# Novel tools for multi-omic characterization of subnuclear RNA-scaffolded structures in development and disease

Evan Erik Kania

A dissertation

submitted in partial fulfillment of the

requirements for the degree of

Doctor of Philosophy

University of Washington

2023

Reading Committee:

David Shechner, Chair

Shao-En Ong

William Catterall

Program Authorized to Offer Degree:

Pharmacology

University of Washington

## Abstract

Novel tools for multi-omic characterization of subnuclear RNA-scaffolded structures in development and disease

Evan Erik Kania

Chair of the Supervisory Committee:
David Shechner
Department of Pharmacology

RNA molecules are increasingly recognized as fundamental regulators of subcellular organization. In the nucleus, RNAs scaffold numerous structures that mediate diverse and essential genomic functions. These range from small-scale chromatin interactions that co-regulate handfuls of genes, to subnuclear organelles that collectively control cellular gene-expression, epigenetic, and metabolic programs. RNA-scaffolded structures are characteristically dysregulated in, and thought to be causally linked to, a plethora of human diseases, including developmental and neurodegenerative disorders, retroviral infections, and cancer. Nuclear architectural RNAs may therefore represent a large, untapped pool of novel therapeutic targets. However, elucidating and characterizing these targets—identifying the proteins, RNAs, and genomic loci with which a given RNA interacts—remains challenging.

The studies presented in this thesis represent groundbreaking advancements in technologies I generated to capture and characterize nuclear RNA-scaffolded structures and their associated genomic loci, proteins, and RNAs, even those with low abundance. Employing these innovative techniques on gene silencing long non-coding RNAs (lncRNAs), I identified the biomolecular constituents of the nuclear structures they scaffold. Notably, this work shed light on the molecular constituents of the X-chromosome inactivation center, scaffolded by the nascent Xist transcript, and provided a better understanding of gene silencing lncRNA function. Moreover, I also applied the suite of technologies I developed to investigate a severe form of dilated cardiomyopathy caused by RBM20 mutations (RBM20-DCM), focusing on the nascent TTN transcript that naturally scaffolds nuclear RBM20 foci. These analyses revealed key biomolecules crucial for the native function of RBM20 foci and unveiled aberrations in alternative splicing and mitochondrial reactive oxygen species handling upon loss of RBM20, offering insight into the pathogenesis of RBM20-DCM.

Collectively, this thesis lays the foundation for exploring nuclear RNA interactomes in various contexts, facilitating a deeper understanding of these structures in both native biological and aberrant disease states. The newly developed technologies hold significant potential for broader applications in understanding the composition and localization of other RNA-scaffolded structures, providing invaluable information for future research and potential therapeutic interventions. By delving into the complexities of nuclear RNA-scaffolded structures, these studies not only advance our understanding of cellular regulation but also open up exciting possibilities for targeted therapies against numerous human diseases. These findings represent a significant advancement in the field of RNA biology and lay the foundation for further investigations in the realm of RNA-based therapeutics and precision medicine. Ultimately, the discoveries made here

pave the way for future breakthroughs in deciphering the mechanisms governing RNA-mediated subcellular organization and the promising utilization of RNA scaffolded structures and their constituents as therapeutic targets.

# TABLE OF CONTENTS

# PREFACE

Portions of the text and data from this dissertation are reprinted (adapted) from the following works under fair use and/or with permission under the terms of the CC-BY License:

1. Tsue AF*, Kania EE*, Lei DQ, Fields R, McGann CD, Hershberg E, Deng X, Kihiu M, Ong SE, Disteche CM, Beliveau BJ, Schweppe DK, and Shechner DM. 2023. Oligonucleotide-directed proximity-interactome mapping (O-MAP): A unified method for discovering RNA-interacting proteins, transcripts and genomic loci in situ. bioRxiv 2023.01.19.524825. *denotes co-first authorship.

# LIST OF FIGURES

## CHAPTER 1 FIGURES



## CHAPTER 2 FIGURES



## CHAPTER 3 FIGURES

# ACKNOWLEDGEMENTS

I would first like to thank my PhD advisor, David Shechner, for his support and encouragement throughout graduate school. I am grateful for the opportunity to pursue the questions that excited me the most, to mentor others, and to present my work in Singapore. I would also like to sincerely thank John Scott for welcoming me to his lab meetings to discuss my ongoing experiments where I received invaluable feedback from his lab. Thank you to my committee members, John Scott, Shao-En Ong, William Catterall, and Brian Beliveau for your guidance and expertise over the years. Thank you to the Ong Laboratory, namely Shao-En Ong and Daphnée Marciniak for the fantastic help with the O-MAP-MS experiments. Thank you to the Murry Laboratory, namely Chuck Murry and Aiden Fenix for teaching me how to culture stem cells and differentiate cardiomyocytes. Thank you to my undergraduate advisors, Jack Yalowich, Terry Elton, and Christopher Coss for providing an environment that fueled my passion for science and for equipping me with a solid foundation that enabled me to embark upon my doctoral studies with confidence. Thank you to my colleagues, namely Sophia, Ashley, Diana, and Caitlin, for making meaningful connections, sharing enjoyable experiences, and for lending each other a helping hand or shoulder to lean on. Thank you to my family, Jennifer, Erik, Hannah, Harrison, Terri, Logan, Shelly, Penny, and George for your unwavering support throughout my time in graduate school. Especially to my parents who ingrained in me resilience, a habit of helping others, and the courage to relentlessly pursue my aspirations. Lastly, I want to express my heartfelt gratitude to my husband Justin for his patience, friendship, and boundless unconditional love. Thank you for the countless adventures we have shared over the years. You ground me in the present moment and remind me to slow down and cherish the little things in life. You bring out the best in me and make every day a joy.

# DEDICATION

To my beloved husband and cherished family,

For the unwavering love and boundless support,

This thesis is dedicated to you.

With all my heart, thank you.

# Chapter 1. INTERROGATING NUCLEAR RNA-SCAFFOLDED STRUCTURES

## 1.1 ABSTRACT

RNA molecules play a critical role in regulating subcellular organization. In the nucleus, RNAs scaffold numerous structures that mediate diverse and essential genomic functions. These cellular processes range from local gene co-regulation to global gene-expression control, epigenetic modifications, and metabolic programs. Dysregulation of these structures is linked to numerous human diseases, presenting RNA-scaffolded structures as potential therapeutic targets. Currently, three main methods are employed to study RNA-scaffolded structures: biochemical fractionation, oligonucleotide-mediated pulldown, and proximity labeling techniques. However, each method faces limitations in terms of native cellular context, RNA targeting specificity, and overall accuracy. To address these challenges, we introduce Oligonucleotide-directed proximity interactome MAPping (O-MAP), a novel multi-omic approach for studying RNA-scaffolded structures. O-MAP enables precise targeting of low-abundance RNA species within their native cellular environment, offering superior biological context and flexibility.

First, O-MAP demonstrated its versatility by precisely probing various RNA targets with diverse lengths and expression levels. It efficiently captured even low-abundance transcripts, relying on longer biotinylation labeling times and larger primary probe pools. Subsequently, O-MAP's capabilities were expanded to transcriptomic analysis. O-MAP-Seq effectively provided an in-depth identification of nucleolar transcripts. Likewise, the O-MAP toolkit was further diversified for genomic analysis. O-MAP-ChIP enabled comprehensive mapping of genomic interactions within RNA-scaffolded structures with unprecedented resolution and revealed

conserved and variable elements of nucleolar-associated domains (NADs), providing insight on their role in pancreatic ductal adenocarcinoma.

In summary, O-MAP overcomes the limitations of current RNA interaction-discovery methods by offering higher precision, broader biological context, and ease of use. This novel multi-omic approach opens new avenues for exploring RNA-mediated compartmentalization, facilitating groundbreaking discoveries in the field of spatial biology. By precisely characterizing RNA interactions within native cellular environments, O-MAP holds significant promise in advancing our understanding of RNA-scaffolded structures and their relevance to human diseases, thus potentially unlocking novel therapeutic targets.

## 1.2 INTRODUCTION

RNA molecules are increasingly recognized as fundamental regulators of subcellular organization (Farley et al., 2015; Fazal et al., 2019; Rinn & Guttman, 2014; Smith et al., 2020). In the nucleus, RNAs scaffold numerous structures that mediate diverse and essential genomic functions. These range from small-scale chromatin interactions that co-regulate handfuls of genes, to subnuclear organelles that collectively control cellular gene-expression, epigenetic, and metabolic programs (Rinn & Guttman, 2014; Smith et al., 2020). RNA-scaffolded structures are characteristically dysregulated in, and thought to be causally linked to, a plethora of human diseases, including developmental and neurodegenerative disorders, retroviral infections, and cancer (Balas & Johnson, 2018; Bernat & Disney, 2015; Lukong et al., 2008; Osborne & Thornton, 2006). Nuclear architectural RNAs may therefore represent a large, untapped pool of novel therapeutic targets. However, elucidating and characterizing these targets—identifying the proteins, RNAs, and genomic loci with which a given RNA interacts—remains challenging.

