Air2 (A), Trf4 and Trf5 (B). Domains are highlighted with transparent squares (Zinc knuckle domain in green, Trf4 and Trf5 central domain in blue, poly(A) polymerase catalytic domain in red). Residues involved in TRAMP internal interactions are designated by colored circles and squares.

C: Mass spectrometry (LC-MS/MS) analysis of Air1, Air2, Trf4, and Trf5 pull-downs under stringent conditions (1M NaCl). iBAQ (Intensity based absolute quantification) values were calculated and TRAMP component recovery was represented relatively to bait protein iBAQ set to 1. See Supplementary Table 2 for values.

D. Schematic representation of the TRAMP nuclear cofactor and exosome complexes. Proteins subjected to CRAC in this study are highlighted in color: Mtr4 helicase in dark blue, Mtr4Δarch mutant in purple. Air1 and Air2 zinc knuckle domains in green and orange, respectively. Trf4 and Trf5 poly(A) polymerase domains in dark red and yellow, respectively. Rrp44 exo- and endonuclease and Rrp6 exonuclease active sites are represented in dark blue and red, respectively. The exosome barrel is colored in grey.

E: riBAQ (relative molar abundances for each protein), was determined by dividing the iBAQ value for each protein by the sum of all non-contaminant iBAQ values. Enrichment was calculated as the riBAQ ratio between TRAMP protein pull-down and the mock sample. All proteins showing a ratio less than 2 were considered to be not significantly enriched. See Supplementary Table 3 for values.

Figure 2: RNA interactions by TRAMP components

A-B: Distribution of all (A) and non-encoded A-tailed (B) reads mapped to different RNA substrate classes recovered in CRAC datasets. Two biological replicates were averaged for each protein.

C: Correlation matrix displayed as a heat map showing correlation between binding sites recovered with different factors. NOTE: Above the diagonal, the matrix shows the extent to which the factors target the same RNA species as “Co-targeting of RNAs”. Below the diagonal, the matrix shows the extent to which the factors target the closely positioned RNA sites (≥50 nt) as “Colocalization of sites”. See Supplementary Table 4 for values for individual CRAC dataset.

Figure 3: TRAMP interaction sites on the pre-rRNA ITS1 region

A-K. Distribution of non-encoded oligo(A) tailed reads, normalized to millions of total mapped reads, across the ITS1 region of the pre-rRNA. The graph shows the distribution of reads recovered with the indicated TRAMP components (A-E), Mtr4 in strains lacking the indicated

TRAMP components (F-I), Mtr4 lacking the Arch domain (J), Rrp44 (K). Scale is linear. A diagram of the pre-rRNA region is also shown. Two independent replicates are shown in panels as light and dark colored lines. (M-N) Hits per millions oligo(A) tailed reads (M) and total reads (N) encompassing the A2-A3 region of ITS1 extended by 10nt from each side were summed and are exhibited as a bar diagram. Individual replicates were averaged, with standard deviation shown as error bars. (O) Northern blot analysis with probe for the A2-A3 fragment (probe for 5S rRNA was used as loading control). (P) Model for surveillance within ITS1. Trf5 and Air1 are recruited cotranscriptionally to the A2-A3 region. Following A2 cleavage, ribosome synthesis factors are released, freeing site A3. Mtr4 is recruited, probably *via* protein-protein interactions involving Trf5 and an Mtr4 Arch interaction. Following A3 cleavage, Trf5 oligo-adenylates the free 3' end prior to degradation by the exosome.

Figure 4: Metagene analyses of TRAMP components across protein coding genes

Binding by RNA surveillance and degradation factors is strongly enriched close to the TSS on protein coding genes.

A-N: Distribution of individual components indicated, across all mRNAs longer than 500 nt. Each panel shows the hit density, normalized to millions reads mapped to mRNAs. The two lines in each panel represent results from independent CRAC experiments. Reads were aligned with transcription start sites (TSS) and polyadenylation sites (pA).

O-P: Distribution of Rrp44 in strains expressing Mtr4 or Mtr4Δarch across the 200 mRNAs showing the greatest dependence on the Arch for Mtr4 recruitment. Each panel shows the hit density, normalized to millions of reads mapped to mRNA, for all genes longer than 500 nt. The two lines in each panel represent results from independent CRAC experiments. Reads were aligned with transcription start sites (TSS) and polyadenylation sites (pA).

Q: Correlation of relative recovery of mRNAs by Mtr4 compared to Mtr4Δarch and Mtr4 in the *trf4*Δ strain. The top 1,500 mRNA bound by Mtr4 were used in this graph. Ratios were calculated using RPM.

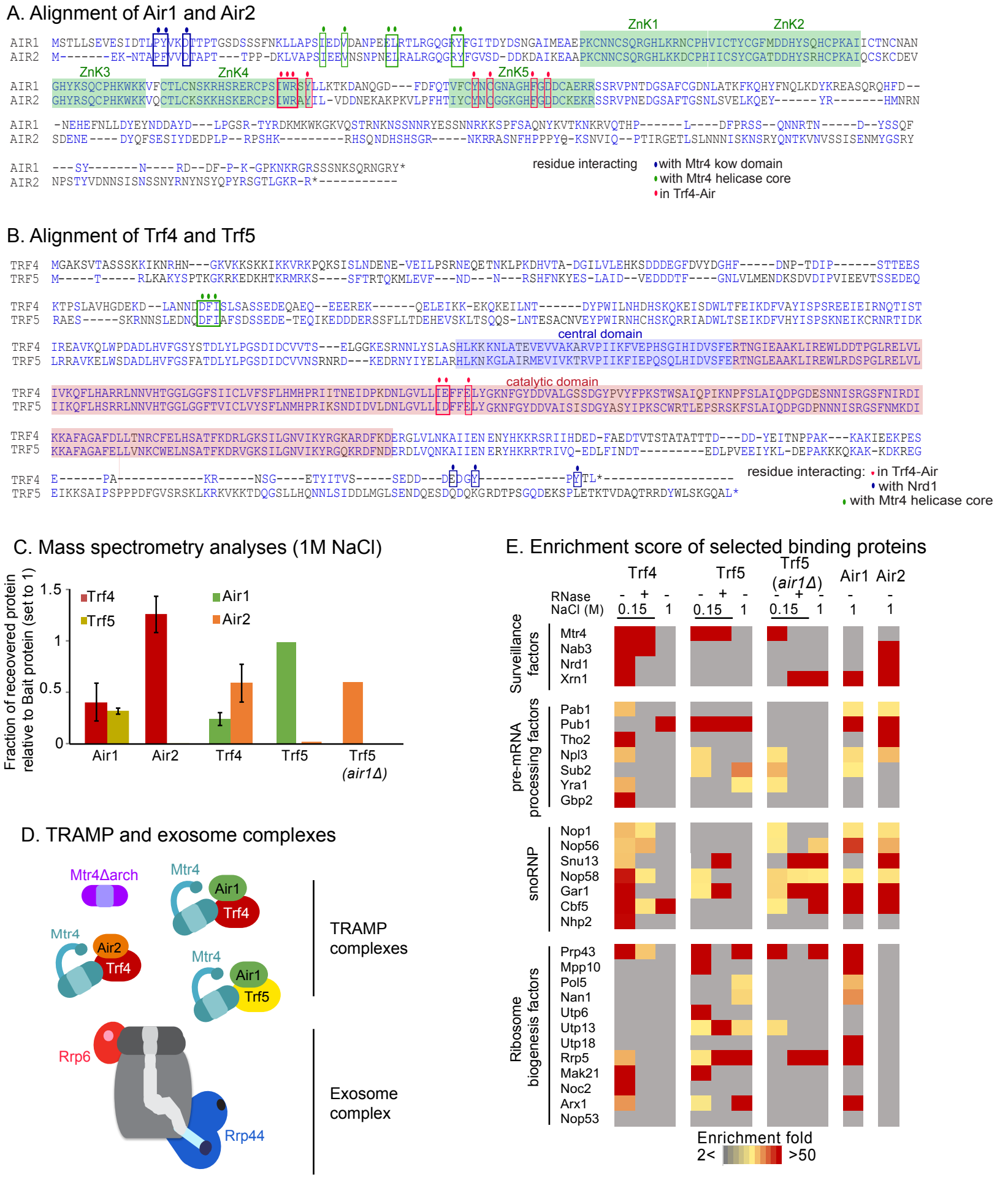
R: Model representing proposed recruitment pathways. Close to the TSS, Nrd1-Nab3-Sen1 mediated early transcription generate TSS proximal ncRNAs and recruits TRAMP4-2 through Trf4 and interactions involving the Arch domain. Following Trf4-mediated oligoadenylation, RNAs are targeted to both Rrp44 and Rrp6 nucleases activity of the exosome. A distinct surveillance mechanism occurs at some pA sites involving specifically TRAMP5-1 and Rrp6.

Figure 5: HEAT maps for distribution of TRAMP components on each mRNA

A. HEAT map showing binding across individual mRNAs of different factors. The top 1,000 protein-coding genes recovered in CRAC with each TRAMP component were selected and combined (2,005 mRNAs in total). Each transcript was divided into 5 bins of equal length from TSS to pA site. Binding across each bin was calculated as a fraction of total binding across individual gene (set to 1). Numbers of reads were averaged between two biological replicates. Only the coverage of Air1, Air2, Trf4, Trf5 and Mtr4 along each transcript was used for Euclidian distance-based clustering. Other data sets shown were not included in the clustering analysis but are sorted according to the clustered list. The data are displayed as heat maps. The RPM for each transcript recovered by each protein was calculated and represented as independent heat maps. See Supplementary Table 5 for cluster composition.
B. Enriched GO-terms [43] and features for clusters of mRNA defined in panel A.

