Nucleic Acids Research, 2025, **53**, gkaf798
https://doi.org/10.1093/nar/gkaf798
RNA and RNA-protein complexes

A time-resolved framework for the recruitment of mRNP processing and assembly factors to a site of transcription

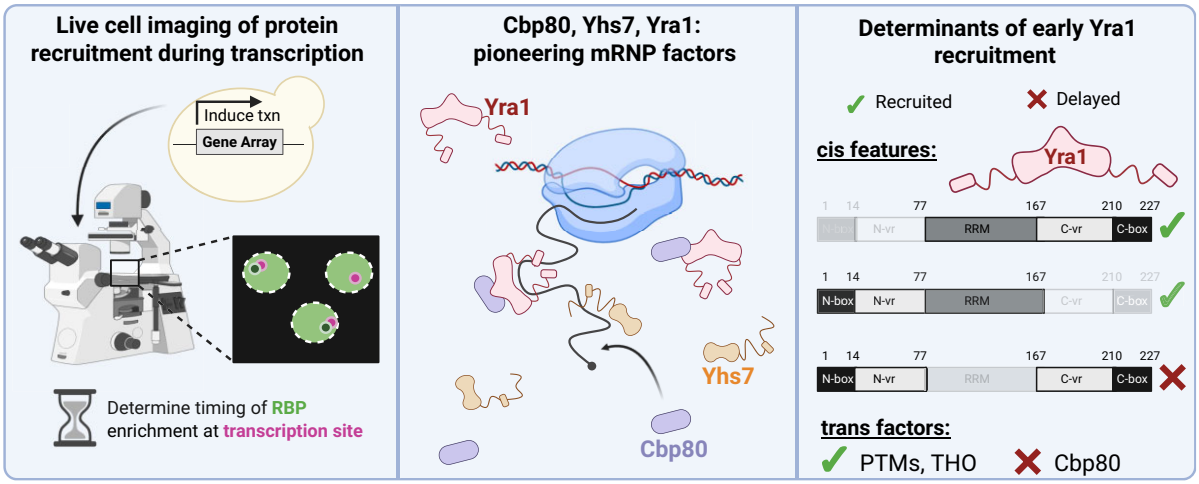
Theresa Wechsler[1,2], Ryuta Asada[1], Andrew N.M. Dominguez[1,2], Rachel Montpetit[1], Julia K. McCormick[1], Kalyn Concepcion[1], Ben Montpetit[1,2,*]

[1]Department of Viticulture and Enology, University of California, Davis, CA 95616, United States
[2]Biochemistry, Molecular, Cellular, and Developmental Biology Graduate Group, University of California, Davis, CA 95616, United States
[*]To whom correspondence should be addressed. Email: benmontpetit@ucdavis.edu

Abstract

Processing and packaging of messenger ribonucleoprotein (mRNP) particles involve complex, coordinated interactions between nascent transcripts, RNA-binding proteins (RBPs), and associated factors. Despite the critical role of co-transcriptional mRNP assembly in gene expression, the temporal dynamics of this process are not well understood. Here, a live cell imaging assay is reported in *Saccharomyces cerevisiae* to detect recruitment of endogenous fluorescently tagged proteins to a transcriptionally active locus. Protein recruitment to an inducible integrated gene array composed of 25 transcriptional units is detected by colocalization with *lacO* repeats. Using arrays with two different promoters and the same coding sequence (*GFA1*), arrival times for a variety of mRNP processing and assembly factors were quantified. These analyses revealed Yra1, Cbp80, and Yhs7 as pioneering mRNP assembly factors. Notably, Yra1 recruitment occurs independently of the THO complex, with early localization supported by Cbp80 and the RNA recognition motif of Yra1. Altogether, this work establishes the first comprehensive temporal framework for understanding protein recruitment during co-transcriptional mRNP assembly, providing mechanistic insights into the dependencies of Yra1 recruitment.

Graphical abstract



Introduction

A defining aspect of eukaryotic gene expression is the physical separation of messenger RNA (mRNA) production in the nucleus and protein translation in the cytoplasm. This necessitates export of mRNA from the nucleus, a highly regulated process requiring precise spatial and temporal control [1–4]. In the nucleus, mRNA transcripts must be processed and packaged into messenger ribonucleoprotein (mRNP) particles for productive export [5]. Processing events (e.g. 5′ capping, splicing, cleavage, and polyadenylation) are coordinated by dynamic interactions between the transcript, RNA-binding proteins (RBPs), non-coding RNAs, and other associated factors [1, 3, 6]. These events are subject to quality control governed by RNA decay pathways to prevent export of aberrant transcripts to the cytoplasm, thereby circumventing production of deleterious protein products [7, 8]. The importance of

Received: April 8, 2025. **Revised:** July 3, 2025. **Editorial Decision:** July 23, 2025. **Accepted:** August 7, 2025

nuclear mRNP processing and export to cellular function is highlighted by widespread dysregulation in disease, including cancer and neurological disorders [9–12].

While much is known about individual mRNA processing steps and the RBPs involved [1, 3, 6], how these events are spatially and temporally coordinated *in vivo* is not well characterized. Furthermore, studies using a wide variety of techniques find both human and yeast nuclear mRNPs are compact particles [13–18], but mechanisms of mRNA compaction *in vivo*, and how packaging is accomplished alongside requisite processing events, remain to be fully understood. Recent works highlight TREX (TRanscription-EXport) complex components as important packaging factors that make critical contacts with RNA and other protein components [14, 15, 18]. TREX is an evolutionarily conserved complex that engages nuclear mRNPs consisting of the THO complex (same name in human), Yra1 (human ALYREF), and Sub2 (human UAP56) [19–22]. THO, a five-subunit complex in *Saccharomyces cerevisiae* consisting of Tho2, Hpr1, Mft1, Tex1, and Thp2, is important for mRNP biogenesis and maintenance of genomic stability [23]. Yra1, an essential protein first identified for its RNA annealing capacity, is critical for mRNP export via interactions with the major export factor Mex67-Mtr2 [24, 25]. Sub2 is a DEAD-box RNA ATPase required for splicing and mRNP export that engages THO and Yra1 and is proposed to act as a molecular switch to regulate mRNA export [19, 26–31].

While TREX components are known to engage in a stable complex, to what extent functions are fulfilled by the full complex or individual subunits is not entirely clear, with evidence suggesting distinct functions (e.g. Sub2 being required for splicing while THO is not) and the presence of super-stoichiometric Yra1 levels in mRNPs relative to THO and Sub2 [14, 18]. Previous work has suggested THO recruits Sub2 and Yra1 to mRNPs during transcription elongation, thus licensing them for nuclear export [32–34]. This model of Yra1 and Sub2 loading via THO is based largely on chromatin immunoprecipitation (ChIP) data, including data showing co-transcriptional association defects of Yra1 and Sub2 in THO mutants; however, ChIP data do not account for cell-to-cell heterogeneity (level of RBP, transcription status of locus, etc.), making the average level of RBPs at gene loci difficult to directly tie to functional outcomes. Furthermore, the wide-ranging impacts of THO mutants on genomic stability, transcription, and mRNA processing leave open many possibilities as to why the association of RBPs with chromatin changes by ChIP. This leaves many unknowns regarding the mechanisms and functions of THO, Yra1, and Sub2 in mRNP assembly, packaging, and export.

More broadly, critical questions remain as to how mRNP assembly and packaging are coordinated *in vivo* with other co-transcriptional events (e.g. capping, 3′ processing, or R-loop formation and resolution). This includes how events are temporally and spatially coordinated, defined by gene-specific features (e.g. promoter, intron, transcript length, etc.), regulated by stress, or perturbed by mutation. These gaps in knowledge exist in part because techniques to follow recruitment of RBPs and other processing factors to transcription sites, with single-cell and temporal resolution, remain limited. Here, for the first time, a live cell imaging assay is employed in *S. cerevisiae* to detect recruitment of a diverse set of mRNP processing and assembly factors to a transcriptionally active locus using gene arrays. Development of this methodology was motivated by previous work in mammalian cells using tandem gene arrays to visualize enrichment of RNA processing factors at active transcription sites [35–37]. The resulting data demonstrate that (i) Yra1, the cap-binding complex (CBC) subunit Cbp80, and YHR127W/Yhs7 (an RBP of unknown function) are amongst the earliest arriving RBPs upon transcriptional activation; (ii) Yra1 is recruited to transcription sites independent of the THO complex; and (iii) early Yra1 recruitment is supported by Cbp80 and the Yra1 RNA recognition motif (RRM) domain. Overall, these data provide a temporal framework and defined dependencies for the recruitment of mRNP processing and assembly factors to a transcription site.

Materials and methods

Reagents

Reagents and specialized equipment used in this study are detailed in Supplementary Table S1.

Biological resources

All plasmids and DNA oligos used in this study are listed in Supplementary Tables S2 and S3, respectively. Yeast strains used in this study were derived from *S. cerevisiae* BY4741, with the exception of yeast two-hybrid experiments, which are derived from L40ccU strain. Strain numbers and genotypes detailed in Supplementary Table S4. Yeast strains and plasmids available upon request to the corresponding author.

Yeast growth and transformations

All yeast culture was conducted at 25°C in YPD media unless otherwise noted. All yeast transformations were done using the LiOAc based method [38]. C-terminal tagging was carried out using PCR-based integration strategy with plasmids and DNA oligos used listed in Supplementary Tables S2 and S3. N-terminal tagging was conducted using plasmid constructs specific to each gene being targeted (Supplementary Table S2), linearized for integration. Gene deletions were generated by transformation of PCR products from genomic DNA of deletion strains. Genomic tagging and deletion strains were validated by colony PCR using one primer within the transformed construct and one up or downstream in the genome to check for proper integration (Supplementary Table S3).

Yeast two-hybrid

Plasmids for yeast two-hybrid were generated with Gateway cloning to insert genes into bait and prey constructs (Supplementary Table S2). Plasmids were transformed into the L40ccU yeast two-hybrid reporter strain, and growth on media lacking histidine was used as a readout for positive two-hybrid interaction. Expression of tagged bait and prey constructs were validated by western blotting with antibodies for LexA DNA-binding domain (bait constructs) and Gal4 activation domain (prey constructs).

Array plasmid cloning

Integrating array plasmids were cloned using the modular cloning (MoClo) Golden Gate yeast toolkit [39]. This system enables modular assembly of parts with predetermined overhang sequences to assemble transcriptional units (TUs) via Golden Gate cloning, which can then be used for hierarchical assembly of multigene plasmids. Parts plasmids used

were a combination of pre-existing ones from the toolkit and newly cloned parts for this study (Supplementary Table S2). New parts were cloned with flanking BsaI restriction sites and overhang sequences enabling their use as one of the predefined parts in the MoClo system. Sequences for endogenous terminators were chosen based on annotated 3′ UTRs on Saccharomyces Genome Database [40, 41]. BsmBI restriction sites necessary for downstream cloning were removed from *GFA1* coding sequence using site-directed mutagenesis to mutate a single base pair while preserving coding sequence. Likewise, a NotI restriction site necessary to digest plasmid DNA for transformation was removed from *Z_3EVpr* and confirmed by sequencing to preserve transcription factor binding sites.

Parts plasmids were assembled into single TU (1× TU) plasmids via BsaI Golden Gate assembly. Five plasmids of each TU were generated with differing "Assembly Connector" parts (as defined in MoClo toolkit) with BsmBI restriction sites and predetermined overhang sequences for their assembly into multi-gene plasmids (Supplementary Table S2 and Supplementary Fig. S1A). Integrating plasmids to serve as entry vectors for these TUs were cloned also via BsaI Golden Gate assembly (Supplementary Table S2). These plasmids contain two different antibiotic resistance markers, kanMX and hphMX, and each contain homology targeting the other for recombination, such that the plasmid with kanMX targets hphMX for integration and vice versa (Supplementary Fig. S1). These integrating plasmids also contain a GFP dropout, which is replaced by multigene assembly with BsmBI Golden Gate, enabling screening for positive assemblies by colony color (Supplementary Table S2). Single TU-containing plasmids were assembled into five-copy integrating plasmids using BsmBI Golden Gate assembly (Supplementary Fig. S1A).

Yeast array strain construction

The starting strain used for array construction (BMY1051, Supplementary Table S4) had the genotype *ura10Δ::kanMX* [42], providing the homology for the first round of integration at the *URA10* locus. For *Z_3EVpr-GFA1* array, the Z_3EV transcription factor was integrated at the *URA3* locus (already nonfunctional in this strain for *URA3* auxotrophic marker use) using an integrating plasmid targeting the locus (Supplementary Table S2) prior to *Z_3EVpr - GFA1* array integrations to enable activation of the *Z_3EVpr* via addition of β-estradiol. The transcription factor sequence was derived from Addgene plasmid #69099. For both promoter systems, each successive transformation to build arrays used alternating markers (i.e. targeting kanMX and depositing hphMX followed by targeting hphMX and depositing kanMX) (Supplementary Fig. S1B). Plasmids were digested with NotI, and 5 μg of plasmid DNA was used per transformation. Digested plasmid was co-transformed along with 1 μg of plasmid containing Cas9 and a guide RNA targeting the site of integration to increase recombination efficiency and minimize mis-targeting [43]. Integrations were screened by growth on selective media for proper antibiotic resistance marker and no growth on selective plate for opposite marker (Supplementary Fig. S1B). Integrations were evaluated for increased expression by *GFA1* RNA fluorescence in situ hybridization (FISH) and western blotting (Supplementary Fig. S2A and B).

A cluster of *lacO* repeats was integrated adjacent to 25× gene arrays using a previously published strategy, where a *HIS3* marker is first integrated at the desired location as a landing pad and then replaced by *lacO* repeats with a *LEU2* selectable marker [44]. Repeats were integrated ∼0.15 kb upstream of *URA10* (see Supplementary Table S3 for primer sequences), with no disruption to any neighboring genes. This integration was done following *GFA1* array construction to minimize potential for recombination events leading to loss of *lacO* repeats during iterative strain construction. LacI fluorescent fusion proteins were integrated at the endogenous *HIS3* locus (nonfunctional in this strain for auxotrophic marker use) by transformation of a linearized construct (Supplementary Table S2).

