Development and application of proteomic and genomic methods in RNA biology

Thesis by

Drew D. Honson

In Partial Fulfillment of the Requirements for

the Degree of

Doctor of Philosophy

CALIFORNIA INSTITUTE OF TECHNOLOGY

Pasadena, California

2025

(Defended December 3, 2024)

Drew Honson
ORCID: 0000-0002-4654-8974

ACKNOWLEDGEMENTS

The work in this thesis, like all scientific progress, was a collective effort. Each chapter has a section describing the individual contributions of colleagues to the work, but I wanted to begin by thanking Mario Blanco, without whom this thesis would not have been possible. Mario is a senior staff scientist in the Guttman lab who has mentored me since my rotation. His steady hand and problem-solving abilities were immeasurably helpful.

I also want to thank the students I have had the opportunity to mentor: Aubrey Stevens, Yu Wang, and Alex Burr. On a personal level, I am very privileged to have worked with such bright, dedicated, and talented young scientists. On a scientific level, mentoring such thoughtful people is a serious test of expertise. Many times, one of them has asked a seemingly basic question and I have paused, stumped, and ran to look up the answer. I appreciate the accountability they've brought, and I hope my mentorship has made them more confident scientists in turn.

I next want to thank the many faculty members who have helped me during my time at Caltech. First, of course, is my advisor Mitch Guttman. I could mention many qualities in Mitch, but in preparing this thesis I have been most reminded of his relentless drive to make our science as accessible as possible through clear writing and data visualization. I hope I can emulate that focus on clarity and concision. I also want to thank my committee members, Michael Elowitz, Kata Fejes-Tóth, and Magdalena Żernicka-Goetz, for their advice throughout the years. Kata in particular has been endlessly supportive of me since I first came to Caltech for my interview, and I have greatly valued her support and our conversations during my time here. I also want to thank the professors whom I assisted in undergraduate courses: Bruce Hay, Paul Sternberg, and Andres Collazo. I'm very glad to have learned from such great teachers. Lastly, I want to thank Joanna Jachowicz, former Guttman lab postdoc and current faculty member at IMBA in Vienna. Joanna invited me to come to Vienna for three weeks to help her group learn RAP-MS 2.0, and it was such a phenomenal experience.

Finally, I want to thank the friends and family who have been so supportive of me during grad school. Thank you to everyone in the Guttman lab for the scientific and personal support, especially Prashant Bhat, Paulomi Bhattacharya, Linlin Chen, Noah Epstein, Olivia Ettlin, Dev Majumdar, Andrew Perez, and Mackenzie Strehle. I am also grateful to the other friends I've made at Caltech, especially David Goertsen, Kadina Johnston, Patrick Almhjell, Gracie Zhang, Jonathan White, and Abdullah Farooq. I also want to thank my mother Dana, my father Nathan, and my siblings Erica and Abbie for their love, support, and their valiant attempts to understand what exactly I'm doing. Most importantly, I want to thank my wife Natalie Gehred for her endless love and support. I am so glad to have had this adventure with you and am excited for our future together. And lastly, though she is the least likely to read this, I would like to thank my cat Tarot, the most affectionate and mercurial animal on earth.

ABSTRACT

This thesis contains three interrelated projects. Chapter 1 describes the development of a novel RNA-proteomics method: RNA-antisense purification followed by mass spectrometry (RAP-MS 2.0). It contains results of a RAP-MS 2.0 study profiling the protein partners of eight RNAs (7SL, 7SK, RMRP, U1, U2, U6, U7, and Xist) as well as a detailed, step-by-step protocol for the new method. Chapter 2 describes a quality control method for Split and Pool Identification of RBP targets (SPIDR). It identifies an underappreciated failure point in SPIDR experiments (the equal loading of antibody-IDs onto beads), and describes a method for monitoring and resolving this issue. Chapter 3 describes the application of SPIDR to ribosome-associated proteins in human cells. The study both validates existing structures and identifies novel interactions between nucleolar proteins and immature ribosomal RNA, and between protein trafficking factors and the large ribosomal subunit.

PUBLISHED CONTENT AND CONTRIBUTIONS

Honson, D.D. et al. (2025). RAP-MS 2.0 reveals substoichiometric components of RNA-protein complexes. *Manuscript in preparation.*

I executed all experiments and wrote the paper.

Honson, D.D. et al. (2025) SPIDR: A high-throughput approach to map RNA-binding proteins to diverse ribosome structures. *Manuscript in preparation.*

I executed all experiments and wrote the paper.

Bhat, P., Honson, D.D., and Guttman, M. (2021) Nuclear compartmentalization as a mechanism of quantitative control of gene expression. In: *Nature Rev. Mol. Cell Bio.* 22 (10), pp. 653-670. doi: 10.1038/s41580-021-00387-1

I contributed to researching topics for the review and writing and editing the manuscript.

Banerjee, A.K., Blanco, M.R. et al. (2020) SARS-CoV-2 disrupts splicing, translation, and protein trafficking to suppress host defenses. In: *Cell* 183 (5), pp. 1325-1339 e21. doi: 10.1016/j.cell.2020.10.004

I performed immunofluorescence imaging and optimized in-vitro translation assays.

TABLE OF CONTENTS

PREFACE

This thesis describes three related projects that center around a methodological question: how should scientists measure RNA-protein complexes? RNA-protein interactions are central to numerous critical cellular functions. Ribonucleoprotein complexes (RNPs) such as 7SK and XIST regulate transcription, snRNPs excise introns from nascent transcripts, and export proteins guide the newly transcribed mRNA into the cytoplasm where they are translated by perhaps the most complex RNA-protein machine, the ribosome. Even these critical functions do not capture the full significance of RNA-protein interactions, as additional RNPs control the maturation and activity of most of the RNP catalysts described above. As such, enumerating the protein binding partners of RNAs in the cell, mapping the specific binding sites of these partners, and interrogating the roles of these interactions are central problems of RNA biology.

Chapter 1 describes a novel approach for enumerating proteins bound to specific RNAs: RNA antisense purification followed by mass spectrometry 2.0 (RAP-MS 2.0). In this method, RNAs of interest are purified from UV-crosslinked lysate using antisense oligonucleotide probes. Covalently linked proteins are then eluted and analyzed by liquid chromatography mass spectrometry (LCMS). RAP-MS 2.0's major advantage compared to previous methods is its stringency. Advancements in the capture strategy greatly reduce the number of false positives and allow for substoichiometric components of RNPs to be confidently identified. Using this approach, we identified novel components of the RNPs of the lncRNA Xist, the exosome component RMRP, and the snRNA U1.

Chapter 2 describes a method for improving reliability of a method previously developed in the Guttman lab: Split and Pool Identification of RBP targets (SPIDR). SPIDR uses a pool of antibody-bound, oligonucleotide-barcoded beads to purify dozens of RNA binding proteins in parallel and map them to their bound RNAs. While using this method, I identified a crucial failure point in bead preparation that dramatically reduced the amount of RNA recovered that could be assigned to a target protein. To address this, I developed a quality control protocol that makes SPIDR experiments less risky and more consistent between replicates.

Chapter 3 describes the application of SPIDR to ribosome-associated proteins. SPIDR both confirmed RNA-protein interactions that were previously observed through structural methods and uncovered interactions between ribosomal RNA and RNA-binding proteins that had not previously been observed. Not only does this study provide a valuable atlas of proteins over ribosomal RNA, but it provides a blueprint for future high-throughput studies of ribosome-associated proteins.

Through these efforts, I hope to have contributed insights both into the interactions between specific RNAs and their protein partners.

Chapter 1: *RAP-MS 2.0 reveals substoichiometric components of RNA-protein complexes*

ABSTRACT

RNA-protein complexes are essential factors in development, homeostasis, and disease. RNA proteomics methods are essential for characterizing these complexes but suffer from high levels of background. This high background hinders identification of RNP components, especially those outside the stable, core RNP. Here, we present RAP-MS 2.0, an adaptation of our original RAP-MS protocol. RAP-MS 2.0 has reduced background signal, lower cost, and allows lysate to be reused to capture multiple RNAs. We demonstrate that RAP-MS 2.0 recapitulates known core and non-core RNPs for 7SL, 7SK, RMRP, U1, U2, U6, U7, and Xist. Additionally, we use RAP-MS 2.0 to identify novel RNA-protein interactions between Xist and TREX components, RMRP and the nucleolar protein Nepro, and U1 with FET family transcriptional regulators.

INTRODUCTION

RNA-protein complexes are central to normal cell function. Even before transcription begins, the ribonucleoprotein (RNP) 7SK regulates RNA polymerase II phosphorylation, thereby controlling transcriptional licensing.[1] During transcription, RNPs and RNA binding proteins (RBPs) such as spliceosomal components, the exon-junction complex, and cleavage and polyadenylation factors help process and stabilize the nascent transcript.[2] For non-coding trasncripts such as small nuclear RNAs (snRNAs), RBPs escort the RNA to cellular compartments that assemble the mature RNP.[3] If the transcript is cytoplasmic, export factors transport the RNA out of the nucleus.[2] Coding RNAs then interact with the largest and most abundant RNP: the ribosome. The eukaryotic ribosome is a massive molecular machine that may interact with hundreds of different RNAs and proteins over its lifetime. The centrality of RNA-protein complexes to life underscores the importance of biological methods to characterize these interactions.

RNA proteomics methods were developed to enumerate the components of RNA-protein complexes. The family of methods employs liquid chromatography mass spectrometry (LCMS) to identify protein binding partners of specific RNAs. The proteins can be isolated by purifying the target RNP, selectively degrading the target RNA to release bound proteins, or employing proximity biotinylation in living cells.[4,5] These methods have elucidated the protein components of numerous RNPs that were inaccessible to older methods.[6–10] Despite this progress, however, RNA proteomics is plagued by high material requirements and high levels of background noise.

Background noise is particularly nettlesome for characterizing RBPs outside the stable, mature RNP. Core RNP components usually bind stably and crosslink efficiently to their

RNA partners. As such, they are robustly detected by RNA proteomics methods and stand out clearly from environmental and experimental contaminants. Most proteins in these datasets, however, are detected at much lower levels, and current methods are unable to distinguish non-specific background from bona fide RBPs in the lower abundance population. Nonetheless, these "non-core RBPs" can be functionally significant. Proteins involved in RNA biogenesis, coordination of RNPs in larger molecular machines, and regulation of RNP activity are among these important factors.

Here, we present RAP-MS 2.0, an adaptation of our original RAP-MS method optimized to reduce background and improve identification of non-core RBPs. RAP-MS 2.0 requires lower input than RAP-MS 1.0 and allows for recovery of multiple RNA species from a single lysate. The innovations of RAP-MS 2.0 include switching from biotinylated antisense probes to probes covalently linked to a paramagnetic bead, incorporating a capture-elute-recapture strategy to increase specificity, and purifying peptides with carboxylate beads to reduce loss (Figure 1a-b). We demonstrate that these modifications improve unambiguous identification of known core and non-core RBPs while reducing cost and the amount of biological material required. Additionally, we show that a single lysate can be reused for several RAP captures without reducing the fidelity of individual RAP-MS 2.0 results. Finally, we identify novel components of the Xist, RMRP, and U1 RNPs that suggest mechanisms for previously unexplained features of these key molecular complexes.

RESULTS

Comparison between RAP-MS and RAP-MS 2.0

In our original RAP-MS protocol, two cell cultures are prepared: one labeled with heavy amino acids (SILAC) and one unlabeled (Figure 1a).[11] After UV crosslinking and lysis, two different RNA species are captured with biotinylated probes: one from the SILAC cells, and one from the unlabeled condition. The biotinylated probes are recovered with streptavidin, the beads are washed, and the proteins are eluted with benzonase. After peptide digestion and desalting, the two elutions are mixed and analyzed by LCMS. Proteins that disproportionately contain heavy peptides are assigned to the RNA captured from the SILAC lysate, while those that disproportionately contain light peptides are assigned to the RNA captured from unlabeled lysate. If a protein has equal representation of heavy and light, it is considered background.

The higher stringency of RAP-MS 2.0 removes the need for comparing to SILAC, so only a single, unlabeled culture is grown. Crosslinking and lysate preparation are effectively identical to RAP-MS 1.0, but the targeted RNA is captured with probes covalently linked to a bead rather than with biotin-streptavidin. After the RNA is captured and the beads are washed, the RNA-protein complexes are eluted with heat. Then, the RNA is recaptured from the first eluent with fresh beads, the beads are washed again, and the RNA-protein complexes again eluted. The second capture is the key innovation for increasing stringency and is discussed in more detail in the next section.

In addition to its higher stringency, RAP-MS 2.0 is dramatically less expensive than RAP-MS 1.0. Eliminating SILAC, switching from biotinylated to unmodified probes, and moving from streptavidin to oligonucleotide-bound beads all reduce cost. We hope that these

reductions will lower the barrier for future researchers to replicate this work and use the method for their own targets of interest.

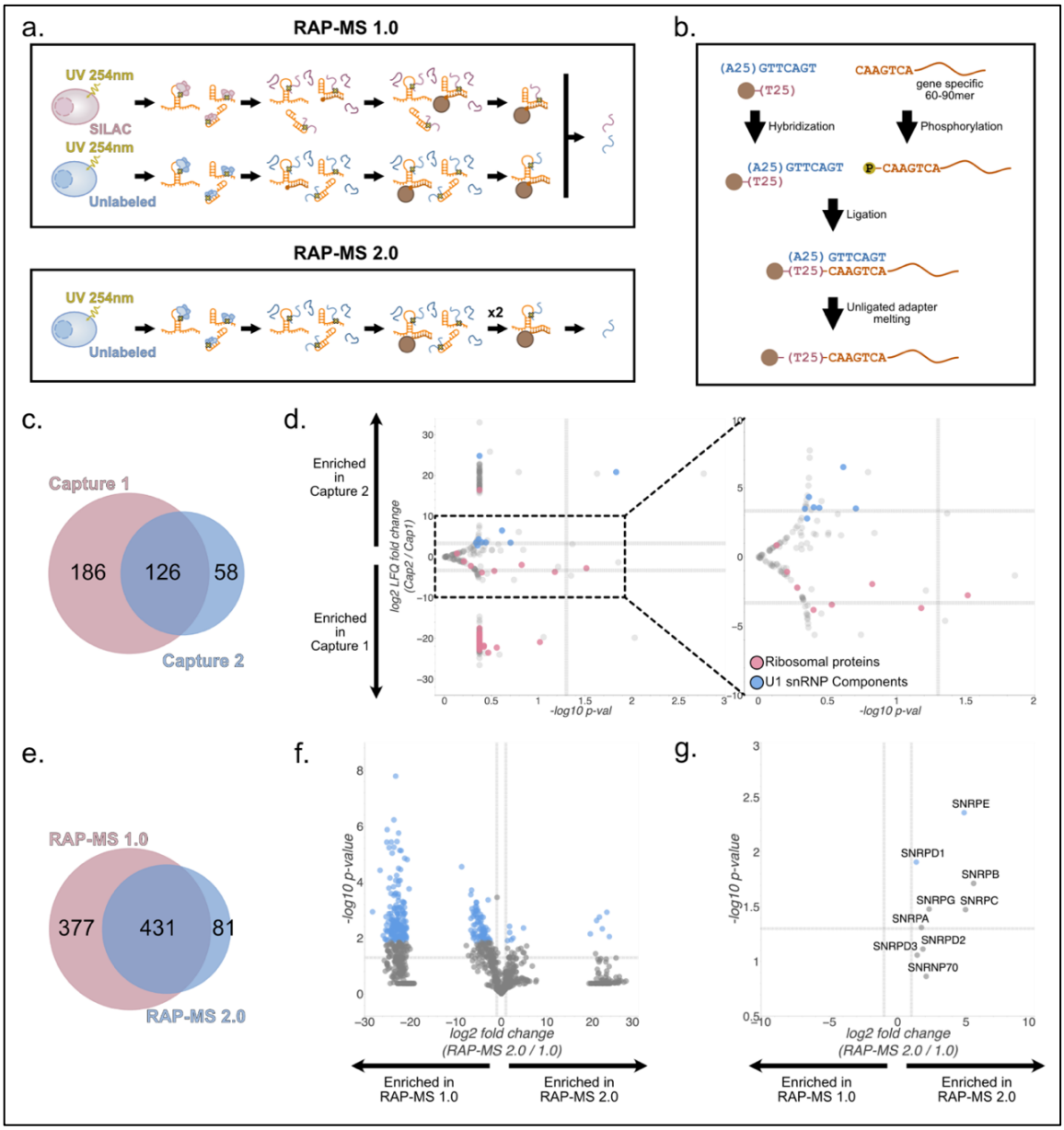

Figure 1: RAP-MS 2.0 uses multiple captures to improve specificity. a, Schematics of RAP-MS 1.0 and 2.0. **b**, Schematic of probe-bead preparation. **c**, Unique and shared proteins identified in capture 1 (pink) and 2 (blue) of U1 snRNA RAP-MS 2.0. **d**, Volcano plot of proteins enriched in capture 1 and 2 of U1 snRNA RAP-MS 2.0. Horizontal dotted lines represent 2-fold or -2-fold changes in LFQ intensity. Ribosomal proteins (background) are

indicated in red, and U1 snRNP components (signal) are in purple. The right panel is an inset of the left focusing only on proteins identified in both captures. **e**, Unique and shared proteins identified in RAP-MS 1.0 (pink) and 2.0 (blue) for U1 snRNA. **f**, Volcano plot of proteins enriched in RAP-MS 1.0 and 2.0 of U1 snRNA. Vertical dotted lines represent 2-fold and -2-fold changes in LFQ intensity, and the horizontal dotted line represents Students T-test p = 0.05. Proteins that differed significantly ($p < 0.05$) by more than 2-fold are indicated in blue. **g**, Inset of panel g showing only U1 snRNP components. Seven of the ten components that differed significantly ($p < 0.05$) by more than 2-fold are indicated in blue. All that significantly differed were enriched in RAP-MS 2.0.

RAP-MS 2.0's two-capture strategy increases stringency

To improve RAP-MS 2.0's specificity and stringency, we developed a multiple capture RAP strategy. In the original RAP-MS protocol, proteins are eluted using benzonase to digest both the targeted RNA and the RAP probes.[11] This approach is incompatible with multiple captures as it destroys the intended target. An alternative elution strategy is to use heat under low salt conditions to melt the RNA-probe hybrid. RAP-MS 1.0, however, uses biotinylated probes bound to streptavidin beads. This causes two issues. First, heating the beads causes streptavidin to shed into the eluent, increasing the level of background. Second, heating under low salt can release the biotinylated probe from the streptavidin. If the probe then binds to shed streptavidin once the eluent cools, it cannot be recaptured. To avoid these complications, we designed a strategy to covalently link unmodified probes to commercial oligo dT(25) beads (Figure 1b).

The bead-probe linkage is accomplished through ligation. All RAP-MS 2.0 probes have a short common sequence (CAAGTCA) at the 5'-end. This 5'-adapter is complementary to the 5'-end of a polyA splint ($TGACTTGA_{25}$). The splint is hybridized to oligo d(T)25 beads

and the probes are 5'-phosphorylated using phosphonucleotide kinase (PNK). Phosphorylated probes are then added to the bead-splint mixture along with DNA ligase. After ligation, the beads are boiled in Tris-EDTA (TE) buffer to remove any unligated probe. This leaves the beads coated with single stranded oligos containing a 5'-oligo d(T)25 and a 3'-specific RAP probe.

The modified oligo dT(25) beads allowed us to use heat instead of enzymatic digestion for elution. After the first capture, the beads are incubated at 95°C in TE buffer to elute bound RNAs. Fresh beads are then added to the eluent and the capture and eluent is repeated. To test the efficacy of this strategy, we performed two replicates of RAP-MS 2.0 on the U1 snRNA and used 10% of capture 1 and 90% of capture 2 for LC-MS analysis. Approximately half as many proteins were identified in capture 2 as in capture 1 in spite of the higher percent aliquot analyzed for capture 2 (Figure 1c). 186 proteins were exclusive to the first capture, suggesting that the second capture removed non-specific noise (Figure 1c). 58 of the 184 proteins identified in capture 2 were exclusive to the second capture, possibly because their relative abundance in capture 1 was too low to be detected.

To determine whether reduced complexity was the result of increased stringency, we examined the identities of the proteins in captures 1 and 2 (Figure 1d). In total, we identified eight of the ten core U1 snRNP members (U1-70K, U1-A, U1-C, Snrpb, Snrpd1, Snrpd2, Snrpd3, and Snrpg).[12–14] Of those, Snrpd1 was exclusively identified in capture 2, indicating that the second capture improved coverage of the U1 snRNP. Six of the seven U1 snRNP factors found in both captures 1 and 2 were enriched by greater than ten-fold in capture 2, again indicating better representation of known U1 core snRNP components in the second capture.

To evaluate background, we analyzed the enrichment of ribosomal proteins in each capture (Figure 1d). Background in RAP-MS experiments is defined as proteins that are not covalently bound to the target RNA, but nonetheless appear in the final eluent. As U1 is not translated, any ribosomal proteins in the sample can be considered non-specific noise. Overall, 76.3% (29 of 38) of ribosomal proteins identified were exclusively found in capture 1. Only one was specific to capture 2, and it was only found in one replicate. 21.1% (8 of 38) ribosomal proteins were identified in both captures, and all but one were enriched in capture 1. The enrichment for U1 core proteins and depletion of ribosomal contaminants in capture 2 confirms that the two-capture strategy improves signal-to-noise ratio in RAP-MS 2.0.

To determine whether RAP-MS 2.0 improves specificity over RAP-MS 1.0, we next captured U1 snRNA using both methods. We speculated that the presence of the free dT stretch on RAP-MS 2.0 beads and the larger amount of probe could lead to more background than the biotin-streptavidin strategy. This was not the case. 42.4% (377 of 889) identified proteins were exclusive to RAP-MS 1.0 compared to 9.1% (81 of 889) exclusive to RAP-MS 2.0 (Figure 1e). Additionally, the RAP-MS 1.0 specific proteins had much stronger LCMS signals than those specific to RAP-MS 2.0, resulting in lower p-values (Figure 1f). As such, not only did RAP-MS 1.0 have higher numbers of background proteins, but the background represented most of the total RAP-MS 1.0 peptides. This high background necessitates the SILAC-unlabeled comparison in our original protocol.

Most important to the comparison of RAP-MS 1.0 and 2.0 are the ten core components of the U1 snRNP. Both methods identified all ten members of the U1 snRNP, but all were greater than two-fold enriched in RAP-MS 2.0 relative to 1.0, and two differed significantly (Student's T-test $p < 0.05$, Benjamini-Hochberg FDR = 0.05, Figure 1g). These results

indicate that RAP-MS 2.0 successfully reduces background compared to our original protocol, increasing the likelihood of finding bona fide RNA-protein interactions.