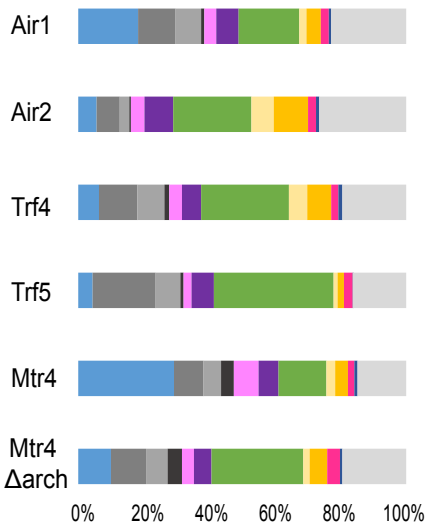
Figure 6: Correlations between protein interaction sites on RNAPII transcripts
A. Correlation matrix displayed as a heat map, showing correlations between binding sites recovered with different factors across RNA polymerase II transcripts. NOTE: Above the diagonal, the matrix shows the extent to which the factors target the same RNA species as "Co-targeting of RNAs". Below the diagonal, the matrix shows the extent to which the factors bind closely positioned RNA sites (≥50 nt) as "Colocalization of sites".
B-M. Binding of the indicated proteins relative to Air1. CRAC peaks for Air1 across the genome were selected and used as a reference to align peaks for other proteins. Data is shown for the combined top 100 RNAPII transcripts bound by Npl3 and Air1 (B-G, 127 genes in total) or top 100 mRNAs (H-M, 147 genes in total). Two independent replicates were combined in each panel.

Figure 7: Alterations in mRNA abundance in strains lacking Trf5
A: Changes in Mtr4 binding by CRAC in strains lacking Trf5, relative to the wild-type.
B: Changes in mRNA abundance by RNAseq in strains lacking Trf5, relative to the wild-type. Genes are ordered by ranking in frequency of Trf5 binding (RPKM) and bars show fold change (log2) in strains lacking Trf5 relative to the wild-type. Note that these are normalized data, so the total recovery of Mtr4 across all genes is constant.
C, D: RNAseq distribution across mRNAs, normalized to the total binding value across the mRNA subset, in wild-type (4 independent replicates averaged) and *trf5*Δ (3 independent replicates averaged). Reads were aligned with the polyadenylation sites (pA). Panel C shows genes grouped according to clusters identified in Fig. 5A, plus all mRNAs. In panel D, the 2,005 genes analyzed in Fig. 5 are grouped by Trf5 binding in the 3' region (bin 5 in Figs. 5A and S6).

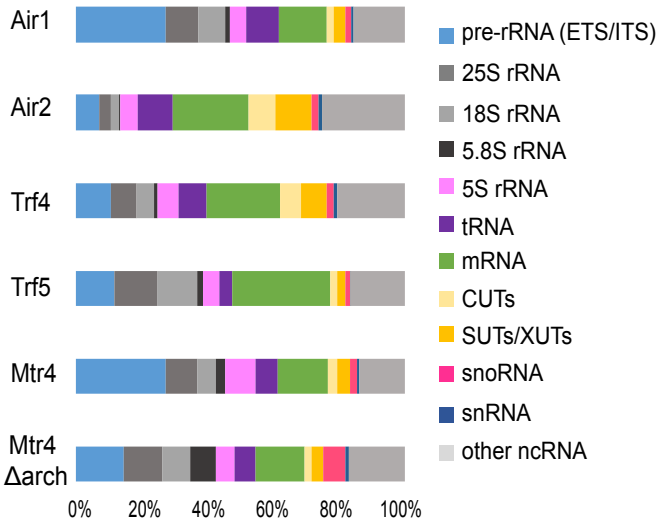


Delan-Forino et al, Figure 1

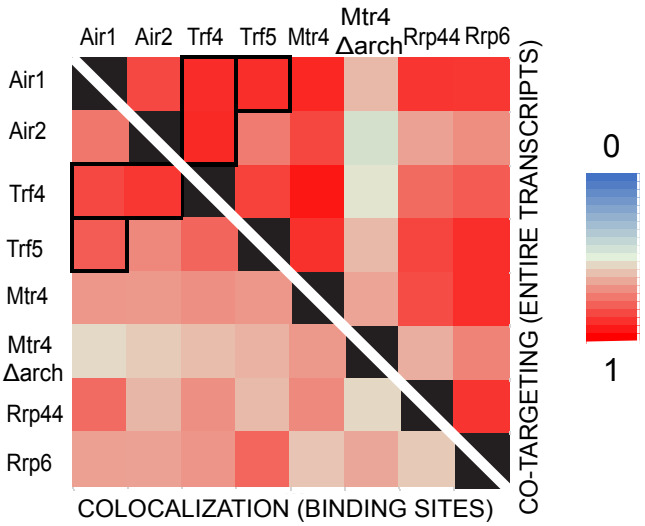
A. Total CRAC hits distribution among RNA classes



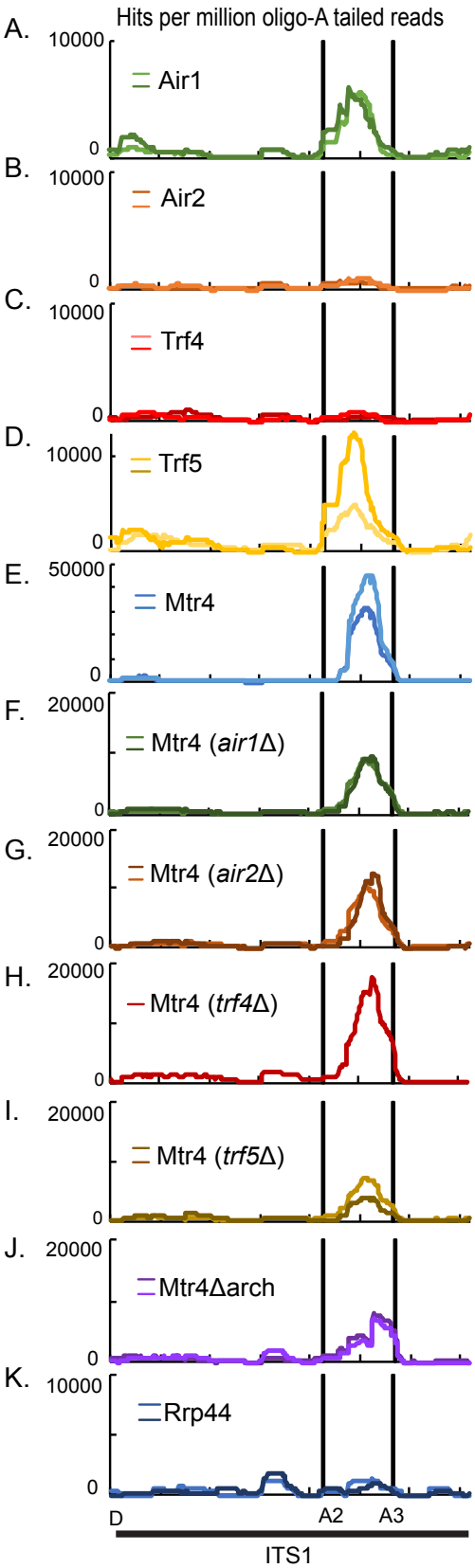
B. Oligo-A tailed CRAC hits distribution among RNA classes

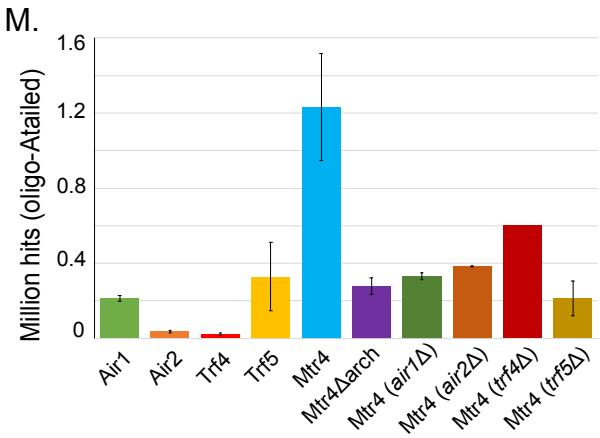


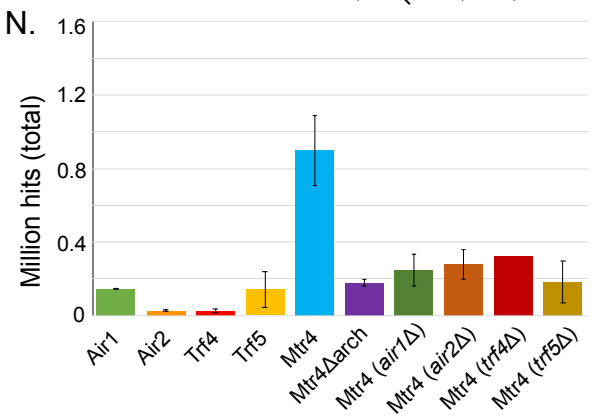
C. Correlation of Oligo-A tailed CRAC hits