Historically, RNA-scaffolded structures have broadly been studied from three perspectives: biochemical fractionation, oligonucleotide-mediated pulldown, and proximity labeling methodologies (Cao et al., 2019; Mitrea et al., 2022; Németh et al., 2010). Canonical biochemical fractionation techniques have been successful for identifying the contents of larger, biochemically well-behaved RNA-scaffolded structures such as nucleoli (Németh et al., 2010; Roeder & Rutter, 1970). The approach has also had some success for analyzing smaller and more heterogeneous RNA-scaffolded structures *en-masse,* including interchromatin granule clusters that constitute nuclear speckles and paraspeckles, among others (Mintz et al., 1999). Oligonucleotide-mediated pulldown methodologies hybridize single-strand DNA oligonucleotides antisense to the RNA of interest via canonical Watson-Crick base pairing. The oligonucleotides are appended with a biochemical handle, such as biotin, which is used to pulldown the RNA-oligonucleotide complex from a native cell lysate, enriching for material directly bound to the RNA of interest (Chu et al., 2012; Engreitz et al., 2013; Simon, 2013). Oligonucleotide-mediated pulldown strategies have been successful for the identification of proteins, RNA, and genomic loci directly bound to RNA (McDonel & Guttman, 2019; Ramanathan et al., 2019; Simon & Machyna, 2019). Lastly, proximity labeling technologies send an enzyme to the RNA-scaffolded structure, to covalently tag nearby proteins with a chemical handle, such as biotin. Enzyme targeting can be achieved either *in vivo,* by genetically appending a protein known to localize to the RNA-scaffolded structure with the enzyme, or *in situ,* via an enzyme-linked antibody against a protein that localizes to the RNA-scaffolded structure (Bar et al., 2018; Fazal et al., 2019; Kaewsapsak et al., 2017; Yi et al., 2020). Following proximity labeling, the cells are then lysed, and the chemical handle is pulled down from crude lysate. Proximity labeling technologies have the advantage of being able

to identify the holistic content of RNA-scaffolded structures, both direct and indirect interactions, in a manner more specific than biochemical fractionation.

All these technologies have key limitations. Biochemical fractionation and oligonucleotide-mediated pulldown methods lack native cellular context since the cell is lysed before being fractionated and before oligos are directed against the RNA of interest (Chu & Chang, 2018; van Koningsbruggen et al., 2010). Moreover, biochemical fractionation and protein-directed proximity labeling techniques have poor portability to different RNAs, as neither technique is RNA-directed, and not every RNA-scaffolded structure has a known proteome that can be used to design the targeting modality (Bar et al., 2018; Mintz et al., 1999; Saitoh et al., 2004). Finally, all three techniques commonly lack specificity for a single RNA-scaffolded structure. Biochemical fractionation schemes commonly isolate mixed populations of RNA-scaffolded structures in each fraction, meaning very few compartments can be accurately fractionated (Huber et al., 2003). Moreover, biochemical purification schemes often fail due to the dynamic and fragile nature of RNA-scaffolded structures. Oligonucleotide-mediated pulldown techniques suffer from the inability to vet the specificity of the oligonucleotides for their target RNA, leading to the unintended capture of off-target transcripts and ultimately, to low signal-to-noise (Machyna & Simon, 2017). Protein-directed proximity labeling methods also often suffer from low signal-to-noise since the protein targeted is rarely exclusively localized to the nuclear compartment of interest (Binder et al., 2014). There is often a sizeable pool of the targeted protein elsewhere in the cell that results in background labeling. For example, proteins must undergo translation in the cytosol prior to being transported to the nuclear RNA-scaffolded structure of interest. This leads to unwanted, diffuse cytosolic and nucleoplasmic labeling. Proximity labeling methods that directly target RNA *in vivo*, including MS2-tagging to pulldown direct RNA-interactors or

CRISPR-based RNA-United Interacting System (CRUIS) which targets a proximity ligase to RNA via dCas13a, require tedious and time-consuming transgenic approaches (Yoon et al., 2012; Z. Zhang et al., 2020). MS2-tagging also suffers from its inability to capture higher-order interactions, while CRUIS is limited by sgRNA optimization (Z. Zhang et al., 2020). Crucially, both CRUIS and MS2-tagging can affect the structure of the targeted RNA, which can drastically alter the RNA's interactome.

To address these challenges, new tools are needed that probe RNA-scaffolded structures within a native cellular context with high specificity. During the course of this thesis, the Shechner lab developed Oligonucleotide-directed proximity interactome mapping (O-MAP) for imaging and proteomic mapping (O-MAP-MS) of RNA-scaffolded structures built on high-abundance RNA species, such as nucleoli and 7sk transcriptional hubs. In this chapter, the core O-MAP imaging technology was optimized for portability to lower abundant RNA species and was further adapted for identifying the transcriptome (O-MAP-Seq) and genomic loci (O-MAP-ChIP) of RNA-scaffolded structures built on both high- and low-abundance RNAs.

## 1.3 RESULTS

### O-MAP IS BROADLY APPLICABLE TO DIFFERENT RNA TARGETS

Oligonucleotide-directed proximity-interactome mapping (O-MAP) is a straightforward and flexible method for applying proximity-labeling to individual RNA targets in genetically unmodified samples. O-MAP utilizes the same peroxidase/tyramide chemistry used in APEX-based proximity-omics approaches. (Kaewsapsak et al., 2017; Rhee et al., 2013), but it relies on programmable oligonucleotide probes, rather than transgenic expression, to deploy biotinylating enzymes to endogenous target RNAs (Fig. 1.1, a). In O-MAP, cells are chemically fixed and pools of antisense DNA probes are hybridized to a target RNA, similar to high-resolution RNA-FISH

(Kishi et al., 2019). These probes then recruit a secondary oligo that is linked to Horseradish Peroxidase (HRP). Upon addition of its biotin-tyramide substrate, HRP generates freely diffusing biotin radicals that affinity-tag nearby proteins, RNAs, and DNA loci, enabling their isolation by simple streptavidin pulldown (Y. Chen et al., 2018; Rhee et al., 2013). To examine O-MAP's portability to different RNAs, an array of transcripts with diverse lengths, expression levels, sequence composition, biogenesis pathways, and localization were targeted. For each new target, probes were designed using the OligoMiner pipeline (Beliveau et al., 2018), and the optimal *in situ* biotinylation time was determined empirically via a labeling time course. Encouragingly, in each case, O-MAP yielded prominent *in situ* biotinylation that recapitulated the target's known subcellular localization. Validation of signal localization was performed using the O-MAP Probe Validation Assay (Fig 1.1, b). Derived from a standard RNA-FISH method (Raj & Rinn, 2019), our O-MAP Probe Validation Assay first groups probes into "odd" and "even" sub-pools. Each sub-pool is outfitted with a distinct landing pad that recruits either an HRP-conjugated or fluorescent secondary oligo, enabling O-MAP and RNA-FISH to be performed simultaneously. Off-targeting (i.e. non-colocalizing) probes are then readily identified and eliminated. The O-MAP Probe Validation Assay further confirmed the RNA-targeting accuracy of the probe pool: in all cases *in situ* biotinylation and RNA-FISH signals exhibited a high degree of overlap (Fig 1.1, c). This was observed with highly abundant transcripts like the chromatin-regulatory lncRNAs MALAT1 and NEAT1, with modestly expressed targets like the WDR7 mRNA, and with low-abundance RNAs like Firre. Obtaining robust O-MAP at low-abundance targets required large primary probe pools (100–500 oligos) to bolster the number of HRP molecules recruited to the target transcript, as well as longer labeling times (60–120 minutes).

## O-MAP SEQ FOR PROBING RNA-PROXIMAL TRANSCRIPTS

Having assessed the validity of our O-MAP probe sets via imaging, we sought to expand the technique to transcriptomic analysis—mapping the transcripts localized near a target RNA. However, because tyramide-radical chemistry is markedly less efficient at labeling nucleic acids than proteins (Y. Chen et al., 2018), we anticipated that the direct capture of in situ-biotinylated RNA would be challenging, and require large-scale input cell growths. Therefore, we adopted a strategy based on APEX-RIP, in which formaldehyde crosslinks are retained during cell lysis and enrichment, and RNAs within the target compartment are captured by pulling down the biotinylated proteins to which they are covalently bound (Fazal et al., 2019) (Fig. 1.2, a). The high capture efficiency of this approach enabled us to precisely map RNA-proximal transcriptomes from as few as $8.3\times10^6$ cells, approximately 12-24–fold lower than that required by oligo-capture-based methods (Engreitz et al., 2014; Simon & Machyna, 2019).