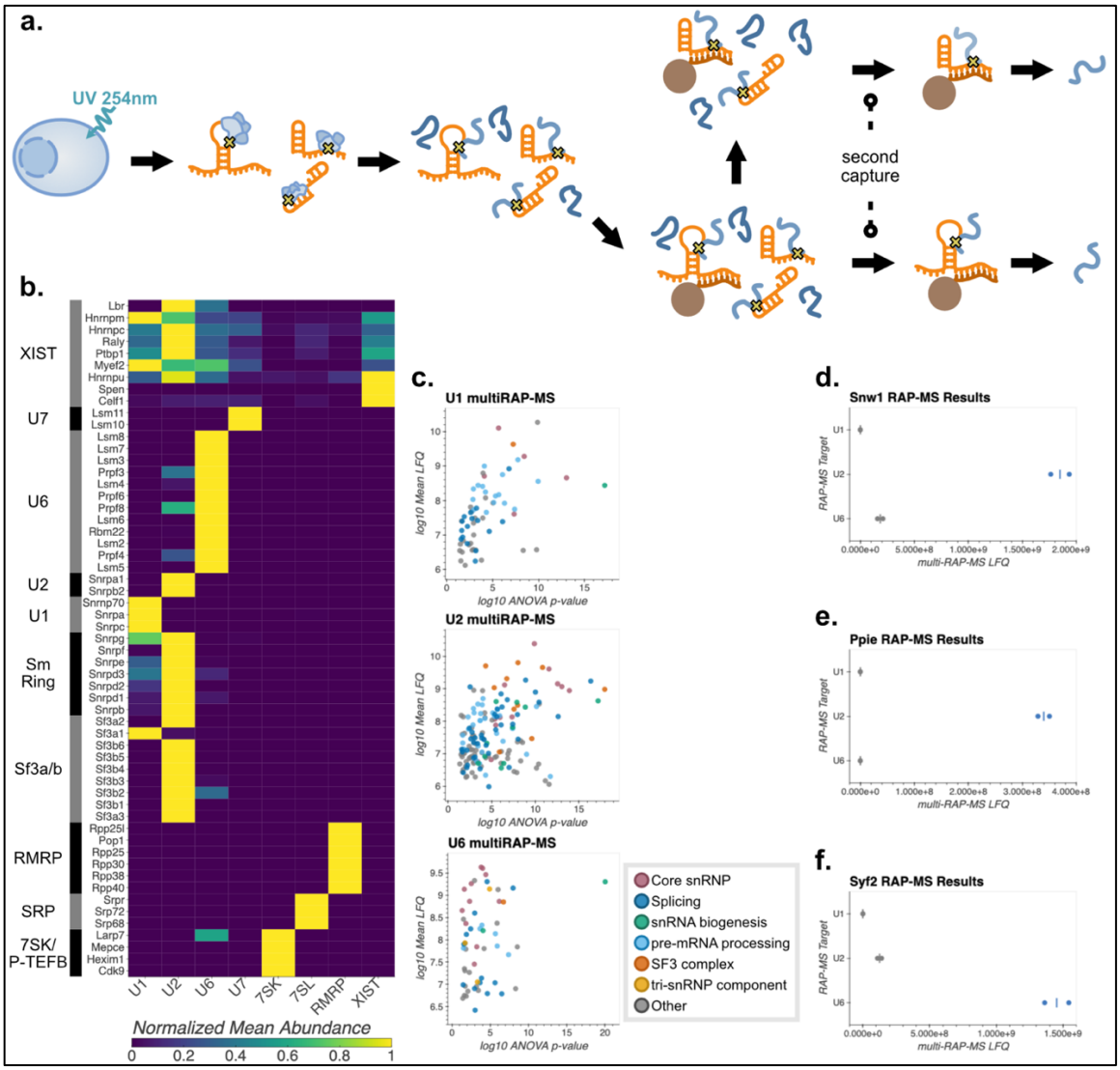

Figure 2: Mulitplexed RAP-MS 2.0 recapitulates known RNP. a, Schematic of multiplexed RAP-MS 2.0 (multi-RAP-MS 2.0). **b**, Heatmap of core RNP components. Individual proteins grouped by RNP are indicated on the y-axis while multi-RAP-MS 2.0 target is on the x-axis. Color indicates the LFQ intensity values normalized per protein. **c**, U1, U2, and U6 snRNA multi-RAP-MS 2.0 colored by functional categories. The x-axis indicates the log10 ANOVA p-value calculated by comparing all RNAs captured. The y-axis the log-transformed mean LFQ between two replicates. **d-f**, Multi-RAP-MS 2.0 enrichments

for spliceosomal components. Dots represent individual replicates and vertical bars represent the mean of the replicates. The RNA annotated to associate with the spliceosomal component is colored in blue.

RAP-MS 2.0 permits reusing lysate for multiple targets

RAP-MS experiments use large amounts of biological material, from 25 million cells for an abundant RNA such as U1, to 100 million or even higher for less abundant targets such as Xist.[15] As such, we wondered whether we could reuse the flow-through from the first capture of RAP-MS 2.0 as the input for other experiments. To do this, we captured single RNA species on-bead, cleared the lysate completely on a magnet, then used the flow-through for additional RNA captures (Figure 2a). For brevity, we refer to this application of the method as multi-RAP-MS 2.0.

In the first experiment, we captured a total of eight RNA species in groups of four: Xist, 7SL, 7SK, and U1 in the first set; U7, RMRP, U6, and U2 in the second. These RNAs were selected based on their high abundances and well characterized core RNPs.

All RAP-MS experiments showed highly specific identification of the members of core RNPs (Figure 2b).[15–23] In addition to the expected complexes, examination of core RNPs revealed some reasonable but unexpected interactions. The SF3a and b complexes are typically considered part of the U2 snRNP.[24,25] RAP-MS confirmed that most SF3 components were specific to U2, but also indicated interactions between U1 and SF3a1, and U6 and SF3b2 (Figure 2b). This is consistent with previous research showing that SF3a1 binds to U1 stem-loop 4 in pre-spliceosomal complexes and SF3b2 stabilizes U6 on pre-mRNA.[26,27]

Multi-RAP-MS 2.0 reveals biogenesis factors for snRNPs

In addition to proteins in the core RNP, RAP-MS for U1, U2, and U6 revealed factors involved in numerous steps of snRNP biogenesis (Figure 2c). Both U1 and U2 showed binding to Gemin5, a component of the SMN complex essential for Sm ring assembly on immature snRNAs.[28] Additionally, U2 RAP-MS revealed three Cajal body proteins (Coilin, Ice2, and Zc3h8) and five H/ACA snoRNA components (Dck1, Gar1, Nhp2, Nop10, and Nop56). Cajal bodies are the site of snRNA pseudouridylation and scaRNAs, a subclass of snoRNAs found in Cajal bodies, catalyze this post-transcriptional modification. U6 RAP-MS recovered two proteins associated with U6 biogenesis: Tut1 which uridylates the 3'-end of immature U6, and Mettl16, which methylates U6 adenosine residues.[29–32] U6 RAP-MS additionally revealed binding to Sart3. After U6 is ejected from the post-spliceosomal complex, Sart3 binds to the snRNA and facilitates regeneration of the U4/6 di-snRNP.[33] These results demonstrate that RAP-MS identifies RBPs outside core RNPs and can provide insights into the life cycles of functional RNAs.

Multi-RAP-MS 2.0 correctly identifies spliceosomal subcomplexes

A challenge of RNA-proteomics is distinguishing between direct and indirect RNA-protein interactions. This problem is especially acute in highly complex molecular structures that contain multiple intimately associated RNPs. The major spliceosome is a classic example of the complexity problem; it undergoes multiple conformational and compositional rearrangements and contains as many as five snRNPs in a single assembly (U1, U2, U4, U5, and U6 in the early pre-catalytic B complex).[34] To examine whether RAP-MS could correctly assign RNA-protein interactions known from cryo-EM structures, we examined three non-

snRNP spliceosomal components that act in the transition from the pre-catalytic (B) to catalytic (B*) spliceosome: Snw1 (Skip), Ppie (CycE) and Syf2.[35] This transition requires at least three intermediate structures: the early, mature, and late active spliceosomes (B^{act}). The precise orchestration of this process is required to position the 2'-hydroxyl group of the branchpoint adenosine for nucleophilic attack on the 3'-hydroxyl of the splicing donor, leading to formation of the lariat.[36]

Snw1 is a splicing factor that enters the spliceosome during the transition from the pre-catalytic B complex to the early B^{act} complex.[35] Existing structures of the B^{act} complexes have substantial missing densities for Snw1, but it interacts specifically with U2 in the C* spliceosome, the structure that catalyzes the second transesterification reaction.[35] As predicted, RAP-MS exclusively identified Snw1 as binding to U2.

Ppie is a peptidylprolyl isomerase also known as cyclophilin E enters the spliceosome in the mature B^{act} complex.[35] It remains in the spliceosome during the first transesterification reaction in the B* complex and becomes closely associated with the U2 snRNP in the resulting C complex.[37,38] Consistent with this interaction, RAP-MS exclusively showed an interaction between U2 and Ppie.

Syf2 is a member of the NineTeen Complex (NTC), which enters the spliceosome in the B-complex and guides spliceosomal conformational changes until the completion of splicing.[39] In existing structures, Syf2 binds closely to the U2/U6 duplex in the C complex, but both RAP-MS showed a specific enrichment of Syf2 over U6.[38] This may indicate that Syf2's association with U6 is more stable than with U2, which could be functionally significant for stabilizing spliceosomal structures.

Multi-RAP-MS 2.0 reveals novel ncRNA-protein interactions

In addition to known RNA-binding factors, RAP-MS 2.0 revealed previously uncharacterized RBP-RNA interactions. we chose six examples for further study: Chtop and Alyref binding to Xist, Nepro binding to RMRP, and Fus, Ewsr1, and Taf15 (FET) binding to U1. These proteins were selected because their RAP-MS signals were comparable to those of core RNP components, and because their annotated functions suggested they might augment the activities of the RNAs they bind.

Many lncRNAs, including Xist, share molecular features with mRNAs such as introns, m7G-caps, and polyA tails.[40] Unlike mRNAs, however, not all lncRNAs are destined for the cytoplasm. An open question in lncRNA biology is how the cell determines which lncRNAs remain in the nucleus and which are exported. Xist is a classic example of a nuclear lncRNA; it remains bound to the inactive X-chromosome for its entire lifecycle. Unexpectedly, multi-RAP-MS 2.0 revealed that Xist strongly and specifically interacts with two members of the transcription export (TREX) complex: Alyref and Chtop (Figure 3a).[41] To confirm this interaction, we reanalyzed published iCLIP data for Alyref and Chtop (Figure 3b).[42] As expected, both proteins bound robustly to Xist compared to a negative control. TREX has functions beyond nuclear export such as influencing splice-site and polyadenylation site choice. The complex could be performing these roles on Xist, but the question remains why TREX does not then cause Xist to be exported. Future research may provide insight into specific roles of the TREX complex on Xist and other nuclear RNAs.

We next examined the interaction between the nucleolar protein Nepro and the RNA component of mitochondrial RNA processing endoribonuclease (RMRP) (Figure 3c). RMRP has RNA cleavage roles in both the mitochondria and nucleolus. In yeast, RMRP promotes

the formation of 5.8S rRNA through cleavage of 45S-ITS1.[43–45] Nepro is a protein found in the nucleolus. In humans, mutations in Nepro lead to ribosomopathy, but its precise function remains unknown.[46,47] To map Nepro binding sites, we performed Covalent Linkage Affinity Purification (CLAP) in mESCs.[48,49] In short, Halo-tagged Nepro was expressed from a plasmid in mESCs. The cells were UV crosslinked and lysed, and the lysate was coupled to Halo ligand agarose resin. After stringent washes, RNA was eluted from the agarose with proteinase K and sequenced. CLAP confirmed that Nepro binds to RMRP but did not reveal a clear binding site (Figure 3d). On ribosomal RNA, however Nepro had a strong peak near the 3'-end of ITS1. Future functional research is necessary to identify the functional consequences of this interaction, but the CLAP suggests a model in which Nepro may guide RMRP to cleave specific sites on pre-rRNA.

Our next targets were Fus, Ewsr1, and Taf15 over the U1 snRNA. Fus, Ewsr1, and Taf15 comprise the FET family of RNA-binding transcriptional activators (Figure 3e).[50] Loss of U1 and FET proteins share many phenotypes: reduced transcriptional activation, changes to alternative splicing, and premature cleavage and polyadenyation.[50–52] Consistent with these shared functions, previous research has suggested that the U1 snRNP interacts with both Fus and Taf15.[53–57] To our knowledge, U1's interaction with Ewsr1 has not been previously reported.

To map FET protein binding sites to U1, we reanalyzed our previously published data from Split-Pool Identification of RBP Targets (SPIDR) (Figure 3f).[58] SPIDR revealed binding of Taf15 near the 5'-splice site recognition motif, and of Fus and Ewsr1 to stem-loop 3. Notably, the Fus binding-site matches the results of a recent paper that employed CLIP and NMR to solve a partial Fus-U1 structure.[55] Recent research has also indicated that the interaction

between Fus and U1 may be relevant in the development of Amyloid Lateral Sclerosis (ALS).[55,59] Future research may evaluate if U1 interactions with Taf15 and Ewsr1 have similar clinical relevance.

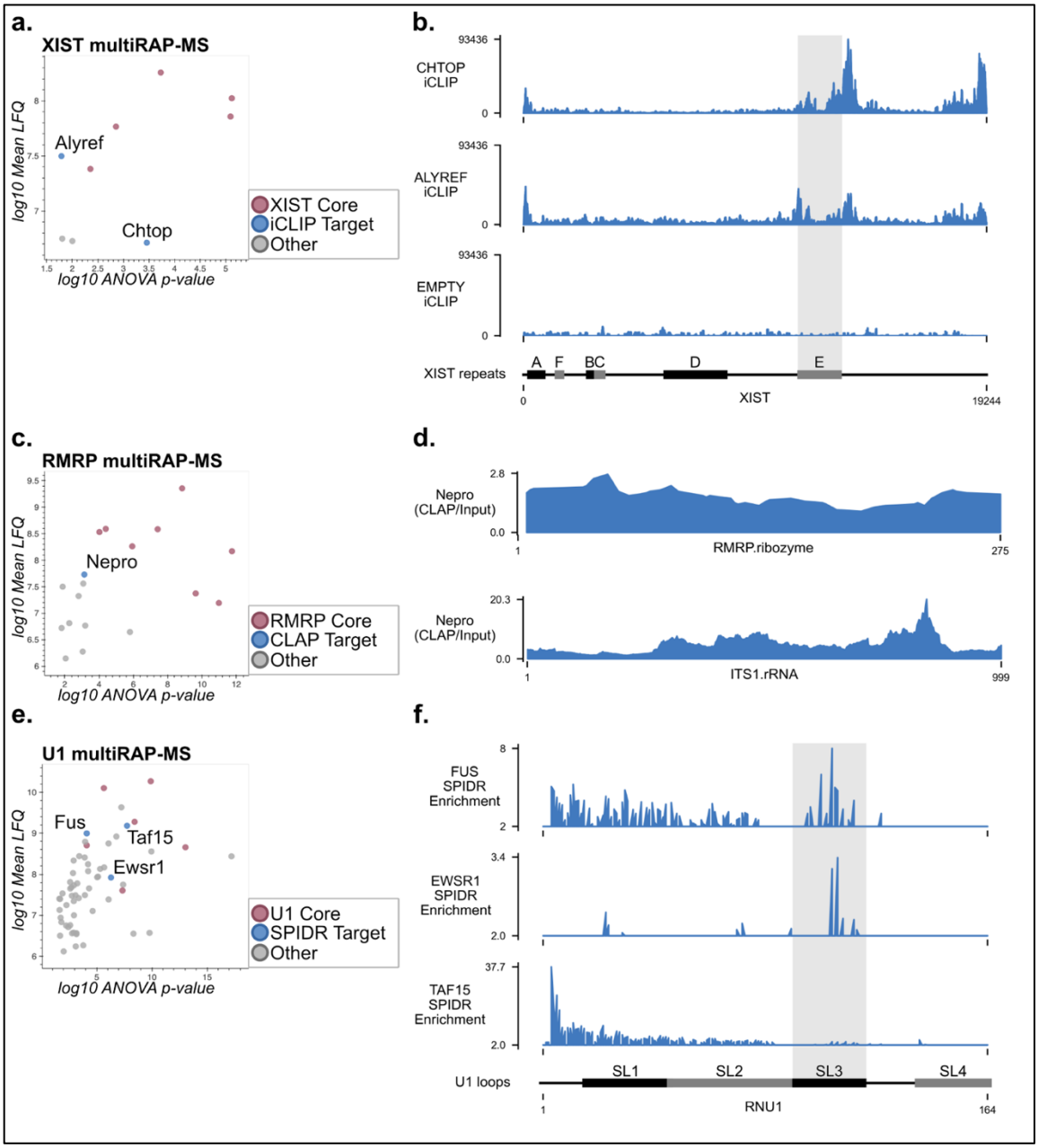

Figure 3: Multi-RAP-MS 2.0 reveals novel RNA-protein interactions. a, Intensity vs ANOVA p-value for significantly differing Xist-bound proteins. In addition to known core proteins, multi-RAP-MS 2.0 reveals TREX components Alyref and Chtop. **b**, anti-FLAG

iCLIP data for Chtop-FLAG, Alyref-FLAG, and Empty-FLAG from Viphakone et al, *Mol Cell* 2019 over human XIST.[42] Plotted data are from a bedgraph with 1nt bins. Both TREX components show 5'-, 3'- and E repeat enrichment. **c**, Intensity vs ANOVA p-value for significantly differing RMRP-bound proteins. In addition to known core proteins, multi-RAP-MS 2.0 reveals the nucleolar protein Nepro. **d**, CLAP data for Nepro over mouse RMRP and 45S-ITS1. Plotted data are from a bedgraph with 1nt bins. **e**, Intensity vs ANOVA p-value for significantly differing U1 snRNA-bound proteins. In addition to known core proteins, multi-RAP-MS 2.0 reveals the FET family of transcriptional regulators: Fus, Ewsr1, and Taf15. **f**, SPIDR enrichment of Fus, Taf15, and Ewsr1 over human U1 snRNA. Data plotted are second read truncations indicating UV crosslinking sites. Fus and Ewsr1 show enrichment over stem-loop 3 (SL3) and Taf15 shows enrichment over the 5'-splice site recognition motif.

DISCUSSION

Enumerating protein components of RNA-protein complexes has long been a challenge in RNA biology. Although many methods have been developed to solve this problem, stringency and cost have remained barriers to discovery. RAP-MS 2.0 overcomes these limitations. The multiple capture strategy improves stringency by reducing recovery of off-target RNAs and proteins. This insight has implications for other antisense purification strategies such as poly-A enrichment for RNA sequencing, in which ribosomal RNA contamination remains an issue due to its abundance and short adenosine stretches. The increased stringency conferred by multiple captures is relevant for other biochemical purifications as well. For example, purification of metabolically labeled RNA or proximity biotinylated proteins with streptavidin also often has high levels of background signal. We have integrated the multiple capture strategy in our 5EU RNA sequencing protocol by eluting

biotinylated RNA from streptavidin beads and then repeating the capture.[60] As with antisense oligos, multiple captures of biotin reduce the level of background, non-nascent RNA and improve detection of nascent transcripts.

The background reduction in RAP-MS 2.0 has allowed us to more confidently call non-core components of RNA-protein complexes. We demonstrated this advantage by identifying snRNA-binding proteins that represent the whole of the snRNA lifecycle, from their biogenesis to the specific snRNA-protein binding dynamics in the spliceosome, and finally through their recycling after release from the spliceosome. Additionally, we identified novel RNA-protein interactions that were previously obscured by background peptides. These novel interactions may provide insights into RNA regulation and human disease. Additionally, future application of RAP-MS 2.0 to recalcitrant RNAs such as repeats (such as LINEs, SINEs, and satellites), vault RNAs, and lncRNAs may deepen our understanding of the biology of these molecules.

LIMITATIONS

A consequence of the reduced background of RAP-MS 2.0 may be a small but significant reduction in signal. This may explain the failure to comprehensively identify all RNP components in all replicates. Care should be taken to optimize cell input counts, RNA fragmentation, and probe selection to limit this error.

A challenge that remains in RNA-proteomics is the identification of protein partners of very low abundance RNAs, such as nascent mRNAs and many lncRNAs. This limitation is largely due to crosslinking efficiency. UV light covalently links RNAs and proteins at a very low rate, and more efficient crosslinkers such as formaldehyde generally do not exclusively link

direct RNA-protein interactions. RAP-MS 2.0 does not solve this problem, but the high stringency of RAP-MS 2.0 will be required when better crosslinking approaches are developed. Each RNA will carry more protein, meaning that non-specific RNA contamination will lead to even higher levels of off-target proteins in the MS data. RAP-MS 2.0's two-capture strategy should reduce the contamination problems that higher crosslinking efficiency would introduce.

DATA AVAILABILITY

All data used in this study are available at Caltech Data, doi 10.22002/x0b3m-r4r34, in the zip folder 20241001_rapMsData.zip. All code to regenerate figures is available on Github at https://github.com/dhonson-lncrna/20241001_RAP-MS-ChapterFigures.git.

ACKNOWLEDGEMENTS

Drew Honson, Mario Blanco, Dev Majumdar, Weixian Deng, James Wohlschlegel, and Mitch Guttman conceived of the project. Ali Palla was involved in early methods development. Weixian Deng developed the peptide purification protocol and performed all LCMS runs and protein quantification. Drew Honson optimized the RAP-MS protocol, performed all RAP-MS experiments, and wrote this report. Mario Blanco and Dev Majumdar provided feedback on drafts of the manuscript.

METHODS

Cell culture

Mouse embryonic stem cell culture

RAP-MS and CLAP experiments were performed on female mouse embryonic stem cells heterozygous for dox-inducible *Xist* (TX1072; gift from E. Heard lab) as previously described.[61] Briefly, TX1072 mESCs were grown on gelatin-coated plates in serum-containing ES cell medium (high glucose DMEM (Gibco, Life Technologies), 15% FBS (Omega Scientific), 2mM L-glutamine (Gibco, Life Technologies), 1mM sodium pyruvate (Gibco, Life Technologies), 0.1mM β-mercaptoethanol, 1000 U/ml leukemia inhibitory factor (LIF, Chemicon), and 2i (3 μM Gsk3 inhibitor CT-99021, 1 μM MEK inhibitor PD0325901). Cell culture medium was replaced every 24 hours.

For RAP-MS experiments targeting *Xist*, TX1072 cells were treated with 2μg/mL doxycycline (Sigma) for 72 hours. Media containing doxycycline was replaced every 24 hours.

Overexpression of Halo-tagged constructs

For CLAP, N-terminal Halo tagged Nepro was expressed from a plasmid containing a CAG promoter. TX1072 cells were lifted, pelleted, and counted. 400,000 cells for each transfection were transferred to 10mL ES cell medium in gelatinized 10-cm plates. Lipofectamine 2000 was used to transfect the cells immediately after plating. 18μg plasmid DNA in 450μl Opti-MEM was mixed with 36μl Lipofectamine 2000 in 450μl OptiMEM and incubated for 5 minutes at room temperature for each set of duplicates. 450μl of the reaction was added to each 10-cm plate. Cells were collected approximately 24 hours after transfection.

UV crosslinking

Cells were removed from the incubator and washed with ice-cold PBS. The PBS was aspirated, the lids of the plates removed, and the plates transferred to a UV crosslinking chamber. For CLAP, cells were treated with $2.5x10^5$ μJ/cm^2 265nm UV light. For RAP-MS, cells were treated with $6.0x10^5$ μJ/cm^2 265nm UV light. Cells were immediately removed from the crosslinker and ice-cold PBS was added to the plates. Cells were scraped and collected in conical tubes, then spun 3 minutes, 330g, 4°C. The supernatant was removed, cells were resuspended in 1mL PBS per 50M cells, and 1mL cell suspension was transferred to Eppendorf tubes. Cells were spun 3 minutes, 1000g, 4°C. The supernatant was removed, cells were snap-frozen in liquid nitrogen and stored at –80°C until use.