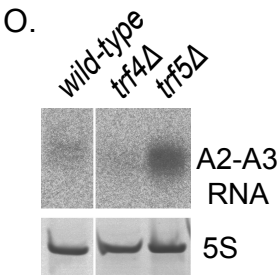


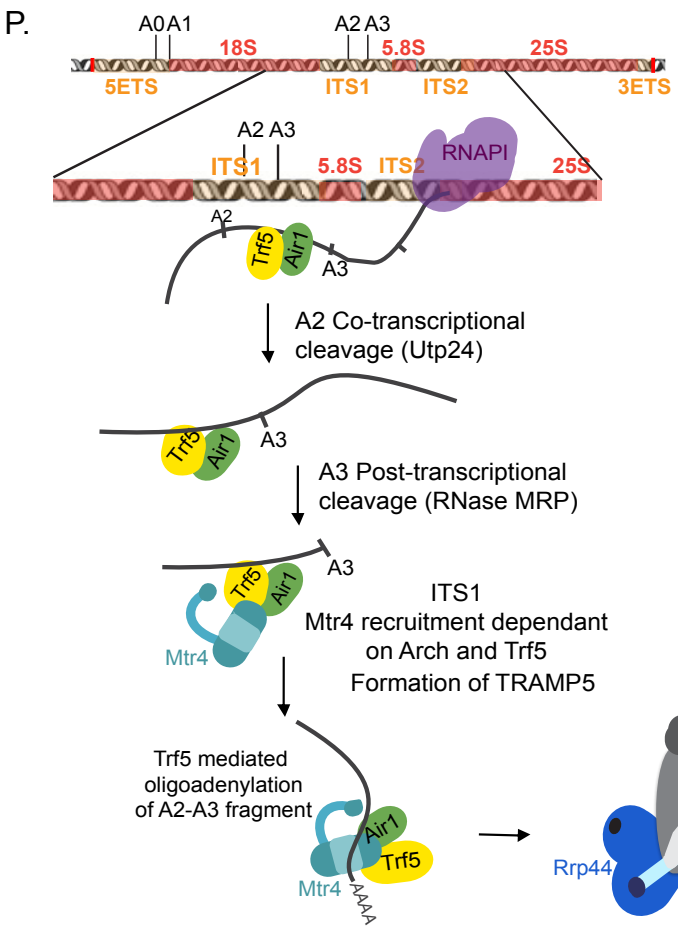
Delan-Forino et al, Figure 2



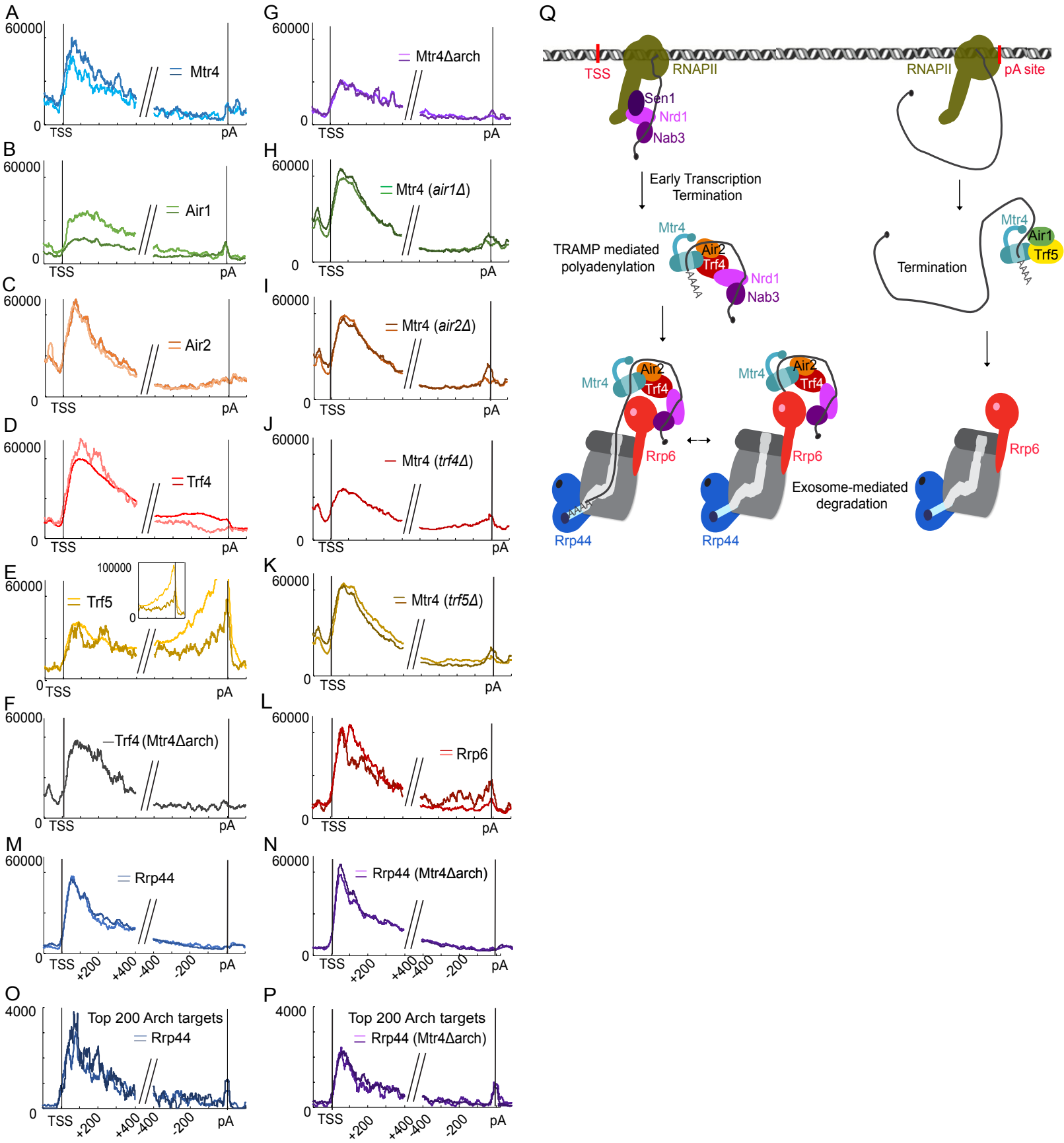






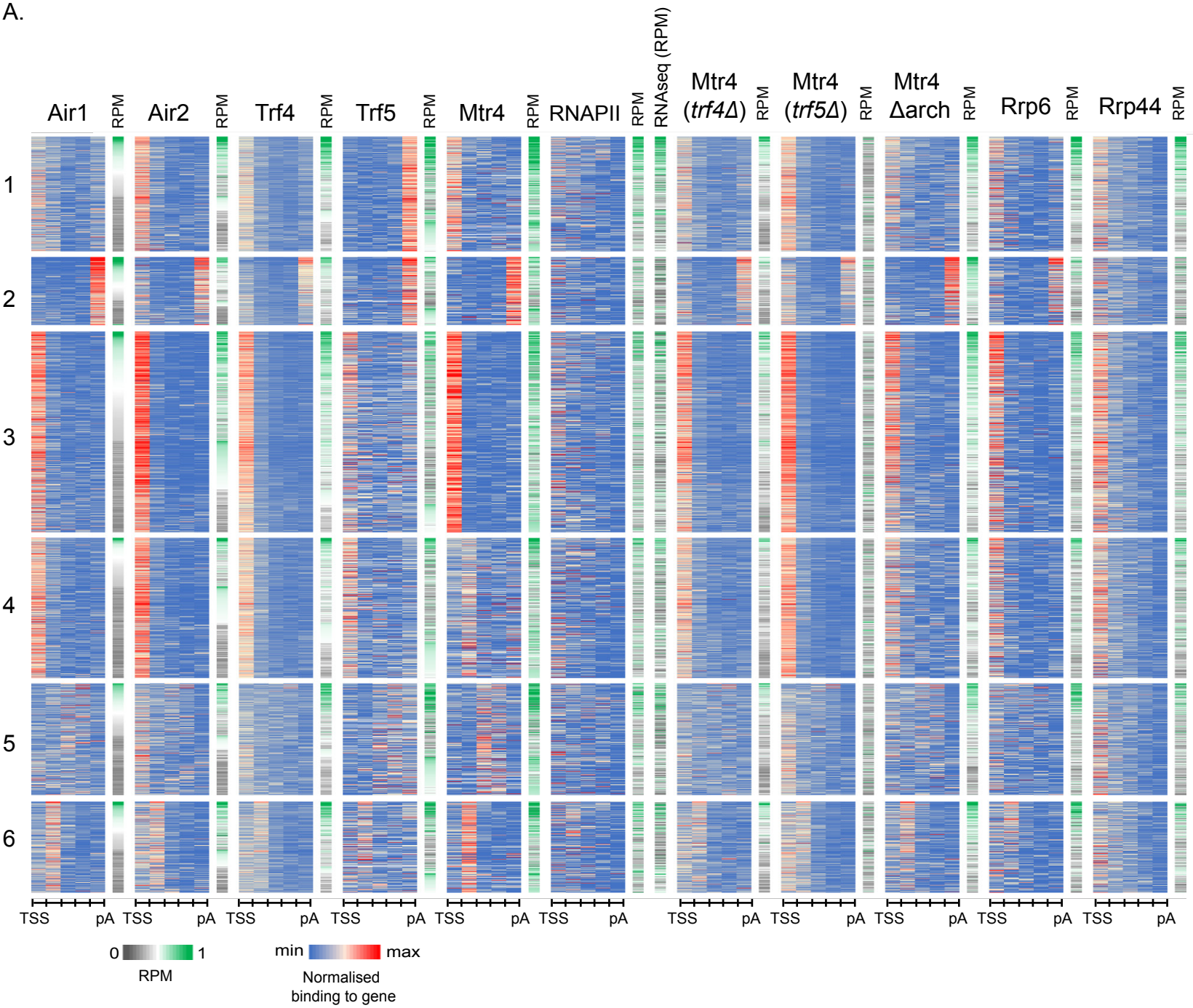


Delan-Forino et al, Figure 3



Delan-Forino et al, Figure 4

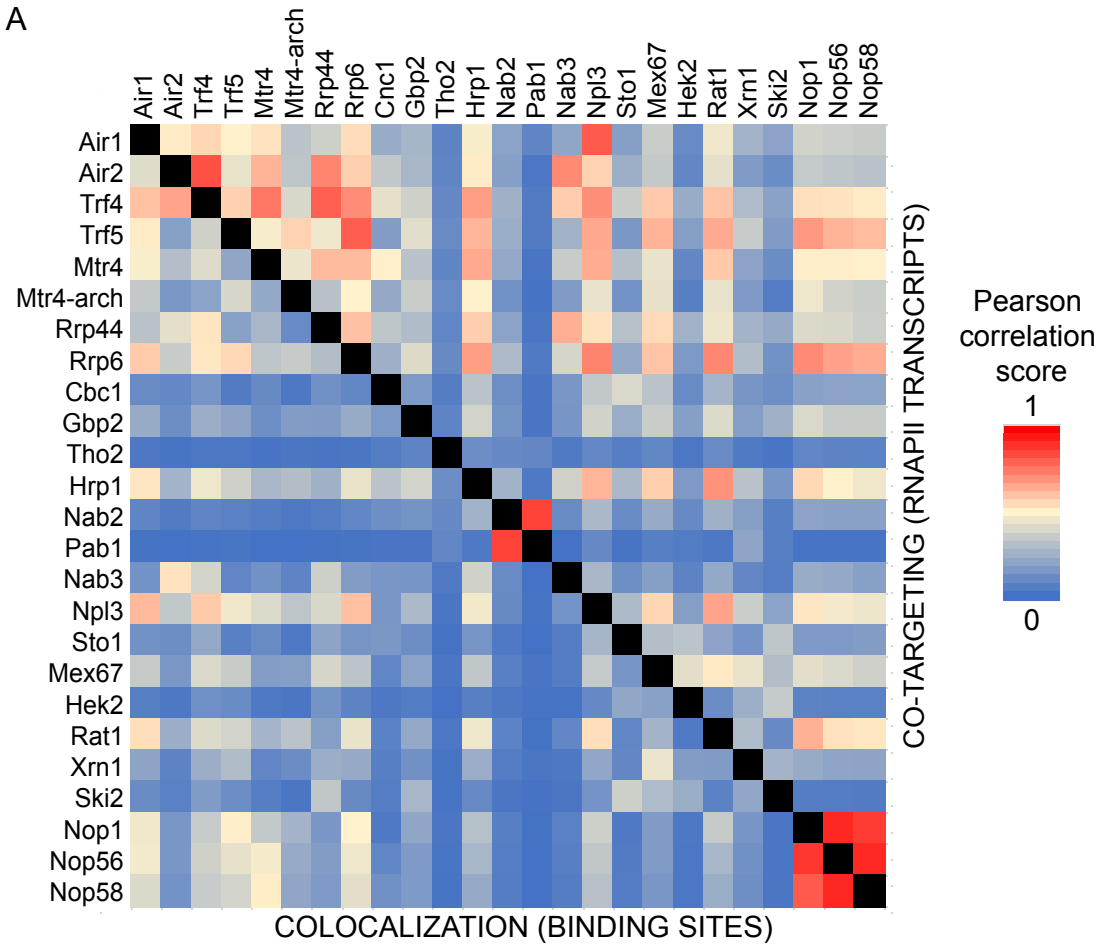
A.

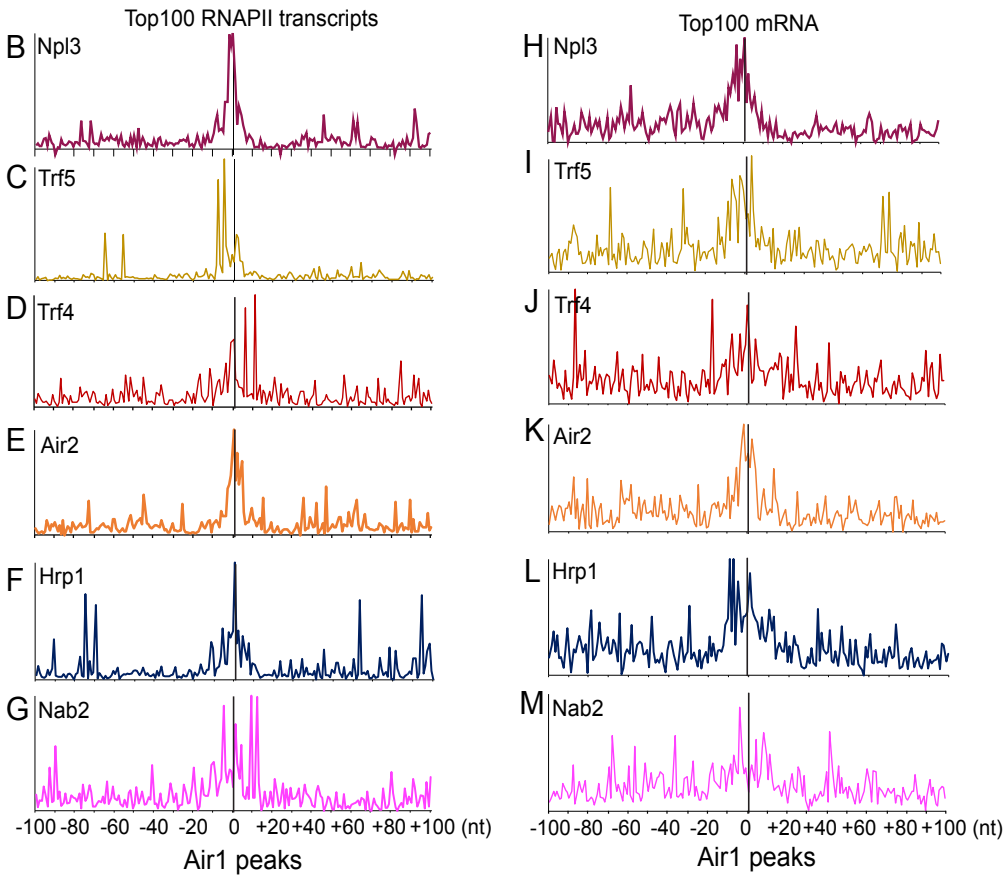


B

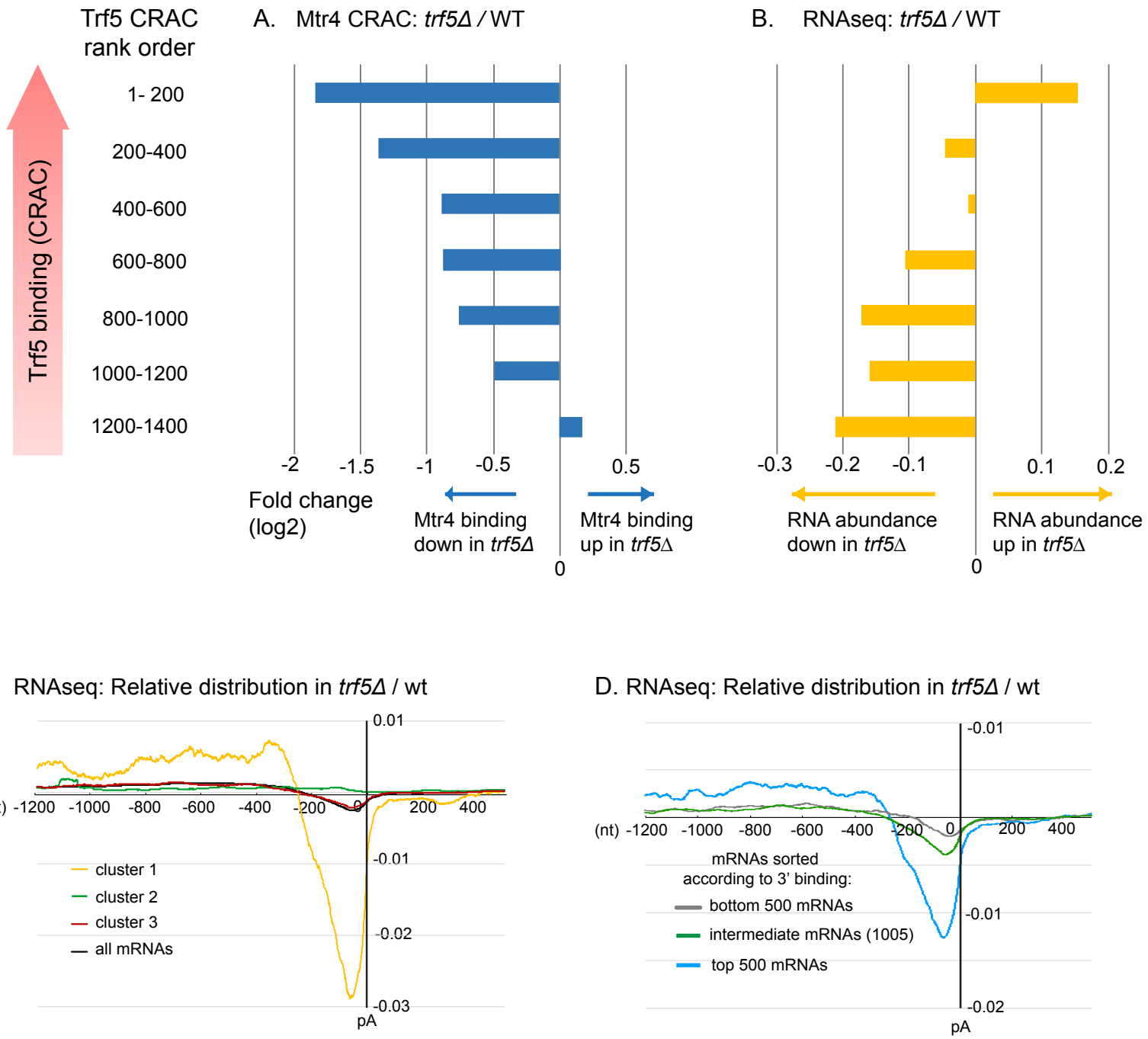
	number of mRNAs	Enriched GO-term	Enrichment pValue	Enriched features
cluster 1	282	process : amide biosynthetic process / translation component : ribosome	5.33E-10 / 5.58 E-10 1.85E-07	
cluster 2	186	process : cellular response to DNA damage component : spliceosomal complex	2.92E-05 5.2E-05	low abundance RNA
cluster 3	591	process : small molecule biosynthetic process / negative regulation of gene expression	1.29E-04 1.45E-04	
cluster 4	385	process : intracellular signal transduction component : membrane	7.45E-04 7.16E-04	
cluster 5	300	process : cytoplasmic translation component : ribosomal subunit	9.48E-08 1.29E-09	
cluster 6	261	x	x	putative proteins feature-containing intron

Delan-Forino et al, Figure 5





Delan-Forino et al, Figure 6



Delan-Forino et al, Figure 7

SUPPLEMENTARY FIGURE LEGENDS:

Figure S1: Mtr4Δarch construct and phenotype.