As a first test case, we applied O-MAP-Seq to the HeLa nucleolus, using a probe set targeting a domain on the 47S pre-ribosomal RNA that is exclusively nucleolar in localization (Padovan-Merhar et al., 2015). The nucleolar transcriptome is thought to be predominantly noncoding (Pederson & Politz, 2000), comprising RNAs that regulate ribosome biogenesis (e.g., small nucleolar RNAs, snoRNAs) (Bachellerie et al., 2002), transcripts with putative nucleolar-architectural and chromatin-regulatory roles (Caudron-Herger et al., 2016), and a large cohort of long noncoding RNAs (lncRNAs) with yet-undefined functions (Fazal et al., 2019). Collectively, these transcripts arise from all three RNA polymerases and span lengths from 20 nucleotides to several kilobases (Bai et al., 2014). We attempted to capture these diverse species by using a non-poly(A)-selective, rRNA-depletion-based library preparation strategy similar to that established previously (Adiconis et al., 2013). This yielded a catalog of 47S-proximal RNAs that recapitulated

much of the known nucleolar transcriptome, underscoring O-MAP-Seq's precision and accuracy (Figs. 1.2, b–c). Of note, we observed conspicuous enrichment of every snoRNA detectable in our libraries (40 total, average 1.98 fold enrichment, p=0.011; FDR 0.05), the "transcribed spacer" domains within the 47S pre-rRNA itself (Raap et al., 1995), and nucleolar lncRNAs like SLERT (Fig 1.3) (M. Wu et al., 2021). As predicted, coding genes were broadly de-enriched, with only 940 (4.64% of detectable genes) exhibiting significant nucleolar enrichment (Fig. 1.2, b–c; average $p_{adj}$ = 0.009; FDR 0.05; p-values calculated by DESeq2 (Love et al., 2014)). Moreover, isoform-level analysis revealed that most of these nucleolar-enriched, nominally protein-coding transcripts were in fact noncoding isoforms—most prominently, transcripts with retained introns (∼46% of enriched transcripts, compared to ∼4% of the de-enriched pool, Fig. 1.2, d). A similar intron-retention phenomenon has been recently observed in other subnuclear compartments (Barutcu et al., 2022), suggesting that our observations represent bona fide biological regulation and not artifacts of the O-MAP-Seq pipeline. Likewise, the genomic localization and domain architecture of 47S O-MAP-Seq-enriched transcripts suggest their validity as nucleolar RNAs. Nearly half (49.1%) of these transcripts are expressed from genes located within Nucleolar Associated chromatin Domains (NADs, Fig. 1.2, e), compared to 7.7% of all genes. A similar enrichment of NAD-encoded transcripts was observed previously (Fazal et al., 2019), suggesting that these RNAs may be localized to the nucleolus co-transcriptionally. Nucleolar transcripts were also enriched for nearly every family of transposable element (TE), with the AluSg4 family of Alu SINEs being the most significant (Fig. 1.2, g, padj = $2.22\times10^{-30}$; FDR 0.05). This is consistent with previous work demonstrating that Alu repeats are particularly enriched in nucleolar- and lamina-associated RNAs (Fazal et al., 2019), and that these RNAs may play a role in stabilizing nucleolar architecture (Caudron-Herger et al., 2015). Finally, our 47S O-MAP-Seq data also identified 3186 putative

novel nucleolar transcripts. We confirmed by RNA-FISH that one such transcript, ENSG00000286147.1 (9.85-fold nucleolar enrichment; $p_{adj}$ = 0.048), exhibited nucleolar and perinucleolar localization (69.77%, n=43; Fig. 1.2, f). This further validates O-MAP-Seq's accuracy and demonstrates its ability to discover novel RNAs within a target subcellular compartment.

**O-MAP CHIP FOR PROBING RNA-PROXIMAL GENOMIC LOCI**

We next examined if O-MAP could map the chromatin loci within a target transcript's subnuclear compartment. As with O-MAP-Seq, our strategy relied on formaldehyde crosslinking and the capture of *in situ* biotinylated proteins to enrich nearby DNA, similar to Chromatin Immunoprecipitation (ChIP; Fig. 1.5, a). We next sought to use O-MAP-ChIP to profile nucleolar-chromatin interactions in HeLa cells, using the same 47S pre-rRNA-targeting probe set as above. Mammalian nucleoli are surrounded by megabase-scale chromatin structures termed Nucleolar Associated Domains (NADs) which comprise nearly half of all heterochromatin, and which are central to epigenetic programming (Matheson & Kaufman, 2016). Although NADs have been characterized by isolating and sequencing intact nucleoli (van Koningsbruggen et al., 2010), this demanding approach has been challenging to apply in more than a handful of human cell lines. In contrast, 47S O-MAP-ChIP enabled comprehensive, high-resolution maps of HeLa NADs from only $4\times10^6$ cells, with approximately five days' hands-on time (Fig. 1.5, b). These O-MAP-ChIP data largely recapitulated the NAD architecture mapped by biochemical fractionation (van Koningsbruggen et al., 2010), with nearly 72% overlap between the two data sets ($p=1.21 \times 10^{-62}$, Fisher's exact test), even though they were acquired from different cell lines (Fig. 1.5, c). O-MAP-ChIP also appeared markedly less noisy than fractionation-sequencing, with higher agreement between replicates (Fig. 1.5, b and Fig 1.6).

NAD architecture is almost universally remodeled in cancer, potentially driving epigenetic and transcriptional changes that facilitate oncogenesis (Derenzini et al., 1998; Farley et al., 2015; Matheson & Kaufman, 2016). Yet, the functional impact of this dysregulation has been difficult to assess without robust methods for characterizing NAD architecture across cancer types. To demonstrate how O-MAP might facilitate this analysis, we applied 47S O-MAP-ChIP across four Pancreatic Ductal Adenocarcinoma (PDA) cell lines, systematically interrogating NAD organization across both the "classical" (ASPC1 and SUIT2 lines) and "basal" (8988T and Panc3.27 lines) PDA subtypes (Diwakarla et al., 2017) (Fig. 1.5, d). Although nucleolar morphology differs markedly between PDA subtypes (Diwakarla et al., 2017), 47S O-MAP-ChIP revealed a set of invariant nucleolar-genomic interactions (160 NADs) conserved across all cell lines, suggesting a core PDA NAD architecture. These conserved domains comprise the most abundant NAD class (approximately 55% of NADs), though sizeable groups of cell line-specific domains (30–52 NADs; 10–18%), or domains uniquely absent from a cell line (4–32 NADs; 1.4–11%) were also observed. We also observed NADs unique to each PDA subtype—15 and six NADs in classical and basal, respectively (Figs. 1.5, d and e). These variable domains may be particularly important for PDA, in which epigenetic regulation is thought to play a key role during oncogenesis and subtype specification (Lomberk et al., 2018). ChromHMM analysis was used to identify cell line specific NAD epigenomic signatures (Ernst & Kellis, 2012). This showed clear differences in zinc-finger/repeats (ZNF/Repeats) and heterochromatin domains between NADs of different cell types, with a profound loss of ZNF/Repeats in 8988T NADs (Fig. 1.5, f). To our knowledge, any one of these data sets would represent the highest-resolution map of human NADs reported to date. Furthermore, these data— combined with our nucleolar proteomic (Tsue, Kania et. al, 2023) and transcriptomic analyses (Fig 1.2)—demonstrate O-MAP's capacity for

parallelized, “multi-omic” dissection of RNA-scaffolded compartments using a common workflow.

## 1.4 DISCUSSION

Here, we have adapted the former O-MAP protocol to optimize it for imaging low abundant RNA species and further expanded the O-MAP toolkit for elucidating the transcripts (O-MAP-Seq) and genomic loci (O-MAP-ChIP) near an RNA of interest. We believe this toolkit holds considerable advantages over established RNA interaction-discovery methods, offering superior precision, biological context, flexibility, ease of use, and cost. Moreover, we anticipate that O-MAP’s ability to probe higher-order interactions within a transcript’s subcellular “neighborhood” will enable unprecedented analysis of RNA-mediated compartmentalization, a powerful new approach for spatial biology.

The most effective RNA interaction-discovery methods use antisense oligonucleotides to pull down a target RNA *ex vivo* (Chu & Chang, 2018; McDonel & Guttman, 2019; Ramanathan et al., 2019; Simon & Machyna, 2019), or transgenic manipulation to localize proximity-labeling enzymes to that RNA in live cells (S. Han et al., 2020; Y. Han et al., 2019; Ramanathan et al., 2018). Compared to oligo pulldown-based approaches (e.g. ChIRP, ChART, RAP) (Chu & Chang, 2018; McHugh & Guttman, 2018; Simon & Machyna, 2019), O-MAP offers more precise optimization, higher sensitivity, and broader biological context. A chief limitation of these pulldown-based techniques is their inability to optimize the targeting precision of the capture-oligo pool (Simon et al., 2011; Simon & Machyna, 2019), leading to off-targeting noise. O-MAP overcomes this using a straightforward validation assay that rapidly identifies and eliminates off-targeting probes. Furthermore, oligo pulldown techniques capture their target RNA in crude cell lysates, where artifactual interactions with abundant, promiscuous RNA-binding proteins can

confound analysis (Mili & Steitz, 2004). Overcoming this issue typically eliminates all but the target RNA's most direct interactors (Simon & Machyna, 2019), thus eliminating higher-order biological context. Moreover, the inefficiency of target capture, typically necessitates large inputs ($>10^8$ cells per replicate) (Chu & Chang, 2018; McHugh & Guttman, 2018; Simon & Machyna, 2019). In contrast, because O-MAP enzymatically amplifies the number of biotins per oligo probe, and enables interacting molecules to be directly captured (rather than co-purified by pulldown), it is markedly more efficient. Of note, our multi-omic analysis of HeLa nucleoli collectively required fewer cells than would a single replicate of CHIRP (Chu & Chang, 2018). Moreover, unlike those methods, proximity-biotinylation approaches like O-MAP are uniquely suited to probe dynamic, higher-order interactions that are too transient or fragile to survive a pulldown.