Gene array imaging

Yeast growth for imaging

When grown for imaging, strains with iRFP-tagged protein had media supplemented with 10 μg/ml biliverdin throughout the duration of growth to enhance folding of iRFP and improve fluorescence. For imaging, 5 ml of YPD was inoculated from fresh grown plates and cultured for ∼6–8 h at 25°C. Cultures were then diluted into 5 ml of fresh YPD and grown overnight to reach desired OD_{600} of ∼1.0 the next morning. Dilution of slower-growing mutants (e.g. *mft1Δ)* were adjusted accordingly to account for longer doubling times. After overnight growth, cultures were then diluted to OD_{600} of 0.25 in copper-free minimal media (Formedium, CYN0901) and grown for 1–2 h before imaging. Cultures were transferred to glass bottom 384-well plate pre-coated with Concavalin A, left for ∼10 min to promote adherence, media exchanged for copper-free media without biliverdin, and finally spun down by pulsing centrifuge for 10 s. A pre-warmed (25°C) climate-controlled imaging chamber was used for imaging. Two independent biological replicate strains were imaged for each factor, and experiments were repeated at least twice as technical replicates. When replicate data were pooled, all experiments had $n > 200$ cells.

Microscopy

Imaging was conducted using a 63 × 1.47 N.A. oil objective connected to an Andor Dragonfly microscope and dual EM-CCD cameras, using the spinning disk confocal mode with 40 μM pinhole. Induction of transcription was performed by adding media with 2× concentration of inducing agent in an equal volume to that of the media in the well, reaching the desired final concentration (1 μM for β-estradiol and 500 μM for $CuSO_4$). Time series images were collected every 10 s starting 20 s post-induction for 30 total time points. For each time point, images were collected simultaneously from dual EM-CCD cameras. Images were acquired as z stacks with step size 0.2 μM for 2.5 μM total depth. Exposure of 100 ms, EM gain of 300, and Power Density 2 were used for all images. For all mNG tagged proteins and LacI-GFP, a 150 mW 488 nm excitation laser was used at 10% laser power with a 525/50 nm filter. For iRFP tagged proteins a 140 mW 637 nm excitation laser was used at 50% laser power with a 700/75 nm filter. For LacI-iRFP the same excitation laser and filter were used but with laser power at 80%.

Image processing and analysis

Time series images were deconvolved using Fusion (Oxford Instruments) software presets. Deconvolved image series were made into maximum intensity projections using FIJI. All pertinent scripts to image analysis workflow are available on

https://doi.org/10.5281/zenodo.15126956. LacI spots were tracked over time course using FIJI TrackMate plugin [45]. Trackmate parameters are detailed in supporting information on Zenodo.

Track information was exported as FIJI regions of interest (ROIs), and a jython script was run through FIJI to export images from ROIs. Shell script was used to organize image directories, sorting track images into their own folders and making merged stack of all time points. Exported cells were then analyzed for time of visual enrichment through visualization of time series stack on FIJI. Time of recruitment was defined as the first frame where enrichment is observed for at least two consecutive frames (i.e. first detected in frames 3 and 4, time of recruitment recorded as frame 3). Any cells for which enrichment was not determined for at least two frames were excluded, as well as any cells containing two LacI foci or cells where LacI moved from mother to bud during course of imaging.

mRNP-SiMPull

Yeast cells were cultured at 30°C, with growth and cryogrinding conducted as described previously [18]. Cryogrindate was used for mRNP pulldown via Cbp80, and then the tagged RBP of interest (Sub2 or Yra1) was captured on slide for TIRF imaging as described previously [18]. Spot detection was conducted using CoSMoS analysis software imscroll, and photobleaching step analysis for stoichiometry was performed using AGATHA CPS program, as described previously [18].

Spot assays

Overnight cultures of yeast cells grown in YPD were diluted to OD_{600} of 0.5 in water, and tenfold serial dilutions were prepared for spotting on YPD plates. Plates were incubated at 25°C or 37°C for ∼48 h, and representative images taken.

Fluorescence *in situ* hybridization

Sequences of *GFA1* FISH probes used in this study are listed in Supplementary Table S5. Strains were grown as described for live imaging, with ∼3.5 ODs of cells collected from log phase growth pre- and post-induction at time points indicated. FISH experiments were conducted as described previously [46]. Briefly, samples were fixed with formaldehyde overnight at a final concentration of 3%, washed with buffer B (1.2 M sorbitol, 0.1 M potassium phosphate, 0.5 mM $MgCl_2$), and the cell wall digested using 50 μg Zymolyase 20T and 5 μl of 200 mM vanadyl ribonucleoside complex (VRC) at 30°C for 30–40 min. Samples were then washed with buffer B and incubated in 70% ethanol for ∼4 h at room temperature. After washing samples with buffer B, hybridization was carried out in 50 μl mix containing 1× saline sodium citrate (SSC), 0.34 mg/ml *Escherichia coli* transfer RNA, 20% formamide, 0.2 mg/ml bovine serum albumin, 11% dextran sulphate, 4 mM VRC, with oligo-dT and *GFA1* probes at 37°C for ∼16 h. Final concentrations the Quasar 570 labeled *GFA1* probe and fluorescein isothiocyanate labeled oligo dT probe were 100 and 25 nM, respectively. After hybridization, samples were washed with 1× SSC and 15% formamide followed by application onto slides treated with poly-L-lysine. After washing with 1× phosphate bufferedsaline (PBS), mounting medium with 4′,6-diamidino-2-phenylindole (DAPI) was applied to each sample and a coverslip was affixed. All reagents for FISH were made with DEPC-treated water. Images were collected using a 60 × 1.4 N.A. oil objective connected to an Andor Dragonfly microscope and EMCCD camera. Maximum intensity projections of z-stacks were generated in FIJI.

Transcription site intensities were determined by thresholding in FIJI to generate mask of ROIs, filtering ROIs for 0.15–2.0 μm^2 and 0.25–1.0 circularity using the "Analyze Particles" function, then measuring and plotting integrated density of the ROIs. Transcription site frequency was determined by manually scoring 50 cells per image using two images per biological replicate for $n = 200$ cells per condition.

Western blotting

Cell lysates were prepared by collecting 2 ODs of cells, washing with DI water, and resuspending in 100 μl of 2× SDS buffer. Cells were lysed by bead beating and boiling at 95°C. Samples were incubated 10 min at 95°C and spun down to remove cell and bead debris before loading onto gel. Samples were run on 10% sodium dodecylsulphate–polyacrylamide gel electrophoresis (SDS–PAGE) gels for HA-Gfa1 and iRFP-Yra1 blots, 8% SDS–PAGE gel for LexA-Cbp80 blot, and 12% SDS–PAGE gel for Gal4-Yra1-RRM blot due to protein size differences. All blots were transferred to nitrocellulose membrane via wet tank transfer (95 V for 1 h) and blocked with 2.5% milk in PBS for 1 h at room temperature. All blotting was done in 2.5% milk in PBS-T. Primary antibodies were used at the following concentrations: anti-HA 1:2500, anti-Dbp5 1:5000, anti-LexA 1:1000, anti-Gal4 1:1000, and anti-Yra1 1:2500, all at 4°C overnight. Anti-mouse DyLight 650 and Anti-Rabbit 488 secondary antibodies were used at 1:10 000 at room temperature for 1 h. Full western blot images are deposited at https://doi.org/10.5281/zenodo.15786809.

RNA-sequencing

Two biological replicates each of parental and *pCUP1-GFA1*$_{25}$ strains were used for RNA collection. Briefly, 5 ml of YPD media was inoculated from fresh grown plates and grown for ∼8 h. Cultures were diluted to OD_{600} of 0.002 in 100 ml of YPD and grown overnight to reach OD_{600} of ∼1.0. Cells were harvested, washed with copper-free minimal media (Formedium, CYN0901), and resuspended into 200 ml copper-free media such that $OD_{600} = 0.25$. Cells were grown for ∼2 h, baseline samples were collected, and then cultures were induced with 500 μM $CuSO_4$. Samples were collected 5, 10, 15, 20, 25, 30, 45, and 60 min post-$CuSO_4$. Samples were snap-frozen with liquid nitrogen and kept at −80°C for RNA extraction. Samples taken in parallel at 5, 15, 30, and 60 min were processed for smFISH (conducted as described above) to confirm proper induction of *GFA1* array expression prior to sending samples for sequencing.

Total RNA was prepared using an RNA extraction kit (Zymo research), excluding the kit's DNase I treatment process. RNA was treated with Turbo™ DNase for 30 min at 37°C and cleaned by column purification (Zymo research). RNA samples were resuspended in molecular-grade water and 3′-Tag-seq library preparation was carried out using Lexogen QuantSeq kit. Samples were sequenced at the UC Davis Genome Center using a NextSeq 500 sequencer. From sequencing data, unique molecule identifier (UMI) sequences at the 5′ end were extracted and the 4 nt fixed and 12 nt random primer sequences were trimmed using UMI-tools. Adapter and

poly A tail sequences at 3′ ends were also trimmed using bbduk. The ribosomal RNA sequence reads were removed with bbsplit. Mapping of trimmed sequence reads on *S. cerevisiae* genome (GCF_000146045.2_R64) was performed using STAR. After removing PCR duplicates with UMI-tools, the read counts in each gene were determined using htseq-counts with the mode intersection-nonempty. Differential expression analysis was performed using DESeq2 with the cut-off of $Log_2FC > 1$ or < -1 and FDR < 0.001. All pertinent scripts used for RNA-sequencing analysis are available on https://doi.org/10.5281/zenodo.15127156. RNA-seq data have been deposited in GEO with accession number GSE293820. Tables with differentially expressed genes and GO terms are available in Supplementary File S1.

Statistical analyses

For live imaging, two biological replicate strains were used for each tagged factor, and experiments were run at least twice as technical replicates. When pooled cell number was >200 with exact cell (n) numbers indicated on figures. Non-parametric Kolmogorov–Smirnoff (K–S) two-sided tests were used for statistical comparisons of distributions, with *P*-values indicated on figures. mRNP-SiMPull experiments were run in triplicate with two biological replicates of each sample. For statistical comparisons of spot number, Student's *t*-test was used and *P*-value indicated on figure. For statistical comparison of stoichiometry distributions, non-parametric Kolmogorov–Smirnoff (K–S) two-sided tests were used and *P*-values indicated on figure.

Novel programs, software, and algorithms

Scripts associated with study pertaining to RNA sequencing (described earlier) are available at https://doi.org/10.5281/zenodo.15127156. Scripts pertaining to image analysis (described earlier) are available at https://doi.org/10.5281/zenodo.15126956.

Results

Construction and validation of gene arrays in *S. cerevisiae*

To generate a strong pulse of semi-synchronous transcription for visualizing the temporal recruitment of RBPs to a locus of transcription, gene arrays were built with repeating copies of the same TU. Each TU contains an inducible promoter, open reading frame (ORF), and terminator (Fig. 1A). The ORF includes a C-terminal HA tag for protein detection by western blot and two MS2 hairpins inserted in the 3′ UTR. To build gene arrays, integrating plasmids with five copies of the same TU were assembled using a Golden Gate cloning strategy (Supplementary Fig. S1A) [39]. The 5× plasmids were designed for alternating transformations with integrations targeting two different antibiotic resistance markers to build up gene array copies (Supplementary Fig. S1B). Using this strategy, gene array strains were constructed that carry twenty-five copies of a TU encoding the *GFA1* gene from *S. cerevisiae* with the promoter from the endogenous *CUP1* gene (*pCUP1*, activated in response to copper) or a synthetic promoter (Z_3EVpr, activated in response to β-estradiol) [47–49]. These two versions of the array, $pCUP1\text{-}GFA1_{25}$ and $Z_3EVpr\text{-}GFA1_{25}$, serve as a point of comparison for assembly of mRNPs with the same coding sequence under two different promoters. *GFA1* was selected because it is categorized as a "canonical mRNA" based on RBP binding profiles [50], is of average length in yeast (2.15 kb), does not have an intron (~96% of yeast genes), and was used previously to study mRNP export kinetics [51]. To allow tracking of the gene array locus independent of transcriptional activation, *lacO* repeats were integrated next to the location of the gene array for visualization using a fluorescent LacI fusion protein (LacI-FP) (Fig. 1A).

In strains harboring integrated $pCUP1\text{-}GFA1$ and $Z_3EVpr\text{-}GFA1$ arrays, transcriptional output increases with TU copy number, as demonstrated by FISH (Supplementary Fig. S2A and B). In $pCUP1\text{-}GFA1_{25}$ and $Z_3EVpr\text{-}GFA1_{25}$ arrays at 5 min post induction, robust transcription sites are seen (Fig. 1B and Supplementary Fig. S2A), with both systems showing a comparable frequency of transcription sites (Supplementary Fig. S2C), indicating robust activation of the arrays across the population. By 30 min, *GFA1* mRNA is abundant in the cytoplasm with no evidence of nuclear accumulation, confirming that transcripts generated by the arrays are exported (Supplementary Fig. S2A). Gene arrays driven by both promoters show similar *GFA1* transcript localization over the time course, one notable exception being the $Z_3EVpr\text{-}GFA1$ transcription sites persist more uniformly at 30 min, whereas $pCUP1\text{-}GFA1$ transcription sites become weaker (Supplementary Fig. S2A). This is consistent with transcriptional downregulation of endogenous *CUP1* transcription at later times points following activation by copper [52], versus the synthetic Z_3EVpr transcription factor that directly and continuously activates transcription [48]. Additionally, there is no change in the localization of total polyadenylated RNA by dT FISH at any time point post-induction, indicating array activation does not impact bulk nuclear mRNA export (Supplementary Fig. S2A). Using an antibody for the HA epitope to detect Gfa1p produced specifically from the array, protein expression is observed following transcriptional activation, confirming translation of array-produced transcripts (Supplementary Fig. S2D). Spot growth assays show there are no associated growth defects with the integration of the $pCUP1\text{-}GFA1_{25}$ or $Z_3EVpr\text{-}GFA1_{25}$ arrays with *lacO* (Supplementary Fig. S2E), indicating strain fitness is not significantly changed.