A. Main domains of Mtr4 and Mtr4Δarch proteins.

B. Serial dilution (1:5) growth complementation tests of strains expressing Mtr4 and Mtr4Δarch, on galactose medium for expression of wild-type Mtr4, or on glucose to deplete Mtr4.

C. Analysis of 5.8S +30 pre-rRNA, a characterized target requiring the Mtr4 Arch domain for processing, in strains expressing Mtr4Δarch during depletion of wild-type Mtr4. Left panel; SybrSafe stained,10% polyacrylamide, denaturing gel. Right panel; northern blot.

D-E. Western blot analysis assessing abundance of (D) Mtr4 in wild-type strain and strains deleted for TRAMP component and (E) Trf4 in strains expressing wild-type Mtr4 or Mtr4Δarch depleted of wild-type Mtr4 for 7 or 30 h. The same strains were used for CRAC analyses.

Figure S2: Correlations between protein interaction sites over all genes

Correlation matrix for protein binding in CRAC analyses, showing correlations between binding sites. NOTE: Above the diagonal, the matrix shows the extent to which the factors target the same RNA species as “Co-targeting of RNAs”. Below the diagonal, the matrix shows the extent to which the factors target the closely positioned RNA sites (≥50 nt) as “Colocalization of sites”. See Supplementary Table 3 for values for individual CRAC dataset. As Figure 6 but including all annotated genes.

Figure S3: TRAMP interaction sites on the pre-rRNA ITS1 region.

A-L. Distribution of total reads across the ITS1 region of the pre-rRNA, normalized to millions of mapped reads and recovered with the indicated TRAMP components (A-E), Mtr4 in strains lacking the indicated TRAMP components (F-I), Mtr4 lacking the Arch domain (J), Rrp44 (K), Rrp44 in strain expressing Mtr4 lacking the Arch domain (L). Scale is linear.

Figure S4: Recovery of pre-mRNA intronic sequences in CRAC analyses

The analysis was performed using read per kilobase (sum of intron hits/ sum of all mRNA hits) on collapsed CRAC dataset in which low complexity reads were filtered out.

Figure S5: Effects of altered Trf5 structure on ITS1 binding

A-B. Schematic representation of major protein domains of Trf4 and Trf5 (A) and Trf5 chimeric constructs used in CRAC analysis (B). NTD: N-terminal domain; CTD: C-terminal domain; cen: central domain; FPH: 3xFlag - Pre-scission protease site - Histidine x 6 Tag.

C. Distribution of all reads and A-tailed reads mapped to different RNA substrate classes recovered in CRAC datasets. At least two biological replicates were averaged for each protein. D. Distribution of A-tailed reads across the ITS1 region of the pre-rRNA, normalized to millions of mapped reads and recovered with the different Trf5 constructs. The lines in each panel represent results from independent CRAC experiments.

Figure S6: Quantitation of total reads in cluster analysis

A-F: Distribution of total hits across mRNAs in each cluster defined in Figure 5. Each transcript was divided in 5 bins of equal length from TSS to pA site. Binding across each bin was calculated as a fraction of total binding across individual gene (set to 1). For each transcript, the average between two biological replicates was used. In each cluster, the values of all bins were added to obtain a binding profile.

G-L: Numbers of reads (RPM) for the indicated factor in each cluster.

Figure S7: Metagene analyses of Air1 around pA sites

A-C. Comparison of Air1 distribution around the pA sites of all mRNAs (A) or in cluster 1 (B) or cluster 4 (C), defined in Figure 5. Each panel shows the hit density, normalized to millions of reads mapped to mRNAs. The two lines in each panel represent results from independent CRAC experiments. Reads were aligned with polyadenylation sites (pA).

Figure S8: Effects of altered Trf5 structure on mRNA binding

A-B. Hits per millions oligo(A) tailed reads encompassing the TSS region (TSS +150 nt) (A) or poly(A) site (pA -150 nt) (B) were summed and exhibited as a bar graph. Individual replicates were averaged, with standard deviation shown by error bars.

C. HEAT map showing binding across individual mRNAs from clusters defined in Figure 5. Each transcript was divided into 5 bins of equal length from TSS to pA site. Binding across each bin was calculated as a fraction of total binding across individual gene (set to 1). Numbers of reads were averaged between two biological replicates. The data are displayed as heat maps. Colors reflect relative binding between different bins of the same mRNA, so the lighter color for the Trf5-ΔCTD panel indicates a more even distribution.

Figure S9: Correlations between protein interaction sites at the nucleotide level

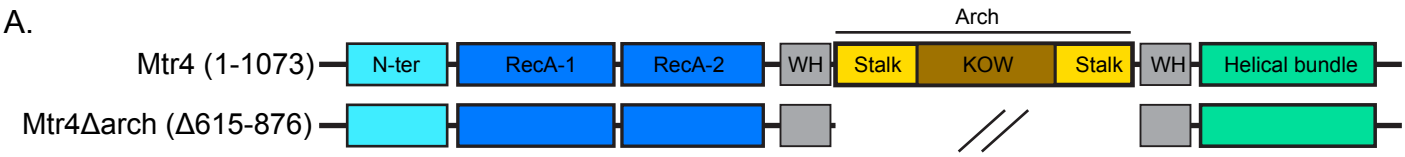
As Figures 6B-6M, except the distribution of binding across each gene was calculated as a fraction of total binding across the gene (set to 1).

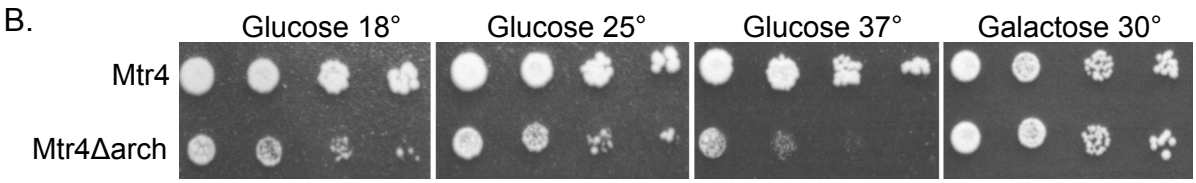
A-L: Binding of the indicated proteins relative to Air1. CRAC peaks for Air1 across the genome were selected and used as a reference to align peaks for other proteins. Data are

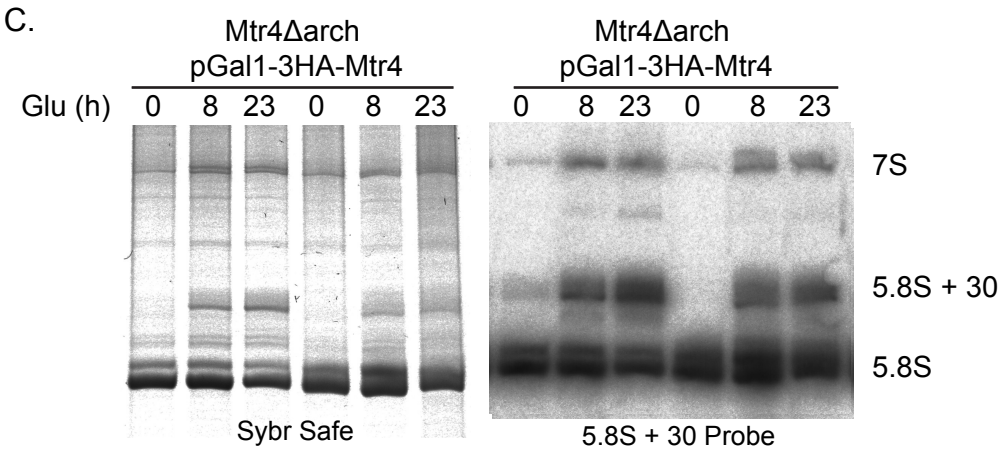
shown for the combined top 100 RNAPII transcripts bound by Npl3 and Air1 (A-F, 127 genes in total) or top 100 mRNAs (G-L, 147 genes in total). Two independent replicates were combined in each panel.

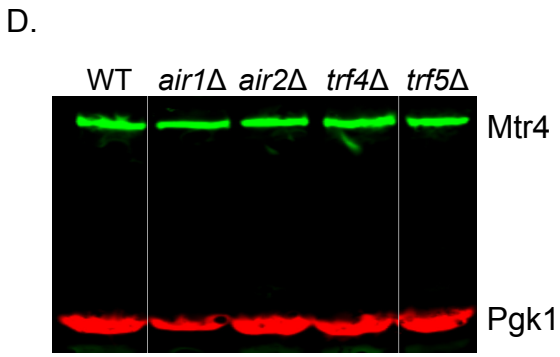
Figure S10: Altered mRNA abundance in *trf5*Δ and correlations of RNAseq samples over mRNA genes

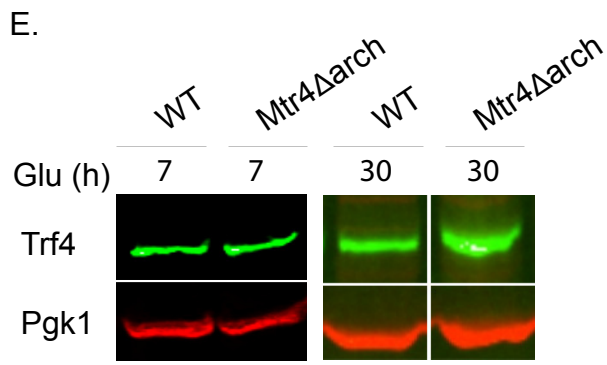
A. Gene expression fold change between wild-type and *trf5*Δ on selected genes. Technical triplicates for each independent biological replicates (2 for wild-type and 3 for *trf5*Δ) were submitted to Real-time PCR amplification. Primer efficiencies were calculated and cycle threshold (Ct) values were averaged between triplicates for each RNA sample. Gene expression fold change were determined as previously reported [55]. For qPCR raw data, primer efficiencies, and fold change calculation on wild-type and *trf5*Δ strains, see Table S7.
B: Correlation matrix of mRNAs recovered between RNAseq replicates (4 wild-type (WT) and 3 *trf5*Δ samples) used for Figure 7.