Compared to RNA-targeted live-cell proximity-labeling methods, O-MAP is more spatially precise, more applicable across transcripts and specimens, and is simpler to design and execute. Live cell approaches (e.g., RapID, MS2/Cas13-APEX, CRUIS) assemble artificial complexes between the labeling enzyme and its RNA target, by transgenically overexpressing these components fused to localization sequences or binding motifs (S. Han et al., 2020; Y. Han et al., 2019; Ramanathan et al., 2018; Z. Zhang et al., 2020). Although such transgenic approaches have had some success, they must contend with substantial background labeling from unbound enzymes (S. Han et al., 2020; Y. Han et al., 2019) and with artifacts from overexpressing the target RNA outside of its native context (Bassett et al., 2014). These issues are particularly problematic with low-abundance transcripts. In contrast, O-MAP doesn't require transgenic manipulation to control enzyme localization. The sequences, abundances, and hybridization parameters of its HRP-targeting oligos can be explicitly controlled to minimize off-targeting (Kishi et al., 2019), and residual, mislocalized probes can be removed by stringent washing. This strategy enabled

endogenous RNAs—even low abundance transcripts—to be biotinylated with nearly undetectable background (Fig 1.2).

O-MAP is conceptually similar to other methods that use affinity reagents, rather than transgenic expression, to localize proximity-labeling enzymes. Several established 'omics tools (e.g., EMARS, SPPLAT, BAR, TSA-Seq, and TSA-MS) (Bar et al., 2018; Y. Chen et al., 2018; Dopie et al., 2020; S. Jiang et al., 2012; Rees et al., 2015) use antibodies to deploy HRP to a target protein in situ, essentially adapting an immunofluorescence workflow for novel interaction discovery. A similar tactic underlies RNA-TRAP (Carter et al., 2002) and HyPro (Yap et al., 2022), which use Digoxigenin (DIG)-modified oligos to recruit HRP to a target RNA—either via an anti-DIG antibody-conjugate (Carter et al., 2002) (Fig. 1.4) or linked to a custom-made DIG-binding protein (Yap et al., 2022). In our experience, these hapten-recruitment strategies are markedly noisier than O-MAP's "landing-pad" approach, and their lack of modularity precludes strategies like our Probe Validation Assay (Fig. 1.1, b), a vital optimization tool. Moreover, unlike the current implementation of HyPro, O-MAP enables ChIP-like genomic interaction discovery (Fig 1.5 and 1.6). O-MAP is also considerably cheaper, using inexpensive, chemically unmodified primary oligos, in lieu of DIG-conjugated probes. And critically, O-MAP uses entirely off-the-shelf parts, and doesn't require the synthesis of custom proteins.

Our work also illustrates O-MAP's unique ability to discover compartment-level interactions that are opaque to current methods. O-MAP-ChIP for example enabled a parallelized analysis of NAD structure across cell lines, revealing both conserved and variable elements of NAD architecture (Fig 1.5 and 1.6). Extending this analysis to other lines, and complementing it with O-MAP-MS and O-MAP-Seq, will enable an unprecedented molecular characterization of nucleolar architectural remodeling during oncogenesis (Derenzini et al., 1998).

We anticipate that O-MAP will be applicable to nearly any target RNA, and in any biological setting, for which RNA-FISH can be performed. However, we note some of the technique's current limitations. O-MAP will likely be challenging with difficult FISH targets (e.g., small, low-abundance RNAs), requiring more advanced probe designs to amplify signal. For example, the number of HRP moieties recruited can be increased by extending the number of landing-pad concatemers of the primary oligonucleotide pool (Kishi et al., 2019). Moreover, the choice of permeabilizing reagent can drastically alter the biomolecules enriched (Schnell et al., 2012). For example, the use of Triton permits permeabilization of the nucleus but also results in a loss of cytoplasmic material. Hence, the use of alternative permeabilization strategies will be needed for the effective recovery of RNA-scaffolded structures existing in non-nuclear subcellular regions. Finally, like all peroxidase-based proximity-labeling strategies, O-MAP's labeling chemistry identifies all molecules within its labeling radius, including both the target's direct and nearby interactors. Future development of O-MAP's probe design and labeling chemistry will enable more precise control over its labeling radius, allowing "shells" of RNA-interactions to be probed independently. Regardless, in its current implementation we believe that O-MAP's ability to elucidate RNA interactions *in situ* overcomes a technical roadblock that has long challenged the RNA community. Given the extraordinarily broad scope of cellular functions performed by RNA (Cech & Steitz, 2014), this advancement will catalyze fundamental discoveries into countless biological phenomena.

## 1.5 METHODS

**Cultured cells, tissue sections and organoids**

HeLa (gift from Dr. Y. Sancak, UW), Patski cells and female mouse embryonic fibroblasts (fMEFs, both gifts from Dr. C. Disteche, UW), and SUIT2 cells (a gift from Dr. S. Kugel, Fred

Hutch) were cultured in High Glucose DMEM with Pyruvate (Thermo Fisher; 11995073), supplemented with 10% (v/v) Fetal Bovine Serum (FBS, Thermo Fisher; 26140079), 100 units/mL penicillin and 100 µg/mL streptomycin (Thermo Fisher; 15140122), and 1x GlutaMAX (Thermo Fisher; 35050061). For Transgenic (CxG)n and (G4C2)n U-2 OS cells (Jain & Vale, 2017) (gifts from A. Jain, MIT) qualified tetracycline-free FBS (Gibco; 26140079) was used. For these lines, transgenic RNA expression was induced by addition of doxycycline to a final concentration of 1 µg/mL for 24 hours prior to cell fixation. ASPC1, and Panc 3.27 cells (gifts from Dr. S. Kugel) were cultured in RPMI 1640 media (Thermo Fisher; 11875093), supplemented with 10% (v/v), FBS, 100 units/mL penicillin, and 100 µg/mL streptomycin. Patient-derived lymphoblast cell lines (Course et al., 2020) (a gift from Dr. P. Valdmanis, UW) were cultured in IMDM (Thermo Fisher; 12440053), supplemented with 20% (v/v) FBS, 100 units/mL penicillin, 100 µg/mL streptomycin, and 2.5 µg/mL amphotericin B. In all cases, cells were maintained at 37°C, under 5% CO2. Cell lines were authenticated by STR testing (ATCC), when possible.

For imaging experiments, cells were cultured in two-well Lab-Tek borosilicate glass #1.0 chambers (Thermo Fisher; 155380). For biochemistry, proteomic, and high-throughput sequencing experiments, cells were cultured in six-well plates. When necessary, material from multiple wells was harvested and merged into a single lysate, as described below.

**O-MAP probe design and synthesis**

Probes targeting the human 47S pre-rRNA ITS1 domain were taken from (Padovan-Merhar et al., 2015). All other probes were designed using OligoMiner pipeline34 using the following settings. The blockParse script was run using the settings: -l 30 -L 37 -t 42 -T 47 -s 390 -F 40. Bowtie2 was used with settings: -U --no-hd -t -k 100 --very-sensitive-local -S. The outputClean script was run with the -u argument; the structureCheck script was run with the settings: -F 40 -s 390 -m dna1998

-T 42. K-mer filtering was performed in Jellyfish version 2.2.10, using a Jellyfish file for the corresponding genome (human or mouse), and using the kmerFilter function with the -m 18 and -k 1 arguments. Jellyfish files were generated for each genome assembly (hg38 for human; mm10 for mouse), using a hash size set to the appropriate size of the genome assembly34. For example, the command -s 3300M -m 18 -o hg38_18.jf --out-counter-len 1 -L 2 hg38.fa was used to generate the 18mer dictionary for hg38. For most targets, all probes that passed this final filtering step—typically 10–150 probes per target—were used. For Kcnq1ot1, a set of 200 k-mer-filtered probes were used (Supplementary File 1).

For the O-MAP Probe Validation Assay (Fig. 1d), probes were divided into sub-pools along the length of the target RNA, appended on their 3′ termini with SABER1 or SABER2 "Landing-pad" sequences (Kishi et al., 2019) (Supplementary File 1). Once probes were validated, the complete pool was reformulated, appended with SABER1 and used for O-MAP alone. Probe sets consisting of fewer than 20 probes were ordered as individual oligos (Sigma; 0.025–0.05 μg synthesis scale, standard desalting), and further purified from preparative polyacrylamide gels, as described previously (Shechner & Bartel, 2011). Purified oligos were resuspended in nuclease-free water, quantified by UV-vis spectroscopy, pooled to a final aggregate concentration of 20 μM and stored at –20°C. Probe pools requiring more than 20 oligos were purchased as oPools (IDT; 50 pmol/oligo scale, unmodified), and dissolved to approximately 100 μM in nuclease free-water. 20 μL were desalted using the Oligo Clean & Concentrator Kit (Zymo Research; D4060), following the manufacturer's instructions. Pools were quantified by UV-vis spectroscopy, diluted to 5 μm and stored at –20°C. Fluorophore-conjugated secondary probe used in RNA-FISH ("SABER2–AF647", Supplementary File 1) was purchased from IDT (100 nmol scale; HPLC purification), resuspended to 100 μM in nuclease-free water, and stored in a light-tight container at –20°C. The

HRP-conjugated secondary probe ("SABER1–HRP," Supplementary File 1) was purchased from Bio-Synthesis, resuspended to 10 μM in resuspension buffer (10mM NaH2PO4, 150 mM NaCl, pH 7.2), allotted into 10 μL single-use aliquots, flash-frozen and stored at – 20°C.