To assess potential impact of gene array induction on the transcriptome, a bulk RNA sequencing (3′-TagSeq) time course was performed comparing $pCUP1\text{-}GFA1_{25}$ cells to wild type up to an hour post-induction (Supplementary Fig. S3). At each time point, *GFA1* and *LEU2* were the most highly upregulated genes, consistent with induced *GFA1* expression from the array and use of *LEU2* as an auxotrophic marker in gene array containing strains. There were only 75 differentially expressed genes (out of >6000 total genes) shared between two or more time points (34 up- and 41 downregulated; see Supplementary Fig. S3A and File S1), indicating induction of the gene array does not have a large or specific impact on cellular transcription. Taken together, these experiments demonstrate the fidelity of the constructed gene array strains for tightly controlled, robust transcriptional activation leading to productive export and translation without acute detrimental effects on the cell.

To determine if a protein of interest could be detected at a gene array upon transcriptional induction, $pCUP1\text{-}GFA1_{25}$ strains with the CBC subunit Cbp20 or RNA polymerase II subunit Rpb7 endogenously tagged with mNeonGreen

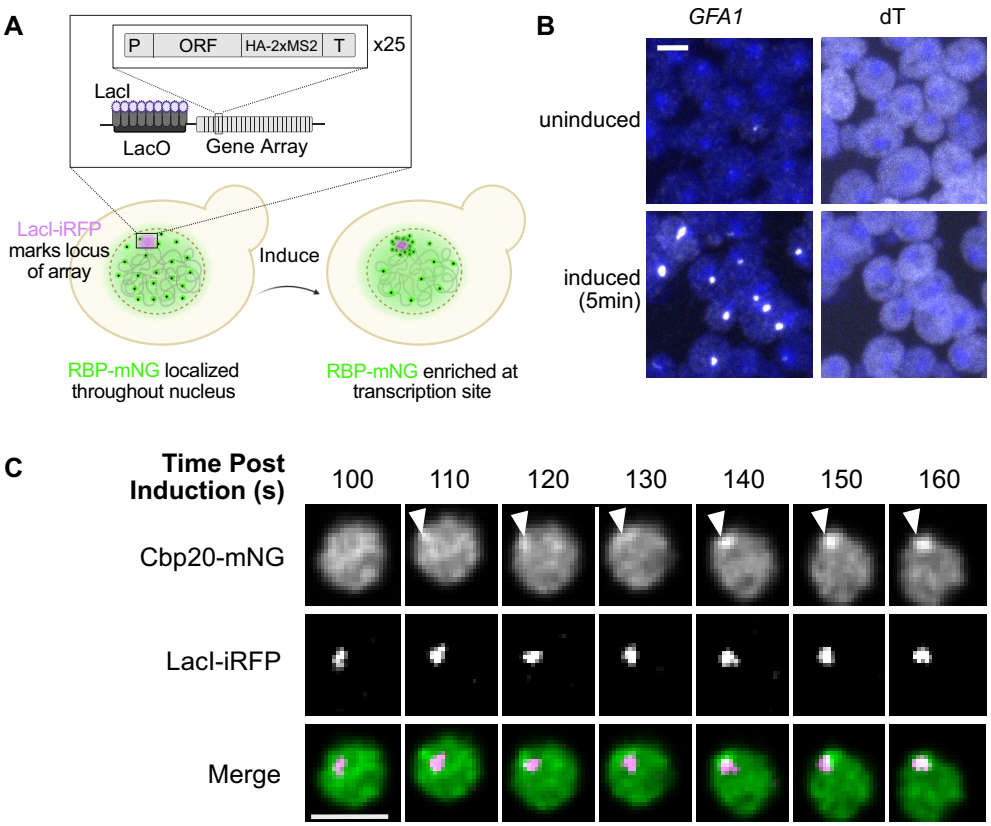


Figure 1. "Gene recruitment assay" for live imaging of co-transcriptional protein recruitment in *S. cerevisiae*. (**A**) Gene arrays are integrated into yeast genome with 25 copies of TU next to LacO repeats for visualization with fluorescent LacI reporter (LacI-FP). When transcription of array is induced, strong initial pulse leads to enrichment of endogenously tagged protein of interest at the locus, which can be visualized colocalizing with LacI. (**B**) Representative FISH at 5 min post induction of *pCUP1-GFA1*$_{25}$. Probe for *GFA1* shows array transcription sites (left) and dT probe shows bulk polyA localization (right). Probes shown in grayscale and DAPI in blue. Scale bar is 2 μM. (**C**) Representative image of protein (Cbp20-mNG) recruitment to array transcription site. Arrows indicate enrichment colocalizing with LacI foci. Scale bar is 2 μM.

(mNG) were generated. Following array induction, protein colocalization with LacI-iRFP is observed for both factors (Fig. 1C and Supplementary Fig. S4A). A 2-min pre-treatment of cells with the transcriptional inhibitor thiolutin [53] reduced detectable enrichment of Rpb7-mNG with Laci-iRFP from 75 ± 9% to 32 ± 5% of cells at 200 s post induction, demonstrating a transcriptional dependence for enrichment of Rpb7-mNG with the gene array (Supplementary Fig. S4B).

With successful visualization of Cbp20-mNG and Rpb7-mNG recruitment to *pCUP1-GFA1*$_{25}$, it was determined if additional factors with diverse roles in nuclear mRNP metabolism would also be visibly recruited to an activated gene array (Supplementary Fig. S5). These factors can be grouped into general categories based on function and presumed timing of recruitment: (i) transcription initiation and 5′ capping, (ii) transcription elongation, splicing, and mRNP packaging, and (iii) cleavage and polyadenylation (Supplementary Fig. S5). Factors were chosen to represent the processes and protein complexes acting across co-transcriptional mRNA processing and packaging for mRNA export. Selected factors were endogenously tagged with mNG or iRFP for visualization, with the tagged version being the sole source of the gene product. Fitness of tagged strains was assessed by spot assays at 25°C and 37°C with only one (Npl3-iRFP) of the 15 modified strains showing an observed growth defect at 25°C, while all others had no observable growth defects at either temperature (Supplementary Fig. S6); however, it remains an unavoidable caveat of this system that tagging may impact protein function in a manner not observed at the level of growth that alters transcriptional activation or protein recruitment timing. Of these factors, 13 of the 15 showed enrichment at *pCUP1-GFA1*$_{25}$ 5 min post-induction, with representative images showing localization of each tagged factor in uninduced and induced cells shown in Supplementary Fig. S4.

As expected, one factor that did not localize to *pCUP1-GFA1*$_{25}$ was the U2 splicing factor Snu17, which is consistent with the lack of an intron in *GFA1*. The second factor that did not localize with the gene array was the nuclear RNA exosome component Rrp6 (Supplementary Fig. S4D). The lack of Rrp6 enrichment indicates there is not a high level of surveillance and decay occurring at *pCUP1-GFA1*$_{25}$ when activated, consistent with *GFA1* export (Supplementary Fig. S2A). While intended for use in live imaging, the included MS2 hairpins did not produce reliable enrichment of the MS2-CP, which may be due to the low number of hairpins, inaccessibility of loops in the nascent RNA conformation, and/or the rapid transcript release following transcription of the 3′ UTR. Overall, these data demonstrate both specificity (i.e. not all nuclear factors associated with mRNA metabolism are recruited to an activated gene array) and a broad utility (i.e. factors across processing and mRNP packaging can be observed) for gene arrays in characterizing co-transcriptional events in live cells.

Recruitment kinetics of transcription, mRNA biogenesis, and export factors to gene arrays

With the establishment of the gene array and imaging methodology, recruitment kinetics of transcription, mRNA biogenesis, and export factors to *pCUP1-GFA1*$_{25}$were determined using an imaging time series. Images were collected starting 20 s after addition of copper and then taken every 10 s thereafter for 5 min. Imaging was performed using a dual-camera setup to simultaneously image the position of the *lacO* repeats (LacI-FP focus) as a marker of gene array location and the tagged factor of interest. Automated tracking followed the LacI-FP focus across the imaging time course within each single cell; cells with a discontinuous LacI track were excluded from data analysis. A time of recruitment was determined for every cell analyzed, defined as the first time at which visual enrichment of the tagged factor was detected at the gene array in two consecutive imaging frames. Detection and calling of enrichment at gene array loci could not be automated due to challenges with signal to noise between the overall nuclear signal and gene array focus, but development of such image analysis pipelines using current data as a training set is a future goal. Arrival times extracted from hundreds of individual cells were used to generate population level distributions of recruitment timing for each characterized factor.

Across the factors analyzed, different distributions emerged, with recruitment events predominantly taking place in the first 200 s for all factors examined (Supplementary Table S6). In line with expectations for a transcription factor, Cup2 shows the earliest and tightest distribution of all factors with median recruitment time of 40 s and the upper quartile of recruitment times at 60 s, indicating >75% of events occur in the first minute post–copper addition (Supplementary Table S6 and Fig. 2A). Rpb7, a subunit of RNA polymerase II, shows a recruitment distribution shifted slightly later than that of Cup2 with median arrival of 50 s (Fig. 2A). In contrast, Hrp1—part of the Cleavage Factor 1 complex that binds nascent RNA upstream of the cleavage site to direct cleavage and 3′ processing [54–56]—exhibited the latest median arrival (100 s) with the widest distribution (Fig. 2A and Supplementary Table S6). Arrival data was additionally visualized as cumulative distribution curves, which highlights the different recruitment behaviors for these factors, in agreement with their known functions (Fig. 2B). These data indicate the transcriptional activation of the gene array produces an initial wave of transcription from which recruitment can be observed and an order of events established. With these data reflecting known transcription events as benchmarks, time series data for the other factors were collected using the same imaging parameters, which are grouped by known functions and described below.

Transcription initiation and 5′ capping factors

Once RNA polymerase II starts transcribing, capping enzymes synthesize a 5′ m_7G cap on the growing RNA that the nuclear CBC engages [57, 58]. CBC subunits Cbp80 and Cbp20 bind the cap synergistically, with Cbp20 contacting m_7G [58, 59]. Cbp80 stabilizes binding and acts as an interaction platform for other processing factors [58, 59]. For example, data suggests interaction of Cbp80 and Npl3 promotes proper transcription, with Rai1 also binding Npl3 to recruit the exonuclease for 5′ to 3′ degradation in the case of improper capping [60, 61]. Of these factors, Cbp80 exhibits early recruitment with a median time of 40 s, the same as Cup2 (Supplementary Table S6 and Fig. 3A), while Cbp20 shows a distribution shifted later than that of Cbp80, with median arrival of 60 s and upper quartile of 100 s (Supplementary Table S6 and Fig. 3A). This suggests that Cbp80 can be recruited independently of Cbp20 to the transcription site, and that Cbp20 alone or in complex with another copy of Cbp80, is recruited after initial Cbp80 recruitment. Npl3 exhibits a median recruitment time of 60 s and upper quartile of 90 s, very similar to Cbp20 (Fig. 3A). Likewise, Rai1 exhibits a distribution similar to Npl3 and Cbp20 (Supplementary Table S6 and Fig. 3A). These data identify Cbp80 as a pioneering mRNP component that is recruited during transcription initiation prior to Cbp20, Npl3, or Rai1.

Transcription elongation and mRNP packaging factors

Mechanisms of mRNP packaging and compaction are not well elucidated, but recent work highlights the importance of TREX components THO, Sub2, and Yra1 in these events [14, 15, 18]. Models suggest THO binding during transcription elongation facilitates recruitment of Sub2 and Yra1 [19, 34]. As such, THO subunits (Mft1, Hpr1), Sub2, and Yra1 were all analyzed for recruitment timing to *pCUP1-GFA1*$_{25}$. THO subunits, Hpr1 and Mft1, showed similar distributions to one another, with median recruitment of 50 and 60 s, respectively, which is distinct from Cup2 and Hrp1 (Supplementary Table S6 and Fig. 3B). These distributions are very similar to Cbp20, Rai1, and Npl3, suggesting these components are all recruited within a short window of time. Likewise, the DEAD-box ATPase Sub2 that associates with the THO complex exhibits median recruitment of 60 s, like Mft1, with a cumulative distribution curve that closely matches both THO components. In contrast, Yra1 shows an unexpectedly early recruitment, with a median arrival of 40 s, the same as Cup2 (Supplementary Table S6 and Fig. 3B). The distribution is also tighter than THO components, with the upper quartile at 70 s, the same as Cbp80 with only Cup2 being earlier (Supplementary Table S6 and Fig. 3B). This difference is also apparent in the Yra1 cumulative distribution being closer to Cup2 and not overlapping with other TREX components, which implies initial Yra1 recruitment to the transcription site is not mediated by the THO complex or 3′ processing machinery [32–34, 62]. These data identify Yra1, like Cbp80, as a pioneering mRNP component that is among the earliest recruited to the transcription site.