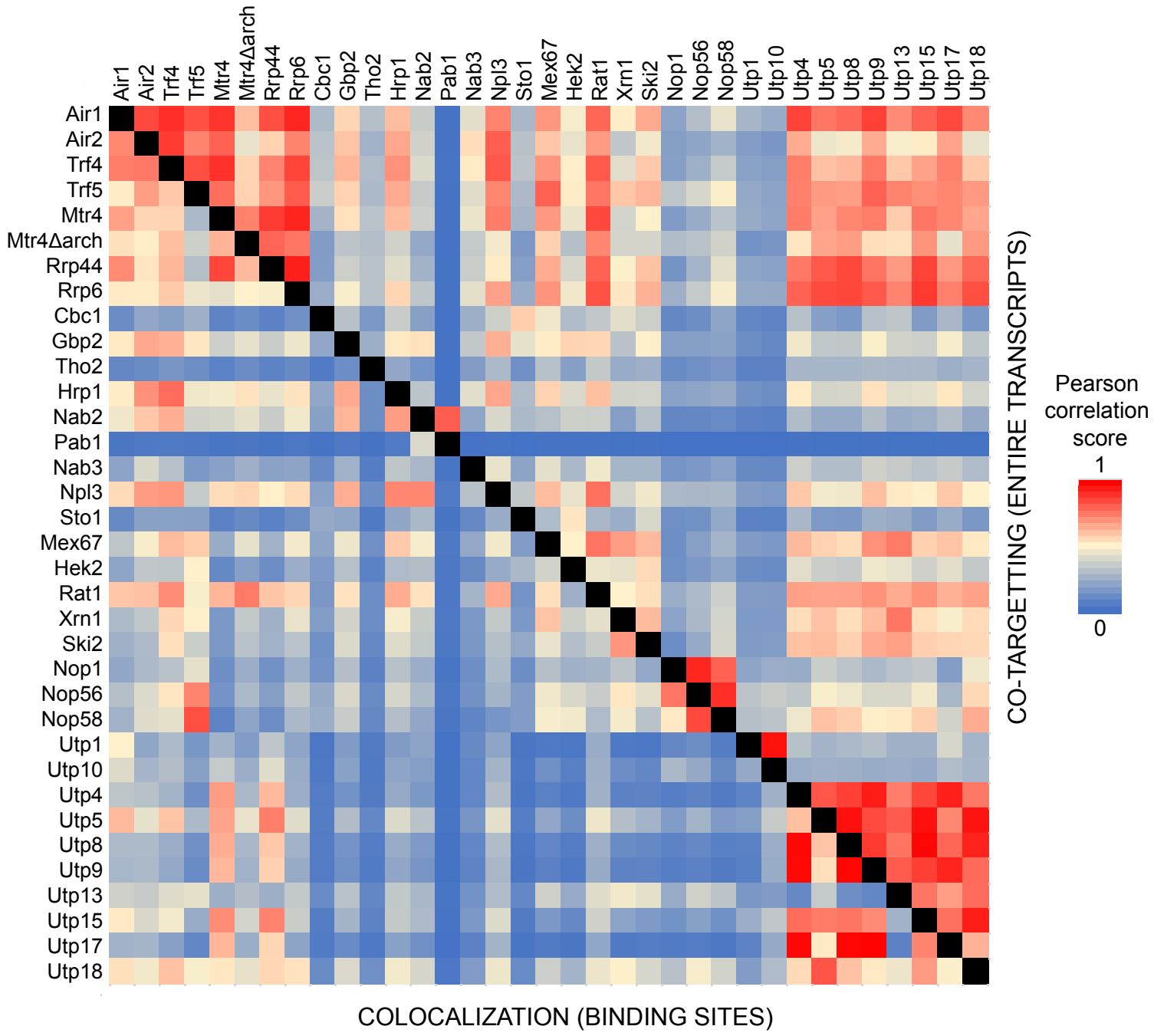




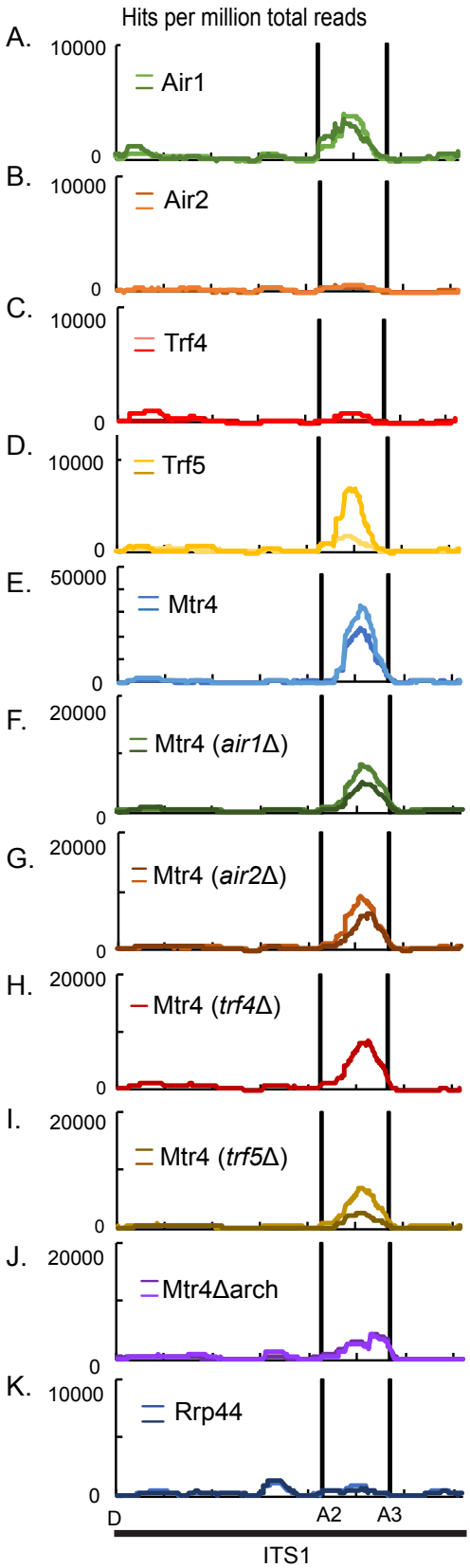




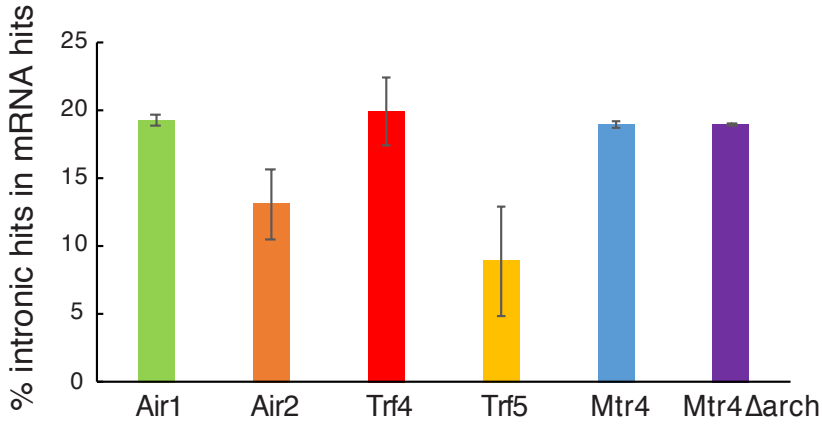
Delan-Forino et al, Supplementary Figure S1