RAP-MS 2.0

Probe design

A Python package for probe design is available through pip, which contains a link to the documentation (https://pypi.org/project/probeutils/). Probes for short RNAs (U1, U2, U6, 7SK, 7SL, and RMRP) were 60nt long (53nt antisense plus 7nt of adapter) and for Xist were 90nt long (83nt antisense plus 7nt adapter). FASTA files for each RNA were downloaded from NCBI nucleotide. All probes had a 5'-CAAGTCA adapter. For Xist, non-specific probes were removed by querying BLAT (maximum of 25 off-target matches in mm39) or Dfam (probes containing any *Mus musculus* annotated repeats). No such filtering was performed for the short RNAs, as many have unannotated pseudogenes in mm39 that are not causes for concern. Complete probe sequences are available online. All probes were ordered from Integrated DNA Technologies (IDT) with standard desalting.

Bead preparation

Probes for each RNA were pooled in equimolar ratios to 100µM final. For each RAP, 10 nmol probes were phosphorylated with phosphonucleotide kinase (NEB) in 150µl reactions for 30 minutes at room temperature.

500µl oligo d(T)25 beads (NEB) were washed twice in 500µl Oligo dT Wash Buffer (50 mM HEPES pH 7.5, 300 mM NaCl, 2.5 mM EDTA, pH 8.0, and 0.1% Triton-X 100) then resuspended in 250µl Oligo dT Binding Buffer (50 mM HEPES, pH 7.4, 500 mM LiCl, 2.5 mM EDTA, pH 8.0, and 0.1% Triton-X 100). 10 nmol 1mM polyA bottom ($TGACTTGA_{25}$) was added to the beads and the mixture was incubated 30 minutes, 25°C, shaking at 1000rpm. The beads were then washed three times in 500µl Oligo dT Wash Buffer.

The final wash was removed and the beads were resuspended in the 150µl phosphorylated probe reaction plus 150µl 2x Quick Ligation Buffer (2.5x NEB 5x Quick Ligation Buffer, 0.625x NEB 2x Instant Sticky-End Master Mix, and 18.75% 1,2-propanediol (Sigma)). The ligation was incubated 30 minutes, 25°C, shaking at 1250rpm. The beads were then washed three times in 500µl Oligo dT Wash Buffer.

To remove unligated probe and excess adapter, beads were resuspended in 500µl TE Elution Buffer (20 mM Tris-HCl, pH 7.4, 1 mM EDTA, pH 8.0, and 0.1% sodium dodecyl sulfate) and incubated 2 minutes, 95°C, shaking at 1250rpm. This wash was repeated for a total of three TE washes. After the final wash, beads were washed three times in 4M Urea Hybridization Buffer (4M urea, 10 mM Tris-HCl, pH 7.4, 5 mM EDTA, pH 8.0, 500 mM LiCl, 0.5% Triton-X 100, 0.2% sodium dodecyl sulfate, and 0.1% sodium deoxycholate). After the final wash, beads were resuspended in 250µl 4M Urea Hybridization Buffer and left at room temperature until the first capture.

Lysate preparation

50M cell pellets were removed from -80°C and thawed on ice for 10 minutes. For U1 RAP-MS experiments comparing captures 1 and 2 and comparing RAP-MS 1.0 and 2.0, 50M cells per replicate were used. For multi-RAP-MS experiments, ~85M cells per replicate were used. Each pellet was resuspended in 1mL RAP-MS Lysis Buffer (10 mM Tris-HCl, pH 7.4, 5 mM EDTA, pH 8.0, 500 mM LiCl, 0.5% Triton-X 100, 0.2% sodium dodecyl sulfate, and 0.1% sodium deoxycholate) and incubated on ice for an additional 10 minutes. Pellets were homogenized by triturating the lysate with an 18 gauge needle, then spinning through Qiashredder columns (Qiagen). Lysates were sonicated with a Branson probe sonicator for 42 cycles, 0.7 seconds on, 2.3 seconds off, 4-5W power. After sonication, lysates were spun 20 minutes, 16000g, 4°C and supernatants were transferred to clean tubes. 1mL of 8M Urea Hybridization Buffer (8M urea, 10 mM Tris-HCl, pH 7.4, 5 mM EDTA, pH 8.0, 500 mM LiCl, 0.5% Triton-X 100, 0.2% sodium dodecyl sulfate, and 0.1% sodium deoxycholate) was added to each milliliter of lysate for 4M urea final. Lysates were then combined and pre-warmed to 42°C for the first capture.

First capture

125µl prepared probe-beads were added to the lysate and the mixture was incubated 30 minutes, 42°C, shaking at 1000rpm. The beads were then magnetically separated. For single RAP-MS experiments, the supernatant was discarded. For multi-RAP-MS experiments, the supernatant was transferred to a new tube, magnetically separated for a second time, then transferred to another clean tube. The supernatant was then placed back at 42°C for 5 minutes, the next set of beads was added, and the incubation was repeated.

After removing the supernatant, beads were washed twice in 4M Urea Hybridization Buffer, twice in SDS Wash Buffer (50mM HEPES, pH 7.4, 10% SDS, and 10mM EDTA), and twice in Oligo dT Wash Buffer. Each wash was performed with 500µl buffer for 2 minutes, 37°C, shaking at 1100rpm. The beads were then washed twice in 500µl TE Elution Buffer at room temperature then resuspended in 150µl TE Elution Buffer.

To elute RNA, beads in TE Elution Buffer were heated 3 minutes, 95°C, shaking at 1350rpm. The beads were briefly spun to collect condensation then magnetically separated. The supernatant was transferred to a clean tube and the beads were again suspended in 150µl TE Elution Buffer. The heat elution was repeated, 3 minutes, 95°C, shaking at 1350rpm. After brief spinning and magnetic separation, the second eluent was pooled with the first. For capture 1 and 2 comparisons, 30µl of this eluent was reserved as the capture 1 sample.

Second capture

The first capture elution was adjusted to 500mM LiCl with 8M LiCl, then 1 volume of 8M Urea Buffer was added. 125µl prepared probe-beads were added to the lysate and the mixture was incubated 30 minutes, 42°C, shaking at 1000rpm. The beads were then magnetically separated. Washes in 4M Urea Hybridization Buffer, SDS Wash Buffer, Oligo dT Wash Buffer, and TE Elution Buffer were performed as in the first capture. RNA was eluted from the beads as described for the first capture.

Peptide preparation and LCMS

5µl hydrophobic and 5µl hydrophilic Sera-Mag carboxylate-coated, paramagnetic beads (Cytiva) were washed three times in 500µl LCMS grade water (Fisher) then resuspended in

10µl water. 300µl elutions were reduced and alkylated with 5mM TCEP and 1.2mM IAA 20 minutes, 25°C, shaking at 1200rpm. 10µl washed beads were added to each sample along with 350µl 100% ethanol. The mixture was incubated on a rotator 10 minutes, room temperature to precipitate proteins onto the beads. The beads were then washed three times with 500µl 80% ethanol. After the final wash was removed, the beads were allowed to dry and then were resuspended in 30µl 50mM TEAB. 1µl Lys-C and 2µl Trypsin were added to the beads, and proteins were digested overnight at 37°C.

To recover the peptides, 1mL acetonitrile (ACN) was added to the beads and the suspension was incubated on a rotator 10 minutes, room temperature. The beads were then washed three times in 500µl ACN. After the final wash was removed, the beads were resuspended in 30µl 2% DMSO to elute the peptide. The beads were magnetically separated and the supernatant transferred to a clean tube. Peptides were lyophilized in a speed-vacuum, then resuspended in 10µl 5% formic acid for LCMS injection.

LCMS was performed on an Orbitrap Eclipse and peptides were analyzed using MaxQuant.

Data analysis

In all analyses, known contaminants such as keratins and trypsin that were flagged by MaxQuant were removed. For the capture 1/capture 2 and RAP-MS 1.0/2.0 analyses (Figure 1c-g), label-free quantification (LFQ) values were compared using a Student's T-test. For examining correct assignment of core RNP components (Figure 2b), RNPs were manually annotated using existing literature.[15–23] The data were filtered to only include these components and mean LFQ values between replicates were collected. For each protein, the

mean LFQ signal was normalized per protein by dividing by the maximum mean LFQ signal for that protein.

To assign distinct RNPs to each of the multi-RAP-MS targets, the LFQ values were analyzed by ANOVA. Significantly different proteins were selected using the Benjamini-Hochberg procedure ($p < 0.05$, FDR = 0.05). After removing non-significant proteins, protein were assigned to individual RNAs based on whether the mean LFQ was overrepresented for a given RNA target. The mean LFQ value across replicates was collected for each protein. These means were summed and the fraction of each mean was calculated (e.g., mean LFQ value of Snrnpg for U1 divided by the sum of the mean LFQ values of Snrnpg for all targets). If the protein was overrepresented in a given target, it was assigned to that RNA (e.g., for 8 targets, if the fraction of Snrnpg for U1 was greater than 1/8 of the summed mean signal for Snrnpg, Snrnpg would be assigned to U1). After the proteins were assigned to targets, any proteins not found in both replicates for that target were removed. The log10 ANOVA p-values and log10 mean LFQ values were used for Figures 2c and 3.

RAP-MS 1.0

Probe design

U1 probes were identical to those used in RAP-MS 2.0 except that the 5'-CAAGTCA was replaced with a 5'-biotin. Probes were ordered from IDT with standard desalting.

RNA antisense purification followed by mass spectrometry (RAP-MS)

RAP-MS was performed as previously described.[11] The only modification was that TCA precipitation, HiPPR purification, and HPLC desalting were replaced with the RAP-MS 2.0 Sp3-bead purification method.

Covalent linkage affinity purification (CLAP)

Plasmid construction

Entry vectors for the targeted ORFs were ordered from DNASU and cloned into CLAP destination vectors (PYPP-CAG-Halo-V5-FLAG) using LR clonase. Correct inserts were validated through Sanger sequencing and validated plasmids were prepared with the ZymoPURE II Plasmid Midiprep Kit.

CLAP CLAP was performed as previously described.[48,49] Briefly, cells were lysed, and lysates were coupled to an agarose Halo resin. The resin was washed vigorously at 90°C under a variety of denaturing conditions. RNAs were freed from the resin using Proteinase K (NEB) and prepared into Illumina Sequencing Libraries. Libraries were sequenced on an Illumina NextSeq.

A custom mm10 genome containing repetitive elements and multicopy short RNAs was used for analysis as previously described.[62] Reads were aligned using STAR and fold change over input was calculated using deepTools's bamCompare tool.[63,64]

iCLIP Reanalysis

Data analysis

FASTQ files for Alyref-FLAG, Chtop-FLAG, and Empty-FLAG iCLIP were pulled from GEO GSE113953.[42] Reads were aligned using STAR to a previously described modified hg38 genome containing separate chromosomes for rRNAs and RMRP.[58,64]

SPIDR Reanalysis

Data analysis

Differentially enriched second read truncations from published SPIDR data for Fus, Ewsr1, and Taf15 were plotted over the U1 snRNA gene body.

RAP-MS 2.0 PROTOCOL

INTRODUCTION

The following protocol describes RNA Antisense Purification followed by Mass Spectrometry 2.0 (RAP-MS 2.0), developed by Drew Honson, Weixian Deng, Mario Blanco, Dev Majumdar, and Mitch Guttman. RAP-MS 2.0 is a method to identify proteins that bind directly to an RNA of interest.

In short, cells expressing the RNA of interest are crosslinked with UV light. The cells are lysed and the RNA is purified using antisense oligonucleotide probes covalently linked to a paramagnetic bead. After RAP, the proteins are desalted using carboxylate beads, digested into peptides, and analyzed by liquid chromatography mass spectrometry (LCMS). If a single RNA was targeted, bona fide RNA binding proteins are called based on the strength of their Label Free Quantification (LFQ) intensity and reproducibility between replicates. If multiple

RNAs were recovered from a single lysate, ANOVA is used on LFQ values to identify proteins specific to each RNA.

BUFFERS

All buffers should be made with RNAse-free water. The product used in the Guttman lab can be found in Appendix A. For easy buffer calculations, a downloadable spreadsheet is linked here.

RAP-MS Lysis Buffer

Storage conditions: 4°C for up to 2 weeks without inhibitors. Always add inhibitors fresh.

Formula:

10 mM Tris-HCl, pH 7.5

5 mM EDTA, pH 8.0

500 mM LiCl

0.5% Triton-X 100

0.2% sodium dodecyl sulfate

0.1% sodium deoxycholate

1x c0mplete EDTA-free protease inhibitor

1% Ribolock RNase Inhibitor

8M Urea Hybridization Buffer

Storage conditions: –20°C for up to 1 month.

Formula:

8M urea*

10 mM Tris-HCl, pH 7.5

5 mM EDTA, pH 8.0

500 mM LiCl

0.5% Triton-X 100

0.2% sodium dodecyl sulfate

0.1% sodium deoxycholate

*Note: Not all solid urea is easily soluble to 8M. Consult section 1.8.2 for the product used in the Guttman lab.

4M Urea Hybridization Buffer

Storage conditions: –20°C for up to 1 month.

Formula:

1 volume RAP-MS Lysis Buffer

1 volume 8M urea hybridization buffer

SDS Wash Buffer

Storage conditions: Room temperature for up to 1 month.

Formula:

50 mM HEPES, pH 7.4

10 mM EDTA, pH 8.0

10% sodium dodecyl sulfate

Oligo dT Binding Buffer

Storage conditions: Room temperature for up to 1 month.

Formula:

50 mM HEPES, pH 7.4

500 mM LiCl

2.5 mM EDTA, pH 8.0

0.1% Triton-X 100

Oligo dT Wash Buffer

Storage conditions: Room temperature for up to 1 month.

Formula:

50 mM HEPES pH 7.4

300 mM NaCl

2.5 mM EDTA, pH 8.0

0.1% Triton-X 100

TE Elution Buffer

Storage conditions: Room temperature for up to 1 month.

Formula:

20 mM Tris-HCl, pH 7.5

1 mM EDTA, pH 8.0

0.1% sodium dodecyl sulfate

2x Quick Ligation Buffer

Storage conditions: –20°C for several months.

Formula:

2.5x NEB 5x Quick Ligation Buffer

0.625x NEB 2x Instant Sticky-End Master Mix

18.75% 1,2-propanediol

Amine-free Lysis Buffer (for UV crosslinking test)

Storage conditions: 4°C for up to 2 weeks without inhibitors. Always add inhibitors fresh.

Formula:

50mM HEPES, pH 7.4

5mM EDTA, pH 8.0

500mM LiCl

0.5% Triton-X 100

0.2% sodium dodecyl sulfate

0.1% sodium deoxycholate

1x c0mplete EDTA-free protease inhibitor

1% Ribolock RNase Inhibitor

Quenching Buffer (for UV crosslinking test)

Storage conditions: Room temperature for several months

Formula:

500mM Tris-HCl, pH 7.5

10mM EDTA

PROBE DESIGN

Oligonucleotide probes for RAP-MS 2.0 should be 60-90 nucleotides long and cover at least 50% of the target RNA. Ideally, probes would tile 100% of the target, but in practice repetitive elements and cost constraints impose limitations. A Python package for probe design is distributed through pip, and complete documentation is available at Drew Honson's website. For RAP-MS 2.0, the "adaptseq" parameter should be set to "CAAGTCA" and "biotin" to "False". We strongly recommend setting both "blat" and "dfam" to "True", unless a repetitive element is being intentionally targeted.

CELL CULTURE AND FIXATION

Protocol

1. Grow 20-100M cells depending on target abundance. See notes for additional details.
2. For adherent cells grown on plates:
    a. Aspirate media and wash the cells in an appropriate volume of ice-cold PBS.
    b. Aspirate the PBS and immediately transfer the plates a UV crosslinker. Remove the plate lids and crosslink with 265nm UV light, $6x10^5$ $\mu J/cm^2$.
    c. Add appropriate volume of PBS to the plates. Use a cell scraper to detach cells and transfer the cell suspension to a conical tube on ice.

Rinse the plate with ice-cold PBS and pool the wash with the suspension. If working with many plates, keep plates awaiting scraping on ice.

3. For adherent cells grown in T flasks, or three-dimensional cultures (see notes):
    a. Detach and pellet cells as is appropriate for the cell line.
    b. Resuspend cells in 15mL ice-cold PBS per 20M cells and distribute into 15-cm plates, 15mL suspension per plate.
    c. Transfer the plates a UV crosslinker. Remove the plate lids and crosslink with 265nm UV light, $6x10^5$ μJ/cm^2.
    d. Transfer the crosslinked cell suspension into conical tubes on ice. Rinse plates with ice-cold PBS and pool the wash with the suspension.
4. For suspension cells:
    a. Pellet cells as is appropriate for the cell line.
    b. Resuspend cells in 15mL ice-cold PBS per 20M cells and distribute into 15-cm plates, 15mL suspension per plate.
    c. Transfer the plates a UV crosslinker. Remove the plate lids and crosslink with 265nm UV light, $6x10^5$ μJ/cm^2.
    d. Transfer the crosslinked cell suspension into conical tubes on ice. Rinse plates with ice-cold PBS and pool the wash with the suspension.
5. Pellet cell suspensions 3min, 500g, 4°C.
6. Remove the supernatant. Resuspend cells in 1mL ice-cold PBS per 50M cells and distribute into 1.7mL Eppendorf tubes, 1mL per tube.
7. Pellet cell suspensions 3min, 500g, 4°C.

8. Remove the supernatant and flash-freeze the pellets in liquid nitrogen. Transfer the tubes to –80°C for storage. Pellets are stable for many months at –80°C.

Notes

1. 20-50M cells is sufficient for high abundance targets such as most snRNAs, but 100M is recommended for lower abundance targets such as dox-induced Xist (~2000 molecules per cell). Proteomes of lower abundance targets than dox-induced Xist have not been successfully characterized with RAP-MS 2.0. Exceeding 200M cells per replicate is not recommended, as the lysate generally becomes too viscous to easily manage.
2. UV crosslinking is blocked by plastic, so T flasks cannot be crosslinked directly. Additionally, the efficiency of UV crosslinking can be diminished by three-dimensional structures (e.g., organoids) or pigments. When optimizing the protocol, we strongly recommend performing the UV crosslinking test in section 1.8.6 on a 1–10M cell sample to confirm that the crosslinking is working as expected.

UV CROSSLINKING TEST (OPTIONAL)

Lysis and Capture

1. Prepare 1-10M cell pellets as described in the Cell Culture and Fixation Section. Two samples should be crosslinked and two should be uncrosslinked controls.
2. Resuspend cells in 200µl ice-cold Amine-free Lysis Buffer. Incubate on ice, 10 minutes.

3. Spin sample 10 min, 15000g, 4°C. Transfer supernatant to a clean tube and discard pellet.
4. Measure protein concentration of each sample using either Qubit or a BCA Assay. For BSA assays, 200µl reactions using a 1:10 dilute sample has worked well.
5. Aliquot equal masses of protein for each sample (0.3-0.6mg is sufficient) and bring the sample volumes to 300µl with ice-cold Amine-free Lysis Buffer. Additionally, reserve a 10% input for each sample.
6. Magnetically separate 50µl NHS beads per sample.
7. Remove the supernatant and resuspend the beads in 500µl ice-cold 1mM HCl. Vortex for 15 seconds.
8. Magnetically separate the beads and remove the supernatant. Wash the beads twice in 500µl ice-cold PBS.
9. Divide the beads into tubes for each sample, 50µl bead equivalents per tube. Magnetically separate the beads and remove the supernatant, then immediately add 300µl normalized lysate on top of the beads.
10. Vortex the samples for 15 seconds, then seal the tubes with Parafilm and incubate on a rotator overnight, 4°C.

Washes and Elution

1. Magnetically separate the beads and reserve the flow through. Immediately resuspend the beads in 500µl Quenching Buffer.
2. Incubate tubes on a rotator 45 min, room temperature.
3. Magnetically separate the beads and discard the supernatant.

4. Wash the beads twice in each of the following buffers. Washes should be performed 2 min, 50°C, shaking at 1000rpm.
    a. 4M Urea Buffer
    b. SDS Wash Buffer
    c. Oligo dT Wash Buffer
    d. TE Elution Buffer
5. After the final wash, magnetically pellet the beads and discard the supernatant.
6. Resuspend the beads in 5µl 10x Turbo DNase Buffer + 5µl Turbo DNase + 2µl Ribolock RNase Inhibitor + 38µl water. Prepare the input sample in the same way, adjusting the water volume to account for the sample. Incubate 30 min, 37°C, shaking at 1000rpm to digest any residual genomic DNA.
7. Add 40µl TE Elution Buffer + 10µl Proteinase K. Incubate Proteinase K reactions 1 h, 37°C, shaking at 1000rpm to digest protein.
8. Clean RNA using the Zymo RNA Clean and Concentrate Kit and elute in 25µl water.

Read Out

1. Measure RNA concentrations by Qubit or Nanodrop. Additionally, measure RNA integrity by RNA HS TapeStation or a similar electrophoresis-based method.
2. A typical result is that the crosslinked samples will recover 1–5% of the input RNA. In contrast, the uncrosslinked samples will usually have undetectable amounts of RNA or yields substantially below 1%.

3. If the crosslinked condition yielded no RNA, check the RNA integrity of the flow through. If flow through RNA is highly fragmented then the issue may be RNA deterioration due to RNase contamination or alkaline buffers rather than crosslinking.

Notes

1. To thoroughly test a UV crosslinker, a titration using 0, 200, 400, 600, and 800 mJ/cm^2 can be performed. We usually see an increase in crosslinking efficiency from 200–600 followed by a plateau at 800.
2. To test a new crosslinking chamber, use a cell culture that forms epithelia such as HEK293 or NIH3T3 as three-dimensional structures can inhibit crosslinking.
3. To crosslink three-dimensional structures such as organoids, consider dissociating the cells before crosslinking. Resuspend the dissociated cells in PBS or another isotonic buffer and add to a cell culture plate such that the depth of the PBS never exceeds 0.5cm. A crosslinking titration can then be performed.
4. This protocol is also compatible with certain chemical crosslinkers. Be aware, however, that crosslinkers incorporating an NHS ester (such as DSG or DSP) and heat reversible crosslinkers (such as formaldehyde) are incompatible with this approach.

BEAD PREPARATION

Probe Phosphorylation

1. Reaction (per 500µl oligo dT beads)

100µl 100µM equimolar probe pool

15µl 10x PNK buffer

7.5µl PNK

7.5µl 100mM ATP

20µl RNase-free water

2. Incubate 30min, 25°C, shaking at 1000rpm.

PolyA Adapter Hybridization

1. Magnetically separate 500µl oligo-dT beads per 100M cells and remove supernatant.
2. Wash the beads twice with 500µl Oligo dT Wash Buffer. Add the wash buffer to beads, rotate the tube 180° to pull the beads through the wash. Repeat the rotation 2-3 times to thoroughly wash the beads, then magnetically separate them and remove the supernatant.
3. Resuspend the beads in 100µl Oligo dT Binding Buffer, then add 10µl 1mM polyA bottom.
4. Incubate 30 min, 25°C, shaking at 1000rpm.
5. Magnetically separate the beads and wash three times with 500µl Oligo dT Wash Buffer. Keep beads in 500µl Oligo dT Wash Buffer until ligation.