Cleavage and polyadenylation factors

At the end of the transcription cycle, the multiprotein cleavage and polyadenylation factor and cleavage factor complexes bind and cleave the 3′ end of the nascent RNA, the poly(A) polymerase adds a non-templated poly(A) tail, and RNA Pol II terminates downstream of the poly(A) site [63]. The poly(A) tail is bound by the nuclear poly(A) binding protein Nab2, which is also important for determining poly(A) tail length and proper mRNA export [64]. As described above, Hrp1 exhibits the latest recruitment distribution, in line with its role in 3′ cleavage. Yet, counter to expectations of timing based on binding to the poly(A) tail, Nab2 shows a median

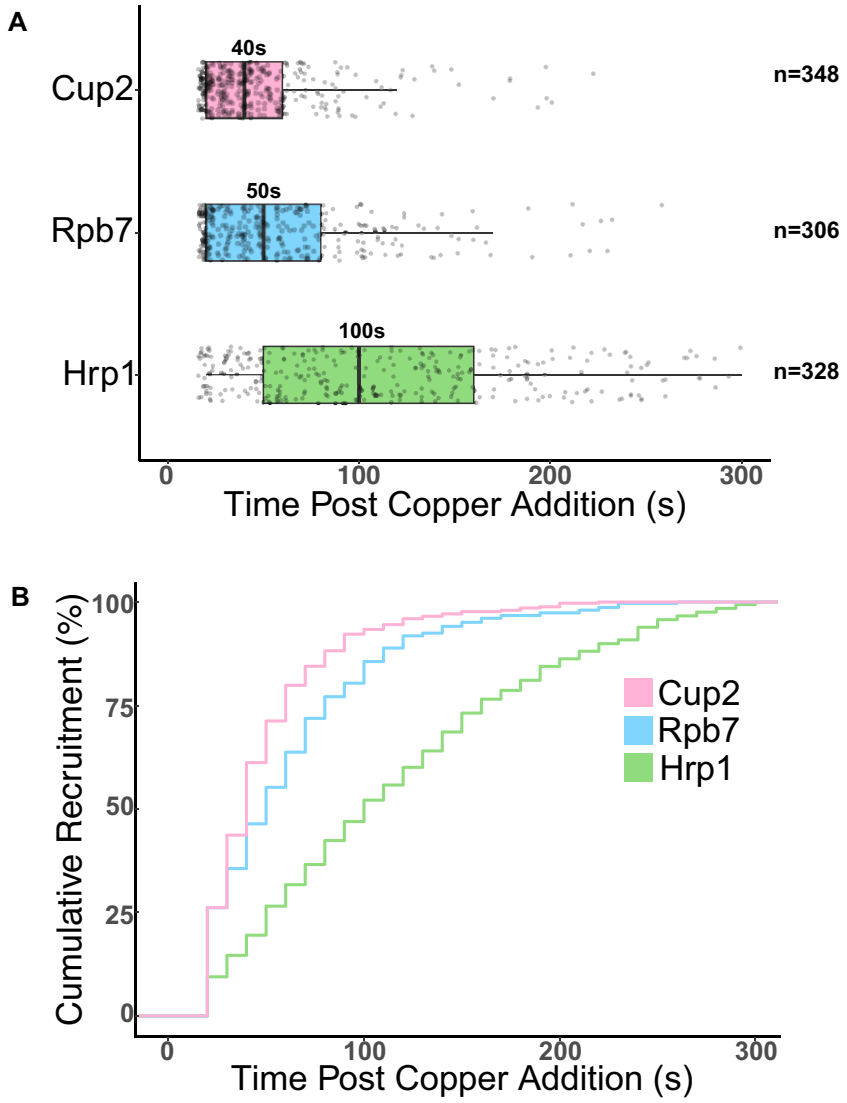


Figure 2. Framework of protein arrival to *pCUP1-GFA1*$_{25}$. (**A**) Distribution of arrival times for transcription factor (Cup2), RNA polymerase II subunit (Rpb7), and cleavage factor (Hrp1). Each dot represents an individual cell with points jittered for visualization. Box represents interquartile range (IQR) from 25th percentile to 75th percentile, line in box represents median, and whiskers represent ±1.5×IQR. Median value is also displayed above box and whiskers. Data are from two biological replicates collected over two technical replicate experiments each. Cell numbers are indicated on the graph. (**B**) Data from panel (A) represented as cumulative distribution curves where the cumulative % of cells displaying recruitment of total cells is plotted at each time point (10 s intervals).

arrival of 50 s, which is half that of Hrp1 (Supplementary Table S6 and Fig. 3C). A previous transcriptome-wide analysis of mRNA binding sites showed a strong 5′ signal for both Nab2 and Hrp1, which was linked to the surveillance of products arising from premature transcription termination [50]. The data presented here demonstrates early Nab2 recruitment that is not matched by Hrp1; furthermore, enrichment of the nuclear RNA exosome component Rrp6 is not observed (Supplementary Fig. S4), suggesting there are not high levels of surveillance at the gene array. Taken together, this provides evidence of Nab2 arrival at the transcription site prior to cleavage and polyadenylation and suggests early Nab2 recruitment is not solely linked to premature termination. These findings highlight the benefit of locus-specific temporal data, as in this instance different co-transcriptional recruitment trends are observed for two factors with very similar binding patterns at the transcriptome level [50].

Characterization of YHR127W/Yhs7

While the arrival time of factors with well characterized roles in mRNP processing and packaging provides insight into a temporal order of events, the same arrival information can be used to infer possible functions of factors with uncharacterized roles in mRNP biogenesis. One such factor is Yhs7 (YHR127W), which was previously found to be associated with nuclear mRNP components and nuclear pores by mass spectrometry [65, 66]. Recently published crosslinking mass spectrometry (XL-MS) in *S. cerevisiae* further supports Yhs7 inclusion in nuclear mRNPs with extensive crosslinking to Yra1 [14]. Furthermore, Yhs7 was found to be structurally similar to, and have RNA annealing capacity like, Yra1 [14]. This motivated live imaging with Yhs7 to determine timing of recruitment to *pCUP1-GFA1*$_{25}$. Strikingly, the observed recruitment distribution of Yhs7 is nearly identical to that of Yra1 with the same median and upper

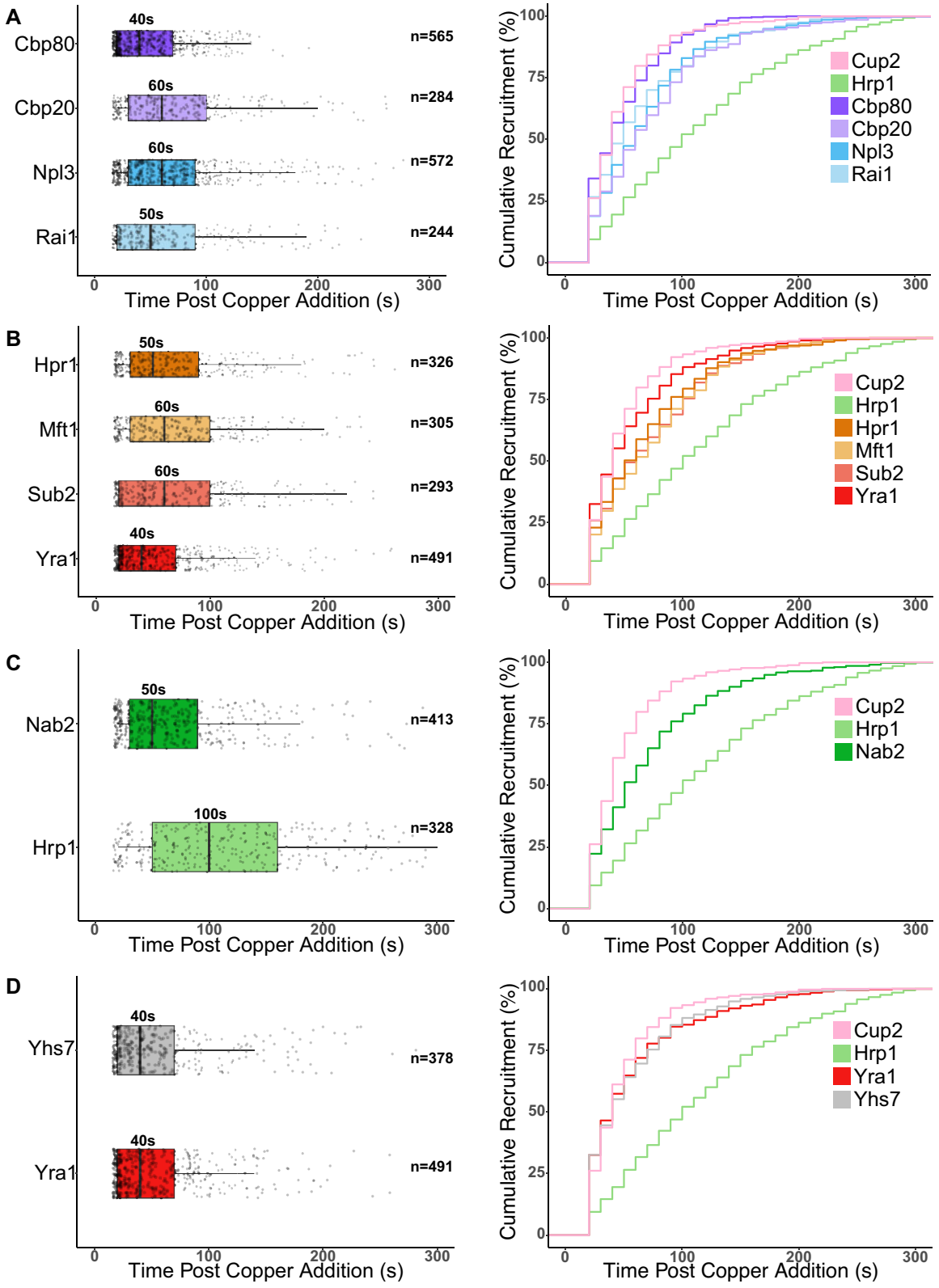


Figure 3. Recruitment times of diverse mRNP processing factors to *pCUP1-GFA1*$_{25}$. Arrival times of (**A**) transcription initiation and 5′ capping factors, (**B**) transcription elongation and mRNP packaging factors, (**C**) cleavage and polyadenylation factors, and (**D**) YHR127W/Yhs7. Data are plotted as box-and-whisker plot (left), with each dot representing an individual cell with points jittered for visualization. Box represents IQR from 25th percentile to 75th percentile, line in box represents median, and whiskers represent ±1.5×IQR. Median value is also displayed above box and whiskers. Data are also represented as a cumulative distribution (right), where the cumulative % of cells displaying recruitment of total cells is plotted at each time point (10 s intervals). Data come from two biological replicates collected over two technical replicate experiments each. Cell numbers are indicated on the graph.

quartile (Supplementary Table S6 and Fig. 3D). Given the limited published data on Yhs7, there was little expectation for how it functions in the biogenesis of mRNPs beyond associating in the nucleus. This recruitment data strongly suggests it is a pioneering mRNP component functioning early in mRNP assembly and packaging, along with Yra1 to which it is in proximity within mRNPs.

Recruitment timing is similar with an engineered promoter

With the unexpected recruitment times for a subset of factors, there is a possibility that aspects of recruitment are due to a pathway of co-transcriptional assembly unique to copper-responsive genes. As such, select factors were tested for recruitment timing using a second gene array, *Z$_3$EVpr-GFA1*. This array is driven by an engineered zinc-finger transcription factor responsive to β-estradiol [48], which is not subject to the same upstream regulation as *pCUP1* [52, 67–69]. Arrival data is in general agreement between the two promoters (Supplementary Table S6 and Supplementary Fig. S7). Given the data is not normally distributed, a non-parametric Kolmogorov–Smirnoff (K–S) two-sided test was deployed to test differences in distribution (Supplementary Fig. S7). The K–S test is sensitive to any differences in distribution, and some *P*-values reach significance, but visual comparison of the distributions shows an overall close resemblance. As such, there may be differences in recruitment timing to the two arrays, but the general trends found with *pCUP1-GFA1$_{25}$* do not appear unique to copper-responsive genes. This suggests the observed temporal framework of recruitment provides generalizable insights into co-transcriptional events and mRNP packaging.

Loss of THO complex function differentially impacts Yra1 and Sub2

Given the observed recruitment differences between Yra1 and other TREX components Sub2 and THO, the relationship between these components in recruitment timing and stability within mRNPs was more closely examined. First, the THO complex mutant *mft1Δ*, which disrupts THO dimerization *in vivo* [20, 70], was used to examine the role of the THO complex in Yra1 and Sub2 recruitment to a transcription site. Deletion of *MFT1* did not display a fitness defect at 25°C in iRFP-Yra1 imaging strains, as shown in Supplementary Fig. S8A. In line with recruitment data placing Mft1 later than Yra1 (Fig. 3B), *mft1Δ* did not impact the timing of Yra1 recruitment to *pCUP1-GFA1$_{25}$* (Fig. 4A). In contrast to iRFP-Yra1, when *mft1Δ* is combined with Sub2-mNG, the resulting strains displayed fitness defects at 25°C (Supplementary Fig. S8B), including cell cycle and nuclear morphology abnormalities that led to an increased frequency of cell exclusion from analysis (e.g. two LacI-FP spots in a cell). Still, the timing of Sub2 recruitment was unchanged in *mft1Δ* for cells analyzed (Fig. 4B), indicating a fully functional THO complex is not necessary for Sub2 recruitment to the *pCUP1-GFA1$_{25}$* gene array. This data demonstrates that both Yra1 and Sub2 can be recruited to a transcription site independently of a fully functional THO complex with a timing that is unchanged.