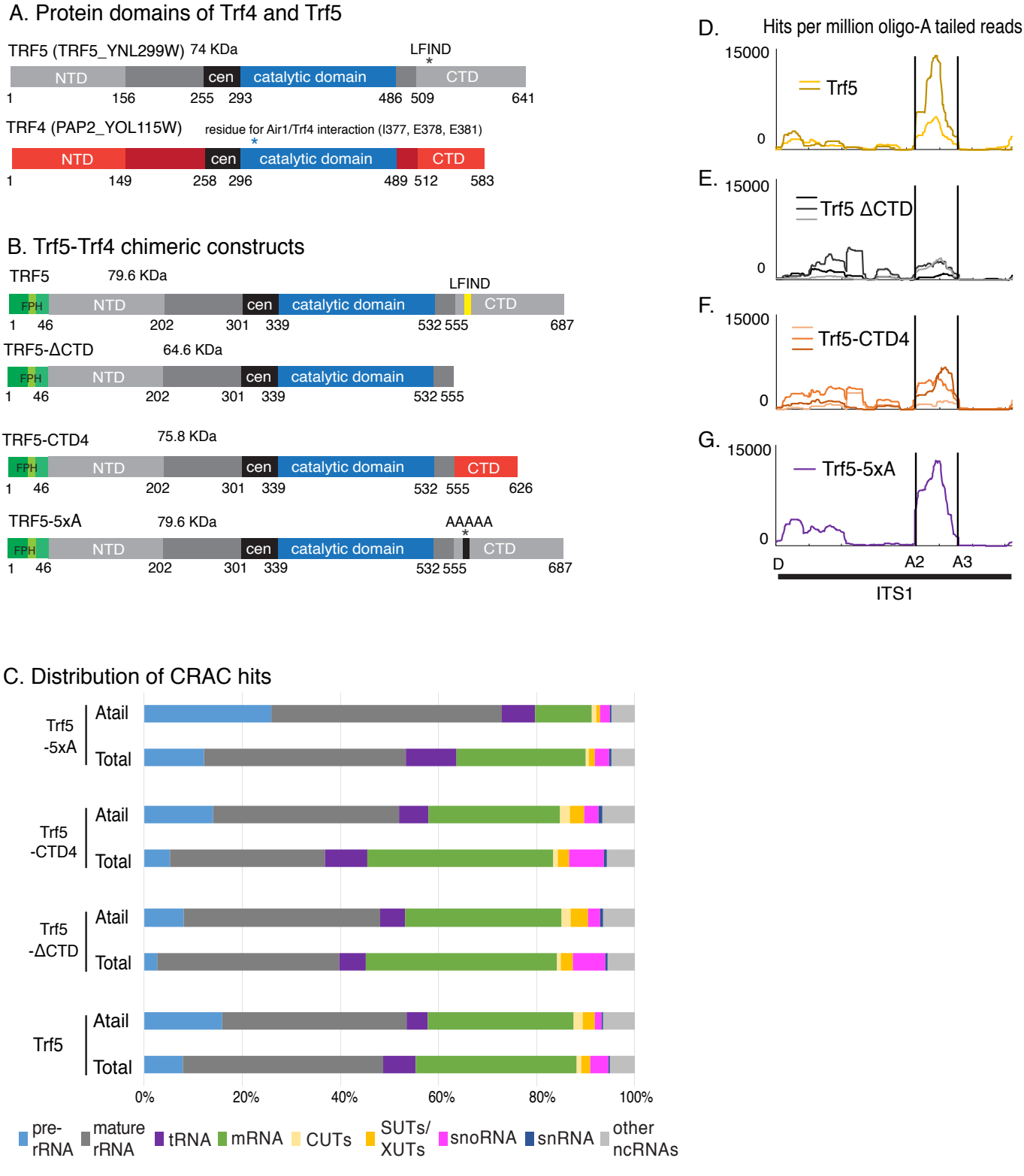
Delan-Forino et al, Supplementary Figure S2



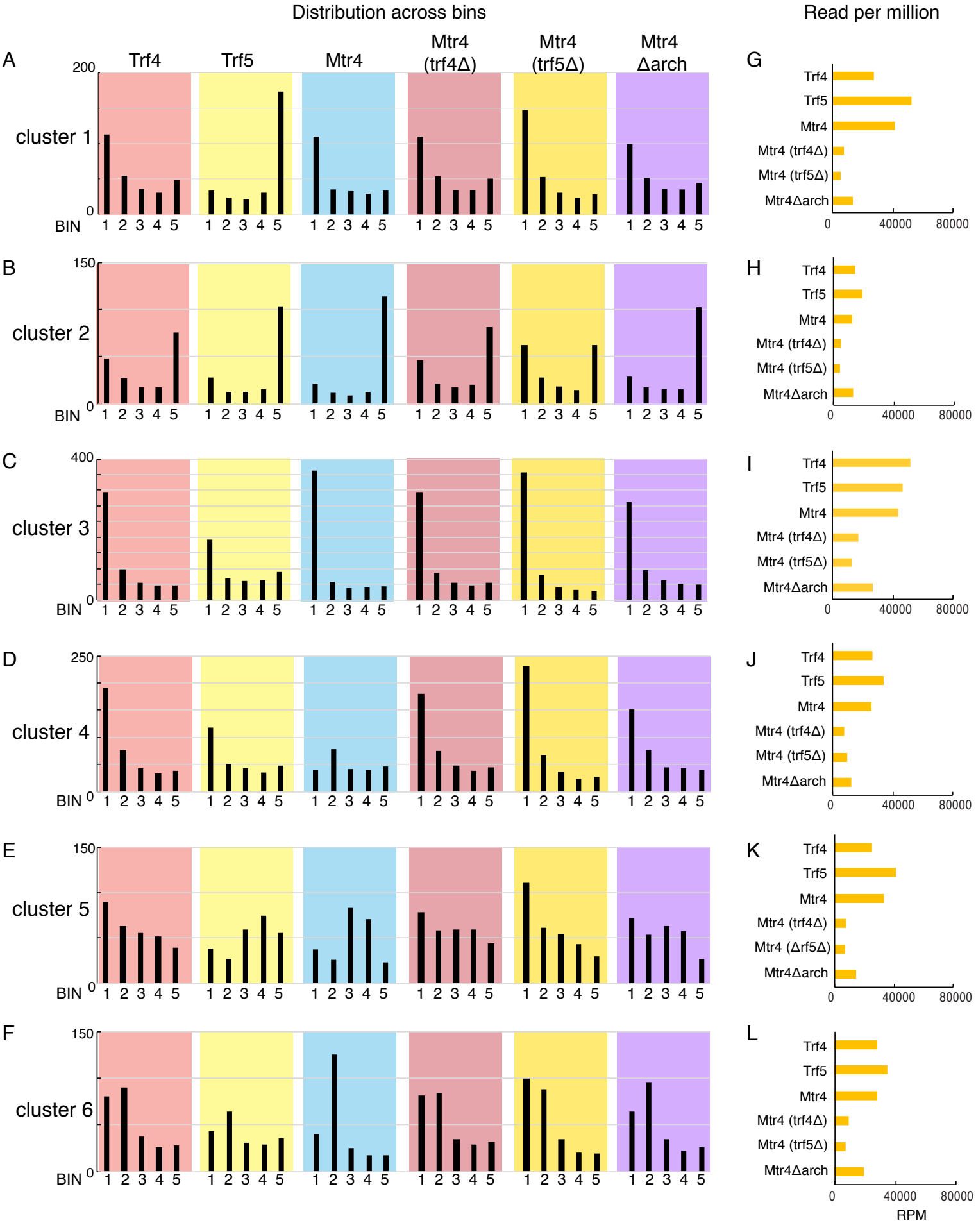
Delan-Forino et al , Supplementary Figure S3



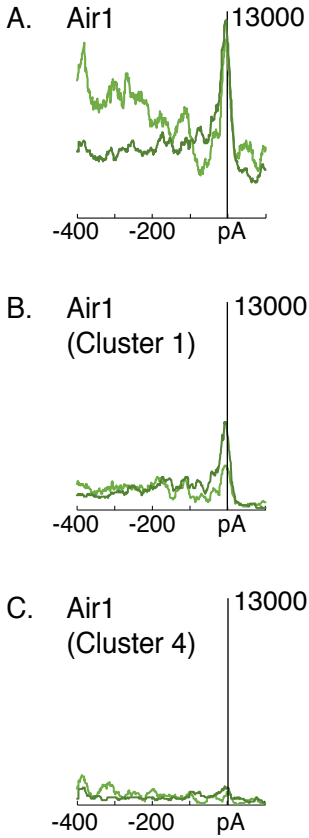
Delan-Forino et al, Supplementary Figure S4



Delan-Forino et al, Supplementary Figure S5

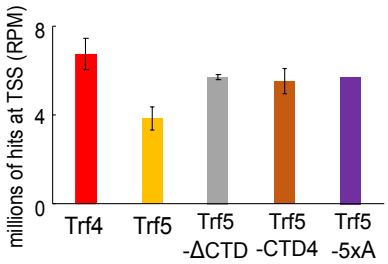


Delan-Forino et al, Supplementary Figure S6

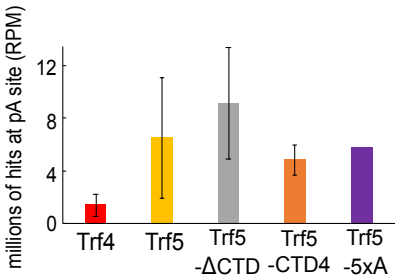


Delan-Forino et al, Supplementary Figure S7

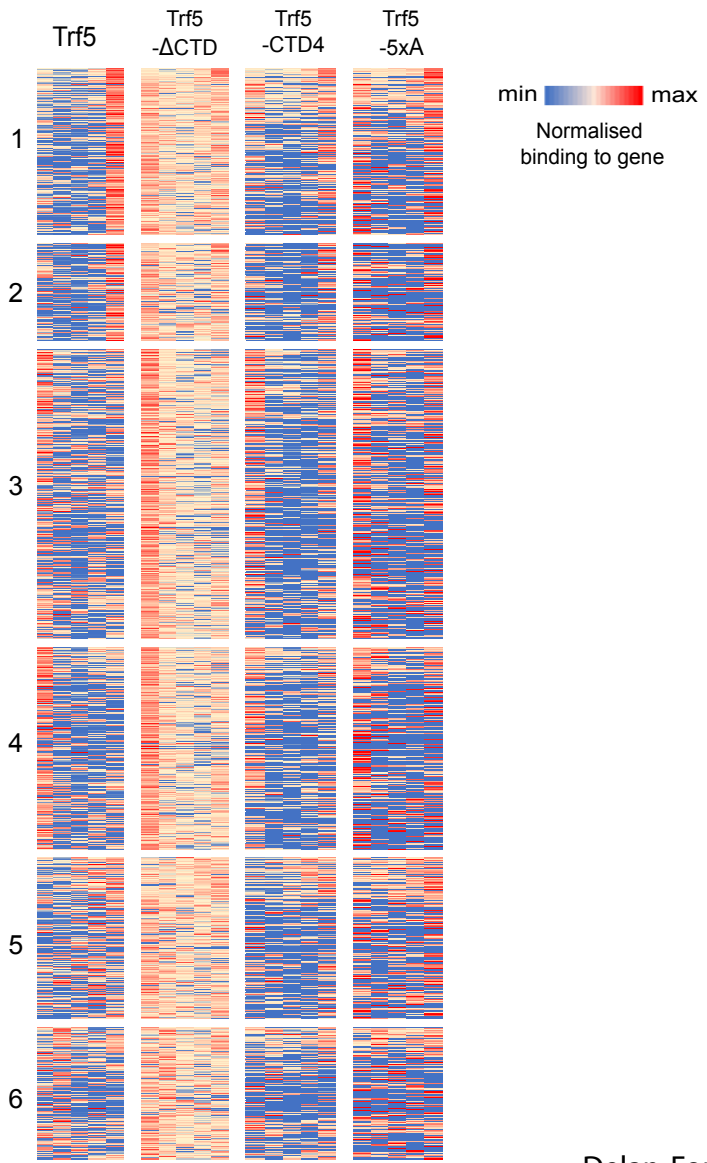
A.