Probe Ligation

1. Mix 150µl probe phosphorylation reaction with 150µl 2x Quick Ligation Buffer.
2. Magnetically separate polyA-bottom hybridized beads and remove supernatant.

3. Immediately resuspend beads in probe-ligase mix. Incubate 30 min, 25°C, shaking at 1200rpm.
4. Wash the beads three times with 1mL Oligo dT Wash Buffer.
5. Resuspend the beads in 500µl TE elution buffer, then transfer the tubes to a block and incubate 2 minutes, 95°C, 1350rpm. This removed unligated probe.
6. Magnetically separate the beads and remove the supernatant. Repeat steps 4-6 for a total of two washes.
7. After removing the final wash, resuspend the beads in 250µl 4M Urea Hybridzation Buffer and keep at room temperature until hybridization.

Notes

1. We have measured the capacity of oligo-dT beads to be approximately 20pmol oligo per 1µl beads.
2. Beads should be prepared fresh the day of the RAP experiment.
3. Removal of unligated probe also exposes the dT(25) covalently linked to the bead. Although this could theoretically enrich for polyA RNAs, the hybridization temperatures are high enough that we have not encountered this problem.

LYSATE PREPARATION

Protocol

1. Prepare RAP-MS Lysis Buffer with RNase and protease inhibitors and chill on ice until cold.
2. Remove cell pellets from -80°C and thaw on ice, 10 min.

3. Resuspend pellets in 1mL ice-cold RAP-MS Lysis Buffer and incubate on ice, 20 minutes.
4. To solubilize chromatin, apply 500µl lysate to a Qiashredder column and spin 1 min, 21000g, 4°C. Repeat with the remaining lysate and pool shredded lysates.
    a. For long RNAs, sonication may be beneficial at this stage. See notes for details.
5. Add 1 volume of 8M urea hybridization buffer to the lysate and pipette up and down to mix.
6. Transfer tubes to a centrifuge and spin 10 min, 14000g, room temperature to pellet insoluble debris. Transfer supernatant to clean tubes.
7. Collect 0.1-1% input of lysate for RNA analysis. Store at -80°C until ready for use.

Notes

1. Fragmenting RNA through sonication may improve capture for some long RNAs, though this has not been thoroughly tested. We recommend fragmenting RNA to 300-2000nt fragments to test this for individual RNAs. The appropriate treatment will vary greatly based on sonicator and cell type and will need to be empirically determined for new types of samples.

RNA ANTISENSE PURIFICATION

First Capture

1. Preheat all wash buffers to 37°C and lysate to 42°C.

2. Add 125µl probe-beads (50% of the total prepared) to the lysate. Incubate 30 min, 42°C, shaking at 1000rpm.
3. Magnetically separate the beads.
4. If capturing multiple RNA species:
    a. Transfer the supernatant to a clean tube and magnetically separate a second time. Resuspend the beads from the first capture in 500µl 4M Urea Hybridization Buffer and keep on ice until washes.
    b. Repeat steps 1-4 for subsequent RNA captures, then proceed to washes with all first captures.
5. Perform 500µl washes, 2 min, 37°C, 1000rpm.
    a. Twice in 4M Urea Hybridization Buffer
    b. Twice in SDS Wash Buffer
    c. Twice in Oligo dT Wash Buffer
6. Wash beads twice in 500µl TE Elution Buffer at room temperature and transfer beads to clean tube.
7. Magnetically separate the beads, discard the supernatant, and resuspend beads in 150µl TE buffer.
8. Elute RNA by heating 3 min, 95°C, 1350rpm.
9. Immediately pop-spin the tubes, magnetically separate the beads, and transfer the supernatant containing RNA-protein complexes to a new tube.
10. Resuspend the beads in 150µl TE buffer, then repeat steps 6 and 7. Pool the second elution with the first. After the second elution, the beads can be discarded.

Second Capture

1. Adjust the eluent to 500mM LiCl by adding 18.75µl 8M LiCl.
2. Add 320µl 8M Urea Buffer and pre-heat the solution to 65°C.
3. Add 125µl probe-beads (the remaining 50%) to the lysate. Incubate 30 min, 65°C, shaking at 1000rpm
4. Perform 500µl washes, 2 min, 37°C, 1000rpm.
    a. Twice in 4M Urea Hybridization Buffer
    b. Twice in SDS Wash Buffer
    c. Twice in Oligo dT Wash Buffer
5. Wash beads twice in 500µl TE Elution Buffer at room temperature and transfer beads to clean tube.
6. Magnetically separate the beads, discard the supernatant, and resuspend beads in 150µl TE buffer.
7. Elute RNA by heating 3 min, 95°C, 1350rpm.
8. Immediately pop-spin the tubes, magnetically separate the beads, and transfer the supernatant containing RNA-protein complexes to a new tube.
9. Resuspend the beads in 150µl TE buffer, then repeat steps 6 and 7. Pool the second elution with the first. After the second elution, the beads can be discarded.
10. Reserve a 10% aliquot for RNA analysis. Store the remaining 90% at -20°C until ready to proceed to protein purification.

Notes:

1. Capture of multiple RNA species has been successfully performed for up to four RNAs. Theoretically it is possible to do more, but that has not been tested.

RNA CLEAN-UP AND ANALYSIS

Protocol

1. Follow instructions for Zymo RNA Clean and Concentrate using a IIC column for both input and 10% second capture samples. Elute in 40µl RNase-free water.
2. DNase Reaction

   40µl sample

   5µl 10x Turbo DNase Buffer

   3µl Turbo DNase

   2µl ExoI

   1µl Ribolock RNase Inhibitor
3. Incubate 45 min, 37°C, shaking at 1000rpm.
4. Clean using Zymo RNA Clean and Concentrate using a IC column. Elute in 30µl RNase-free water.
5. Use 50% of the RNA for a cDNA reaction using random priming. Perform RT-qPCR for the target of interest, a ribosomal RNA (such as 18S or 28S), and an abundant mRNA (such as GAPDH or Beta-actin).
6. The RT-qPCR results can be analyzed with a modified ΔΔCt analysis

a. Calculate ΔCt (target – control) for each target and control in input and capture 2. For example, if in the input U1 had a Ct value of 20 and 18S had a Ct value of 15, then ΔCt for the input would be 5.
b. Calculate the fold expression difference between target and controls by taking $2^{\Delta Ct}$. Continuing the above example, you would interpret ΔCt = 5 to indicate a 32-fold difference between U1 and 18S in the input.
c. Subtract fold expression difference ($2^{\Delta Ct}$) in the input from the control, then divide by fold expression difference of the control to calculate the background depletion value. For example, if $2^{\Delta Ct}$ for input was 32 and $2^{\Delta Ct}$ in the capture was 1, then the background depletion is (32 – 1)/32 = 0.97 or 97% depletion.

Notes

1. If Ct values for the target are suspiciously low in capture 2 (for example, if they indicate greater than 100% yield), the RT reaction may have probe contamination. This can result from incomplete removal of unligated probe during bead preparation or incomplete DNase during RNA preparation.
2. RNA-seq can also be performed to analyze the enrichment, which has the advantage of detecting aberrant enrichment of RNAs with homology. Any standard RNA-seq library prep should be compatible with the post-DNase purified RNA, but the RNA may require additional fragmentation to reach appropriate insert sizes.

3. Depletion values should be >95% for any off-target RNA examined. >99% depletion is not unusual, especially for abundant mRNAs which are often undetectable by RT-qPCR in capture 2.

PROTEIN PURIFICATION AND DESALTING

Protocol

1. Mix 5µl hydrophilic and 5µl hydrophobic Sp3 carboxylate beads. Magnetically separate beads and wash three times with 500µl LCMS-quality water, then resuspend in 10µl LCMS-quality water.
2. Add 7.5µl 200mM TCEP and 7.25µl 500mM IAA to each sample and incubate 20 min, 25°C, shaking at 1200rpm.
3. Add beads to the sample, then add 350µl 100% ethanol. Incubate on a rotator 10 min, room temperature.
4. Magnetically separate beads and discard supernatant.
5. Wash beads three times with 500µl 80% ethanol. After removing the final wash, pop-spin the tube, magnetically separate the beads, and remove any residual ethanol from the bottom.
6. Allow the beads to dry, usually 3–5 minutes.
7. Resuspend the beads in 30µl 50mM TEAB, then add 1µl Lys-C and 2µl Trypsin to each sample. The beads should stay in the sample during protease digestion.
8. Incubate 37°C, overnight.

9. Add 1mL acetonitrile (ACN) and incubate on a rotator 10 min, room temperature.
10. Magnetically separate the beads and discard the supernatant.
11. Wash the beads three times with 500µl ACN. After removing the final wash, pop-spin the tube, magnetically separate the beads, and remove any residual ACN from the bottom.
12. The beads should dry almost instantaneously, but double check that they have dried before proceeding to elution.
13. Elute peptides by resuspending beads in 30µl 2% DMSO. Magnetically separate the beads and transfer the supernatant to a fresh tube.
14. Use a speed vacuum to lyophilize the peptides, then resuspend in 10µl 5% formic acid or another suitable solvent for LCMS injection.

Notes

1. A mixture of hydrophobic and hydrophilic beads is required to purify all protein species.
2. Alternative reduction and alkylation methods are likely compatible but have not been tested.

LIQUID CHROMATOGRAPHY MASS SPECTROMETRY

After lyophilization, the peptides are desalted and free of detergents. We have successfully used Orbitrap Fusion and Orbitrap Eclipse mass spectrometers, but other set-ups are likely compatible with the protocol. We use MaxQuant analysis software for peptide search and the resultant Label Free Quantification (LFQ) values for analysis.

DATA ANALYSIS

Begin data processing by filtering out known contaminants (keratins, trypsins, etc.). Given the low peptide yields of RAP-MS 2.0 (generally not measurable except by HPLC), most of the total signal will be represented by environmental contaminants. Next, remove all proteins that were not detected in all replicates. This harsh filtering prevents spurious calling of replicate-specific background. Next, examine the distribution of LFQ values. Typically, the top 10-20% of hits will represent most of the total signal. If the distribution is fairly even across all hits, the experiment may have failed due to poor crosslinking or weak recovery of the target RNA. These are the high confidence hits, whereas the remaining 80-90% of proteins may be genuine but are more likely to be abundant, non-specific RNA or ssDNA binding proteins. These non-specific proteins may be highly reproducible because of their affinity to probes, beads, or RNA, but are unlikely to be of biological interest.

For multi-RAP-MS, the above analysis can be applied to each target individually, but proteins specific to each RNA can be identified by comparing hits across targets. To do this, perform the contaminant and reproducibility filtering described above, then compare LFQ values across targets using ANOVA. Apply the Benjamini-Hochberg procedure with the desired FDR (1-10% FDR gave similar results in our multi-RAP-MS experiment). Remove any hits below the desired significance level. To assign proteins to individual RNAs, take the mean LFQ value of the replicates for each protein. Divide this by the sum of the means of all RNA targets for that protein. If a protein is disproportionately represented in RAP-MS for a given target (i.e., the LFQ fraction is greater than one over the total number of targets), assign it to that RNA. We have also used post-hoc Tukey tests for protein assignment, but this is problematic for proteins that are shared between two targets (e.g., Sm ring proteins for U1

and U2). The proteins assigned by this method will represent the unique RNP components for each target, but will exclude bona fide RNP components that do not differ significantly across the targets.

REFERENCES

1. C. Quaresma, A. J., Bugai, A. & Barboric, M. Cracking the control of RNA polymerase II elongation by 7SK snRNP and P-TEFb. *Nucleic Acids Res.* **44**, 7527–7539 (2016).

2. Darnell, J. E. Reflections on the history of pre-mRNA processing and highlights of current knowledge: A unified picture. *RNA* **19**, 443–460 (2013).

3. Su, Y. *et al.* Spliceosomal snRNAs, the essential players in pre-mRNA processing in eukaryotic nucleus: From biogenesis to functions and spatiotemporal characteristics. *Adv. Biol.* **8**, 2400006 (2024).

4. Gräwe, C., Stelloo, S., van Hout, F. A. H. & Vermeulen, M. RNA-Centric Methods: Toward the interactome of specific RNA transcripts. *Trends Biotechnol.* **39**, 890–900 (2021).

5. Dodel, M. *et al.* TREX reveals proteins that bind to specific RNA regions in living cells. *Nat. Methods* **21**, 423–434 (2024).

6. Munschauer, M. *et al.* The NORAD lncRNA assembles a topoisomerase complex critical for genome stability. *Nature* **561**, 132–136 (2018).

7. Schmidt, N. *et al.* The SARS-CoV-2 RNA–protein interactome in infected human cells. *Nat. Microbiol.* **6**, 339–353 (2021).

8. Gandhi, M. *et al.* The lncRNA lincNMR regulates nucleotide metabolism via a YBX1 - RRM2 axis in cancer. *Nat. Commun.* **11**, 3214 (2020).

9. Flynn, R. A. *et al.* Discovery and functional interrogation of SARS-CoV-2 RNA-host protein interactions. *Cell* **184**, 2394-2411.e16 (2021).

10. Wang, X. *et al.* The long non-coding RNA CYTOR drives colorectal cancer progression by interacting with NCL and Sam68. *Mol. Cancer* **17**, 110 (2018).

11. McHugh, C. A. & Guttman, M. RAP-MS: A Method to Identify Proteins that Interact Directly with a Specific RNA Molecule in Cells. in *RNA Detection: Methods and Protocols* (ed. Gaspar, I.) 473–488 (Springer, New York, NY, 2018). doi:10.1007/978-1-4939-7213-5_31.

12. Lührmann, R., Kastner, B. & Bach, M. Structure of spliceosomal snRNPs and their role in pre-mRNA splicing. *Biochim. Biophys. Acta BBA - Gene Struct. Expr.* **1087**, 265–292 (1990).

13. Pomeranz Krummel, D. A., Oubridge, C., Leung, A. K. W., Li, J. & Nagai, K. Crystal structure of human spliceosomal U1 snRNP at 5.5 Å resolution. *Nature* **458**, 475–480 (2009).

14. Stark, H., Dube, P., Lührmann, R. & Kastner, B. Arrangement of RNA and proteins in the spliceosomal U1 small nuclear ribonucleoprotein particle. *Nature* **409**, 539–542 (2001).

15. McHugh, C. A. *et al.* The Xist lncRNA interacts directly with SHARP to silence transcription through HDAC3. *Nature* **521**, 232–236 (2015).

16. Luirink, J. & Sinning, I. SRP-mediated protein targeting: Dtructure and function revisited. *Biochim. Biophys. Acta BBA - Mol. Cell Res.* **1694**, 17–35 (2004).

17. Jarrous, N. Roles of RNase P and its subunits. *Trends Genet.* **33**, 594–603 (2017).

18. Schümperli, D. & Pillai, R. S. The special Sm core structure of the U7 snRNP: Far-reaching significance of a small nuclear ribonucleoprotein. *Cell. Mol. Life Sci. CMLS* **61**, 2560–2570 (2004).

19. Brogie, J. E. & Price, D. H. Reconstitution of a functional 7SK snRNP. *Nucleic Acids Res.* **45**, 6864–6880 (2017).

20. Barrandon, C., Bonnet, F., Nguyen, V. T., Labas, V. & Bensaude, O. The transcription-dependent dissociation of P-TEFb-HEXIM1-7SK RNA relies upon formation of hnRNP-7SK RNA complexes. *Mol. Cell. Biol.* **27**, 6996–7006 (2007).

21. Zhang, Z. *et al.* Molecular architecture of the human 17S U2 snRNP. *Nature* **583**, 310–313 (2020).

22. Scofield, D. G. & Lynch, M. Evolutionary diversification of the Sm family of RNA-associated proteins. *Mol. Biol. Evol.* **25**, 2255–2267 (2008).

23. Montemayor, E. J. *et al.* Architecture of the U6 snRNP reveals specific recognition of 3′-end processed U6 snRNA. *Nat. Commun.* **9**, 1749 (2018).

24. Lin, P.-C. & Xu, R.-M. Structure and assembly of the SF3a splicing factor complex of U2 snRNP. *EMBO J.* **31**, 1579–1590 (2012).

25. Sun, C. The SF3b complex: Splicing and beyond. *Cell. Mol. Life Sci.* **77**, 3583–3595 (2020).

26. Sharma, S., Wongpalee, S. P., Vashisht, A., Wohlschlegel, J. A. & Black, D. L. Stem–loop 4 of U1 snRNA is essential for splicing and interacts with the U2 snRNP-specific SF3A1 protein during spliceosome assembly. *Genes Dev.* **28**, 2518–2531 (2014).

27. Parker, M. T. *et al.* m6A modification of U6 snRNA modulates usage of two major classes of pre-mRNA 5' splice site. *eLife* **11**, e78808 (2022).

28. Gubitz, A. K., Feng, W. & Dreyfuss, G. The SMN complex. *Exp. Cell Res.* **296**, 51–56 (2004).

29. Pendleton, K. E. *et al.* The U6 snRNA m6A methyltransferase METTL16 regulates SAM synthetase intron retention. *Cell* **169**, 824-835.e14 (2017).

30. Yamashita, S., Takagi, Y., Nagaike, T. & Tomita, K. Crystal structures of U6 snRNA-specific terminal uridylyltransferase. *Nat. Commun.* **8**, 15788 (2017).

31. Mellman, D. L. *et al.* A PtdIns4,5P2-regulated nuclear poly(A) polymerase controls expression of select mRNAs. *Nature* **451**, 1013–1017 (2008).

32. Trippe, R. *et al.* Identification, cloning, and functional analysis of the human U6 snRNA-specific terminal uridylyl transferase. *RNA* **12**, 1494–1504 (2006).

33. Bell, M., Schreiner, S., Damianov, A., Reddy, R. & Bindereif, A. p110, a novel human U6 snRNP protein and U4/U6 snRNP recycling factor. *EMBO J.* **21**, 2724–2735 (2002).

34. Wahl, M. C., Will, C. L. & Lührmann, R. The spliceosome: Design principles of a dynamic RNP machine. *Cell* **136**, 701–718 (2009).

35. Zhang, X. *et al.* Structure of the human activated spliceosome in three conformational states. *Cell Res.* **28**, 307–322 (2018).

36. Wachutka, L., Caizzi, L., Gagneur, J. & Cramer, P. Global donor and acceptor splicing site kinetics in human cells. *eLife* **8**, e45056.

37. Bertram, K. *et al.* Structural insights into the roles of metazoan-specific splicing factors in the human step 1 spliceosome. *Mol. Cell* **80**, 127-139.e6 (2020).

38. Zhan, X., Yan, C., Zhang, X., Lei, J. & Shi, Y. Structure of a human catalytic step I spliceosome. *Science* **359**, 537–545 (2018).

39. Hogg, R., McGrail, J. C. & O'Keefe, R. T. The function of the NineTeen Complex (NTC) in regulating spliceosome conformations and fidelity during pre-mRNA splicing. *Biochem. Soc. Trans.* **38**, 1110–1115 (2010).

40. Mattick, J. S. *et al.* Long non-coding RNAs: Definitions, functions, challenges and recommendations. *Nat. Rev. Mol. Cell Biol.* **24**, 430–447 (2023).

41. Katahira, J. mRNA export and the TREX complex. *Biochim. Biophys. Acta BBA - Gene Regul. Mech.* **1819**, 507–513 (2012).

42. Viphakone, N. *et al.* Co-transcriptional loading of RNA export factors shapes the human transcriptome. *Mol. Cell* **75**, 310 (2019).

43. Henry, Y. *et al.* The 5′ end of yeast 5.8S rRNA is generated by exonucleases from an upstream cleavage site. *EMBO J.* **13**, 2452–2463 (1994).

44. Mitchell, P., Petfalski, E. & Tollervey, D. The 3' end of yeast 5.8S rRNA is generated by an exonuclease processing mechanism. *Genes Dev.* **10**, 502–513 (1996).

45. Lygerou, Z., Allmang, C., Tollervey, D. & Séraphin, B. Accurate processing of a eukaryotic precursor ribosomal RNA by ribonuclease MRP in vitro. *Science* **272**, 268–270 (1996).

46. Robertson, N. *et al.* A disease-linked lncRNA mutation in RNase MRP inhibits ribosome synthesis. *Nat. Commun.* **13**, 649 (2022).

47. Narayanan, D. L. *et al.* An emerging ribosomopathy affecting the skeleton due to biallelic variations in NEPRO. *Am. J. Med. Genet. A.* **179**, 1709–1717 (2019).

48. Guo, J. K. *et al.* Denaturing purifications demonstrate that PRC2 and other widely reported chromatin proteins do not appear to bind directly to RNA *in vivo*. *Mol. Cell* **84**, 1271-1289.e12 (2024).

49. Banerjee, A. K. *et al.* SARS-CoV-2 Disrupts splicing, translation, and protein trafficking to suppress host defenses. *Cell* **183**, 1325-1339.e21 (2020).

50. Schwartz, J. C., Cech, T. R. & Parker, R. R. Biochemical properties and biological functions of FET proteins. *Annu. Rev. Biochem.* **84**, 355–379 (2015).

51. Kaida, D. *et al.* U1 snRNP protects pre-mRNAs from premature cleavage and polyadenylation. *Nature* **468**, 664–668 (2010).

52. Mimoso, C. A. & Adelman, K. U1 snRNP increases RNA Pol II elongation rate to enable synthesis of long genes. *Mol. Cell* **83**, 1264-1279.e10 (2023).

53. So, B. R. *et al.* A complex of U1 snRNP with cleavage and polyadenylation factors controls telescripting, regulating mRNA transcription in human cells. *Mol. Cell* **76**, 590-599.e4 (2019).

54. Leichter, M. *et al.* A fraction of the transcription factor TAF15 participates in interactions with a subset of the spliceosomal U1 snRNP complex. *Biochim. Biophys. Acta-Proteins Proteomics* **1814**, 1812–1824 (2011).

55. Jutzi, D. *et al.* Aberrant interaction of FUS with the U1 snRNA provides a molecular mechanism of FUS induced amyotrophic lateral sclerosis. *Nat. Commun.* **11**, 6341 (2020).

56. Jobert, L. *et al.* Human U1 snRNA forms a new chromatin-associated snRNP with TAF15. *EMBO Rep.* **10**, 494–500 (2009).

57. Yu, Y. & Reed, R. FUS functions in coupling transcription to splicing by mediating an interaction between RNAP II and U1 snRNP. *Proc. Natl. Acad. Sci. U. S. A.* **112**, 8608–8613 (2015).

58. Wolin, E. *et al.* SPIDR: A highly multiplexed method for mapping RNA-protein interactions uncovers a potential mechanism for selective translational suppression upon cellular stress. *BioRxiv Prepr. Serv. Biol.* 2023.06.05.543769 (2023) doi:10.1101/2023.06.05.543769.

59. Yu, Y. *et al.* U1 snRNP is mislocalized in ALS patient fibroblasts bearing NLS mutations in FUS and is required for motor neuron outgrowth in zebrafish. *Nucleic Acids Res.* **43**, 3208–3218 (2015).

60. Bhat, P., Honson, D. & Guttman, M. Nuclear compartmentalization as a mechanism of quantitative control of gene expression. *Nat. Rev. Mol. Cell Biol.* **22**, 653–670 (2021).

61. Schulz, E. G. *et al.* The two active X chromosomes in female ESCs block exit from the pluripotent state by modulating the ESC signaling network. *Cell Stem Cell* **14**, 203–216 (2014).