In some cases, the RNA-FISH signal was amplified by extending the FISH probe subpool with concatemers of additional "SABER2" Landing-pads (Supplementary File 1). These were enzymatically added via the Polymerase Exchange Reaction (PER), essentially as described (Kishi et al., 2019). Briefly, pooled probes (5 μM, aggregate) were incubated with 0.5 μM Template hairpin and 0.1 μM Clean.G DNA hairpin (IDT; Supplementary File 1), in 10 mM MgSO4, 0.6 mM each ATP, CTP, and TTP, 4 U Bst 2.0 DNA Polymerase (NEB; M0537), 1x PBS, in a final volume of 50 μL. Reactions were incubated at 37°C for two hours in a thermocycler, heat-inactivated at 80°C for 20 minutes, and chilled on ice. The resulting PER-extended oligos were then purified with OligoClean and Concentrator Kits, eluting into nuclease-free water, and their length was examined on denaturing 10% Polyacrylamide TBE-Urea gels, stained with SYBR-Gold (Thermo Fisher; S11494).

**O-MAP core protocol**

The following protocol was used for omics-scale O-MAP, using cells grown in six-well dishes ($3.5\times10^5$ cells/well; plated one day prior to harvest). For imaging-only experiments, cells were plated at $7\times10^4$ cells per well, in two-chamber LabTeks. In all cases, RNase-free reagents and manipulations were used throughout.

The core O-MAP workflow is split over two days. The first day comprises fixation, permeabilization, peroxidase inactivation, and primary probe hybridization; the second day comprises secondary probe hybridization, an optional endogenous biotin blocking step, and *in situ*

biotinylation. Thereafter, the protocol varies depending on the endpoint assay—imaging, RNA-Seq, or ChIP-Seq.

**O-MAP Day 1**

All manipulations were performed at room temperature, unless noted. Cells were washed briefly with 1x Ca- and Mg-free DPBS (Thermo Fisher; 14190250) and fixed with freshly prepared 2% (v/v) formaldehyde (Electron Microscopy Sciences; 15710) in 1x PBS (Sigma; 6506), for 10 minutes without agitation. The formaldehyde solution was aspirated and the crosslinking reaction terminated by two washes with 250mM glycine in 1x PBS, five minutes each, with gentle rocking (3 rpm on a platform rocker). Cells were briefly washed with DPBS, permeabilized with 0.5% (v/v) triton-X 100 in PBS (10 min; gentle rocking), and washed three times with DPBS. Next, to inactivate endogenous peroxidases, samples were treated with 0.5% (v/v) $H_2O_2$, in 1x PBS, for 10 minutes with gentle rocking, and washed twice with PBS. Samples were then equilibrated in Formamide Wash Buffer (10–40% (v/v) deionized formamide (Thermo Fisher; AM9342); 2x SSC (Thermo Fisher; AM9765); 0.1% (v/v) Tween-20), for five minutes with gentle rocking. The formamide concentration was matched to the primary probe hybridization mix, as determined by the binding parameters of the primary probe pool (Supplementary File 1). This buffer was then aspirated, and each sample was treated with 115 µL of Probe Mix (0.1 µM primary oligo probe pool, in 1x Hybridization Buffer: 10–40% deionized formamide; 2x SSC; 0.1% (v/v) Tween-20; 10% (w/v) dextran sulfate (SIGMA; D8906); in nuclease-free water) and this mix was gently spread over the sample by covering with a clean, 30 mm diameter #1.5 thickness glass cover slip (Bioptechs; 30-1313-03192). A 2x SSC-soaked kimwipe was placed between the wells to maintain humidity during hybridization. Plates were then sealed with Parafilm and incubated without agitation for 8 hours at 37°C or 42°C, depending on the probe set (Supplementary File 1).

**O-MAP Day 2**

Following primary hybridization, coverslips were removed and cells were washed three times with pre-warmed 30% Formamide Wash Buffer, 10 minutes per wash, at 37°C with gentle rocking. For imaging experiments, this was followed by a blocking step to mask endogenous biotinylated proteins, as described below (see "O-MAP Imaging"). In all cases, subsequent manipulations were carried out in the dark, to avoid photooxidation of the HRP conjugate. Each well was treated with 115 µL O-MAP Secondary Probe Mix (100 nM SABER1–HRP oligo (Supplementary File 1), in 30% Formamide Hybridization Buffer), and covered with a clean coverslip. Samples were incubated for 1 hour at 37°C, with gentle rocking. Coverslips were then removed, and samples were washed four times with PBST (0.1% (v/v) Tween-20 in 1x PBS), 15 minutes per wash, with gentle rocking. Buffer was aspirated, and in situ biotinylation induced by addition of 1 mL Labeling Solution (0.8 µM biotinyl-tyramide (Sigma; SML2135), 1 mM $H_2O_2$, 1x PBST), and incubation at room temperature. Labeling times (Supplementary File 1) varied between RNA targets and were determined empirically using the O-MAP Imaging assays described below. In all cases, biotinylation was halted by addition of sodium azide and sodium ascorbate (10 mM each, final) in 1x PBST, for three washes of five minutes each.

**O-MAP imaging**

For imaging experiments, the background signal from endogenous biotinylated proteins was blocked after the secondary probe hybridization step. Briefly, samples were washed three times in 1x PBST and incubated in pre-blocking solution (1% (w/v) nuclease-free BSA (VWR; 97061-420) in 1x PBST) at room temperature for 30 min with gentle rocking. Samples were then blocked with 1 mL of Neutravidin Blocking Solution (10 µg/mL neutravidin (Thermo Fisher; 31000), 1% (w/v) nuclease-free BSA, in 1x PBST) for 15 min with gentle rocking at room temperature, and washed

three times with PBST. To saturate free streptavidin binding sites, samples were next treated with 10 µg/mL D-biotin (Thermo Fisher; B20656) in 1x PBST, for 15 minutes with gentle rocking, followed by three washes with room temperature PBST. Thereafter, in situ biotinylation and quenching proceeded as described above, using 50 µL volumes for primary and secondary hybridization buffers. After biotinylation and quenching, samples were stained with 1 mL 1 µg/mL neutravidin-DyLight 550 conjugate (Thermo Fisher; 84606), in 1% BSA pre-blocking solution, for one hour at room temperature with gentle rocking, followed by three washes with 1x PBST. Samples were counterstained with DAPI (5 µg/mL, in 1x PBST) and imaged immediately, or were mounted in Vectashield (Vector Labs; H-1900-10) and stored at 4°C.

**O-MAP–Seq**

O-MAP-labeled cells (approximately $9\times10^6$ cells—one six-well dish—per replicate; three biological replicates per experimental condition) were harvested by scraping into PBSTq (1x PBST, supplemented with 10 µM sodium azide, 10 µM sodium ascorbate) and pelleted by centrifugation at 3,000 x g for 10 minutes. Buffer was aspirated and cells were flash frozen in liquid nitrogen and stored at –80°C until use. Pellets were thawed on ice and resuspended by gentle pipetting in 1000 µL ice cold RIPA Buffer (50 mM Tris-HCl pH 7.5, 150 mM NaCl, 0.1% (w/v) SDS, 0.5% (w/v) sodium Deoxycholate, 1% (v/v) Triton X-100, 5 mM EDTA, 0.5 mM DTT), supplemented with 1x EDTA-Free Halt Protease Inhibitor Cocktail, 0.1 U/µL RNase-OUT (Thermo Fisher; 10777019)), and 10mM sodium azide, and rocked end-over-end at 4°C for five minutes. Samples were then sheared using a Branson Digital Sonifier 250 outfitted with a double stepped microtip, at 10–12 Watts for 30 s intervals (0.7 s on; 1.3 s off), with 30 s resting steps between intervals, seven intervals total. Samples were held in ice-cold metal thermal blocks throughout sonication. Lysates were then clarified by centrifugation at 15,000 x g for 10 min at

4°C, moved to fresh tubes and diluted with 1 mL Native Lysis Buffer (NLB: 25 mM Tris-HCl pH 7.5, 150 mM KCl, 0.5% (v/v) NP-40, 5 mM EDTA, 0.5 mM DTT), supplemented with 1X Halt Protease Inhibitor Cocktail, 0.1 U/µL RNase-OUT and 10 mM sodium azide. For each sample, 5% was removed as "input," and to the remainder was added 100 µL of Pierce streptavidin magnetic bead slurry that had been equilibrated by two washes in 1:1 RIPA:NLB supplemented with 10mM sodium azide, 0.1 U/µL RNase-OUT, and 1X Halt Protease Inhibitor Cocktail. Samples were incubated for 2 hr at room temperature with end-over-end agitation. Beads were then washed with the following series of buffers (1 mL each, 5 min per wash at room-temperature with end-over-end agitation). All buffers were supplemented with 1x EDTA-Free Halt protease inhibitor cocktail, 0.1 U/µL RNase-OUT, and 0.5 mM DTT, unless otherwise noted: (1) RIPA, supplemented with 10 mM sodium azide; (2) RIPA alone (3) High Salt Buffer (1 M KCl, 50 mM Tris-HCl pH 7.5, 5 mM EDTA) (4) Urea Buffer (2 M Urea, 50mM Tris-HCl pH 7.5, 5 mM EDTA) (5) 1:1 RIPA:NLB, without protease inhibitors (6) NLB, without protease inhibitors, (7– 8) two washes in TE buffer (10 mM Tris-HCl pH 7.5, 1 mM EDTA), without protease inhibitors.