Next, it was considered if the loss of THO complex function impacts downstream loading of Yra1 and Sub2 into mRNPs, which is expected to occur following recruitment to the transcription site. These data were generated by mRNP-SiMPull, a single molecule imaging method that characterizes protein occupancy and stoichiometry within purified mRNPs [18]. This data shows Cbp80-isolated mRNPs in *mft1Δ* cells contain Yra1 at the same frequency as wildtype (Fig. 4C and Supplementary Fig. S9); however, copy number of Yra1 within these mRNPs is reduced (Fig. 4D). These results are consistent with the impact of another THO complex mutant, *tho2Δ*, on Yra1 stoichiometry in mRNPs [18]. In contrast to Yra1, mRNP-SiMPull experiments showed the fraction of Cbp80-isolated mRNPs that contain Sub2 is reduced by ∼70% in *mft1Δ* compared to control (Fig. 4E). In addition, for those minority of mRNPs in *mft1Δ* cells that still contain Sub2-mNG, Sub2 copy number significantly decreases (Fig. 4F). Together, these data indicate that the THO complex is not required for the recruitment of Yra1 or Sub2 to the transcription site, nor the frequency of Yra1 association with mRNPs. Rather, the THO complex appears to [1] play a central role in the ability of Sub2 to stably engage mRNPs and [2] promote mRNP structures that contain multiple copies of both Yra1 and Sub2. These findings align well with recent studies placing Yra1/ALYREF at the mRNP core with a role in RNA packaging [14, 15, 18], while Sub2 engages mRNPs and Yra1/ALYREF in a manner that is regulated by THO as part of the mRNP maturation pathway leading to export [30, 31].

Determinants of Yra1 recruitment

THO complex disruption did not impact recruitment of Yra1 to the transcription site, so further work was performed to identify factors important to early Yra1 recruitment. This involved deletion of both *trans* acting factors and domains within Yra1 to assess impact on Yra1 recruitment to the *pCUP1-GFA1$_{25}$* gene array (fitness of mutants shown in Supplementary Fig. S8A). First, post-translational modifiers of Yra1 (i.e. ubiquitination by Tom1 and methylation by Hmt1) were examined [71, 72]; however, neither a *tom1Δ* or *hmt1Δ* significantly altered Yra1 recruitment timing (Fig. 5A). In contrast, deletion of other identified pioneering factors Cbp80 and Yhs7 delay Yra1 recruitment with the medians and upper quartiles of recruitment distribution shifted 10 and 20–30 s later, respectively (Fig. 5A). These mutants do not show reduced transcription site intensity via *GFA1* FISH to control cells, suggesting similar strength of array transcription (Supplementary Fig. S8C and D). Of note, there is a reduction in the overall population of *cbp80Δ* cells with detectable transcription sites at 5 min, suggesting reduced response across cells population-wide (Supplementary Fig. S8E). Together, these data suggest a relationship amongst the pioneering factors Yra1, Yhs7, and Cbp80 that may involve establishing an environment conducive to efficient mRNA processing and mRNP packaging.

In terms of possible Yra1 domain(s) directing co-transcriptional recruitment, Yra1 is a 227 amino acid hnRNP-like protein in the RNA and export factor (REF) family consisting of a central RRM domain, helical N- and C-terminal boxes (N-box, C-box), and two positively charged disordered regions (N-vr, C-vr) (Fig. 5B) [73]. The RRM domain is not required for RNA binding, while the N-vr and C-vr confer RNA annealing capacity [14, 74]. The N-vr region is also shown to interact with the phosphorylated RNA Pol II CTD [75], while the C-box contacts Sub2 in yeast mRNPs [14]. Structural data further supports both box motifs

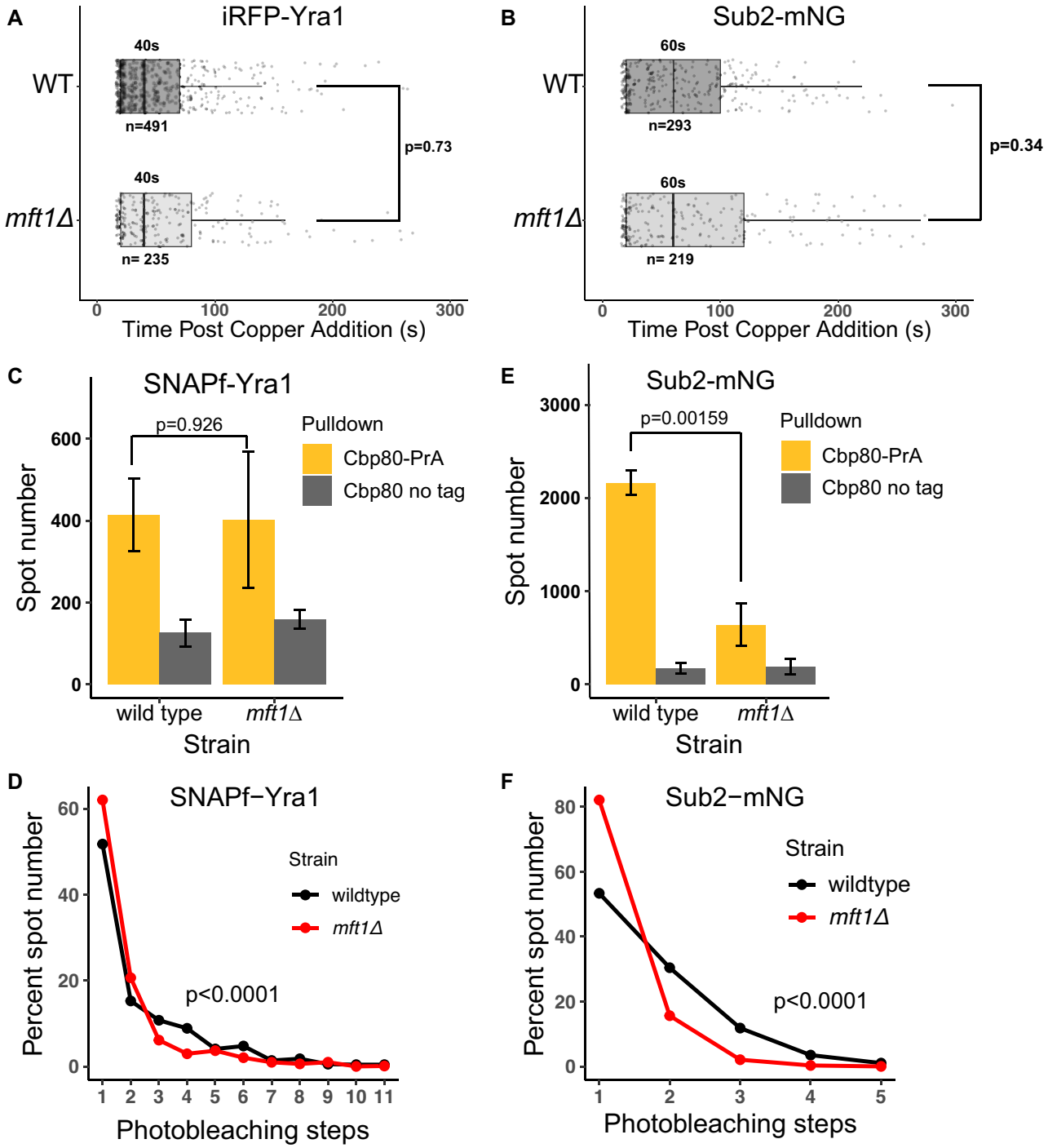


Figure 4. Impacts of THO complex subunit deletion on TREX component co-transcriptional recruitment and mRNP composition. Distribution of arrival times for Yra1 (**A**) and Sub2 (**B**) to *pCUP1-GFA1*$_{25}$ are not changed in *mft1Δ*. For box plots, each dot represents an individual cell with points jittered for visualization. Box represents IQR from 25th percentile to 75th percentile, line in box represents median, and whiskers represent ±1.5×IQR. Median value is also displayed above box and whiskers. Data come from two biological replicates collected over two technical replicate experiments each. Cell numbers are indicated on the graph. *P*-values are calculated by a non-parametric Kolmogorov–Smirnoff (K–S) two-sample test. Bar graphs show mean detected spot number by mRNP-SiMPull from triplicate experiments for SNAPf-Yra1 (**C**) and Sub2-mNG (**E**) comparing wild type to *mft1Δ* in Cbp80-PrA pull-down samples. Cbp80 no tag serves as negative control showing background spot levels. *P*-values are calculated by Student's *t*-test. Line graphs show uncorrected raw mean photobleaching step data from triplicate mRNP-SiMPull experiments for SNAPf-Yra1 (**D**) and Sub2-mNG (**F**) comparing wild type to *mft1Δ* in Cbp80-PrA pull-down samples. *P*-values are calculated by a non-parametric Kolmogorov–Smirnoff (K–S) two-sample test.

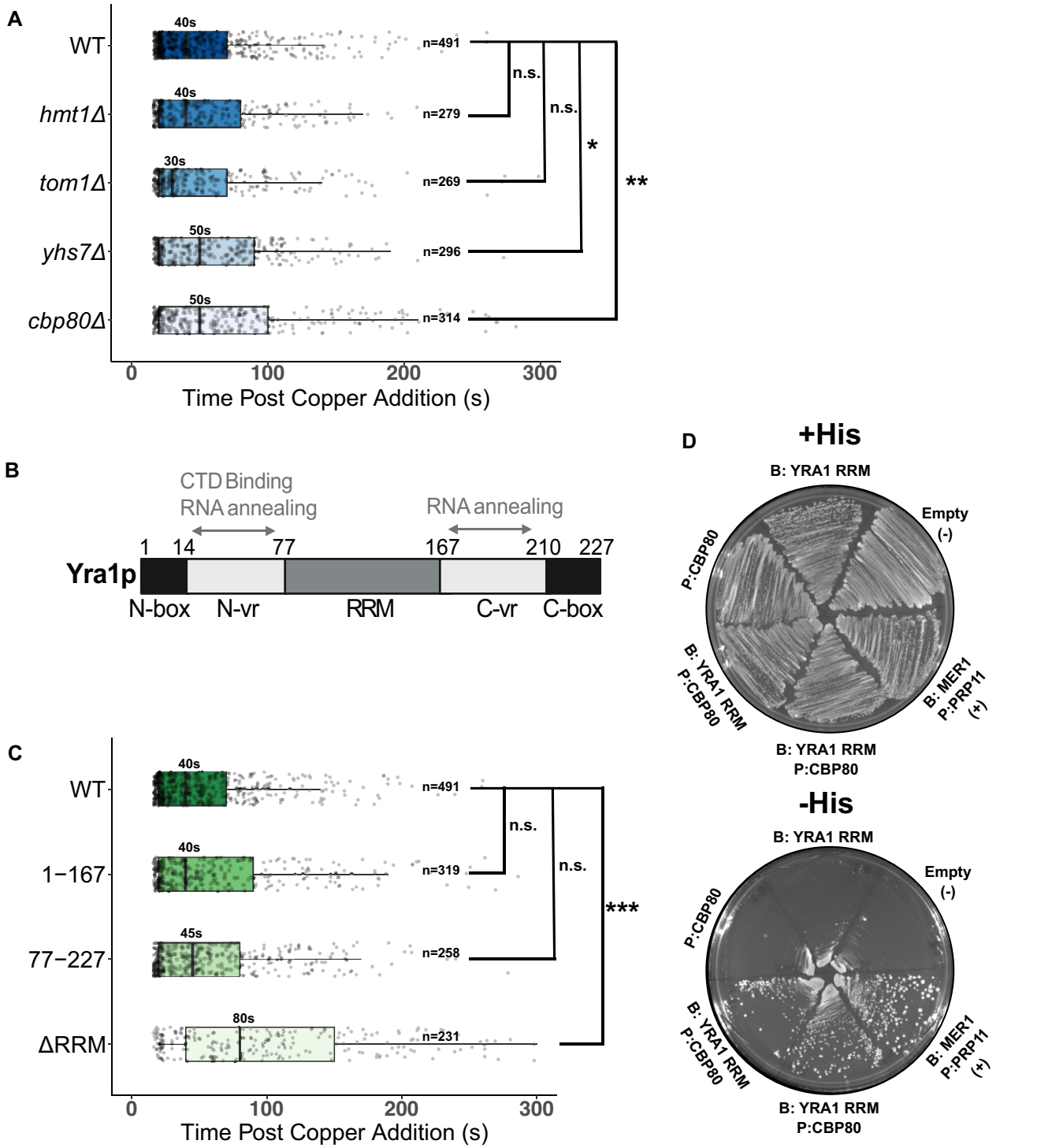


Figure 5. Impacts of deletions and truncations on Yra1 recruitment to *pCUP1-GFA1*$_{25}$. (**A**) Recruitment timing of iRFP-Yra1 in indicated gene deletion mutants. (**B**) Schematic of Yra1 protein domain organization with the regions indicated with arrows that have C-terminal domain (CTD) binding and/or RNA annealing capabilities shown in previous literature. (**C**) Recruitment timing of iRFP-Yra1 truncations. For box plots, each dot represents an individual cell with points jittered for visualization. Box represents IQR from 25th percentile to 75th percentile, line in box represents median, and whiskers represent ±1.5×IQR. Median value is also displayed above box and whiskers. Data come from two biological replicates collected over two technical replicate experiments each. Cell numbers are indicated on the graph. *P*-values are calculated by a non-parametric Kolmogorov–Smirnoff (K–S) two-sample test and $*P = .04$, $**P = .006$, $***P = 2.32 \times 10^{-13}$. (**D**) Yeast two-hybrid interaction of Yra1 RRM and Cbp80. Bait (**B**) and Prey (**P**) constructs contained in strain marked next to plates. Growth on + His (top) represents growth of strains without selection for two-hybrid interaction. Growth on -His (bottom) selects for only strains that activate HIS3 locus expression, indicative of positive two-hybrid interaction. Plate growth conducted at 30°C and images taken on fourth day after plating.

being involved in the Yra1-Sub2 binding based on ALYREF-UAP56 interactions [15]. To test the contribution of Yra1 features in co-transcriptional recruitment, truncations were made that removed the C-terminal region after the RRM (contains residues 1–167), the N-terminal region until the RRM (contains residues 77–end), or the RRM itself (contains 1–77 and 167–end); henceforth referred to as $Yra1_{1–167}$, $Yra1_{77–227}$, and $Yra1_{\Delta RRM}$. Truncations fused to iRFP were integrated at the endogenous *YRA1* locus in *pCUP1-GFA1*$_{25}$strains, which were each viable with the truncation as the sole source of Yra1 (fitness shown in Supplementary Fig. S8A and protein expression in Supplementary Fig. S8F), consistent with previous findings [74].