62. Quinodoz, S. A. *et al.* RNA promotes the formation of spatial compartments in the nucleus. *Cell* **184**, 5775-5790.e30 (2021).

63. Ramírez, F. *et al.* deepTools2: A next generation web server for deep-sequencing data analysis. *Nucleic Acids Res.* **44**, W160–W165 (2016).

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

Chapter 2: *A quality control method for SPIDR bead preparation*

ABSTRACT

Protein-nucleic acid interactions are central to many essential biological processes. Protein binding to RNA and DNA is often mapped using immunoprecipitation-based genomics methods, but most methods do not allow for more than one protein to be assayed at a time. Recently, the Guttman lab released two highly scalable methods that allow for dozens of proteins to be mapped simultaneously: ChIP-DIP for DNA-protein interactions, and SPIDR for RNA-protein interactions. A common failure point in these assays is the preparation of paramagnetic beads carrying an antibody and an oligonucleotide barcode known as an antibody-ID. This chapter outlines the factors that cause bead preparation failure, and describes a protocol for mitigating this common issue.

INTRODUCTION

Most essential cell processes involve cooperation between proteins and nucleic acids. Gene regulation, translation, cell division, and dozens of other biological phenomenon require the precise binding of proteins to RNA and DNA targets. As such, methods for mapping protein binding sites on RNA and DNA are essential in molecular biology. The most common mapping methods involve immunoprecipitation: using an antibody bound to a solid support to purify a protein of interest and its associated nucleic acids. Two of the most common methods are chromatin immunoprecipitation sequencing (ChIP-seq) and cross-linking immunoprecipitation (CLIP).[1–5] ChIP-seq targets DNA-bound proteins that have been covalently linked to chromatin using a chemical fixative such as formaldehyde. ChIP-seq experiments have dramatically expanded our understanding of gene regulation by profiling histone modifications, transcription factors, and structural proteins across the genome. CLIP targets RNA-bound proteins that have been linked to their RNA targets using short-wave UV light. CLIP profiles have provided insights both into the mechanisms of functional, noncoding RNAs and the lifecycles and structures of coding mRNAs. Due to the importance of this information, international consortia such as the Encyclopedia of DNA Elements (ENCODE) have spent decades of collective effort and millions of dollars to generate high-quality ChIP and CLIP datasets for dozens of factors in well-characterized cell lines.[6]

ENCODE is an invaluable resource for the scientific community but is limited by its choice in model systems. Most ENCODE datasets were generated in immortalized cell lines. The choice of these cultures is reasonable because they are so commonly used, but immortalized cells have major shortcomings. These cells typically have large chromosomal abnormalities and do not represent the transcriptional state of any physiological cell population. These traits

lead to two major problems. First, results in immortalized cells may not generalize to cells in living organisms. Although certain general trends are consistent across cell types, such as the enrichment of H3K4me3 over promoters and H3K36me3 over active gene bodies, the precise genes that carry these marks vary widely among different cell types. Second, immortalized lines express a small set of the proteins encoded in the genome. Proteins involved in specialized cell functions can be artificially expressed in immortalized cells, but their localization and cytological effects likely differ from their native, physiological context.

These limitations underscore the need for methods that allow individual research groups to generate consortium-level protein-nucleic acid profiles in their systems of interest. Two methods recently released from the Guttman lab address this need. Chromatin immunoprecipitation done in parallel (ChIP-DIP) allows dozens of protein-DNA maps to be generated in a single experiment and split and pool identification of RBP targets (SPIDR) does the same for protein-RNA maps.[7,8] The underlying technology of both methods is a paramagnetic bead linked to both an antibody and an oligonucleotide carrying a distinct barcode sequence, the antibody-ID. For each target in a ChIP-DIP or SPIDR experiment, a separate bead set is generated, each with a distinct combination of antibody and antibody-ID. All the bead sets are then pooled and mixed with lysate such that all target proteins are immunoprecipitated in a single reaction.

From there, the beads are divided into wells containing oligonucleotide barcodes that can be ligated to both the antibody-ID and the nucleic acid of interest (gDNA or cDNA bound to the target proteins). After the ligation, the beads are pooled, washed, and split for a second round of barcode ligations. This process is repeated until the number of unique barcodes matches or exceeds the number of beads in the experiment. The result is that each antibody-

ID and nucleic acid of interest has a barcode that matches them to a specific bead. Once antibody-IDs and targeted nucleic acids are known to be on the same bead, the antibody-ID is used to assign the nucleic acid to a target protein.

Detection of the antibody-ID is therefore essential to the success of the experiment. Without it, cDNA or gDNA can be assigned to a bead but not to an antibody. The protein bound to these fragments is labeled "ambiguous" or "none" and the sequences are removed from future analysis. To robustly detect all antibody-IDs in a SPIDR or ChIP-DIP experiment, the number of antibody-IDs per bead (antibody-ID density) must be roughly equivalent for all antibody-bead sets at the start of the experiment. Bead sets with abnormally high antibody-ID density can cause "PCR jackpotting" where overabundant sequences are amplified disproportionately in early rounds of PCR, and their overrepresentation is amplified exponentially in subsequent PCR cycles.[9] Jackpotting results in a small number of antibody-IDs dominating the sequencing libraries while the rest are barely detectable. Bead sets with abnormally low antibody-ID density can drop out of the final libraries due to oligo shedding. During barcoding, some of the antibody-ID sheds off the beads. If the initial antibody-ID density is high enough, enough oligo remains on the bead that shedding does not affect detection. If the density is too low, however, shedding result in a bead set with too little antibody-ID to be robustly amplified by PCR.

Although our protocols aim to keep antibody-ID density consistent across all bead sets, we have observed a high level of variability in antibody-ID representation in SPIDR experiments even when beads in different replicates are prepared from the same antibody-ID plate. In this chapter, we determine that the cause of this variability is the gradual, non-uniform increase in concentration of oligos stored at 4°C and -20°C due to evaporation and sublimation. We

identify that the increase in antibody-ID concentration quenches the streptavidin used to link antibody-IDs to beads, resulting in dropout of high concentration wells. Finally, we describe a simple method to measure and control antibody-ID density before adding any antibody, substantially reducing the risk of a failed SPIDR or ChIP-DIP experiment. We recommend that any laboratories that plan to try SPIDR or ChIP-DIP implement this quality control method to make these powerful genomics methods more consistent and reliable.

RESULTS

SPIDR bead clustering is highly sensitive to oligonucleotide concentration

Oligo-antibody beads are prepared in three steps. First, recombinant streptavidin is coupled to biotinylated antibody-ID oligos in a 1:1 molar ratio. Next, streptavidin-coupled antibody-IDs are bound to a biotinylated protein G bead. Finally, antibody is coupled to the same beads for the final antibody-oligo bead used for IP and barcoding. The molar ratio of the streptavidin-oligo linkage is critical for the success of the experiment. Streptavidin has four biotin binding pockets, allowing for five possible assemblies ranging from empty to fully saturated biotin pockets with four oligos. Only two of these assemblies, one and two oligos, are usable for bead labeling. Empty streptavidin has no oligo to be labeled, streptavidin with four oligos has no free binding site to bind to the bead, and with three oligos binds to beads very inefficiently due to steric inhibition by the bulky oligonucleotide tags. As such, maximizing the number of streptavidin-oligo linkages in the one or two oligo states is critical for success of the experiment.

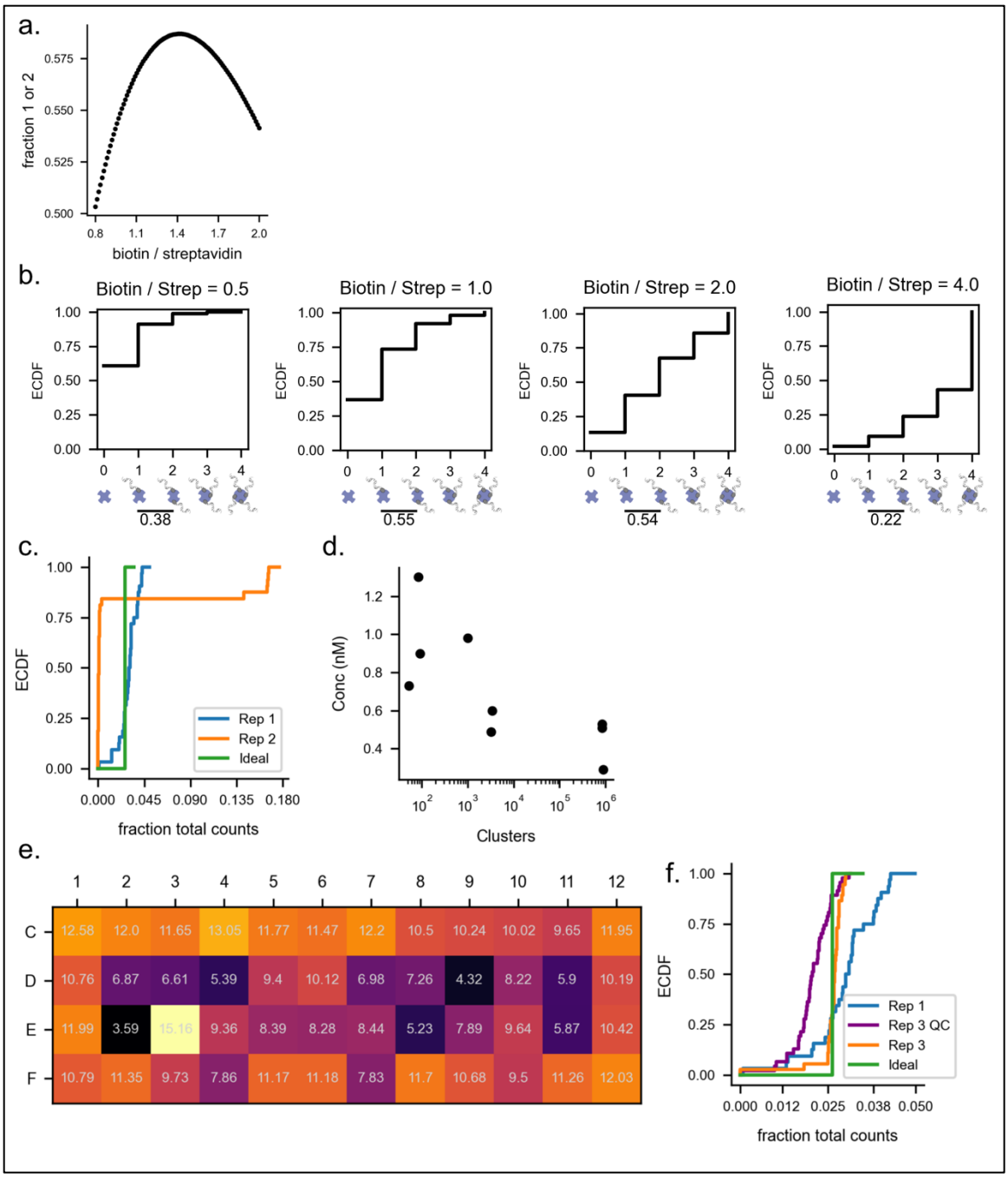

Figure 1: A method for improving antibody-ID representation. a, Poisson modeling of the optimal biotin-streptavidin molar ratio for coupling. Only streptavidin with 1 or 2 biotin-oligos bound is available for bead binding, so maximizing streptavidin in these complexes is essential for high antibody-ID density. **b,** Deviations from optimal biotin-streptavidin ratios rapidly lead to streptavidin quenching. Poisson models suggest that a 4:1 biotin to

streptavidin ratio leads to more than a 50% reduction in usable streptavidin compared to a 1:1 ratio. **c,** ECDFs of SPIDR replicates. Both replicate 1 and replicate 2 used previously validated antibody-ID plates but replicate 2's plate had been stored for around 6 months. In replicate 2, 5 antibody-IDs dominated the oligo library, preventing cDNA assignment. **d,** Concentration of antibody-IDs plotted against SPIDR clusters. Higher antibody-ID concentrations led to lower detection of those wells by SPIDR, suggesting that the higher concentration was quenching streptavidin during bead preparation. **e,** Concentrations of an antibody-ID plate. Rows and columns match those on the plate and concentrations are reported in nM. **f,** ECDFs of antibody-ID clustering for replicate 1 (as in panel c), the quality control run (Rep 3 QC), and the final replicate 3 final library (Rep 3). Note that Rep 3 closely matches the ideal distribution of all antibody-IDs having equal representation.

Modeling the binding with a Poisson distribution suggests that optimal streptavidin loading is achieved between a 1:1 and 1:2 molar biotin to streptavidin ratio (Figure 1a). Going below this optimal range leads to a rapid accumulation of empty streptavidin, whereas going above rapidly leads to streptavidin saturation (Figure 1b). As such, accurate knowledge of the concentration of these reagents and precise pipetting are essential for these experiments.

Antibody-ID representation drifts over time

We previously assumed that if an antibody-ID plate had given approximately uniform antibody-ID density in one experiment that the same plate could be safely used for future experiments even after extended storage. Once plates had been stored for longer periods, however, we found that antibody-ID density became uneven across targets (Figure 1c).

In a replicate prepared from a freshly diluted plate (a 10µM plate diluted from a 100µM stock plate stored at -80°C), all 33 targets yielded similar antibody-ID counts. Some antibody-IDs

were underrepresented, but not enough to dramatically affect cDNA assignment to antibodies. In a separate replicate using an plate that had been stored for about 7 months, 5 of the 33 targets consumed essentially the entire antibody-ID library. None of the remaining 28 targets had sufficient oligo coverage to assign cDNA to antibodies, so those bead sets were effectively wasted.

To determine whether the cause of the poor oligo clustering was related to streptavidin-oligo ratio, we measured the concentrations of oligos on the plate used in the unsuccessful replicate. As equal volumes of the same streptavidin stock were used for coupling, streptavidin was not considered a likely source of the error. We found that the diluted antibody-ID plate, which should have been uniformly 1 nM, instead had more than a four-fold difference in concentration among the wells (Figure 1d). Additionally, wells with the lowest oligo concentrations received nearly four orders of magnitude more clusters than those with the highest concentration, suggesting that the higher concentration wells were saturating the streptavidin and preventing antibody-ID binding to beads. To remedy this problem, we designed a more extensive quality control experiment to empirically measure antibody-ID density in every well used.

Bead sequencing before antibody coupling improves oligo representation

To ensure that streptavidin-oligo coupling was performed in the correct ratio, the concentrations of all 48 biotin oligos in the plate used in the uneven second replicate were measured by Nanodrop immediately before the reaction (Figure 1e). Consistent with previous results, the concentration within the plate varied by about four-fold, from 3.6 nM to 15.2 nM, a wide variation from the expected 10 nM.

Using the empirically determined concentrations, streptavidin coupling was performed with a 1:1 molar ratio of streptavidin to antibody-ID from all 48 oligos. Streptavidin-coupled oligos were then linked to biotinylated protein G beads. Once the antibody-ID bead preparation was complete, 10% of each well was collected and pooled into a single tube. A terminal barcode containing an Illumina i7 overhang was ligated to the antibody-IDs on bead and the ligated oligos were amplified into Illumina sequencing libraries. The libraries were sequenced and the representation of each antibody-ID calculated. Although some variability remained, antibody-ID was much more even in the quality control run than in the SPIDR experiment previously performed from the same plate (Figure 1f). This demonstrates that using antibody-ID concentrations measured immediately before bead preparation makes oligo representation more even. Despite these precautions, however, one oligo (D12) was still nearly undetectable in the library. We had accounted for this eventuality by preparing more oligo-bead samples than we had antibodies for our intended SPIDR experiment. When we proceeded to antibody coupling, we simply removed D12 and used an alternate well.

After antibody coupling, we proceeded with the SPIDR experiment. After IP, barcoding, and sequencing, the oligo distribution was largely unchanged (Figure 1f). This confirmed that correcting the oligo distribution before the addition of valuable antibodies greatly improves oligo representation, and therefore the ability to correctly assign cDNAs, at the end of the SPIDR experiment.

DISCUSSION

The scalability of SPIDR and ChIP-DIP is unmatched by other protein-nucleic acid profiling methods. We previously demonstrated that over 200 antibodies can be used in a single ChIP-

DIP experiment without affecting the data quality of each track.[7] These advantages, however, can only be realized with a high investment in antibodies. For the 33 antibody SPIDR experiments described in this section, a single replicate cost around $2700 in antibody alone. The biotinylated antibody-IDs are costly to order initially (~$4000 for 48 oligos from Integrated DNA Technologies without discounts), but each SPIDR experiment uses a tiny fraction of the minimum order amounts from most manufacturers. As such, the cost in oligos for one experiment is trivial (less that $1 per replicate for 33 antibodies). Because of this, our method for validating oligo distribution before using any antibody or biological material makes attempting SPIDR and ChIP-DIP substantially less risky.

Even in the absence of cost concerns, SPIDR and ChIP-DIP are laborious methods. Most of the labor, however, is in barcoding and library preparation. From start to finish, the bead preparation takes no more than an hour of hands-on work. Because our quality control method takes place before barcoding and all the oligo libraries are prepared in a single pool, the labor involved is minimal. The wet-lab portion of the experiment takes a single day. The sequencing requirements are minimal both in terms of reads (10^5 reads is sufficient to calculate the required metrics) and read length (due to the lack of barcodes, even a 50x50 nt run gives enough coverage). If a sequencing run is immediately available, the time from beginning bead preparation to having the antibody-ID distributions and being able to proceed with immunoprecipitation is around 48 hours. In contrast, over two weeks of intensive lab work and computational analysis went into the replicate that failed due to antibody-ID jackpotting.

In sum, the quality control protocol described in this chapter is a simple approach to derisking SPIDR and ChIP-DIP experiments. We recommend that any new user of the methods

incorporate this step to avoid wasted labor and reagents due to uneven antibody-ID representation.

LIMITATIONS

An open question is how the variability in oligo concentration arises in the first place, and why it increases over time. Human error from pipetting is certainly a possibility but seems improbable for the magnitude of the variation observed. It also fails to explain why a plate that gave excellent data when initially diluted gives very poor data after several months of storage. We believe that sublimation and evaporation are the primary culprits, as evidenced by the general trend that wells near the edge of the plate tend to have higher concentrations than those in the middle (Figure 1e). This is challenging to study systematically, however, given the time span required to observe the effects. Nonetheless, we recommend some best practices for reducing water loss during storage: seal plates thoroughly, do not reuse films after plates are opened, store at –20°C or –80°C rather than 4°C, and use a thermal plate sealer if available.

Another mystery is the difference in magnitude between the oligo concentration and the final oligo clustering (1:4 compared to $1:10^4$). We suspect the difference is largely due to two factors. First, beads that receive few oligos at the beginning of the protocol are highly sensitive to oligo shedding during barcoding. From developing ChIP-DIP we know that beads shed oligos during barcoding. Typically, shedding does not affect the final data quality; beads are loaded with excess oligo to account for the loss. If, however, the initial oligo density on beads is substantially lower than anticipated, shedding could leave the beads with too little oligo to detect by sequencing. Second, uneven oligo representation could lead to

PCR biases. More abundant oligos could be amplified more efficiently than low abundance oligos, exponentially worsening the unequal representation in each PCR cycle. Shedding is not a major concern in the quality control protocol described in this chapter, but PCR biases could be a concern. If the quality control sequencing run shows strong biases towards a small number of oligos, it may be worth preparing fresh oligo-bead conjugates and excluding those wells from the library preparation.

DATA AVAILABILITY

All data used in this study and code to regenerate figures is available on Github at https://github.com/dhonson-lncrna/20241028_SpidrOptFigures.git.

ACKNOWLEDGEMENTS

Drew Honson and Mario Blanco identified the antibody-ID density problem and conceived of the quality control method. Drew Honson performed all SPIDR and quality control experiments; Mario Blanco ran the SPIDR pipeline and quantified antibody-ID distributions. Drew Honson wrote this report.

METHODS

Biotin-streptavidin ratio modeling

Poisson distribution model

Biotin-streptavidin binding was modeled in Python using the Poisson distribution probability mass function (PMF):

$$f(n;\lambda) = \frac{\lambda^n}{n!}e^{-\lambda}$$

where n is the number of streptavidin pockets filled and λ, the mean and variance of the Poisson distribution, is the molar ratio of biotin to streptavidin. The function was evaluated at n = 0, 1, 2, 3 and 4. The model as written allows for numbers of binding events greater than 4, which are not biologically relevant due to the number of streptavidin binding pockets. To collapse results for n > 4 into the saturated streptavidin category (n = 4), the sum of the PMF results for n between 0 and 4 was subtracted from 1 and added to the PMF result for n = 4. With free biotin, streptavidin-binding is cooperative: biotin-streptavidin affinity increases with each pocket filled. Because of steric hindrance from the bulky oligonucleotide, it is unclear whether this cooperativity is relevant in the context of streptavidin-oligo linkage. As such, cooperativity was not included in the model. PMF results were converted to CDF values for the visualization in Figure 1b. In Figure 1a, the same protocol was followed for 100 values between λ = 0.8 and λ = 2.0 and the sum of PMF values for n = 1 and n = 2 was reported.

SPIDR experiments

Cell culture

HEK293 cells were cultured in HEK293 media (high glucose DMEM (Gibco, Life Technologies), 10% FBS (Seradigm), and 1% penicillin/streptomycin (Gibco, Life Technologies). Once confluent, three 10-cm plates were crosslinked for each SPIDR experiment. Cell media was aspirated and cells were rinsed with ice-cold PBS. The PBS was removed and cells were treated with 2.5×10^5 μJ/cm^2 265nm UV light in a UV crosslinking chamber. Ice-cold PBS was immediately added to cover the cells, then cells were scraped,

collected in conical tubes, and spun 3 minutes, 330g, 4°C. The supernatant was removed, cells were resuspended in 1mL PBS and transferred to Eppendorf tubes. Cells were spun 3 minutes, 1000g, 4°C. The supernatant was removed, cells were snap-frozen in liquid nitrogen and stored at -80°C until use.

SPIDR

SPIDR experiments were performed and analyzed as described previously.[8] Four rounds of barcoding with twelve barcodes was performed and 1%, 5%, and 10% aliquots were reserved for sequencing. Sequencing was performed on an Element AVITI instrument with Cloudbreak chemistry and asymmetric 100x200 nt read lengths. A full list of antibodies for the experiments described in this chapter can be found in Appendix B.

Quality control sequencing

Streptavidin-oligo conjugation

Concentrations of oligonucleotide barcodes were measured from a plate meant to be 10µM using a NanoDrop One^{C} spectrophotometer. Concentrations in ng/µl were converted to nM using NEBioCalculator ssDNA Mass to Moles with 19nt DNA length. Using these values, 5pmol oligonucleotide probe was coupled to 5pmol recombinant streptavidin (Biolegend) in PBS to a final volume of 50µl. The coupling reaction was incubated 1 hour, 25°C, shaking at 1250 rpm. The plate was then diluted 1:4 for a final concentration of 227nM.

Oligo-bead preparation

For each oligo, 10µl biotin-protein G beads prepared as described in the SPIDR protocol were washed in PBST then resuspended in 200µl 1x Bind and Wash Buffer (5 mM Tris-HCl, pH 8.0, 0.5 mM EDTA, pH 8.0, and 1 M NaCl). 2.5µl 227nM streptavidin-coupled oligo was added to each 200µl reaction and incubated 1 hour, 25°C, shaking at 1250 rpm. The beads were washed twice in 200µl M2 Buffer (20mM Tris-HCl, pH 7.5, 50mM NaCl, 0.2% Triton-X 100, 0.2% NP-40, and 0.2% sodium deoxycholate), twice in 200µl PBST, then resuspended in 100µl PBST.