RNA was isolated from both input and O-MAP-enriched samples by proteolysis in 100 µL Elution Buffer (2% (v/v) N-lauryl sarcoside, 10 mM EDTA, 5 mM DTT, 200 µg proteinase K (Thermo Fisher; AM2548), in 1x PBS). Reactions were shaken at 700 rpm in a Mixer HC (USA Scientific) for 1 hour at 42°C, followed by 1 hour at 60°C. RNA was then extracted once with 1 volume of phenol pH 4.3, and twice thereafter with an equal volume of absolute chloroform. Reactions were supplemented with 15 µg Glycoblue (Thermo Fisher; AM9515) and NaCl to 300 mM, and ethanol precipitated at –20°C overnight. RNA was harvested by centrifugation at 15,000 x g for 30 minutes at 4°C, washed twice with 70% ethanol, and resuspended in 84.75 µL nuclease free water.

Contaminating DNA was removed by digestion with 5 U RQ1 RNase-free DNase I (Promega; M6101) in 100 µL of the manufacturer's supplied buffer (1x final concentration) at 37°C for 30 min, and this reaction was terminated by addition of EDTA to 15 mM, final. RNA was purified by phenol extraction and ethanol precipitation, as described above, and resuspended in 15 µL nuclease free water. Sample concentration was measured using a Nanodrop One (Thermo Fisher).

**O-MAP-Seq library preparation and sequencing**

Ribosomal RNA was first depleted by RNase-H digestion, using pools of antisense DNA oligonucleotides that targeted mature rRNAs, but not the pre-rRNA "transcribed spacer" domains as described previously (Adiconis et al., 2013) . The antisense oligos used for HeLa cells (Fig. 3a–g), described previously (Adiconis et al., 2013), were synthesized as an oPool (IDT; 50 pmol per oligo, unmodified), desalted using a Zymo OligoClean and Concentrator Kit, following the manufacturer's instructions, and resuspended into nuclease-free water. 1 µg of antisense probes were added to 1 µg of RNA (whole cell, or O-MAP-enriched), in 200 mM NaCl, 100 mM Tris-HCl, pH 7.4, at a final volume 10 µL. This solution was heated to 95°C for 2 minutes and then slowly cooled to 45°C at a rate of –0.1°C/s, using a ProFlex PCR system (Thermo Fisher). Reactions were then supplemented with 10 µL of preheated RNase H mix (10 U Hybridase Thermostable RNase H (Lucigen; H39500), 20 mM MgCl2). Reactions were incubated at 45°C for 30 minutes and placed on ice. RNA was then purified by acidic phenol-chloroform extraction and ethanol precipitation, residual DNA was removed using RQ1 DNase, and RNA was again purified by phenol-chloroform extraction and ethanol precipitation, as described above.

Each sample was quantified on a Nanodrop One (Thermo Fisher). Sequencing libraries were prepared from 300 ng RNA using the NEBNext Ultra II Directional RNA Library Prep Kit and NEBNext Multiplex Oligos for Illumina (NEB; E7760 and E7735), according to the

manufacturer's instructions. Three biological replicates were used for each experimental condition; each library was given a unique index. Libraries were quantified using the NEBNext Library Quant Kit for Illumina (NEB; E7360), following the manufacturer's instructions, and the quality of these libraries was confirmed using an Agilent 4200 TapeStation with an "DNA High Sensitivity" kit (Fred Huch Genomics Core). Libraries were pooled in equimolar concentrations to 20 nM aggregate concentration in nuclease-free water, with no more than 12 libraries per pool. These were subjected to 150 cycles of paired-end sequencing, followed by indexing, on one lane of an Illumina HiSeq 4000 per pool, run in high output mode (Azenta Life Sciences).

**O-MAP-Seq data analysis**

For gene- and isoform-specific expression analyses (Fig. 1.2, c-d), raw RNA-seq FASTA files were aligned to reference genomes using HISAT2 version 2.2.1, in the paired-end setting with default parameters (Kim et al., 2019). For 47S-O-MAP, reads were mapped to a modified GRCh38 genome assembly (courtesy of T. Moss, Université Laval) in which all rDNA repeats are replaced by a single copy of the consensus rDNA locus as an extra chromosome ("GRCh38_rDNA") (Mars et al., 2018). The resulting SAM files were converted to BAM format and sorted using Samtools (H. Li et al., 2009) version 1.15.1. Bigwig files for visualizing strand-specific information were created using deepTools (Ramírez et al., 2014) version 3.5.1 with parameters: --filterRNAstrand forward/reverse --binSize 1 --normalizeusingBPM. Mapped reads were quantified using StringTie (Pertea et al., 2015) version 2.2.1 and the StringTie output was prepared for differential expression analysis using the prepDE.py function. The resulting gene and transcript count matrices were used for differential expression analysis using DESeq2 (Love et al., 2014) with an FDR cutoff of 0.05. Transposable element (TE) expression analysis (Fig 1.2, g) was performed using the TEtranscripts pipelie (Jin et al., 2015). Briefly, raw RNA-seq fasta files were mapped to GRCh38 (for 47S),

using STAR (Dobin et al., 2013) version 2.7.10a, allowing multi-mapped reads with the following settings: --sjdbOverhang 149 -- winAnchorMultimapNmax 200 --outFilterMultimapNmax 100 --outSAMtype BAM Unsorted. TE expression was then quantified using TEtranscripts110 version 2.2.1, with the following parameters: --stranded reverse –mode multi --minread 1 --padj 0.05 -i 100. Differentially expressed TEs were identified using DESeq2 version 1.34.0 with a FDR cutoff of 0.05.

Volcano plots were generated using EnhancedVolcano version 1.12.0. All statistical analysis (Fisher's exact test, hypergeometric distribution test, or Student's t-test, where appropriate) was performed in R or in python using the ggplot2 (https:// https://ggplot2.tidyverse.org/), or seaborn and matplotlib modules.

**O-MAP-ChIP**

O-MAP-labeled cells (approximately $5\times10^6$ per replicate; two biological replicates per experimental condition) were harvested in PBSTq (1x PBST, supplemented with 10 μM sodium azide, 10 μM sodium ascorbate) by scraping, and pelleted by centrifugation at 3,000 x g for 10 minutes. Buffer was aspirated and cells were frozen in liquid nitrogen and stored at –80°C until use. Pellets were thawed on ice and gently resuspended by pipetting in 1 mL CLB (20 mM Tris pH 8.0, 85 mM KCl, 0.5% (v/v) NP-40), supplemented with 1x Halt EDTA-Free protease inhibitor cocktail and 10 mM sodium azide, for 10 minutes. Lysates were then clarified by centrifugation at 3,000 x g for five minutes at 4°C. Supernatant was aspirated, and samples were subjected to another round of CLB extraction, clarification, and supernatant aspiration. Pellets were then lysed by gentle pipetting in 1 mL of NLB (10 mM Tris-HCl pH 7.5, 1% (v/v) NP-40, 0.5% (w/v) sodium Deoxycholate, 0.1% (w/v) SDS) and incubated on ice for 10 minutes. Samples were then sheared using a Branson Digital Sonifier outfitted with a double stepped microtip, at 10–12 Watts over 30

s intervals (0.7 s on; 1.3 s off), with 30 s resting steps between intervals, 18 intervals total. This resulted in an average shearing size of approximately 200 bp, as gauged on an agarose gel. Samples were held in ice-cold metal thermal blocks throughout sonication. Lysates were then clarified by centrifugation at 15,000 x g for 10 minutes at 4°C and supernatants were moved to fresh tubes. For each sample, 10% was removed as 'input'; to the remainder was added 100 µL of streptavidin-coated magnetic bead slurry that had been equilibrated by two washes in NLB. Samples were incubated for 2 hours at room-temperature with end-over-end agitation. Beads were subsequently washed with the following series of buffers (1 mL each, 5 minutes per wash, at room-temperature, with gentle end-over-end agitation): (1) NLB, supplemented with 5 mM EDTA, 10 mM sodium azide and protease inhibitors (1x Halt EDTA-free Protease Inhibitor Cocktail), 150 mM NaCl; (2) NLB, supplemented with 5 mM EDTA, 10mM sodium azide and protease inhibitors, (3–4) two washes in 1 M KCl, 10 mM Tris-HCl pH 7.5, 5 mM EDTA, (5–6) two washes in 2 M Urea, 10 mM Tris-HCl pH 7.5, 5 mM EDTA, (7) 10 mM Tris-HCl pH 7.5, 1% (w/v) SDS, (8–9) 10 mM Tris-HCl pH 7.5, 1 mM EDTA.