Both $Yra1_{1-167}$ and $Yra1_{77-227}$ exhibit no change in recruitment timing (Fig. 5C), demonstrating neither the N-box/N-vr or C-box/C-vr regions are required for early recruitment. This also suggests early recruitment is not dependent on Yra1 binding to the phosphorylated CTD, but instead likely precedes it, given that the N-terminal region is important for CTD binding [75]. In contrast, the arrival of $Yra1_{\Delta RRM}$ is significantly delayed, with the median and upper quartile of recruitment distribution shifted 40 and 80 s later, respectively (Fig. 5C). $Yra1_{\Delta RRM}$ cells exhibit a robust transcriptional response upon copper addition (Supplementary Fig. S8C and E) with transcription sites of slightly reduced intensity compared to control (Supplementary Fig. S8D), suggesting the RRM may also contribute to the efficiency of transcription and/or protection of transcripts from decay.

In XL-MS data the Yra1 RRM domain shows strong crosslinks with Ysh7 and other mRNP protein components, as well as extensive self-crosslinks, suggesting it is a hub of protein–protein interactions within mRNPs [14]. This suggests early Yra1 recruitment is driven by protein–protein interactions, not RNA binding/annealing. In further support of a functional relationship between Cbp80 and Yra1, the mammalian ortholog of Yra1 (ALYREF) is known to bind the CBC, with recent data elucidating the structure of CBC and ALYREF bound to a cap analog via cryo-EM [76, 77], with the Cbp80 ortholog NCBP1 alone being sufficient to interact with ALYREF *in vitro* [77]. While published data does not report a direct interaction of Yra1 and CBC in *S. cerevisiae*, Yra1 occupancy at gene loci measured by ChIP is diminished in a *cbp80Δ* strain, while THO occupancy is unchanged [78]. Given this published data and delayed recruitment of Yra1 in the *cbp80Δ* and $Yra1_{\Delta RRM}$ mutants, a possible interaction between Yra1 RRM domain and Cbp80 was interrogated by yeast two-hybrid. To do so, the LexA DNA-binding domain was fused to Yra1 RRM (bait) and the Gal4 activating domain to Cbp80 (prey), and expression confirmed by western blot (Supplementary Fig. S10). When tested, a positive two-hybrid interaction was observed through *HIS3* reporter gene activation as evidenced by growth on media lacking histidine (Fig. 5D), which was similar in strength to a previously reported positive yeast two-hybrid interaction between splicing components Mer1 and Prp11 [79]. This suggests the RRM domain of Yra1 is sufficient to interact with Cbp80 in the context of a yeast two-hybrid assay. Altogether, the delayed arrival of Yra1 in *cbp80Δ* and $Yra1_{\Delta RRM}$, combined with a Yra1 RRM–Cbp80 two-hybrid interaction, suggest a model where early co-transcriptional recruitment of Yra1 involves protein contacts including Cbp80, supporting functional conservation of the interaction observed with recombinant mammalian proteins [77].

Discussion

This work employs a live cell imaging approach to investigate co-transcriptional recruitment of processing factors and RNA-binding proteins central to mRNP assembly and function. The approach employs a synchronized burst of transcription to detect recruitment of target proteins to an active transcription site (Fig. 1A), as previously done in mammalian cells [35–37]. Here, for the first time, this is adapted for live cell imaging in *S. cerevisiae,* an extremely tractable system from a genetic engineering standpoint [80]. The modular nature of gene array construction allows exchange of gene features (e.g. promoter, intron, UTRs) for direct comparison of arrays with differing characteristics to assess their impacts on co-transcriptional assembly, as done here for two promoters. In addition, previously used stable mammalian cell lines with tandem arrays were generated using a stable transfection technique leading to random insertions, so location and copy number were not controlled, whereas this system allows for targeted integration at a chosen locus in an iterative fashion to build arrays in a controlled manner (Supplementary Fig. S1). Previous work used antibodies or overexpression of a tagged protein of interest to address localization, whereas here the sole copy of proteins of interest are endogenously tagged. This genetic tractability has allowed for exploration of a wide variety of factors using gene arrays with the same gene (*GFA1*) controlled by different inducible promoters (*pCUP1* and *Z$_3$EVpr*) across hundreds of individual cells for each condition. For factors assessed with both promoters, recruitment trends are consistent, suggesting observed timings are not unique to copper-responsive genes and likely define a more generalizable pattern of transcription site dynamics (Fig. 6A). Of course, this may not be the case for all genes, particularly in cases where generated mRNPs have unique characteristics related to gene expression regulation (e.g. translational repression or subcellular localization), which this method could be used to explore.

The observed framework of protein arrival to *pCUP1-GFA1*$_{25}$ shows the *pCUP1* transcription factor (Cup2) having the earliest median arrival and cleavage factor Hrp1 the latest median arrival, serving as useful benchmarks relative to other processing factors (Fig. 2). 5′ processing factors arrived after Cup2, with distributions shifted slightly later than that of the RNA Pol II subunit Rpb7 (Fig. 3A). A notable exception being Cbp80, which showed a distribution closer to Cup2 (Fig. 3A). Cbp80 recruitment prior to the other CBC subunit Cbp20 suggests Cbp80 is first recruited independently of Cbp20, counter to a model of CBC always acting as a dimer. Individual CBC component deletions are viable in *S. cerevisiae* [81], indicating they can carry out some functions independently. The 3′ processing cleavage factor Hrp1 showed the latest arrival, while the nuclear poly(A) binding protein Nab2 is shifted earlier. Knowing Nab2 interacts with Yra1 within mRNPs [14] and plays a role in condensing RNA *in vitro* [82], these data indicate a role for Nab2 upstream of cleavage and polyadenylation in co-transcriptional packaging of mRNPs.

TREX subunits, known to play roles in transcription elongation, mRNP packaging, and nuclear export, generally show recruitment trends similar to 5′ processing factors (Fig. 3B). A noted exception being the early arrival of Yra1, which exhibits a distribution closer to Cbp80 and Cup2 (Fig. 3B). This early arrival of Yra1 was surprising, given its prescribed role in licensing mRNPs for nuclear export. Furthermore, this is

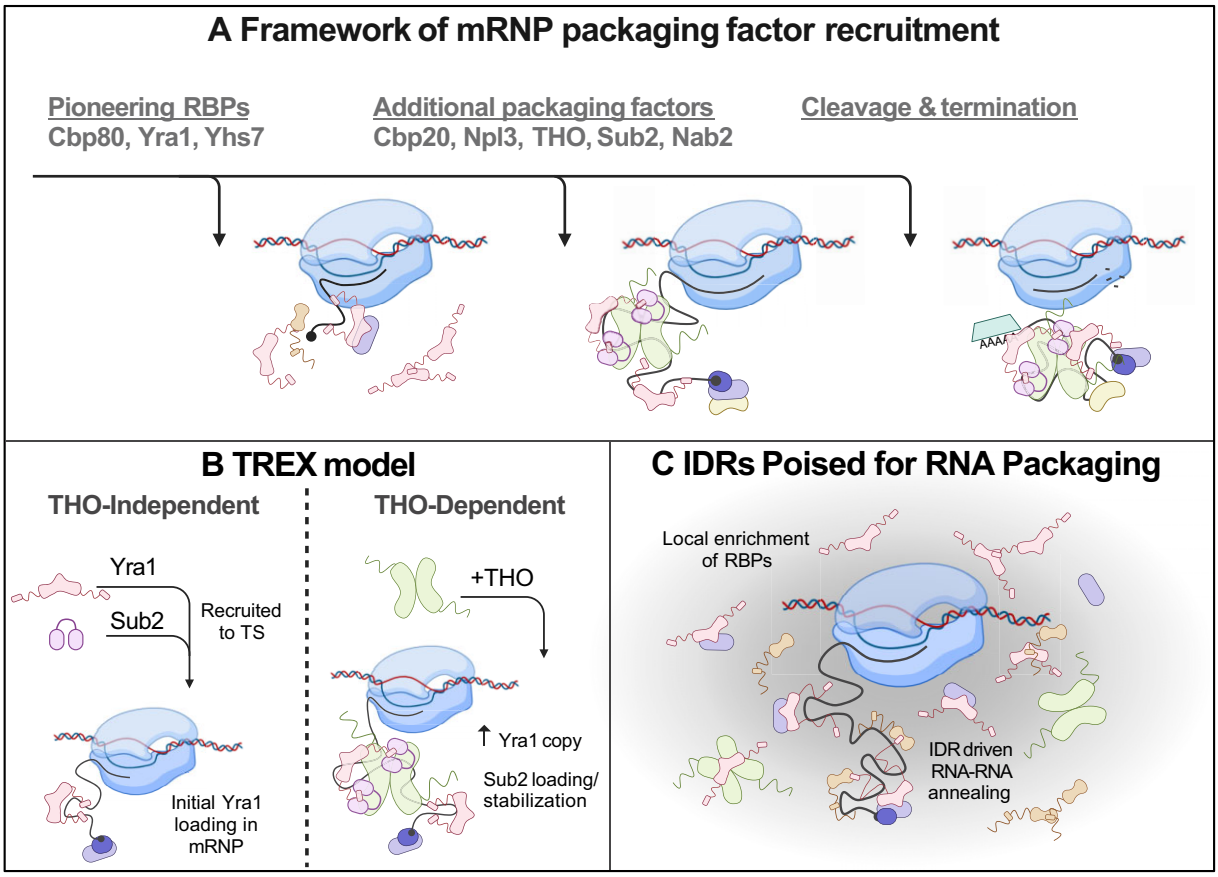


Figure 6. Model of nascent mRNA packaging at a transcription site. (**A**) Based on arrival timing, Cbp80, Yra1, and Yhs7 are pioneering RBPs at an active transcription site, hypothesized to be driven by protein–protein interactions. This is followed thereafter by additional packaging factors recruited very close in time and finally by cleavage and termination factors. (**B**) Recruitment of Yra1 and Sub2 to the transcription site and stable Yra1 engagement in mRNPs are shown to be independent of the THO complex. While Sub2 loading and/or stability within mRNPs and the generation of higher order Yra1 containing mRNPs are shown to depend on the THO complex. (**C**) Early recruitment of RBP and processing factors to an activated transcription site through protein–protein interactions would create a local environment rich with RBPs, many of which have positively charged intrinsically disordered regions (IDRs), including Yra1, Yhs7, and Tho2. These positive IDRs have affinity for both the phosphorylated CTD of RNA polymerase II and the RNA backbone, as well as demonstrated RNA annealing capabilities. Thus, the local environment would make factors immediately available for mRNP assembly with IDRs present to promote RNA compaction, illustrated here through IDRs promoting nascent RNA folding. It is expected this nascent packaging would promote productive elongation as well as R-loop prevention and/or resolution.

counter to common models positing Yra1 is recruited by the THO complex and/or 3′ processing machinery and loaded into mRNPs by THO [25, 33, 62]. Data demonstrated initial Yra1 recruitment (Fig. 4A) and loading into mRNPs (Fig. 4C) are independent of THO (Fig. 6B), while THO is important to supporting formation and/or stability of high Yra1 stoichiometry within mRNPs (Figs 4D and 6B). In contrast to Yra1, data demonstrated Sub2 loading into mRNPs depends on THO, with a dramatic reduction in Sub2-containing nuclear mRNPs in the absence of the THO subunit Mft1 (Figs 4E and 6B). This aligns well with models of human mRNP maturation from recent studies, positing THO on the surface of mRNPs stabilizes UAP56 (human ortholog of Sub2) in an open conformation during early stages of mRNP biogenesis [15, 30]. Subsequent removal of THO leads UAP56 to adopt a closed RNA-bound conformation that acts as a molecular switch licensing mRNPs for nuclear export [15, 30]. As such, the reduced occupancy of Sub2 in mRNPs upon loss of THO may reflect premature maturation and/or reduced Sub2 stability within nuclear mRNPs. Future exploration into the remaining fraction of Sub2-containing complexes in *mft1Δ* cells (Fig. 4E and F) and whether they reflect unstable intermediates or are enriched for RNAs with specific features, such as intron-containing genes that relate to the role of Sub2 in splicing [27, 83], will be of interest.

The early arrival of Yra1 at transcription sites, with the differential role for THO in Yra1 vs Sub2 mRNP occupancy (Fig. 6B), prompted further exploration of Yra1 recruitment dependencies. Of gene deletion mutants and Yra1 truncations tested, including a THO mutant, Yra1 recruitment was delayed only in *yhs7Δ*, *cbp80Δ*, and $Yra1_{\Delta RRM}$ mutants. While the RRM of Yra1 is dispensable for RNA binding and annealing activity *in vitro*, it is shown to be a hotspot for protein–protein interactions in XL-MS data, including prominent interactions with itself, Yhs7, and Nab2 [14]. This suggests protein–protein interactions conferred by the RRM are critical for early recruitment and help to position the RNA annealing IDR regions for packaging of nascent RNA (Fig. 6C), reminiscent of the "IDR Network" model of mRNP packaging put forth by Bonneau *et al.* [14]. This is further supported by delay of Yra1 recruitment in *yhs7Δ*. While Yra1 does not show any crosslinks to Cbp80 in the XL-MS data of nuclear mRNPs, this does not preclude the possibility of a direct Cbp80-Yra1 interaction, given that this interaction may

be transient during mRNP assembly or otherwise may be difficult to detect by crosslinking methodologies. Further supporting the notion Cbp80 is involved in Yra1 co-transcriptional recruitment, Yra1 ChIP occupancy is specifically reduced in the context of *cbp80Δ* [78]. Additionally, a recently reported cryo-EM structure shows mammalian Yra1 ortholog ALYREF binding the CBC via its RRM [77], with other published data suggesting CBC acts as a "landing pad" for ALYREF at the 5′ end of the transcript before loading onto the RNA [84]. This possibility is supported by the data presented here wherein both *cbp80Δ* and $Yra1_{\Delta RRM}$ delay recruitment of Yra1 to the transcription site and Cbp80 and Yra1 RRM interact in a yeast two-hybrid assay.