Library preparation and sequencing

10µl of each oligo-bead set was collected and pooled in a single tube. The beads were magnetically separated, the supernatant was removed, and the beads were suspended in 200µl ligation mix (64µl 2x NEB Instant Sticky-End Master Mix, 24µl 4.5µM TermLig Even barcode, and 112µl water). The reaction was incubated 10 minute, 25°C, shaking at 1350 rpm. The beads were magnetically separated and washed three times in 1mL M2 buffer and three times in 1mL PBST. After the final wash, beads were resuspended in 225µl water.

For library preparation, 10% (22.5µl) of the ligated beads were placed into a 50µl NEB Q5 PCR reaction with 2PUNI_PC50 and Illumina i7 index primers (see SPIDR protocol for sequence details). The library was amplified for 4 cycles with a 69°C annealing temperature and a 15 second extension. The beads were magnetically separated and the supernatant was transferred to a clean tube. 60µl (1.2x) SPRI beads (Bulldog Bio) were added to the supernatant and triturated to mix. The reaction was incubated for 10 minutes at room temperature, then the SPRI beads were magnetically separated. The beads were washed three

times with 200µl 80% ethanol. The ethanol was removed and beads were allowed to dry then resuspended in 23µl water to elute.

A second NEB Q5 PCR was performed using the same Illumina i7 index primer as the first PCR as well as an Illumina i5 index primer. The library was amplified 4 cycles with a 69°C annealing temperature and a 15 second extension, then 4 additional two-step cycles with a 72°C, 30 second annealing/extension step. The 1.2x SPRI clean was repeated and the final library was eluted in 20µl water. The full library was gel-extracted to remove residual primer and its concentration was calculated using Qubit and TapeStation D1000. Sequencing was performed on an Illumina iSeq with symmetric 75nt paired end reads.

Oligos were assigned using the published SPIDR pipeline. After the one aberrant well was identified and removed, SPIDR was performed as described in this chapter.

REFERENCES

1. Johnson, D. S., Mortazavi, A., Myers, R. M. & Wold, B. Genome-wide mapping of in vivo protein-DNA interactions. *Science* **316**, 1497–1502 (2007).

2. Barski, A. *et al.* High-resolution profiling of histone methylations in the human genome. *Cell* **129**, 823–837 (2007).

3. Mikkelsen, T. S. *et al.* Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. *Nature* **448**, 553–560 (2007).

4. Robertson, G. *et al.* Genome-wide profiles of STAT1 DNA association using chromatin immunoprecipitation and massively parallel sequencing. *Nat. Methods* **4**, 651–657 (2007).

5. Ule, J. *et al.* CLIP Identifies Nova-regulated RNA networks in the brain. *Science* **302**, 1212–1215 (2003).

6. Dunham, I. *et al.* An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**, 57–74 (2012).

7. Perez, A. A., Goronzy, I. N., Blanco, M. R., Guo, J. K. & Guttman, M. ChIP-DIP: A multiplexed method for mapping hundreds of proteins to DNA uncovers diverse regulatory elements controlling gene expression. 2023.12.14.571730 Preprint at https://doi.org/10.1101/2023.12.14.571730 (2023).

8. Wolin, E. *et al.* SPIDR: A highly multiplexed method for mapping RNA-protein interactions uncovers a potential mechanism for selective translational suppression upon cellular stress. *BioRxiv Prepr. Serv. Biol.* 2023.06.05.543769 (2023) doi:10.1101/2023.06.05.543769.

9. Marotz, C. *et al.* Triplicate PCR reactions for 16S rRNA gene amplicon sequencing are unnecessary. *BioTechniques* **67**, 29 (2019).

Chapter 3: *SPIDR: A high-throughput approach to map RNA-binding proteins to diverse ribosome structures*

ABSTRACT

The ribosome is among the cell's most important and complex molecular machines. Mapping the structure of the dozens to hundreds of proteins that bind the ribosome is an active area in ribosomal biology, but is hindered by the limited scalability of structural methods. Here, we use Split and Pool Identification of RBP targets (SPIDR) to map the binding of 39 RNA-binding proteins to ribosomal RNA (rRNA). We show that SPIDR accurately maps proteins to pre-rRNA, as well as 18S and 28S mature rRNAs in a variety of structural states. We also observe novel binding events, such as previously unreported binding of nucleolar components PES1, ILF3, and LIN28B to the 45S-5'ETS, and binding sites on 28S that are shared by NAC and translocon components. This study provides an atlas of rRNA binding sites for a diverse set of proteins, as well as a blueprint for future ribosome studies using SPIDR.

INTRODUCTION

The ribosome is among the most intricate molecular machines in the cell. Its centrality to living systems is reflected in its abundance: around 85% of the cell's RNA and a third of its total dry mass is represented by ribosomes.[1] The ribosome's importance is matched by the complexity of its biogenesis and regulation. Throughout their life cycle, ribosomes undergo numerous compositional and conformational changes that reflect their changing functions.

Structural biology has contributed greatly to our understanding of the ribosome. In particular, cryo-electron microscopy (cryo-EM) has provided a wealth of insights on diverse stages of the ribosomal lifecycle such the initial synthesis of the 60S subunit, the assembly of pre-initation complexes, and the trafficking of translating ribosomes to the endoplasmic reticulum (ER).[2–9] Although structural methods are incredibly powerful tools for investigating the ribosome, however, they have several limitations.

First, most structural methods can only evaluate purified complexes. This presents a challenge for non-stoichiometric assemblies or structures that depend on their cellular environment. Specifically, immature ribosomes in the nucleolus interact with proteins such as fibrillarin, nucleolin, and TGS1 that likely function both to maintain nucleolar structure and to facilitate specific steps of ribosome biogenesis.[10] Whether the structure of a purified pre-ribosomal particle containing these factors represents its native configuration in the nucleolus is unclear. Additionally, the need to purify specific complexes reduces the scalability of structural methods. Purifying many individual complexes in a single experiment is impractical, as is purifying the same complex from large numbers of independent biological samples. Given that over 1300 proteins have been identified in the human nucleolus alone, scalability provides a significant challenge for structural methods.[11]

Second, structural methods struggle to resolve highly flexible or disordered regions. Many proteins involved ribosome biogenesis and regulation have extensive disordered regions that cannot be structurally characterized.[11] The ambiguities left by these low complexity domains likely conceal biologically important interactions between proteins and rRNA.

Finally, structural methods do not usually analyze interactions of ribosome components with RNAs outside the ribosome. Regulatory proteins such as LARP1 and SERBP1 bind to the ribosome but also have many mRNA targets.[12–17] Having knowledge of RNA binding both in and outside the ribosome can provide clues to function that ribosomal binding alone cannot.

Previously, we developed the method Split and Pool Identification of RBP targets (SPIDR) that allows dozens of RNA binding proteins (RBPs) to be mapped to their bound RNAs in a single experiment.[18] In brief, SPIDR uses beads carrying an antibody and a distinct oligonucleotide barcode sequence, or antibody-ID. For each target protein, a separate bead set is generated, each with its own antibody and unique antibody-ID. The bead sets are then pooled and used for a single immunoprecipitation from UV-crosslinked cell lysate. The RNA recovered is ligated to an adapter on-bead and reverse-transcribed. The resulting cDNA and antibody-ID contain a shared overhang that can be used for split-pool barcoding. The beads are split into wells that each contain a distinct DNA barcode sequence. The barcode is ligated to both the cDNA and the antibody-ID, then the beads are once again pooled. The process is repeated until the number of unique barcodes exceeds the number of beads. Once libraries are prepared and sequenced, the split-pool barcodes are used to determine which antibody-IDs and cDNAs were on the same bead. From there, the antibody-ID is used to assign cDNA to a protein target.

SPIDR has many features that complement structural methods and address their limitations. SPIDR can accommodate large numbers of antibodies in a single experiment, and adding additional antibodies does not increase the experiment's complexity. As such, testing many putative but unproven ribosome-binding proteins simultaneously is far simpler and more cost-effective with SPIDR than with structural approaches. Additionally, SPIDR can profile multiple biological samples simultaneously by using the first split-pool barcode as a sample identifier. SPIDR can also confidently identify transient or context-dependent interactions. UV crosslinking is performed on live cells, so RNA-protein interactions in nuclear bodies or with disordered protein regions are covalently fixed before lysis. Finally, SPIDR profiles all RNAs that a protein binds in the cell simultaneously. As such, SPIDR can map ribosome-binding proteins not only to rRNAs, but also to other RNAs they may interact with either within or outside the ribosome.

In our original SPIDR paper, we included several ribosomal targets that yielded structural and functional insights. The association of LARP1 with both 18S rRNA and TOP motif-containing mRNAs suggested its role in regulating initiation of these factors. We additionally found that the protein 4EBP binds to the 5'-UTRs of LARP1 mRNA targets specifically under conditions of mTOR inhibition. These observations together led to a model for how specific mRNAs are translationally regulated through mTOR signalling. These results suggested that SPIDR may be a valuable tool for assessing ribosome function more broadly. Here, we perform SPIDR on an expanded antibody panel exclusively targeting ribosome-associated factors. Our panel covers proteins that bind to diverse regions of the ribosome (pre-rRNA, 40S, and 60S) and with diverse subcellular locations (the nucleolus, nucleoplasm, cytoplasm, ER, and mitochondria). Our SPIDR data agrees well with existing

structural data where it is available. Additionally, we identify several novel RNA-protein interactions such as the binding of PES1, ILF3, and LIN28B to the 45S 5'-externally transcribed spacer (45S-5'ETS) and the association of BTF3/NACB and SEC61B to regions of 28S that have not been previously reported. We hope that the data in this study will be a useful resource for ribosome biology, and provide a workflow for future studies to follow.

RESULTS

SPIDR maps dozens of proteins to ribosomal RNA in a single experiment

We selected 39 proteins to be profiled in a single SPIDR experiment (Figure 1a). These proteins were selected based on RNA-proteomics data, discussions with colleagues in the translation field, and antibody availability.[19] These proteins fell into three broad categories: nuclear and nucleolar proteins that likely bound to immature ribosomes, factors associated with the small 40S subunit (either as a pre-initiation complex or within the 80S holoenzyme), and factors associated with the large subunit. As expected, most targets showed specific enrichment over ribosomal RNA regions, though a small number of factors showed little ribosomal signal either due to failure of the immunoprecipitation or because they bound other targets (Figure 1b).

To validate the results, we examined three large subunit (RPL14, 23, and 36) and three small subunit (RPS5, 19, and 25) proteins. As expected, the RPS proteins specifically bound to 18S while the RPL proteins specifically bound to 28S (Figure 1c-d). We next compared the SPIDR enrichments to an existing cryo-EM structure of the human 80S holoenzyme (PDB 4UG0, Figure 1e).[20] The strongest enriched bins agreed well with the structure. Interestingly, RPL23 and RPS19 also had more distant bindings sites in the correct subunit.

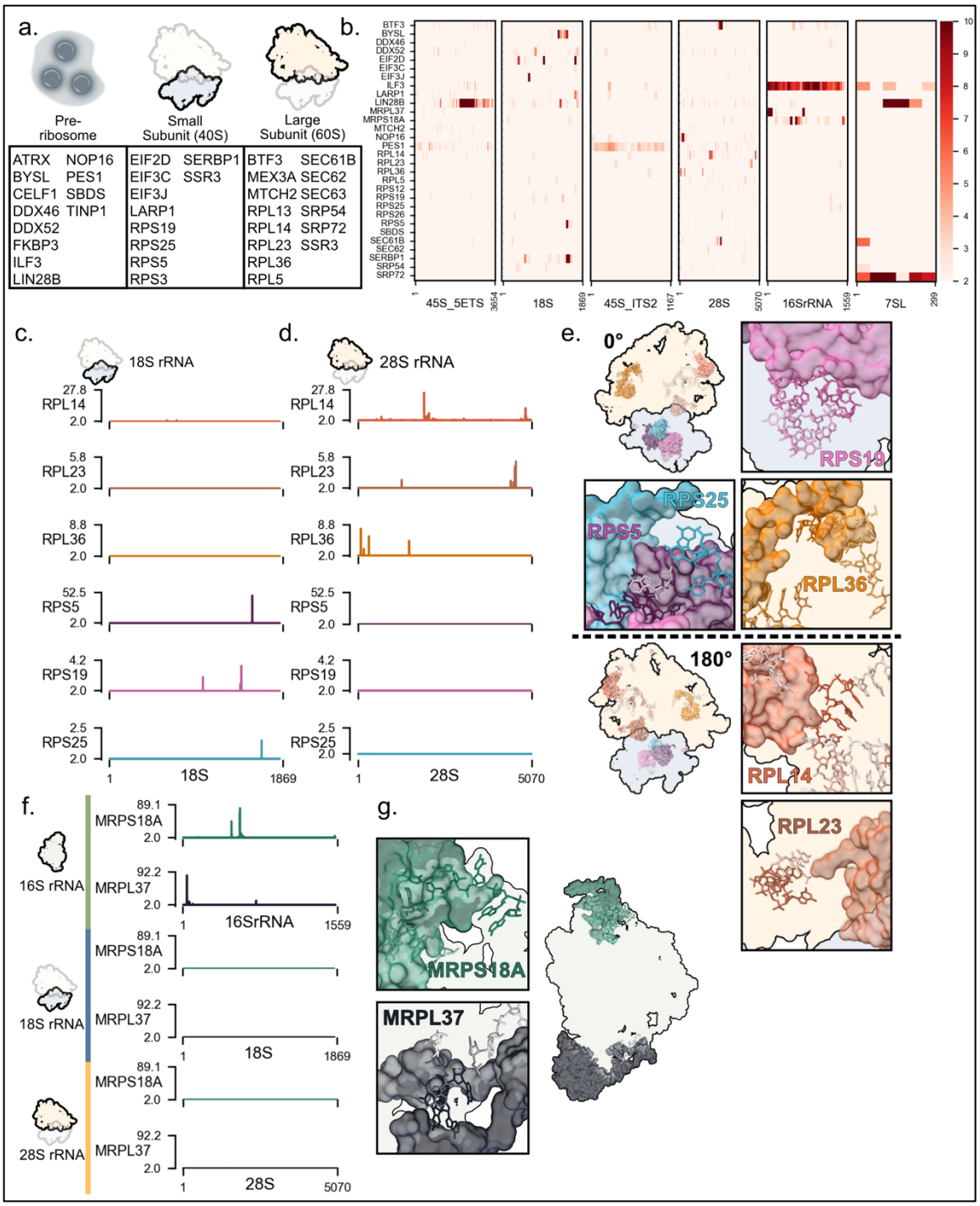

Figure 1: SPIDR maps dozens of RNA binding proteins to rRNA in parallel. a, List of proteins targeted in this study, grouped by ribosomal binding. Omitted are two proteins, MRPL37 and MRPS18A which are detailed in panels f and g. **b,** Heatmap of SPIDR over rRNA regions and 7SL. Second read enrichments floored at 2.0 are plotted in 10 nt max-

normalized bins. **c and d,** RNA tracks of RPL and RPS proteins over 18S and 28S rRNA. Second read enrichments are floored at 2.0 are plotted with no binning. Note the robust signal of RPS proteins over 18S and of RPL over 28S, consistent with ribosome localization. **e,** Mapping of SPIDR data to a cryo-EM structure of human 80S (PDB ID 4UG0). rRNA is visualized as ribbons and proteins as surfaces. rRNA is colored by SPIDR enrichment for each protein in 5 nt max-normalized bins. **f,** SPIDR enrichment of mitochondrial ribosomal proteins MRPS18A and MRPL37 over mitochondrial (16S) and eukaryotic (18S, 28S) ribosomes. Note the strong enrichment over 16S and the absence of signal over 18S and 28S. **g,** Mapping of SPIDR data to a cryo-EM structure of the human mitochondrial ribosome (PDB ID 3J9M). rRNA is visualized as ribbons and proteins as surfaces. rRNA is colored by SPIDR enrichment for each protein in 5 nt max-normalized bins.

This may reflect other ribosomal conformations, such as intermediate states of biogenesis or flexibility of the rRNA structure.

In addition to eukaryotic ribosomal proteins, we included two mitochondrial ribosomal proteins (MRPS18A and MRPL37) as negative controls. Both proteins showed specific enrichment over 16S mitochondrial rRNA, but not over 18S or 28S eukaryotic rRNAs (Figure 1f). The binding sites SPIDR uncovered agreed well with a cryo-EM structure of the human mitochondrial ribosome (PDB ID 3J9M), supporting that our enrichments were valid.[21]

SPIDR uncovers nucleolar and nuclear rRNA-protein interactions

Ribosome biogenesis is a highly controlled process involving dozens of factors and spanning several cellular compartments. The process begins in the nucleolus, the most prominent nuclear compartment. The nucleolus comprises three layers: the fibrillar center (FC), dense

fibrillar center (DFC), and granular component (GC).[10] The GC is the largest compartment and surrounds multiple subcompartments of FC surrounded by a layer of DFC.[10] Although the precise role of these structures is an area of active research, steps of ribosome biogenesis are appear to be partitioned by these layers.[10] rRNA transcription occurs at the interface between the FC and DFC producing an initial 47S fragment that is rapidly digested to 45S. 45S cleavage into 18S, 5.8S, and 28S occurs in the DFC as does base modification (pseudouridylation and 2'-O-methylation) by snoRNAs (Figure 2a).[10,22] Pre-ribosomal particles begin forming in the DFC, but pre-60S and pre-40S particles continue to undergo conformational and compositional changes as they pass through the GC and into the nucleoplasm, and maturation may even continue in the cytoplasm.[3,10,23] Due to the complexity of this process and its reliance on nuclear compartmentalization, we wondered whether SPIDR could reveal novel rRNA binding sites for nucleolar factors.

PES1 is a component of the Pes1, Bop1, and Wdr12 (PeBoW) complex and is found throughout the nucleolus.[24] PeBoW is a critical factor for large subunit assembly and PES1 is one of the few proteins that remains bound to pre-60S particles throughout their biogenesis.[3,24] Due to its role in 60S synthesis, we were unsurprised to see strong SPIDR enrichment of PES1 over 28S and 45S's second internally transcribed spacer (45S-ITS2) (a). More surprising was that PES1 showed similarly high enrichment over 45S's 5–externally transcribed spacer (45S-5'ETS). Because 45S-5'ETS is nearly 3000 nt away from the earliest 45S segment to be loaded into 60S (5.8S rRNA), this binding seems unlikely to contribute to PES1's canonical role in large subunit formation.

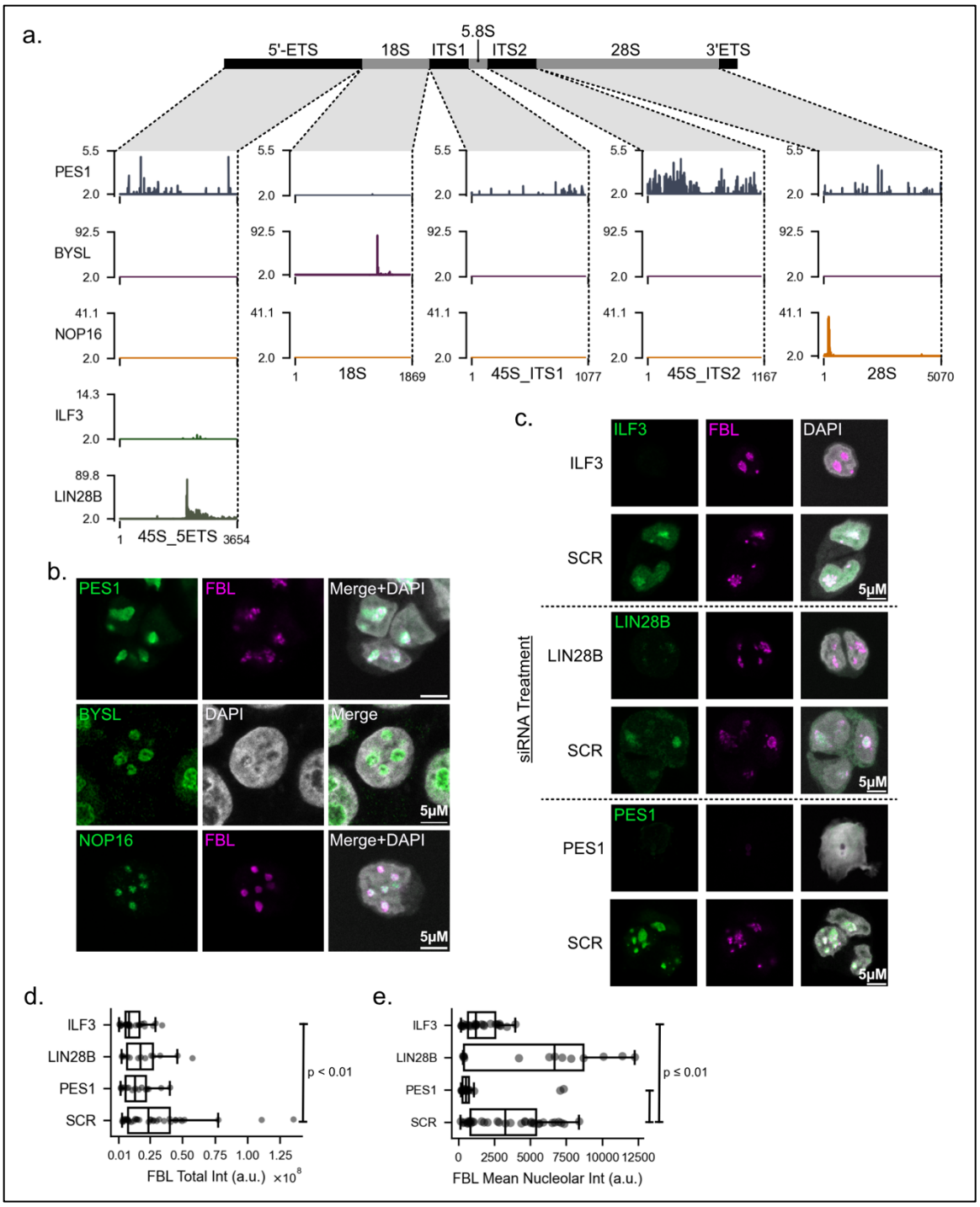

Figure 2: SPIDR maps nucleolar proteins to immature ribosomes. a, Schematic of 45S gene structure and SPIDR enrichment of nucleolar proteins. In addition to the expected binding of PES1 to 45S-ITS1, ITS2 and 28S, we see robust enrichment over the 45S-5'ETS. BYSL is enriched exclusively over 18S, and NOP16 over 28S as expected. ILF3 and LIN28B

were unexpectedly found to bind the 45S-5'ETS. **b,** Immunofluorescence imaging of SPIDR targets and fibrillarin (FBL) in HEK293. PES1 shows enrichment throughout the nucleolus, BYSL appears confined to the GC, and NOP16 appears localized to the FC and/or DFC. **c,** Immunofluorescence imaging of siRNA-treated HEK293s for ILF3, LIN28B, and PES1. Depletion of ILF3 and LIN28B did not show gross morphological defects in nucleoli, whereas loss of PES1 should a profound reduction in FBL compartmentalization. **d and e,** Quantification of immunofluorescence images. PES1 depletion showed a reduction in nucleolar FBL intensity but not in total FBL intensity, suggesting a failure of compartmentalization. ILF3 depletion showed that total nuclear FBL and nucleolar FBL were lost with similar effect sizes, indicating a loss of FBL but not specifically of its compartmentalization. LIN28B depletion did not show significant changes to FBL on any metrics measured.