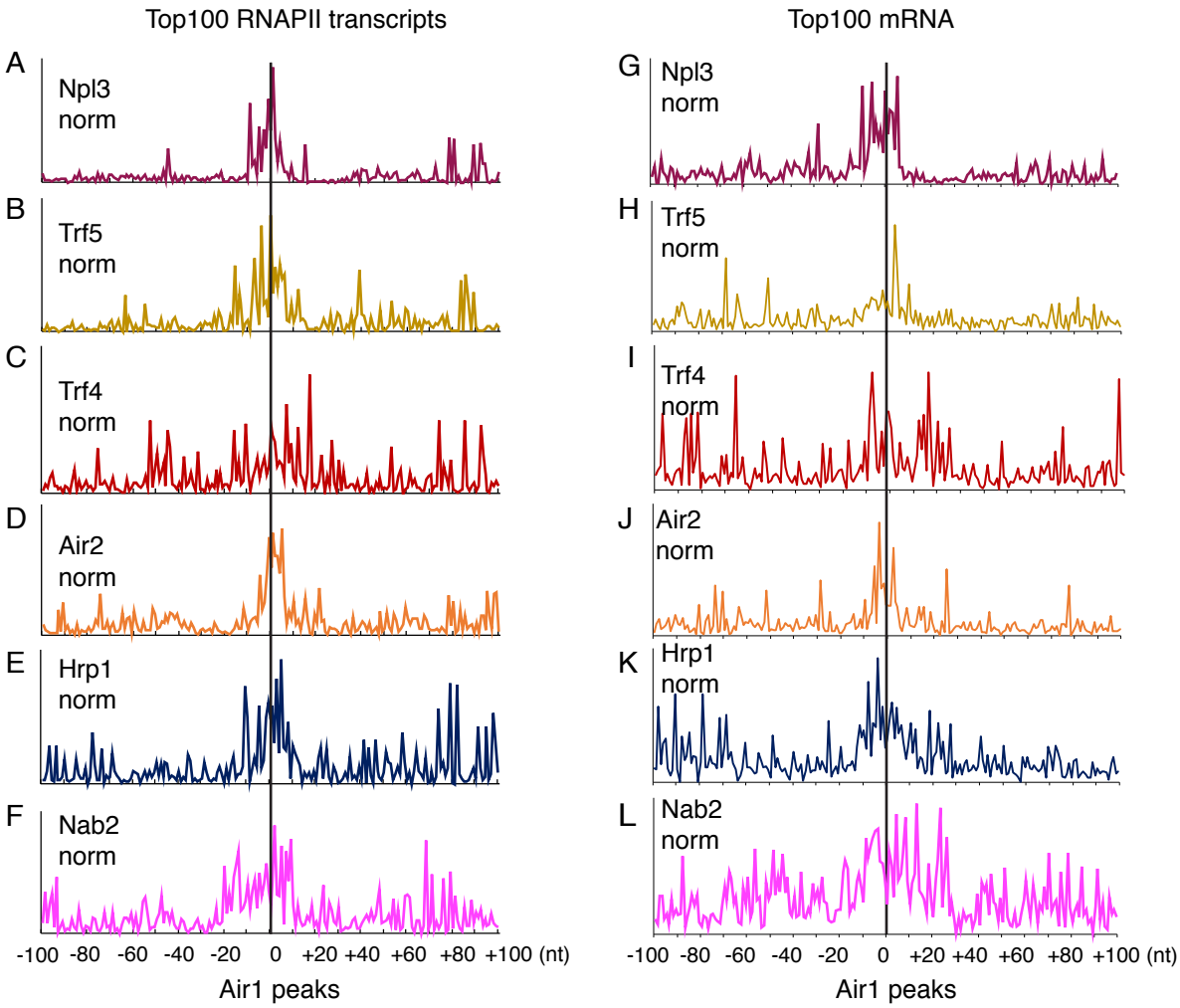
B.



C.

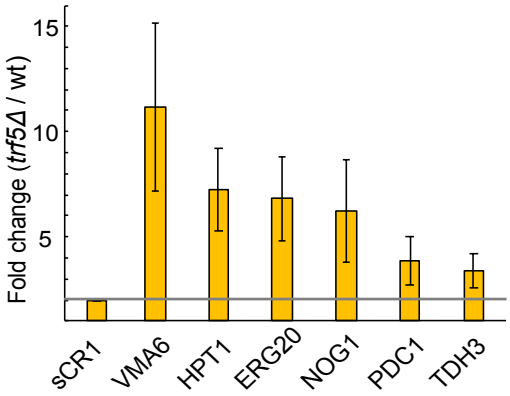


Delan-Forino et al, Supplementary Figure S8

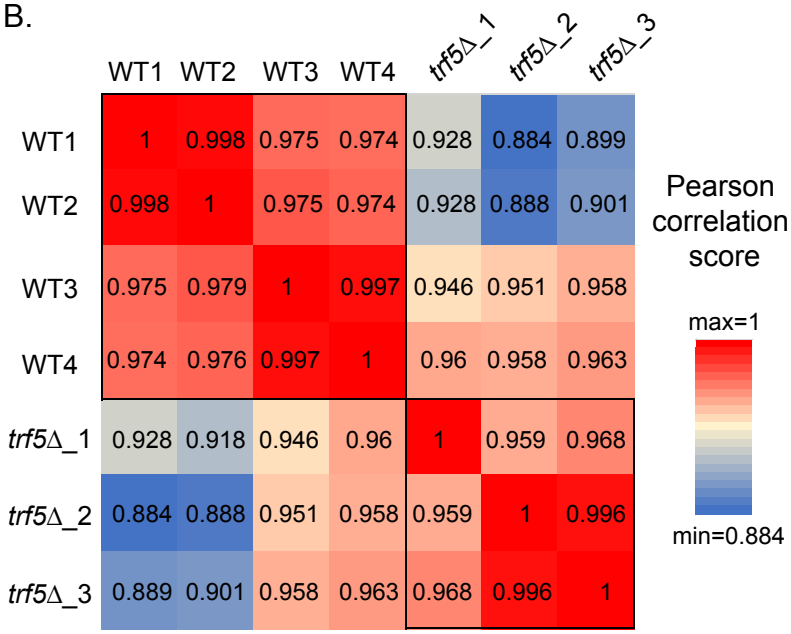


Delan-Forino et al , Supplementary Figure S9

A.



B.



Delan-Forino et al, Supplementary Figure S10

SUPPLEMENTARY TABLES

Table S1: Strains used in this study

Table S2: iBAQ (Intensity based absolute quantification) values for proteins recovered by mass spectrometry (LC-MS/MS) analysis. Air1, Air2, Trf4 and Trf5 were used as bait for pull-downs in buffer containing 1M NaCl. IBAQ values in the mock precipitated sample are shown as the negative control.

Table S3: riBAQ (relative molar abundances for each protein) values for proteins recovered by mass spectrometry (LC-MS/MS) analysis. The first sheet shows rIBAQ ratios, determined by dividing the iBAQ value for each protein by the sum of all non-contaminant iBAQ values. The second sheet shows enrichment, calculated as the riBAQ ratio between TRAMP protein pull-down and the mock sample. All proteins showing a ratio less than 2 were considered to be not significantly enriched.

Table S4: Correlation matrix showing Pearson correlation between binding sites recovered with different factors. The first sheet shows correlations calculated from hits across individual transcript (the extent to which the factors target the same RNA species as "Co-targeting of RNAs"). The second sheet shows correlations calculated from hits across 50 nt windows encompassing the whole genome (the extent to which the factors target closely positioned RNA sites (≥50 nt) as "Colocalization of sites"). Values for individual CRAC datasets are shown.

Table S5: Heat maps for distribution of TRAMP components on each mRNA
Heat map showing binding across individual mRNAs of different factors. The top 1,000 protein-coding genes recovered in CRAC with each TRAMP component were selected and combined (2,005 mRNAs in total). Each transcript was divided into 5 bins of equal length from TSS to pA site. Binding across each bin was calculated as a fraction of total binding across individual genes (set to 1). Numbers of reads were averaged between two biological replicates. Only the coverage of Air1, Air2, Trf4, Trf5 and Mtr4 along each transcript was used for Euclidian distance-based clustering. Other data sets shown were not included in the clustering analysis but are sorted according to the clustered list. The data are displayed as heat maps. The RPM for each transcript recovered by each protein was calculated and represented as independent heat maps. Displayed in Figure 5.

Table S6: Correlation matrix showing Pearson correlation between binding sites recovered with different factors across RNA polymerase II transcripts. The first sheet shows correlations calculated from hits across individual RNAPII transcript (the extent to which the factors target the same RNA species as “Co-targeting of RNAs”). The second sheet shows correlations calculated from hits across 50 nt windows encompassing the parts of genome coding for RNAPII transcripts (the extent to which the factors target the closely positioned RNA sites (≥50 nt) as “Colocalization of sites”). Values for individual CRAC datasets are shown.

Table S7: Raw data, Primer efficiencies, and fold change calculation for Real-time PCR amplification analysis on wild-type and *trf5*Δ strains. Technical triplicates for each Independent biological replicates (2 for wild-type and 3 for *trf5*Δ) were submitted to qPCR. Primer efficiencies were calculated and cycle threshold (Ct) values were averaged between triplicates for each RNA sample. Gene expression fold change were determined as previously reported [55].