Nascent RNA packaging from the very earliest stages of transcription aligns with findings of nuclear mRNPs as densely compact structures [13–15, 17] and the role of TREX components in preventing R-loops [23, 25]. It is envisioned that early recruitment and poising of Yra1 and Yhs7 to promote intramolecular RNA annealing would prevent the nascent RNA from generating deleterious R-loops co-transcriptionally or being recognized by RNA surveillance machinery for decay (Fig. 6C). Furthermore, ALYREF has recently been shown to bind R-loops *in vitro* via its positively charged IDRs [85]. This binding of ALYREF also enhances the previously demonstrated ability of UAP56 to resolve R-loops [85, 86]. This points towards a more active role for TREX components in resolving R-loops co-transcriptionally, maintaining genome stability whilst also supporting productive packaging and maturation of mRNPs for nuclear export.

Taken together, this supports a model where protein–protein interactions involving pioneering factors Cbp80, Yra1 and Yhs7 with other early arriving mRNP network factors (Npl3, Nab2, etc.) would establish a local environment rich in RNA-binding proteins at the transcription site (Fig. 6C). This local interaction network would position proteins with positively charged IDRs (Yra1, Yhs7, Tho2, etc.) to promote nascent RNA annealing and compaction based on their demonstrated RNA annealing capabilities [14, 24, 85], contributing to productive elongation and R-loop prevention. As transcription progresses, other Yra1 interactors (e.g. THO, Sub2, and Pcf11) likely promote continued Yra1 loading into mRNPs for packaging and export [33, 62]. The existence of multiple Yra1 binding partners within mRNPs may offer redundant mechanisms to ensure Yra1 delivery to the mRNP, explaining why mutation of individual factors (e.g. Cbp80 and THO) does not prevent Yra1 recruitment and fulfillment of essential Yra1 functions. Additionally, different mechanisms of Yra1 binding may be required to support the packaging and export of the diverse array of gene transcripts produced by the cell. Differing transcript features (e.g. length, base content, propensity for secondary/tertiary structure) and environmental cues may dictate different levels of Yra1 engagement to mediate the proper compaction levels. This is evidenced by the heterogeneity in Yra1 stoichiometry across global nuclear mRNPs and changes in Yra1 stoichiometry with temperature [18].

In considering this model of mRNP assembly, key questions to address include: (i) Do differing gene features require a varied order of protein recruitment? (ii) What structural conformations do mRNAs take within nuclear mRNPs? (iii) What is the role of Yra1 and other RBPs in facilitating mRNA folding? And (iv) how are these events regulated in the context of stress? Further development of the gene array methodology is expected to refine and advance models of mRNP assembly through characterization arrays with different features (intron-containing, heat shock promoter, lncRNA, etc.), use of three-color imaging for pairwise analysis of factors, and application of microfluidics for the observation of the earliest events. It will also be important to validate and extend findings reported here, recognizing the limitations of the gene array approach; as such, it is expected that future work using a combination of imaging, biochemical, and molecular techniques will further refine and enrich these models. Finally, knowing the widespread changes in RNA metabolism in disease, it will be critical to extend these studies to the impact of disease-linked mutations on co-transcriptional recruitment and mRNP packaging.

Acknowledgements

We are grateful to all present and past members of the Montpetit Lab for their discussion and feedback on this work. We additionally thank members of Atchison, Wozniak, and Shav-Tal groups for feedback. We thank Dr Amir Aharoni for providing the CRISPR/Cas9 plasmids, Dr Sean Burgess for providing L40ccU strain for yeast two-hybrids, and Dr Douglas R. Kellogg for providing Yra1 antibody. We acknowledge Dr Andrew Blandino for advice on statistical testing. We thank Laureline Lequint for aiding with yeast two-hybrid experiments. Select figures were created in BioRender: Montpetit, B. (2025) https://BioRender.com/4iav3ws.

Author contributions: Theresa Wechsler (Conceptualization [equal], Data curation [lead], Formal analysis [lead], Investigation [equal], Methodology [equal], Validation [lead], Visualization [lead], Writing—original draft [lead], Writing—review & editing [lead]), Ryuta Asada (Investigation [supporting], Methodology [supporting], Validation [supporting], Visualization [supporting], Writing—review & editing [supporting]), Andrew Dominguez (Formal analysis [supporting], Investigation [supporting], Methodology [supporting], Software [supporting], Validation [supporting], Visualization [supporting], Writing—review & editing [supporting]), Rachel Montpetit (Investigation [supporting], Methodology [supporting], Resources [supporting], Writing—review & editing [supporting]), Julia McCormick (Investigation [supporting], Methodology [supporting], Resources [supporting]), Kalyn Concepcion (Investigation [supporting], Methodology [supporting], Resources [supporting]), and Ben Montpetit (Conceptualization [equal], Data curation [supporting], Formal analysis [supporting], Funding acquisition [lead], Investigation [equal], Methodology [equal], Project administration [lead], Resources [equal], Software [supporting], Supervision [lead], Validation [supporting], Visualization [supporting], Writing—original draft [supporting], Writing—review & editing [lead]).

Supplementary data

Supplementary data is available at NAR online.

Conflict of interest

None declared.

Funding

Research reported in this publication was supported by the Division of Molecular and Cellular Bioscience of the National Science Foundation under award number 2140761. T.W. was supported by the National Institute of Neurological Disorders and Stroke under award number F31NS131037. T.W. and A.D. were funded by the predoctoral Training Program in Molecular and Cellular Biology at UC Davis via National Institute of General Medical Sciences T32 training grant GM007377. R.A. was supported by an overseas research fellowship from the Japan Society for the Promotion of Science. RNA sequencing was carried out at the DNA Technologies and Expression Analysis Cores at the UC Davis Genome Center, which is supported by National Institutes of Health Shared Instrumentation Grant 1S10OD010786-01. The content is solely the responsibility of the authors and does not necessarily represent the views of the funding agencies. Funding to pay the Open Access publication charges for this article was provided by Grant.

Data availability

RNA sequencing data are available using GEO ascension number GSE293820. Code related to RNA sequencing is available at https://doi.org/10.5281/zenodo.15127156, and code related to image analysis is available at https://doi.org/10.5281/zenodo.15126956. Full western blot images are deposited at https://doi.org/10.5281/zenodo.15786809.

References

1. Kelly SM, Corbett AH. Messenger RNA export from the nucleus: a series of molecular wardrobe changes. *Traffic* 2009;**10**:1199–208. https://doi.org/10.1111/j.1600-0854.2009.00944.x
2. Niño CA, Hérissant L, Babour A *et al.* MRNA nuclear export in yeast. *Chem Rev* 2013;**113**:8523–45. https://doi.org/10.1021/cr400002g
3. Wende W, Friedhoff P, Sträßer K. Mechanism and regulation of co-transcriptional mRNP assembly and nuclear mRNA export. In: Oeffinger M, Zenklusen D (eds.), *Advances in Experimental Medicine and Biology*. Vol. **1203**, Cham, Switzerland: Springer International Publishing, 2019, 8523–45.
4. Bensidoun P, Zenklusen D, Oeffinger M. Choosing the right exit: how functional plasticity of the nuclear pore drives selective and efficient mRNA export. *Wiley Interdiscip Rev RNA* 2021;**12**:e1660.
5. Meinel DM, Sträßer K. Co-transcriptional mRNP formation is coordinated within a molecular mRNP packaging station in *S.cerevisiae*. *Bioessays* 2015;**37**:666–77. https://doi.org/10.1002/bies.201400220
6. Mitchell SF, Parker R. Principles and properties of eukaryotic mRNPs. *Mol Cell* 2014;**54**:547–58. https://doi.org/10.1016/j.molcel.2014.04.033
7. Bresson S, Tollervey D. Surveillance-ready transcription: nuclear RNA decay as a default fate. *Open Biol* 2018;**8**:170270. https://doi.org/10.1098/rsob.170270
8. Garland W, Jensen TH. Nuclear sorting of RNA. *Wiley Interdiscip Rev RNA* 2020;**11**:e1572. https://doi.org/10.1002/wrna.1572
9. Heath CG, Viphakone N, Wilson SA. The role of TREX in gene expression and disease. *Biochem J* 2016;**473**:2911–35. https://doi.org/10.1042/BCJ20160010
10. Hautbergue GM. RNA nuclear export: from neurological disorders to cancer. In: El-Khamisy S (ed.), *Personalized Medicine: Advances in Experimental Medicine and Biology*. Vol. **1007**, Cham, Switzerland: Springer International Publishing, 2017, 89–109.
11. Borden K, Culkovic-Kraljacic B. mRNA export and its dysregulation in disease. In Yang W (ed.), *Nuclear-Cytoplasmic Transport, Nucleic Acids and Molecular Biology*. Vol. **33**, Cham, Switzerland: Springer International Publishing, 2018, 179–204.
12. Nussbacher JK, Tabet R, Yeo GW *et al.* Disruption of RNA metabolism in neurological diseases and emerging therapeutic interventions. *Neuron* 2019;**102**:294–320. https://doi.org/10.1016/j.neuron.2019.03.014
13. Batisse J, Batisse C, Budd A *et al.* Purification of nuclear poly(A)-binding protein Nab2 reveals association with the yeast transcriptome and a messenger ribonucleoprotein core structure. *J Biol Chem* 2009;**284**:34911–7. https://doi.org/10.1074/jbc.M109.062034
14. Bonneau F, Basquin J, Steigenberger B *et al.* Nuclear mRNPs are compact particles packaged with a network of proteins promoting RNA–RNA interactions. *Genes Dev* 2023;**37**:505–17. https://doi.org/10.1101/gad.350630.123
15. Pacheco-Fiallos B, Vorländer MK, Riabov-Bassat D *et al.* mRNA recognition and packaging by the human transcription-export complex. *Nature* 2023;**616**:828–35. https://doi.org/10.1038/s41586-023-05904-0
16. Adivarahan S, Livingston N, Nicholson B *et al.* Spatial organization of single mRNPs at different stages of the gene expression pathway. *Mol Cell* 2018;**72**:727–38. https://doi.org/10.1016/j.molcel.2018.10.010
17. Metkar M, Ozadam H, Lajoie BR *et al.* Higher-order organization principles of pre-translational mRNPs. *Mol Cell* 2018;**72**:715–26. https://doi.org/10.1016/j.molcel.2018.09.012
18. Asada R, Dominguez A, Montpetit B. Single-molecule quantitation of RNA-binding protein occupancy and stoichiometry defines a role for Yra1 (Aly/REF) in nuclear mRNP organization. *Cell Rep* 2023;**42**:113415. https://doi.org/10.1016/j.celrep.2023.113415
19. Sträßer K, Masuda S, Mason P *et al.* TREX is a conserved complex coupling transcription with messenger RNA export. *Nature* 2002;**417**:304–8.
20. Schuller SK, Schuller JM, Prabu JR *et al.* Structural insights into the nucleic acid remodeling mechanisms of the yeast tho-SUB2 complex. *eLife* 2020;**9**:e61467.https://doi.org/10.7554/eLife.61467
21. Pühringer T, Hohmann U, Fin L *et al.* Structure of the human core transcription-export complex reveals a hub for multivalent interactions. *eLife* 2020;**9**:e61503.
22. Xie Y, Clarke BP, Kim YJ *et al.* Cryo-EM structure of the yeast TREX complex and coordination with the SR-like protein Gbp2. *eLife* 2021;**10**:e65699. https://doi.org/10.7554/eLife.65699
23. Luna R, Rondón AG, Pérez-Calero C *et al.* The THO complex as a paradigm for the prevention of cotranscriptional R-loops. *Cold Spring Harb Symp Quant Biol* 2019;**84**:105–14. https://doi.org/10.1101/sqb.2019.84.039594
24. Portman DS, O'Connor JP, Dreyfuss G. YRA1, an essential *Saccharomyces cerevisiae* gene, encodes a novel nuclear protein with RNA annealing activity. *RNA* 1997;**3**:527–37.
25. Infantino V, Stutz F. The functional complexity of the RNA-binding protein Yra1: mRNA biogenesis, genome stability and DSB repair. *Curr Genet* 2020;**66**:63–71. https://doi.org/10.1007/s00294-019-01011-8
26. Zhang M, Green MR. Identification and characterization of yUAP/Sub2p, a yeast homolog of the essential human pre-mRNA splicing factor hUAP56. *Genes Dev* 2001;**15**:30–5. https://doi.org/10.1101/gad.851701
27. Libri D, Graziani N, Saguez C *et al.* Multiple roles for the yeast SUB2/yUAP56 gene in splicing. *Genes Dev* 2001;**15**:36–41. https://doi.org/10.1101/gad.852101
28. Jensen TH, Boulay J, Rosbash M *et al.* The DECD box putative ATPase Sub2p is an early mRNA export factor. *Curr Biol* 2001;**11**:1711–5. https://doi.org/10.1016/S0960-9822(01)00529-2
29. Sträßer K, Hurt E. Splicing factor Sub2p is required for nuclear mRNA export through its interaction with Yra1p. *Nature* 2001;**413**:648–52.