As PES1 is among the most abundant nucleolar proteins, a reasonable concern is that its apparent binding to the 45S-5'ETS is simply due to random collisions in the dense environment of the DFC. If this were true, we would expect to see some level of PES1 localization across the entire 45S rRNA. Instead, PES1 had almost no coverage over 18S despite the rRNA region's proximity to 45S-5'ETS (Figure 2a). Additionally, other proteins with high abundance in the FC and DFC showed highly specific binding to their canonical targets. NOP16 and BYSL are two nucleolar proteins with comparable nucleolar abundance to PES1 (Figure 2b). In contrast to PES1, their binding on rRNA shows very narrow peaks over individual rRNA segments. NOP16 is involved in 60S biogenesis and is exclusively enriched over 28S, consistent with previous cryo-EM structures (Figure 2a).[25–27] BYSL is essential solely for 40S biogenesis.[28] Consistent with this function, SPIDR revealed BYSL binding only on 18S (Figure 2a). These results suggest that PES1 binding to 45S-5'ETS is

specific and distinct from its role in maintaining nucleolar structure. Future research may evaluate the function of this interaction.

Two other factors showed strong enrichment over 45S-5'ETS: ILF3 and LIN28B (Figure 2b). The ILF3 gene encodes two proteins: a long isoform known as ILF3 and a short isoform known as NF90.[29] Our antibody binds a shared regions between the two isoforms, so we likely recover both. ILF3 is a dsRNA binding protein implicated in diverse processes including transcriptional initiation, viral replication, and RNA metabolism.[30–32] Additionally, ILF3 is known to localize to the nucleolus and may help initiate transcription of rDNA.[33,34] Previously, however, direct binding of ILF3 to pre-rRNA has not been reported. Though the function of this interaction is unclear, future research may evaluate how ILF3-rRNA binding affects rDNA transcription and ILF3 localization. LIN28B is best known for its inhibitory role in let-7 miRNA biogenesis.[35] Specifically, LIN28B sequesters pre-let-7 in the nucleolus to prevent its processing by Dicer.[35] Previous research has implicated LIN28B's homolog LIN28A as an important contributor to normal nucleolar structure and ribosome biogenesis.[36] We speculated that LIN28B may serve a similar role in nucleolar structure through its interaction with 45S-5'ETS.

To test if ILF3 or LIN28B help maintain nucleolar structure, we performed siRNA treatments for both proteins as well as for PES1, then stained for nucleoli using fibrillarin immunofluorescence (Figure 2c). PES1 knockdown led to a dramatic loss of nucleolar fibrillarin signal. In contrast, the general morphology and number of nucleoli did not appear to change dramatically in ILF3 or LIN28B knockdown. We quantified fibrillarin intensity from the images to attempt to identify more subtle phenotypes (Figure 2d, e). PES1 showed no significant difference in the total amount of fibrillarin in the nucleus ($p = 0.09$) but a strong

loss of fibrillarin from nucleoli (~60% loss, $p = 0.01$). This is consistent with a structural role for PES1; its loss does not affect the amount of fibrillarin in the cell, but rather its localization. In contrast, ILF3 showed reductions in both the total amount of nuclear fibrillarin (~59% loss, $p = 0.007$) and the average intensity of fibrillarin in the nucleolus (~56% loss, $p < 0.001$). Given the diverse roles that previous work has attributed to ILF3, this effect may be independent of ILF3's 45S-5'ETS binding. Future research may attempt to disentangle the relationship between ILF3 rRNA binding and fibrillarin levels. LIN28B did not show significant changes in either metric. LIN28B may still have roles in ribosome biogenesis, but under the conditions used in this study it does not appear to have the same importance for nucleolar structure as LIN28A.

SPIDR maps 18S binding proteins across diverse small subunit states

Translational initiation depends on the precise binding of eukaryotic initiation factors (eIFs), the initiator tRNA, and the mRNA to the small 40S subunit. 40S initially binds to a complex of eIF1, eIF1a, and eIF3, then recruits the initiator methionine tRNA bound to eIF2.[6] This assembly is the 43S translation pre-initiation complex and is now capable of recruiting an mRNA.[6] A separate eIF complex delivers the mRNA to 43S: eIF4 binds to the m7G cap, PABP binds to the polyA tail, and the two complexes bind together to bend the mRNA into a loop.[6] mRNA binding to 43S forms the 48S complex, which scans the 5'-UTR of the mRNA to identify the start codon.[6] Once the Met-tRNA recognizes the start codon, eIFs are evicted to allow for recruitment of 60S and progression into translational elongation.[6] Once the newly synthesized protein has been released, 60S and 40S separate and recycling factors prepare the small subunit for reuse in subsequent rounds of translation.[6]

In our SPIDR experiment, we included factors that span the translation cycle of the small subunit (Figure 3a). eIF3C and eIF3J are members of the eIF3 complex, one of the first factors to bind to the small subunit.[6,37] eIF3C is found in all eIF3 complexes while eIF3J exists only in a subset of them.[37] Due to its unstable, substoichiometric association with 40S, eIF3J is absent from early structures of eIF3 in complex with 40S.[38] Only recently has a 48S structure containing eIF3J been published.[39] SPIDR data for both eIF3C and eIF3J agrees well with this structure, indicating that SPIDR performs equally well for stoichiometric and substoichiometric ribosomal components (PDB ID 6ZMW, Figure 3a, b).[39] SPIDR may therefore be useful in the future for mapping other rare ribosomal configurations.

SERBP1 is an RNA binding protein that inhibits translation by blocking the mRNA entry channel of elongation factor 2 (eEF2) containing 80S holoenzymes.[16,40] SPIDR profiling of SERBP1 agrees with an existing cryo-EM structure of SERBP1-containing stalled ribosomes (PDB ID 4V6X, Figure 3c, d).[40] This demonstrates that 18S remains accessible to SPIDR during translational elongation as well as initiation, expanding the possible targets for future studies.

Finally, eIF2D is a protein that binds to the 40S subunit after the nascent peptide has been released.[41] eIF2D assists in small subunit recycling both by facilitating the removal of the mRNA and final tRNA from 40S and by promoting the assembly of 43S.[41] The exact role of eIF2D in 43S assembly remains slightly murky; studies have suggested it may assist in initiator tRNA recruitment but also noted that its binding patterns on 40S may overlap with those of other initiation factors such as eIF1 and eIF3.[41,42] SPIDR agrees with existing structural data for eIF2D in complex with 40S (PDB 5OA3, Figure 3e, f).[42]

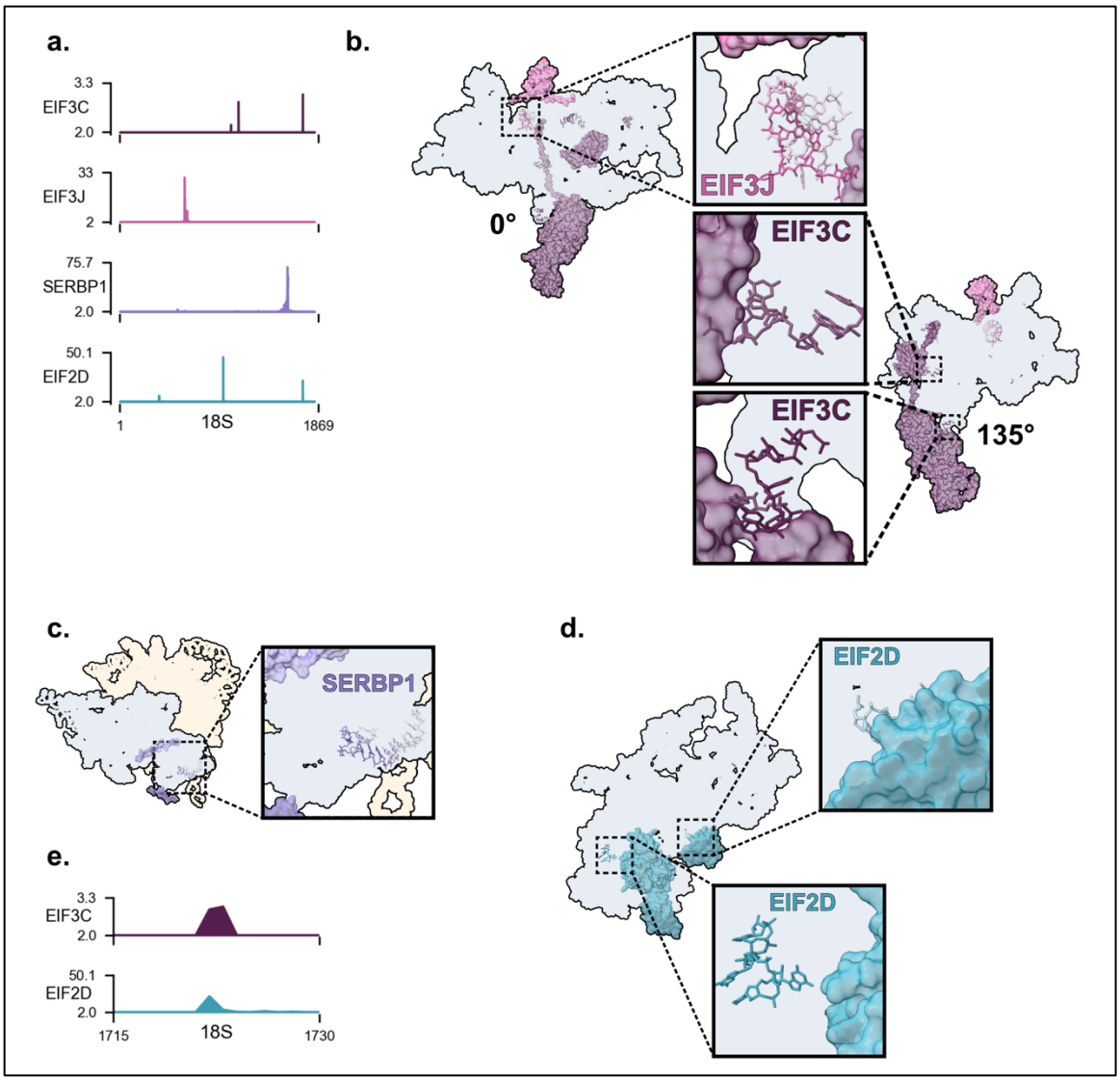

Figure 3: SPIDR captures diverse states of the 40S subunit. a, SPIDR enrichments of 18S-binding proteins over 18S. **b,** Mapping of SPIDR enrichments of eIF3 components to a cryo-EM structure of human 48S (PDB ID 6ZMW). Proteins are visualized as surfaces and RNAs as ribbons. eIF3J is a substoichiometric, unstable component of 48S while eIF3C is a stoichiometric, stable component. SPIDR maps both with similar fidelity to structure. **c,** Mapping of SPIDR enrichments of SERBP1 to a structure of human 80S (PDB ID 4V6X). SERBP1 stalls translating ribosomes by binding near the mRNA entry channel. SPIDR agrees with this localization. **d,** Mapping of SPIDR enrichments of eIF2D to a structure of eIF2D complexed to 40S (PDB ID 5OA3). eIF2D is a recycling factor that binds 40S after

it is released from 80S and prepares the subunit to be assembled into 43S. SPIDR agrees with the existing structure. **e,** Zoom in of a shared peak between eIF3C and eIF2D over 18S. eIF2D has been speculated to facilitate eIF3 recruitment, but the significance of this shared binding site remains unknown.

Additionally, SPIDR shows a highly enriched UV crosslinking site that is precisely shared between eIF2D and eIF3C (Figure 3e). Future research may examine the role of this structural overlap.

SPIDR reveals novel 28S-protein interactions that may regulate trafficking

The large subunit (60S) contains the nascent polypeptide exit tunnel (NPET) and is therefore a major site of regulation for nascent polypeptide folding and ribosome trafficking.[43] Access to the nascent polypeptide is controlled by the nascent polypeptide-associated complex (NAC).[8,9] NAC shields the nascent polypeptide from aberrant recognition by the signal recognition particle (SRP) while allowing SRP to scan the newly translated protein for the signal peptide.[9] Signal peptides are amino acid sequences that mark a protein as being destined for the endoplasmic reticulum (ER) either because they have transmembrane domains or because they require additional post-translational processing in the Golgi apparatus.[9,44] Once SRP recognizes the signal peptide, it stalls translation and guides the ribosome to pore complexes on the ER membrane called translocons.[44] Once SRP hands the ribosome off to the translocon, translation continues with the nascent peptide being extruded through the pore.[44]

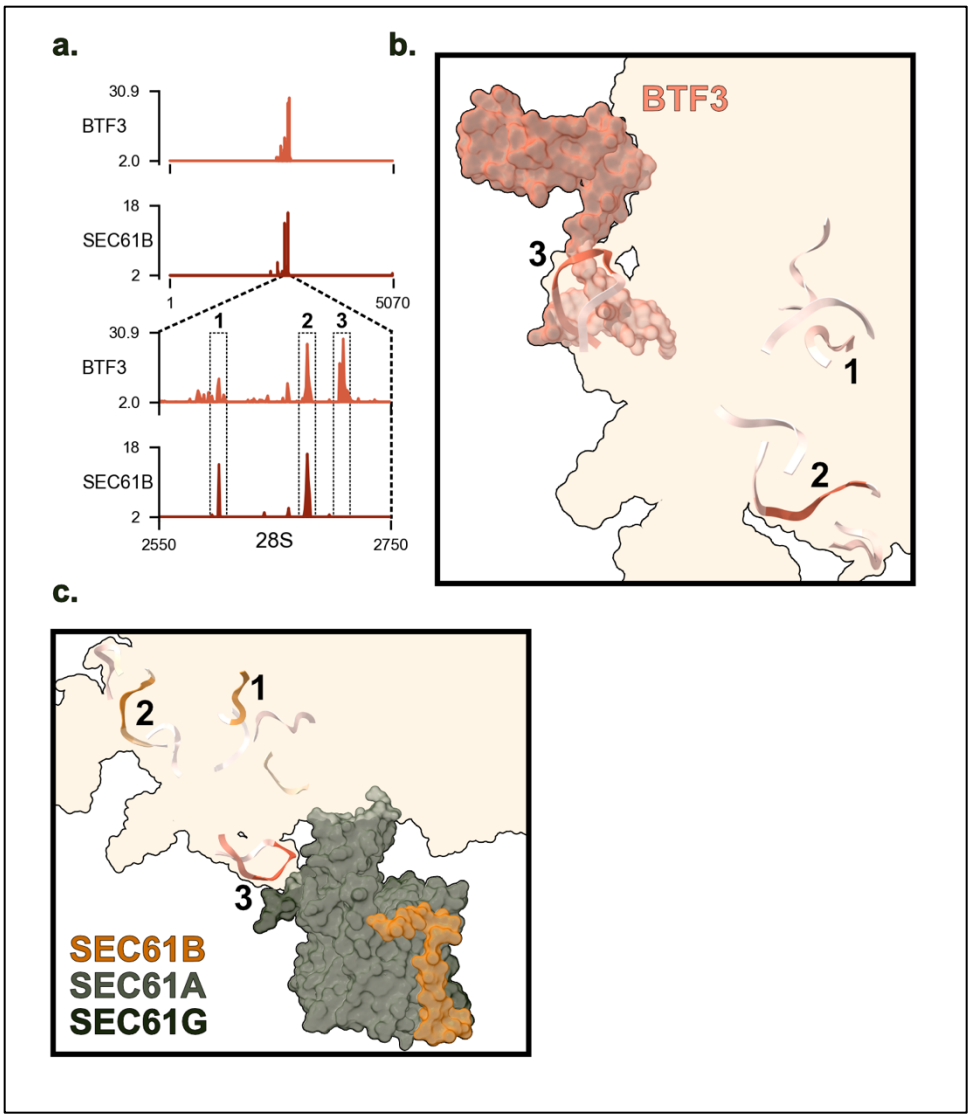

Figure 4: SPIDR reveals novel binding sites of BTF3 and SEC61B over 28S. a, SPIDR enrichments of BTF3 and SEC61B over 28S. BTF3 showed three peaks that we have labeled 1, 2, and 3. Peaks 1 and 2 are shared between BTF3 and SEC61B while peak 3 is BTF3-specific. **b,** Mapping of SPIDR enrichments of BTF3 over a structure of 60S bound to NAC (PDB ID 7QWR). Proteins are visualized as surfaces and RNA as ribbons. SPIDR peak 3 appears to represent BTF3 localization in this structure, but peaks 1 and 2 are more distant, indicating that those peaks may represent a previously unknown conformation of 60S bound to NAC. **c,** Mapping of SPIDR enrichments of BTF3 and SEC61B to a structure of 60S bound to the translocon (PDB ID 3J7Q). Consistent with its localization near the nascent

polypeptide exit tunnel, peak 3 is very near the pore complex. In contrast, peaks 1 and 2 are distant from the pore in general and SEC61B specifically. This may suggest that SEC61B binds to peaks 1 and 2 in an unknown ribosome structure on the ER membrane.

NAC remains bound to the ribosome after the association of SRP and does not dissociate until cotranslational import commences.[9] Its function during the period between handoff to SRP and the reinitiation of translational elongation at the ER, however, remains unclear. We included one member of the NAC complex, BTF3/NACB, in our SPIDR panel. SPIDR revealed three closely clustered peaks over 28S, which we have labeled peaks 1, 2, and 3 (Figure 4a). Peak 3 is localized near the exit channel and agrees well with an existing structure of NAC bound to 60S before handoff to SRP (PDB ID 7QWR, Figure 4b).[9] Peaks 1 and 2, however, do not comport with this structure. Although a structure of NAC bound to 60S after SRP association also exists, it is missing the majority of BTF3, making comparison to SPIDR less meaningful.[9] Also in our SPIDR panel was the translocon protein SEC61B. Strikingly, peaks 1 and 2 from BTF3 map precisely onto the two major SEC61B peaks that we observe over 28S (Figure 1a). We compared SEC61B and BTF3 SPIDR enrichments to a structure of 60S bound to the translocon (PDB ID 3J7Q, Figure 1b).[7] Peak 3 was close to the translocon (specifically to SEC61A), as we would expect due to its proximity to the exit channel. Peaks 1 and 2, however, were distant from SEC61B and the rest of the translocon. These discrepancies might imply that SPIDR captured an intermediate state of BTF3 and SEC61B binding to the ribosome that is not represented in existing structures. From these preliminary data, we propose a model in which BTF3 helps coordinate the ribosome over the translocon. In this model, BTF3 might hand off SRP-bound ribosomes to SEC61B, after which BTF3 would dissociate from 60S. The 28S-SEC61B interaction could briefly stabilize

the ribosome on the ER membrane, facilitating SRP binding to the SRP receptor. At this stage, SRP dissociates, translation resumes, and the ribosome is stably bound to the channel. This model is extremely preliminary, but we hope that the SPIDR binding sites will be a useful resource for future research into this important transitional stage of translation.

DISCUSSION

The ribosome is singularly important for cell function, as effectively all other cellular processes depend on timely and accurate protein synthesis. Consistent with its structural and functional complexity, dozens of proteins interact with the ribosome from the initial transcription of rRNA, to the regulation of translational initiation and elongation, and to the recycling of ribosomal subunits following translation's completion. Structural biology has provided an immense volume of knowledge on these critical cellular machines but is limited by its reliance on purified complexes, its bias towards structurally stable and stoichiometric complexes, and its scalability. Here, we have demonstrated that SPIDR can address these shortcomings. Most trivially, SPIDR provides an in-cell validation for existing structures. Because UV-crosslinking is rapid and performed in live cells, SPIDR can confirm whether the interactions observed in purified complexes represent RNA-protein binding in the cell.

More significantly, SPIDR can profile dozens of RNA binding proteins over rRNA in a single experiment, allowing for high throughput analysis of putative ribosomal binding factors. We demonstrate that SPIDR can profile diverse proteins at diverse stages of the ribosomal life cycle. We identify three proteins (PES1, ILF3, and LIN28B) that bind robustly to the 45S-5'ETS. Few structures of this segment and of pre-40S particles generally have

been generated from human cells. We hope that future research can use our binding profiles to interrogate the function of these RNA binding proteins in ribosome biogenesis.

We further show that SPIDR can capture protein interactions with 18S in various ribosomal states. In the 43S preinitiation complex, we accurately mapped both the stable, stoichiometric component eIF3C as well as the unstable, substoichiometric component eIF3J. In the 80S holoenzyme, we correctly determined the binding site of SERBP1 on 18S in stalled ribosomes. In free 40S subunits after the completion of translation, we correctly profiled the recycling factor eIF3D.

Finally, we show that SPIDR can reveal previously uncharacterized interactions with 28S in the 60S subunit. For the NAC component BTF3, we not only identified its known binding site from existing structures, but also two novel binding sites. In the same SPIDR experiment we mapped the translocon protein SEC61B and were surprised to see that it perfectly overlaps with BTF3's novel binding sites. Not only does this demonstrate the value of SPIDR for identifying new ribosomal structures, but it underscores the value of SPIDR's scalability. Because adding additional antibodies to a SPIDR experiment is trivial, serendipitous observations such as the overlap between BTF3 and SEC61B become more frequent.

We hope that the data generated in this study will be a useful resource for future ribosomal research, and that our approach will be useful for future panels of ribosomal factors and other complex RNA-protein assemblies.

LIMITATIONS

SPIDR as implemented in this study has three major limitations. First, SPIDR cannot determine if interactions are occurring in the same or different ribosomes. This is especially

significant for our nucleolar targets. We cannot determine from SPIDR data alone whether PES1, ILF3, and LIN28B exist on the same 45S molecules simultaneously, or whether these represent independent structures. Second, SPIDR is limited by antibody quality. Several antibodies in our panel, such as CELF1, SEC62, and SSR3, did not show any enrichment over RNA. From SPIDR data alone, we cannot tell if these negative results are due to antibody quality, degradation of the targets in lysate, or simply a legitimate lack of RNA binding. Finally, no functional conclusions can be drawn from SPIDR in the absence of perturbation experiments. As such, all models generated from SPIDR should be considered preliminary until the suggested structures are validated by orthogonal methods and the functional consequences of the RNA-protein interactions have been rigorously tested.

DATA AVAILABILITY

Data and code used to generate sequencing tracks are available on Github at https://github.com/dhonson-lncrna/20241028_RibosomeSpidrFigures.git. Data and code used to plot SPIDR enrichments on cryo-EM structures are available on Github at https://github.com/dhonson-lncrna/20241101_chimeraX-Visualization.git. Raw SPIDR data is available on request pending publication of the manuscript.

ACKNOWLEDGEMENTS

Drew Honson and Mitch Guttman conceived of the project. Drew Honson performed all experiments and wrote this report. Mario Blanco ran the SPIDR data analysis pipeline and Mitch Guttman wrote the Java script for enrichment calculation. Zikun Zhu, Shou-ou Shan, and Jay Brito-Querido provided valuable feedback on the chosen proteins and the SPIDR

results. All imaging was performed on microscopes in the Caltech Biological Imaging Facility.