DNA was isolated from both input and enriched samples by proteolysis in 100uL of Elution Buffer (2% (v/v) N-lauryl Sarcoside, 10 mM EDTA, 5 mM DTT, in 1x PBS, supplemented with 200 µg proteinase K). Samples were shaken for 1 hour at 700 rpm in a Mixer HC at 65°C. Supernatants were transferred to 0.2 mL tubes and incubated at 65°C overnight in a thermocycler. DNA was then extracted with an equal volume of phenol pH 6.6, followed by two extractions in equal volumes of absolute chloroform. Samples were supplemented with 1 µg GlycoBlue and NaCl to 300 mM final, and ethanol precipitated at –20°C overnight. DNA was harvested by centrifugation at 15,000 x g for 30 minutes at 4°C, washed twice with 70% ethanol, and resuspended into 20 uL nuclease free water. To remove residual RNA, each sample was supplemented with 10 µg RNase

Cocktail Enzyme Mix (Thermo Fisher; AM2286) and incubated at 37°C for 1 hour. DNA was then purified by phenol extraction and ethanol precipitation as described above and resuspended in 20 µL nuclease-free water.

**O-MAP-ChIP library preparation and sequencing**

DNA samples were quantified using a NanoDrop One. 300 ng DNA of each sample was used for library preparation, using the NEBNext Ultra II DNA Library Prep Kit and NEBNext Multiplex Oligos for Illumina (NEB; E7645 and E7335), according to the manufacturer's instructions. Two biological replicates were used per experimental condition; each library was given a unique index during synthesis. Library concentrations were measured using the NEBNext Library Quant Kit for Illumina, and the quality of each sample was confirmed using an Agilent 4200 TapeStation with a "DNA High Sensitivity" kit (Fred Hutch Genomics Core). Libraries were pooled in equimolar concentrations to 20 nM aggregate concentration in nuclease-free water, with no more than eight libraries per pool. These were subjected to 150 cycles of paired-end sequencing, followed by indexing, on one lane an Illumina HiSeq 4000 per pool, run in high output mode (Azenta Life Sciences).

**O-MAP-ChIP data analysis**

Deep sequencing reads were trimmed using TrimGalore! (https://www.bioinformatics.babraham.ac.uk/projects/trim_galore/) with parameters -q 30 --phred33, and mapped to the appropriate reference genome using Bowtie2 version 2.4.4 (Langmead & Salzberg, 2012). For 47S-O-MAP-ChIP, reads were mapped to GRCH38_rDNA (Mars et al., 2018). Duplicate reads were removed with the Picard MarkDuplicates function (http://broadinstitute.github.io/picard) before peak calling. O-MAP-ChIP data were normalized to replicate-matched input samples. For 47S-O-MAP, nucleolar associated domains were called by

merging peaks from Enriched Domain Detector (EDD) (Lund et al., 2014) and epic2 (Stovner & Sætrom, 2019). EDD was performed using default settings and— because NADs are enriched for highly repetitive sequences like centromeres—an empty BED file for the blacklist region. Epic2 peaks were called with the settings --bin-size 50000 -g 2. Peaks from EDD and epic2 were first concatenated and then merged with the BEDtools (Quinlan & Hall, 2010) merge function, using the default settings. Further statistical analysis was performed in R or python, as described for O-MAP-Seq, above.

Epigenomic analysis of NADs was performed using ChromHMM version 1.22 (Ernst & Kellis, 2012). The OverlapEnrichment function was called using the E117_25_imputed12marks_hg38lift_segments.bed file from the RoadMap Epigenomics Project and the final NAD calls in BED file format. Raw enrichment of each epigenomic signature for each cell line was plotted as a heatmap using seaborn version 0.10.1.

**RNA-FISH, O-MAP Probe-Validation assays, and immunofluorescence**

For RNA-FISH, cells were fixed in formaldehyde, permeabilized with Triton-X 100, equilibrated in formamide wash buffer, and hybridized to primary probes as described above (see: O-MAP Day 1), but omitting the peroxidase inactivation step. Thereafter, samples were washed three times in 30% Formamide Wash Buffer (five minutes per wash; room temperature) and incubated with 50 µL FISH Secondary Probe Mix (100 nM SABER2– AF647 oligo, in 30% Formamide Hybridization Buffer), and covered with a clean coverslip in a hybridization chamber. Samples were incubated for 1 hour at 37°C in the dark. Samples were washed three times with PBST (five minutes per wash), counterstained with DAPI solution (5 µg/mL, in 1x PBST) and either imaged immediately, or stored sealed in vectashield at 4°C.

For combined O-MAP and RNA-FISH experiments (Fig 1.1, c), including the Probe-Validation Assay (Fig 1.1, b), RNA-FISH signal was dramatically diminished when HRP- and Fluorophore-conjugated secondary oligos were hybridized simultaneously, presumably due to fluorophore damage during the *in situ* biotinylation reaction. To avoid this, O-MAP and RNA-FISH were performed sequentially. O-MAP was performed first, as described above (see: O-MAP Core Protocol), including the endogenous biotin blocking step (see: O-MAP Imaging). After quenching the O-MAP biotinylation reaction, samples were washed three times in 30% Formamide Wash Buffer (5 minutes per wash; room temperature), and incubated with 50 µL FISH Secondary Probe Mix (100 nM SABER2–AF647, in 30% Formamide Hybridization Buffer) for 1 hour at 37°C. Samples were washed four times with PBST (five minutes per wash), stained with neutravidin–fluorophore conjugate, counterstained with DAPI solution for two minutes, and either imaged immediately in 1x PBST or mounted in vectashield.

Fluorescence widefield microscopy was performed on a Leica DM IL, equipped with a HC Fluotar 100x oil immersion objective with a 1.32 numerical aperture and planar correction (Leica; 11506527), a white LED light source (Leica; EL6000) and a DFC365 FX digital camera (Leica; 11547004). The following filter cubes were used: Texas Red (Leica TX2 ET; 11504180; used with Dylight-550 conjugates), Cy5 (Leica Y5 ET; 11504181, used for Alexafluor-647), GFP (Leica GFP ET; 11504174, used for Alexa Fluor-488), and DAPI (Leica DAPI ET; 11504204). Illumination intensity was adjusted using the light source manual control; acquisition times ranged from 40–2000 ms, as controlled by the Leica LASX software. Fluorescence confocal microscopy was performed on a Leica SP8X microscope (UW Keck Imaging Center), outfitted with a HC CS2 63x oil immersion objective, with 1.40 numerical aperture with both planar and apochromatic correction. The average voxel size was 0.06 × 0.06 × 0.346 µm. Samples were illuminated using

a 470–670nm tunable White Light Laser system, with a typical laser power of 0.1% for DAPI, 3% for 550 nm, and 30% for 647 nm. Gain and offset settings were adjusted to avoid pixel saturation. Images were line-averaged twice, with an average pixel dwell time of 1.58 µs. A bit-depth of 8 or 16 was used and zoom factor between 1-3 was used for all images.

**Image processing**

Images were processed using Fiji (Schindelin et al., 2012) and ImageJ (Schneider et al., 2012), and multicolor overlays were made using the screen setting in Adobe Photoshop as previously described (Kishi et al., 2019). Most confocal images are maximum projections of z-stacks; the remainder correspond to single z-slices. Brightness and contrast were adjusted for display purposes using Fiji and ImageJ or Adobe Photoshop. In all cases, contrast adjustment was applied to improve signal visibility, by changing the minimum (black) and maximum (white) values only. Automated despeckling was applied when necessary (e.g. in RNA-FISH images with weak, diffuse speckling in between cells) to reduce residual background signal. Colocalization analysis (Fig. 1.2, f) was performed using Fiji (Gilles et al., 2017).

## 1.6 ACKNOWLEDGMENTS

We thank C. Hsu and C. Anderson for general technical assistance; A. Jain, M. Das, R. Akilesh, and P. Valdmanis for the kind donation of materials; T. Moss for rRNA computational tools; N. Peters, K. Collins, and A. Fenix for imaging assistance; S. Attar and E. Nichols, for help with cell-based models; A. Mishra and R.D. Hawkins for sequencing assistance. This work was supported by National Institutes of Health Grants 1R01GM138799-01 and T32GM007750, a Safeway Albertsons Early Career Award in Cancer Research, a Brotman Baty Institute Catalytic Collaborations Award, the UW Royalty Research Fund (RRF), and the UW Student Technology Fund (STF). E.E.K. was supported by the NSF DEB2016186 and the AHA 902616. Imaging at

the UW Keck Center was supported by NIH S10 OD016240 and the UW Student Technology Fee. NGS data analysis was facilitated through the use of advanced computational, storage, and networking infrastructure provided by the Hyak supercomputer system and funded by the STF at the University of Washington.