30. Hohmann U, Graf M, Schellhaas U. *et al.* A molecular switch orchestrates the nuclear export of human messenger RNA. bioRxiv, https://doi.org/10.1101/2024.03.24.586400, 27 March 2024, preprint: not peer reviewed.
31. Xie Y, Clarke BP, Xie D *et al.* Structures and mRNP remodeling mechanism of the TREX-2 complex. *Structure* 2025;**33**:566–82. https://doi.org/10.1016/j.str.2024.12.019
32. Lei EP, Krebber H, Silver PA. Messenger RNAs are recruited for nuclear export during transcription. *Genes Dev* 2001;**15**:1771–82. https://doi.org/10.1101/gad.892401
33. Zenklusen D, Vinciguerra P, Wyss J-C *et al.* Stable mRNP formation and export require cotranscriptional recruitment of the mRNA export factors Yra1p and Sub2p by Hpr1p. *Mol Cell Biol* 2002;**22**:8241–53. https://doi.org/10.1128/MCB.22.23.8241-8253.2002
34. Abruzzi KC, Lacadie S, Rosbash M. Biochemical analysis of TREX complex recruitment to intronless and intron-containing yeast genes. *EMBO J* 2004;**23**:2620–31. https://doi.org/10.1038/sj.emboj.7600261
35. Janicki SM, Tsukamoto T, Salghetti SE *et al.* From silencing to gene expression: real-time analysis in single cells. *Cell* 2004;**116**:683–98. https://doi.org/10.1016/S0092-8674(04)00171-0
36. Brody Y, Neufeld N, Bieberstein N *et al.* The In vivo kinetics of RNA polymerase II elongation during co-transcriptional splicing. *PLoS Biol* 2011;**9**:e1000573. https://doi.org/10.1371/journal.pbio.1000573
37. Hochberg-Laufer H, Neufeld N, Brody Y *et al.* Availability of splicing factors in the nucleoplasm can regulate the release of mRNA from the gene after transcription. *PLoS Genet* 2019;**15**:e1008459. https://doi.org/10.1371/journal.pgen.1008459
38. Gietz RD, Woods RA. Transformation of yeast by lithium acetate/single-stranded carrier DNA/polyethylene glycol method. *Methods Enzymol* 2002;**350**:87–96.
39. Lee ME, DeLoache WC, Cervantes B *et al.* A highly characterized yeast toolkit for modular, multipart assembly. *ACS Synth Biol* 2015;**4**:975–86. https://doi.org/10.1021/sb500366v
40. Nagalakshmi U, Wang Z, Waern K *et al.* The transcriptional landscape of the yeast genome defined by RNA sequencing. *Science* 2008;**320**:1344–9. https://doi.org/10.1126/science.1158441
41. Cherry JM, Adler C, Ball C *et al.* SGD: Saccharomyces Genome Database. *Nucleic Acids Res* 1998;**26**:73–9. https://doi.org/10.1093/nar/26.1.73
42. Li Z, Vizeacoumar FJ, Bahr S *et al.* Systematic exploration of essential yeast gene function with temperature-sensitive mutants. *Nat Biotechnol* 2011;**29**:361–7. https://doi.org/10.1038/nbt.1832
43. Soreanu I, Hendler A, Dahan D *et al.* Marker-free genetic manipulations in yeast using CRISPR/CAS9 system. *Curr Genet* 2018;**64**:1129–39. https://doi.org/10.1007/s00294-018-0831-y
44. Rohner S, Gasser SM, Meister P. Modules for cloning-free chromatin tagging in *Saccharomyces cerevisae*. *Yeast* 2008;**25**:235–9. https://doi.org/10.1002/yea.1580
45. Tinevez JY, Perry N, Schindelin J *et al.* TrackMate: an open and extensible platform for single-particle tracking. *Methods* 2017;**115**:80–90. https://doi.org/10.1016/j.ymeth.2016.09.016
46. Lari A, Farzam F, Bensidoun P *et al.* Live-cell imaging of mRNP–NPC interactions in budding yeast. In: Shav-Tal Y (ed.), *Imaging Gene Expression: Methods in Molecular Biology*. Vol. **2038**, Cham, Switzerland: Springer International Publishing, 2019, 131–50.
47. McIsaac RS, Silverman SJ, McClean MN *et al.* Fast-acting and nearly gratuitous induction of gene expression and protein depletion in *Saccharomyces cerevisiae*. *MBoC* 2011;**22**:4447–59. https://doi.org/10.1091/mbc.e11-05-0466
48. McIsaac RS, Oakes BL, Wang X *et al.* Synthetic gene expression perturbation systems with rapid, tunable, single-gene specificity in yeast. *Nucleic Acids Res* 2013;**41**:e57. https://doi.org/10.1093/nar/gks1313
49. McIsaac RS, Gibney PA, Chandran SS *et al.* Synthetic biology tools for programming gene expression without nutritional perturbations in *Saccharomyces cerevisiae*. *Nucleic Acids Res* 2014;**42**:e48. https://doi.org/10.1093/nar/gkt1402
50. Tuck AC, Tollervey D. A transcriptome-wide atlas of RNP composition reveals diverse classes of mRNAs and lncRNAs. *Cell* 2013;**154**:996–1009. https://doi.org/10.1016/j.cell.2013.07.047
51. Smith C, Lari A, Derrer CP *et al.* *In vivo* single-particle imaging of nuclear mRNA export in budding yeast demonstrates an essential role for Mex67p. *J Cell Biol* 2015;**211**:1121–30. https://doi.org/10.1083/jcb.201503135
52. Peña MM, Koch KA, Thiele DJ. Dynamic regulation of copper uptake and detoxification genes in *Saccharomyces cerevisiae*. *Mol Cell Biol* 1998;**18**:1121–30.
53. Tipper DJ. Inhibition of yeast ribonucleic acid polymerases by thiolutin. *J Bacteriol* 1973; **116**:245–56. https://doi.org/10.1128/jb.116.1.245-256.1973
54. Kessler MM, Henry MF, Shen E *et al.* Hrp1, a sequence-specific RNA-binding protein that shuttles between the nucleus and the cytoplasm, is required for mRNA 3′-end formation in yeast. *Genes Dev* 1997;**11**:2545–56. https://doi.org/10.1101/gad.11.19.2545
55. Hill CH, Boreikaitė V, Kumar A *et al.* Activation of the endonuclease that defines mRNA 3′ ends requires incorporation into an 8-subunit core cleavage and polyadenylation factor complex. *Mol Cell* 2019;**73**:1217–31. https://doi.org/10.1016/j.molcel.2018.12.023
56. Li J, Querl L, Coban I *et al.* Surveillance of 3′ mRNA cleavage during transcription termination requires CF IB/Hrp1. *Nucleic Acids Res* 2023;**51**:8758–73. https://doi.org/10.1093/nar/gkad530
57. Cramer P. Organization and regulation of gene transcription. *Nature* 2019;**573**:45–54. https://doi.org/10.1038/s41586-019-1517-4
58. Rambout X, Maquat LE. The nuclear cap-binding complex as choreographer of gene transcription and pre-mRNA processing. *Genes Dev* 2020;**34**:1113–27. https://doi.org/10.1101/gad.339986.120
59. Gonatopoulos-Pournatzis T, Cowling VH. Cap-binding complex (CBC). *Biochem J* 2014;**457**:231–42. https://doi.org/10.1042/BJ20131214
60. Wong C-M, Qiu H, Hu C *et al.* Yeast cap binding complex impedes recruitment of cleavage factor IA to weak termination sites. *Mol Cell Biol* 2007;**27**:6520–31. https://doi.org/10.1128/MCB.00733-07
61. Klama S, Hirsch AG, Schneider UM *et al.* A guard protein mediated quality control mechanism monitors 5′-capping of pre-mRNAs. *Nucleic Acids Res* 2022;**50**:11301–14. https://doi.org/10.1093/nar/gkac952
62. Johnson SA, Cubberley G, Bentley DL. Cotranscriptional recruitment of the mRNA export factor Yra1 by direct interaction with the 3′ end processing factor Pcf11. *Mol Cell* 2009;**33**:215–26. https://doi.org/10.1016/j.molcel.2008.12.007
63. Rodríguez-Molina JB, West S, Passmore LA. Knowing when to stop: transcription termination on protein-coding genes by eukaryotic RNAPII. *Mol Cell* 2023;**83**:404–15. https://doi.org/10.1016/j.molcel.2022.12.021
64. Soucek S, Corbett AH, Fasken MB. The long and the short of it: the role of the zinc finger polyadenosine RNA binding protein, Nab2, in control of poly(A) tail length. *Biochim Biophys Acta* 2012;**1819**:546–54. https://doi.org/10.1016/j.bbagrm.2012.03.006
65. Rout MP, Aitchison JD, Suprapto A *et al.* The yeast nuclear pore complex: composition, architecture, and transport mechanism. *J Cell Biol* 2000;**148**:635–52. https://doi.org/10.1083/jcb.148.4.635
66. Oeffinger M, Wei KE, Rogers R *et al.* Comprehensive analysis of diverse ribonucleoprotein complexes. *Nat Methods* 2007;**4**:951–6. https://doi.org/10.1038/nmeth1101
67. Wimalarathna RN, Pan PY, Shen C-H. Chromatin repositioning activity and transcription machinery are both recruited by Ace1p

in yeast CUP1 activation. *Biochem Biophys Res Commun* 2012;**422**:658–63. https://doi.org/10.1016/j.bbrc.2012.05.047
68. Wimalarathna RN, Pan PY, Shen C-H. Co-dependent recruitment of Ino80p and Snf2p is required for yeast CUP1 activation. *Biochem Cell Biol* 2014;**92**:69–75. https://doi.org/10.1139/bcb-2013-0097
69. Mehta GD, Ball DA, Eriksson PR *et al.* Single-molecule analysis reveals linked cycles of RSC chromatin remodeling and Ace1p transcription factor binding in yeast. *Mol Cell* 2018;**72**:875–87. https://doi.org/10.1016/j.molcel.2018.09.009
70. Chen C, Tan M, Wu Z *et al.* Structural and functional insights into R-loop prevention and mRNA export by budding yeast THO-Sub2 complex. *Sci Bull (Beijing)* 2021;**66**:2347–52. https://doi.org/10.1016/j.scib.2021.08.004
71. Yu MC, Bachand F, McBride AE *et al.* Arginine methyltransferase affects interactions and recruitment of mRNA processing and export factors. *Genes Dev* 2004;**18**:2024–35. https://doi.org/10.1101/gad.1223204
72. Iglesias N, Tutucci E, Gwizdek C *et al.* Ubiquitin-mediated mRNP dynamics and surveillance prior to budding yeast mRNA export. *Genes Dev* 2010;**24**:1927–38. https://doi.org/10.1101/gad.583310
73. Stutz F, Bachi A, Doerks T *et al.* REF, an evolutionary conserved family of hnRNP-like proteins, interacts with TAP/Mex67p and participates in mRNA nuclear export. *RNA* 2000;**6**:638–50. https://doi.org/10.1017/S1355838200000078
74. Zenklusen D, Vinciguerra P, Strahm Y *et al.* The yeast hnRNP-like proteins Yra1p and Yra2p participate in mRNA export through interaction with Mex67p. *Mol Cell Biol* 2001;**21**:4219–32. https://doi.org/10.1128/MCB.21.13.4219-4232.2001
75. MacKellar AL, Greenleaf AL. Cotranscriptional association of mRNA export factor Yra1 with C-terminal domain of RNA polymerase II. *J Biol Chem* 2011;**286**:36385–95. https://doi.org/10.1074/jbc.M111.268144
76. Cheng H, Dufu K, Lee C-S *et al.* Human mRNA export machinery recruited to the 5′ end of mRNA. *Cell* 2006;**127**:1389–400. https://doi.org/10.1016/j.cell.2006.10.044
77. Clarke BP, Angelos AE, Mei M *et al.* Cryo-EM structure of the CBC-ALYREF complex. *eLife* 2024;**12**:RP91432. https://doi.org/10.7554/eLife.91432.3
78. Sen R, Barman P, Kaja A *et al.* Distinct functions of the cap-binding complex in stimulation of nuclear mRNA export. *Mol Cell Biol* 2019;**39**:e00540-18. https://doi.org/10.1128/MCB.00540-18
79. Spingola M, Armisen J, Ares M. Mer1p is a modular splicing factor whose function depends on the conserved U2 snRNP protein Snu17p. *Nucleic Acids Res* 2004;**32**:1242–50. https://doi.org/10.1093/nar/gkh281
80. Nielsen J. Yeast Systems Biology: model organism and cell factory. *Biotechnol J* 2019;**14**:e1800421. https://doi.org/10.1002/biot.201800421
81. Fortes P, Kufel J, Fornerod M *et al.* Genetic and physical interactions involving the yeast nuclear cap-binding complex. *Mol Cell Biol* 1999;**19**:6543–53. https://doi.org/10.1128/MCB.19.10.6543
82. Aibara S, Gordon JMB, Riesterer AS *et al.* Structural basis for the dimerization of Nab2 generated by RNA binding provides insight into its contribution to both poly(A) tail length determination and transcript compaction in *Saccharomyces cerevisiae*. *Nucleic Acids Res* 2017;**45**:1529–38. https://doi.org/10.1093/nar/gkw1224
83. Masuda S, Das R, Cheng H *et al.* Recruitment of the human TREX complex to mRNA during splicing. *Genes Dev* 2005;**19**:1512–7. https://doi.org/10.1101/gad.1302205
84. Viphakone N, Sudbery I, Griffith L *et al.* Co-transcriptional loading of RNA export factors shapes the human transcriptome. *Mol Cell* 2019;**75**:310–23. https://doi.org/10.1016/j.molcel.2019.04.034
85. Bhandari J, Guillén-Mendoza C, Banks K *et al.* The molecular chaperone ALYREF promotes R-loop resolution and maintains genome stability. *J Biol Chem* 2024;**300**:107996. https://doi.org/10.1016/j.jbc.2024.107996
86. Pérez-Calero C, Bayona-Feliu A, Xue X *et al.* UAP56/DDX39B is a major cotranscriptional RNA–DNA helicase that unwinds harmful R loops genome-wide. *Genes Dev* 2020;**34**:898–912. https://doi.org/10.1101/gad.336024.119

Received: April 8, 2025. **Revised:** July 3, 2025. **Editorial Decision:** July 23, 2025. **Accepted:** August 7, 2025