METHODS

Cell culture

HEK293 cell culture

HEK293s were used for all experiments in this study. Cells were maintained in standard HEK293 media (high glucose DMEM (Gibco, Life Technologies), 10% FBS (Seradigm), 1% penicillin/streptomycin (Gibco, Life Technologies)) at 37°C with 5% CO_2. Cells were passed using 0.25% trypsin-EDTA (Gibco, Life Technologies) every two to three days when cells reached 80-100% confluence.

SPIDR cell preparation

For each SPIDR experiment, three 10-cm plates of HEK293s were grown to 100% confluence. Media was removed and cells were rinsed with ice-cold PBS. The PBS was immediately removed, and cells were transferred to a UV crosslinking chamber. Cells were treated with $2.5x10^5$ μJ/cm^2 265nm UV light. Ice-cold PBS was immediately added to the plate, then cells were scraped and transferred to a conical tube. Cells were spun 3 minutes, 330g, 4°C. The supernatant was aspirated, and cells were resuspended in 1mL ice-cold PBS and transferred to 1.7mL Eppendorf tubes. Cells were spun 3 minutes, 1000g, 4°C, then the supernatant was removed, and cells were snap-frozen in liquid nitrogen. Cells were stored at -80°C until use.

siRNA Treatments

HEK293 cells were grown in 6-well dishes until approximately 50% confluence. Lipofectamine RNAiMAX (Thermo Scientific) was used to deliver siRNAs. 300µl final RNAiMAX reactions were created containing 9µl Lipofectamine reagent and 30pmol siRNA (see Appendix B). Cells were incubated for 72 hours, then lifted and replated onto coverslips as described in the *Immunofluorescence cell preparation* section.

Immunofluorescence cell preparation

For each immunofluorescence sample, #1.5 12mm coverslips (Warner) were coated with 0.2% gelatin (Millipore Sigma) for 1 hour in the incubator. The gelatin was aspirated and 100,000 HEK293 cells suspended in 500µl media were added to each coverslip. Coverslips were incubated for 3-4 hours to allow cells to adhere to the glass, then the media was removed and 4% formaldehyde in 1x PBS was added. Cells were crosslinked for 10 minutes at room temperature, then the formaldehyde was removed. The cells were rinsed twice in room temperature PBS, then dehydrated through 50%, 70%, and 100% ethanol. Cells were stored at -20°C until use, at which point they were rehydrated in the same manner.

SPIDR

Lysate preparation

Pellets comprising three 10-cm plates were thawed on ice and lysis was performed as described previously. 50µl aliquots of lysate were collected and treated with 1:500, 1:750, 1:1000, and 1:2000 final dilutions of Rnase If (NEB), 10 minutes, 37°C. RNA was immediately column cleaned using Zymo RNA Clean and Concentrate, treated with Dnase

to remove residual genomic DNA, then column cleaned again. The size distributions of the resultant RNA were measured using a TapeStation RNA High Sensitivity tape (Agilent). The optimal Rnase If dilution was scaled up to obtain RNA sizes between 50-1000nt with an average size 300-400nt. After Rnase If treatment, SUPERase·In Rnase Inhibitor (Thermo Scientific) was added to a 1:50 final dilution to inactivate Rnase If.

Bead preparation and quality control

Beads were prepared as previously described. Briefly, paramagnetic protein G beads (Thermo Scientific) were linked to biotin using EZ-Link Sulfo-NHS-Biotin (Thermo Scientific). Biotinylated antibody-ID oligonucleotides (IDT) were coupled to recombinant streptavidin (Biolegend) in 1:1 molar ratios. 10µl biotin-protein G beads were coupled to 0.6 pmol streptavidin-antibody-IDs.

To test antibody-ID density in each bead preparation, 10% of each bead set was removed and pooled. The remaining beads were stored at 4°C. The reserved beads were magnetically separated and resuspended in 200µl 1x Instant Sticky-End Master Mix (NEB) plus 100pmol term-ligates-even SPRITE barcode to add an i7 Illumina overhang. From there, SPIDR antibody-ID library preparation was completed, and the quality control library was sequenced. Any wells that were more than two-fold under- or over-represented in the antibody-ID library were not used for antibody coupling (i.e., if the cluster count was two-fold or more separated from 1 / total number of antibodies).

Once antibody-ID bead wells were selected, 2.5µg antibody was coupled to each bead set, then beads were pooled for immunoprecipitation.

Immunoprecipitation, barcoding, and library preparation

Immunoprecipitation was performed using the bead pool as previously described. 4 rounds of 12-well barcoding was performed and the resulting barcoded beads were divided into 10%, 5%, and 1% aliquots for library preparation. cDNA and antibody-ID libraries were prepared as described previously. All libraries were sequenced on an Element Biosciences AVITI 300-cycle Cloudbreak Kits using 180 nt read 1 x 120 nt read 2.

Data analysis

Bead assignment, cDNA assignment, and genome alignment were performed using the SPIDR pipeline. Second read truncations were used for enrichment analysis using a custom Java script. Briefly, SPIDR enrichments for each target were computed by randomly sampling second read counts with replacement, and computing the observed count over the expected counts from the total set for each randomization. The 95th percentile enrichment values floored at 2.0 were plotted for genomic tracks.

To map SPIDR data to structures, a custom Python script was written. Sequences from the genomic alignment file were aligned to sequences from the Protein Database (PDB) structures using MUSCLE. Nucleotides with SPIDR enrichment values greater than 2 were selected, and their indices were switched from those in the alignment genome to those from PDB. Colors were assigned based on the max-normalized enrichment values and values less than 5% the max value were excluded. A ChimeraX-readable Python script was then generated. Running the script from ChimeraX opened the structure based on PDB ID, deleted chains not desired for visualization, and recolored chains and residues based on the SPIDR data.

Immunofluorescence

Antibody labeling

After recovery from storage in ethanol, cells were treated with PBS + 0.5% Triton-X 100 for 10 minutes at room temperature. The liquid was removed and replaced with ViewRNA Cell Plus blocking buffer (Thermo Scientific). Cells were blocked 30 minutes at room temperature. Coverslips were flipped onto drops of diluted primary antibodies diluted in blocking buffer on parafilm and incubated overnight at 4°C. See Appendix B for antibody dilution factors. After the overnight incubation, coverslips were returned to the 24-well plate and washed three times for 5 minutes with PBST. Secondary antibodies diluted in blocking buffer were added to the plate and cells were incubated 1 hour at room temperature. See Appendix B for antibody dilution factors. Cells were then washed three times for 5 minutes with PBST then mounted in Prolong Gold + DAPI (Thermo Scientific).

Imaging

Coverslips were imaged on a Zeiss LSM 980 or LSM 880 with a 63x/1.4 NA oil objective. Zeiss Blue (for the LSM 980) or Black (for the LSM 880) Smart Setup was used to configure channels for DAPI, AlexaFluor 488 or DyLight 650. Individual cell images were collected at 6x optical magnification.

REFERENCES

1. Kimball, J. W. *3.4: Ribosomes*. (Tufts University and Harvard, 2022).

2. Zhang, Y. *et al.* Visualizing the nucleoplasmic maturation of human pre-60S ribosomal particles. *Cell Res.* **33**, 867–878 (2023).

3. Vanden Broeck, A. & Klinge, S. Principles of human pre-60S biogenesis. *Science* **381**, eadh3892 (2023).

4. Liang, X. *et al.* Structural snapshots of human pre-60S ribosomal particles before and after nuclear export. *Nat. Commun.* **11**, 3542 (2020).

5. Brito Querido, J. *et al.* The structure of a human translation initiation complex reveals two independent roles for the helicase eIF4A. *Nat. Struct. Mol. Biol.* **31**, 455–464 (2024).

6. Brito Querido, J., Díaz-López, I. & Ramakrishnan, V. The molecular basis of translation initiation and its regulation in eukaryotes. *Nat. Rev. Mol. Cell Biol.* **25**, 168–186 (2024).

7. Voorhees, R. M., Fernández, I. S., Scheres, S. H. W. & Hegde, R. S. Structure of the mammalian ribosome-Sec61 complex to 3.4 Å resolution. *Cell* **157**, 1632–1643 (2014).

8. Lentzsch, A. M. *et al.* NAC guides a ribosomal multienzyme complex for nascent protein processing. *Nature* **633**, 718–724 (2024).

9. Jomaa, A. *et al.* Mechanism of signal sequence handover from NAC to SRP on ribosomes during ER-protein targeting. *Science* **375**, 839–844 (2022).

10. Lafontaine, D. L. J., Riback, J. A., Bascetin, R. & Brangwynne, C. P. The nucleolus as a multiphase liquid condensate. *Nat. Rev. Mol. Cell Biol.* **22**, 165–182 (2021).

11. Stenström, L. *et al.* Mapping the nucleolar proteome reveals a spatiotemporal organization related to intrinsic protein disorder. *Mol. Syst. Biol.* **16**, e9469 (2020).

12. Saba, J. A. *et al.* LARP1 binds ribosomes and TOP mRNAs in repressed complexes. *EMBO J.* 1–18 (2024) doi:10.1038/s44318-024-00294-z.

13. Al-Ashtal, H. A., Rubottom, C. M., Leeper, T. C. & Berman, A. J. The LARP1 La-Module recognizes both ends of TOP mRNAs. *RNA Biol.* **18**, 248 (2019).

14. Philippe, L., van den Elzen, A. M. G., Watson, M. J. & Thoreen, C. C. Global analysis of LARP1 translation targets reveals tunable and dynamic features of 5′ TOP motifs. *Proc. Natl. Acad. Sci.* **117**, 5319–5328 (2020).

15. Brown, A., Baird, M. R., Yip, M. C., Murray, J. & Shao, S. Structures of translationally inactive mammalian ribosomes. *eLife* **7**, e40486 (2018).

16. Martini, S. *et al.* A genetically-encoded crosslinker screen identifies SERBP1 as a PKCε substrate influencing translation and cell division. *Nat. Commun.* **12**, 6934 (2021).

17. Heaton, J. H., Dlakic, W. M., Dlakic, M. & Gelehrter, T. D. Identification and cDNA cloning of a novel RNA-binding protein that interacts with the cyclic nucleotide-responsive sequence in the type-1 plasminogen activator inhibitor mRNA *. *J. Biol. Chem.* **276**, 3341–3347 (2001).

18. Wolin, E. *et al.* SPIDR: A highly multiplexed method for mapping RNA-protein interactions uncovers a potential mechanism for selective translational suppression upon cellular stress. *BioRxiv Prepr. Serv. Biol.* 2023.06.05.543769 (2023) doi:10.1101/2023.06.05.543769.

19. Dodel, M. *et al.* TREX reveals proteins that bind to specific RNA regions in living cells. *Nat. Methods* **21**, 423–434 (2024).

20. Khatter, H., Myasnikov, A. G., Natchiar, S. K. & Klaholz, B. P. Structure of the human 80S ribosome. *Nature* **520**, 640–645 (2015).

21. Amunts, A., Brown, A., Toots, J., Scheres, S. H. W. & Ramakrishnan, V. The structure of the human mitochondrial ribosome. *Science* **348**, 95–98 (2015).

22. Morello, L. G. *et al.* The NIP7 protein is required for accurate pre-rRNA processing in human cells. *Nucleic Acids Res.* **39**, 648–665 (2011).

23. Moraleva, A. A. *et al.* Eukaryotic ribosome biogenesis: The 40S subunit. *Acta Naturae* **14**, 14–30 (2022).

24. Rohrmoser, M. *et al.* Interdependence of Pes1, Bop1, and WDR12 controls nucleolar localization and assembly of the PeBoW complex required for maturation of the 60S ribosomal subunit. *Mol. Cell. Biol.* **27**, 3682 (2007).

25. Sanghai, Z. A. *et al.* Modular assembly of the nucleolar pre-60S ribosomal subunit. *Nature* **556**, 126–129 (2018).

26. Sailer, C. *et al.* A comprehensive landscape of 60S ribosome biogenesis factors. *Cell Rep.* **38**, 110353 (2022).

27. Kater, L. *et al.* Visualizing the assembly pathway of nucleolar pre-60S ribosomes. *Cell* **171**, 1599-1610.e14 (2017).

28. Adachi, K., Soeta-Saneyoshi, C., Sagara, H. & Iwakura, Y. Crucial role of Bysl in mammalian preimplantation development as an integral factor for 40S ribosome biogenesis. *Mol. Cell. Biol.* **27**, 2202 (2007).

29. Castella, S., Bernard, R., Corno, M., Fradin, A. & Larcher, J.-C. Ilf3 and NF90 functions in RNA biology. *WILEY Interdiscip. Rev.-RNA* **6**, 243–256 (2015).

30. Reichman, T. W. *et al.* Selective regulation of gene expression by nuclear factor 110, a member of the NF90 family of double-stranded RNA-binding proteins. *J. Mol. Biol.* **332**, 85–98 (2003).

31. Wu, T.-H., Shi, L., Lowe, A. W., Nicolls, M. R. & Kao, P. N. Inducible expression of immediate early genes is regulated through dynamic chromatin association by NF45/ILF2 and NF90/NF110/ILF3. *PLOS ONE* **14**, e0216042 (2019).

32. Wu, T.-H. *et al.* NF90/ILF3 is a transcription factor that promotes proliferation over differentiation by hierarchical regulation in K562 erythroleukemia cells. *PLOS ONE* **13**, e0193126 (2018).

33. Viranaicken, W. *et al.* L-Ilf3 and L-NF90 traffic to the nucleolus granular component: Alternatively-spliced exon 3 encodes a nucleolar localization motif. *PLOS ONE* **6**, e22296 (2011).

34. Tsai, H. *et al.* NF45/NF90-mediated rDNA transcription provides a novel target for immunosuppressant development. *EMBO Mol. Med.* **13**, e12834 (2021).

35. Lee, H., Han, S., Kwon, C. S. & Lee, D. Biogenesis and regulation of the let-7 miRNAs and their functional implications. *Protein Cell* **7**, 100–113 (2016).

36. Sun, Z. *et al.* LIN28 coordinately promotes nucleolar/ribosomal functions and represses the 2C-like transcriptional program in pluripotent stem cells. *Protein Cell* **13**, 490–512 (2022).

37. Sha, Z. *et al.* The eIF3 interactome reveals the translasome, a supercomplex linking protein synthesis and degradation machineries. *Mol. Cell* **36**, 141–152 (2009).

38. des Georges, A. *et al.* Structure of mammalian eIF3 in the context of the 43S preinitiation complex. *Nature* **525**, 491–495 (2015).

39. Brito Querido, J. *et al.* Structure of a human 48S translational initiation complex. *Science* **369**, 1220–1227 (2020).

40. Anger, A. M. *et al.* Structures of the human and Drosophila 80S ribosome. *Nature* **497**, 80–85 (2013).

41. Grove, D. J., Russell, P. J. & Kearse, M. G. To initiate or not to initiate: A critical assessment of eIF2A, eIF2D, and MCT-1·DENR to deliver initiator tRNA to ribosomes. *WIREs RNA* **15**, e1833 (2024).

42. Weisser, M. *et al.* Structural and functional insights into human re-initiation complexes. *Mol. Cell* **67**, 447-456.e7 (2017).

43. Wilson, D. M. *et al.* Structural insights into assembly of the ribosomal nascent polypeptide exit tunnel. *Nat. Commun.* **11**, 5111 (2020).

44. Luirink, J. & Sinning, I. SRP-mediated protein targeting: Structure and function revisited. *Biochim. Biophys. Acta BBA – Mol. Cell Res.* **1694**, 17–35 (2004).

APPENDIX A: Materials Tables for Chapter 1

REAGENTS TABLE

Reagent	Source	Catalog #	Notes
1,2-propanediol	Sigma	398039	
1x PBS	Thermo Scientific	10010023	
200 Proof Pure Ethanol	VWR	V1001TP	For Zymo DNA Wash Buffer
2x Instant Sticky-End Master Mix	NEB	M0370L	
5x Quick Ligation Mix	NEB	B6058S	
Acetonitrile (ACN)	Fisher Scientific	AA47138K2	
ATP, 100mM	NEB	N0437A	
Collection Tubes	Zymo	C1001-50	
cOmplete EDTA-free protease inhibitor cocktail	Sigma	4693159001	
Cytiva Sera-Mag SpeedBeads™ Carboxyl Magnetic Beads, hydrophilic	Fisher Scientific	09-981-121	

Cytiva Sera-Mag SpeedBeads™ Carboxyl Magnetic Beads, hydrophobic	Fisher Scientific	09-981-123	
DMSO	Sigma	D8418	
dNTP Mix, 10mM	NEB	N0447L	
EDTA, pH 8.0, 0.5M	Thermo Scientific	15575020	
Ethyl alcohol (HPLC, Meets ACS)	Fisher Scientific	A995-4	For peptide clean-up
Exonuclease I	NEB	M0293S	
HEPES, pH 7.4, 1M	Teknova	H1030	
IC Columns	Zymo	C1004-50	
IIC Columns	Zymo	C1011-50	
Iodoacetamide (IAA)	Thermo Scientific	35603	
Lithium chloride, 8M	Sigma	L7026-1L	
Lys-C Mass Spec Grade	Promega	VA117A	

Maxima H Minus Reverse Transcriptase	Thermo Scientific	EP0751	
NEB Dynabeads Oligo(dT)25	NEB	S1419S	
Pierce™ NHS-Activated Magnetic Beads	Thermo Scientific	88826	For UV crosslinking test
QIAshredder	Qiagen	79656	
RiboLock RNase Inhibitor	Thermo Scientific	EO0384	
RNA Clean & Concentrator-5	Zymo	R1015	
Sodium chloride	Macron (via VWR)	7581-06	
Sodium deoxycholate	Sigma	D6750	
Sodium dodecyl sulfate, 20%	Thermo Scientific	AM9820	
T4 PNK	NEB	M0201L	
TCEP-HCl	Thermo Scientific	20490	
TEAB	Millipore Sigma	T7408	

Tris-HCl, pH 7.5, 1M	Thermo Scientific	15567027	
Triton X-100	Sigma	T8787	
Trypsin Platinum, Mass Spec Grade	Promega	VA9000	
Turbo Dnase	Thermo Scientific	AM2238	
UltraPure Water	Thermo Scientific	10977015	For RAP-MS buffers
Urea	Sigma	U1250	Pellet format improves solubility
Water, Optima™ LC/MS Grade	Fisher Scientific	W64	For peptide clean-up

APPENDIX B: Materials Tables for Chapters 2 and 3

ANTIBODY TABLE

Target	Manufacturer	Product No	Notes
ATRX (D1N2E) Rabbit mAb	CST	14820S	
BTF3/NACB	Bethyl	A302-319A	Used only in rep3
BYSL Polyclonal antibody	Proteintech	28319-1-AP	1:50 for IF
CELF1 (E8Q4G) Rabbit mAb	CST	95084S	
DDX46 Polyclonal antibody	Proteintech	16927-1-AP	
EIF3C Polyclonal antibody	Proteintech	12733-1-AP	
eIF3J (D21G7) XP® Rabbit mAb	CST	8161S	
EIF2D/LGTN Polyclonal antibody	Proteintech	12840-1-AP	
Fibrillarin Monoclonal antibody	Proteintech	66985-1	IF only, 1:1000
MEX3A Polyclonal antibody	Proteintech	30865-1-AP	
MRPL37 Polyclonal antibody	Proteintech	29522-1-AP	Used in rep3 only

MRPS18A Polyclonal antibody	Proteintech	16235-1-AP	Used in rep3 only
MTCH2 Polyclonal antibody	Proteintech	16888-1-AP	Used in rep3 only
PES1 Polyclonal antibody	Proteintech	13553-1-AP	1:50 for IF
RPL13 Polyclonal antibody	Proteintech	11271-1-AP	
RPL14 Polyclonal antibody	Proteintech	14991-1-AP	
RPL23 Polyclonal antibody	Proteintech	16086-1-AP	
RPL36 Polyclonal antibody	Proteintech	15145-1-AP	
RPL5 (D5Q5X) Rabbit mAb	CST	51345S	
RPS12 Polyclonal antibody	Proteintech	16490-1-AP	
RPS19 Polyclonal antibody	Proteintech	15085-1-AP	
RPS25 Polyclonal antibody	Proteintech	23599-1-AP	
RPS26 Polyclonal antibody	Proteintech	14909-1-AP	
RPS5 Polyclonal antibody	Proteintech	16964-1-AP	

SBDS Polyclonal antibody	Proteintech	17618-1-AP	Used only in rep1
Sec61B (D5Q1W) Rabbit mAb	CST	14648S	
SEC62 Polyclonal antibody	Proteintech	28693-1-AP	Used only in rep1
SEC63 Polyclonal antibody	Proteintech	13978-1-AP	
SERBP1 Polyclonal antibody	Proteintech	10729-1-AP	
SRP54 Polyclonal antibody	Proteintech	11729-1-AP	
SSR3/TRAPG Polyclonal antibody	Proteintech	30851-1-AP	Used in rep3 only
TINP1/NSA2 Polyclonal antibody	Proteintech	16230-1-AP	
NOP16/HSPC111 antibody	Bethyl	A305-125A	1:50 for IF
SRP72 Polyclonal antibody	Bethyl	A304-594A	
LARP1 (D8J4F) Rabbit mAb	CST	14763	
LIN28B (D4H1)	CST	11965	1:100 for IF
DDX52	Bethyl	A303-054A	
ILF3	Bethyl	A303-0651A	1:300 for IF

RPS3	Bethyl	A308-840A	

REAGENTS TABLE

Reagent	Source	Catalog #	Notes
1,2-propanediol	Sigma	398039	
1x PBS	Thermo Scientific	10010023	
200 Proof Pure Ethanol	VWR	V1001TP	For Zymo DNA Wash Buffer
2x Instant Sticky-End Master Mix	NEB	M0370L	
5x Quick Ligation Mix	NEB	B6058S	
ATP, 100mM	NEB	N0437A	
Collection Tubes	Zymo	C1001-50	
cOmplete EDTA-free protease inhibitor cocktail	Sigma	4693159001	
dNTP Mix, 10mM	NEB	N0447L	

Dynabeads™ Protein G for Immunoprecipitation	Invitrogen	10003D	
EDTA, pH 8.0, 0.5M	Thermo Scientific	15575020	
Exonuclease I	NEB	M0293S	
EZ-Link Sulfo-NHS Biotin	Thermo Scientific	21217	
HEPES, pH 7.4, 1M	Teknova	H1030	
IC Columns	Zymo	C1004-50	
IIC Columns	Zymo	C1011-50	
Maxima H Minus Reverse Transcriptase	Thermo Scientific	EP0751	
Q5 High Fidelity 2X Master Mix	NEB	M0492L	
RiboLock RNase Inhibitor	Thermo Scientific	EO0384	
RNA Clean & Concentrator-5	Zymo	R1015	
Silencer® Select (1nmol) siRNA: ILF3 Human	Fisher	s7403	

Silencer® Select (1nmol) siRNA: LIN28B Human	Fisher	s52479	
Silencer® Select (1nmol) siRNA: PES1 Human	Fisher	s23913	
Sodium chloride	Macron (via VWR)	7581-06	
Sodium deoxycholate	Sigma	D6750	
Sodium dodecyl sulfate, 20%	Thermo Scientific	AM9820	
Purified Streptavidin	Biolegend	405150	
T4 PNK	NEB	M0201L	
Tris-HCl, pH 7.5, 1M	Thermo Scientific	15567027	
Triton X-100	Sigma	T8787	
Turbo Dnase	Thermo Scientific	AM2238	
UltraPure Water	Thermo Scientific	10977015	