## 1.7 FIGURES

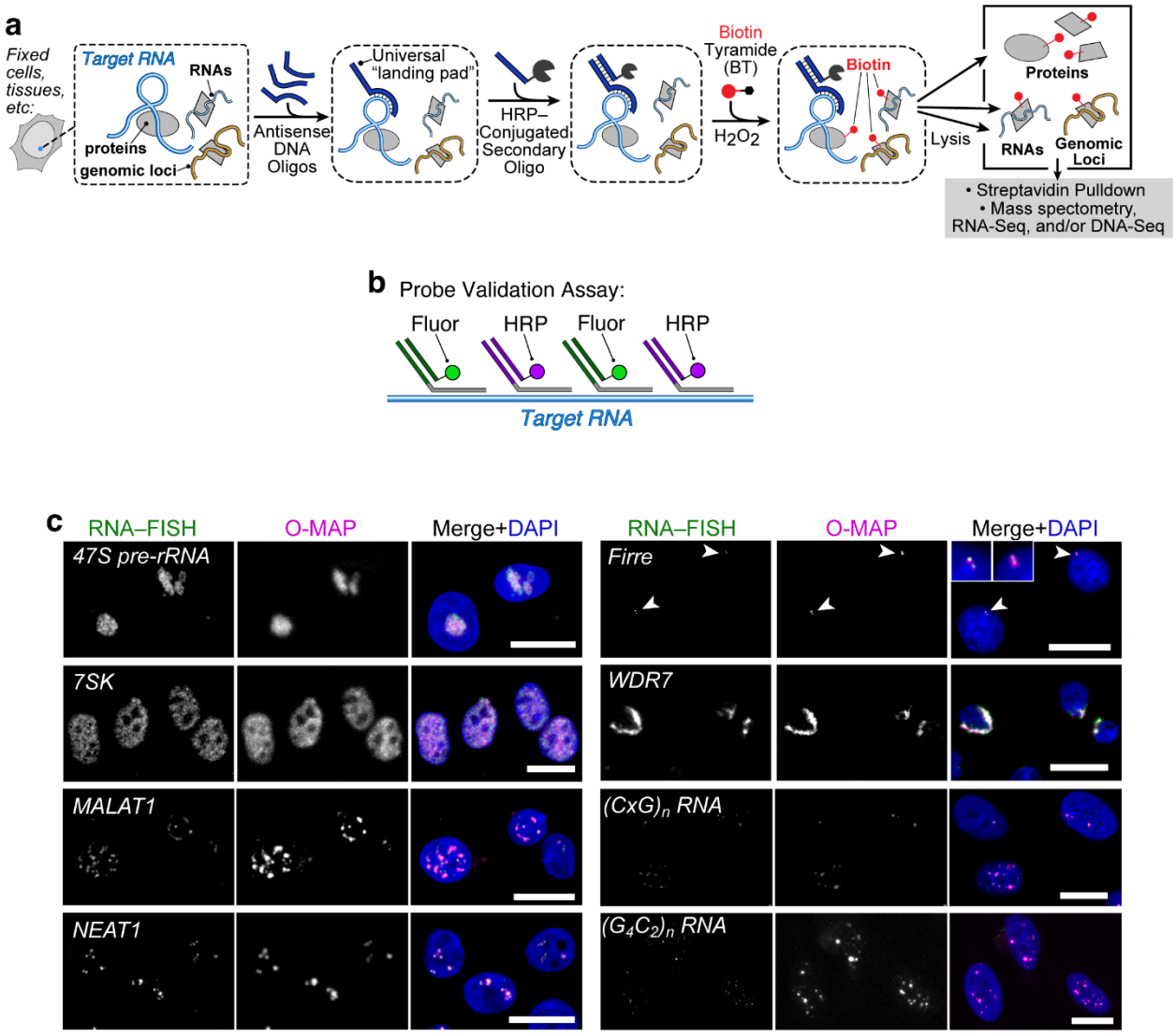



Figure 1.1. **O-MAP is readily portable to different RNA targets. a,** Overview of O-MAP. Specimens are chemically fixed, and pools of antisense DNA probes are hybridized to the target RNA. These probes recruit a common, HRP-conjugated secondary probe that catalyzes in situ proximity-biotinylation. **b,** O-MAP Probe Validation Assay. Primary probes are split into sub-pools that enable O-MAP and RNA-FISH to be performed simultaneously. Lack of co-localization suggests probe off-targeting. **c,** O-MAP Probe Validation Assay applied to a compendium of target transcripts. Note conspicuous overlap between RNA–FISH (green) and O-MAP (magenta) signals. Images from HeLa (47S, 7SK, MALAT1, NEAT1), Patski (Firre), patient-derived fibroblast (Course et al., 2020) (WDR7), and U2OS cells ((CxG)n and (G4C2)n RNAs). Firre is expressed from a single-copy transgene; (CxG)n and (G4C2)n RNAs are artificial constructs under doxycycline-inducible expression (Jain & Vale, 2017); all other targets are endogenous transcripts. Insets show zoomed-in sections of the same images, to highlight signal overlap.

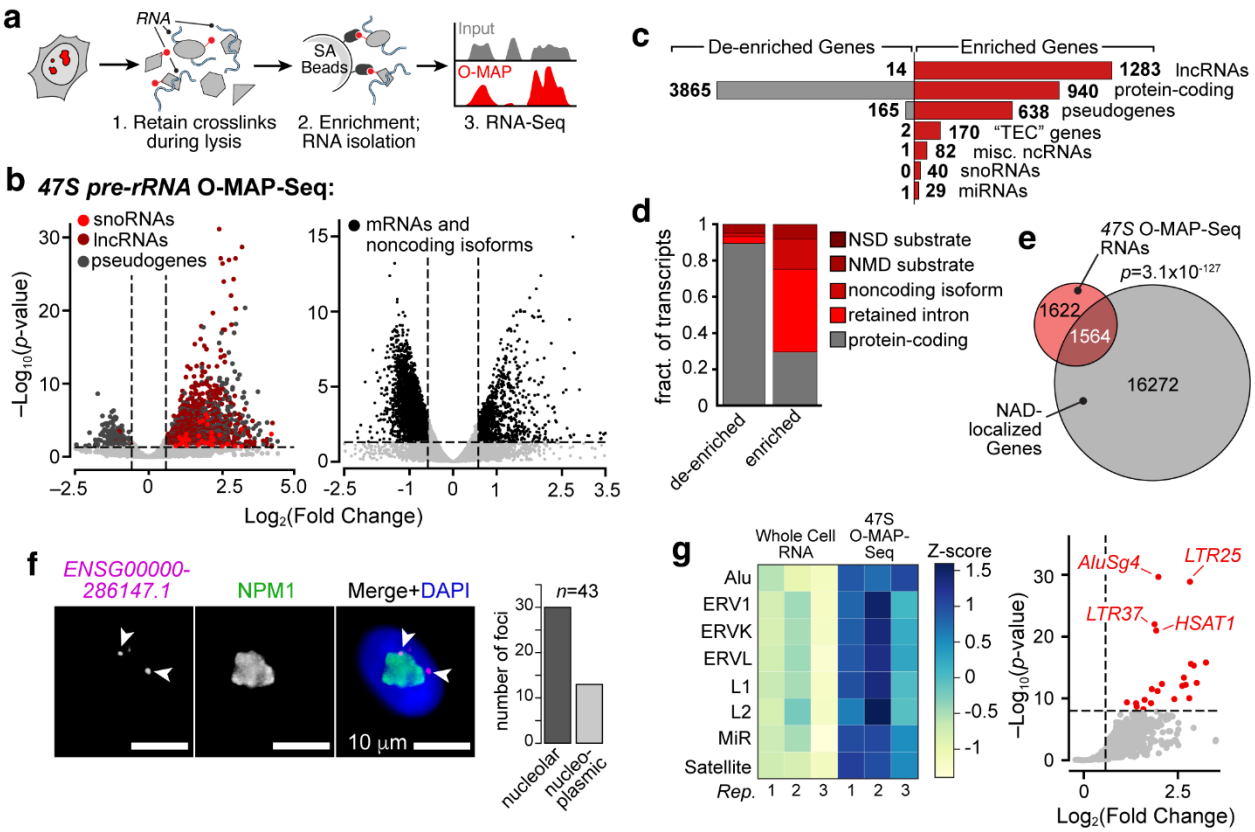



Figure 1.2. **O-MAP-Seq for probing RNA-proximal transcripts. a,** Approach. In situ biotinylated proteins are used to enrich nearby transcripts, which are then quantified by RNA-sequencing. **b–f**, O-MAP-Seq characterization of the HeLa nucleolar transcriptome, **b,** Volcano plots of 47S O-MAP-Seq data, demonstrating enrichment of snoRNAs, lncRNAs, pseudogenes (left), and de-enrichment of mRNAs (right) (n = 3 biological replicates). **c,** Summary of enriched and de-enriched RNA classes in the nucleolar transcriptome. **d,** Nucleolar transcripts expressed from protein-coding genes are predominantly noncoding variants; de-enriched transcripts are predominantly mRNAs. NSD: nonstop decay; NMD: nonsense-mediated decay. **e,** Nearly half of the nucleolar transcriptome is encoded from loci within nucleolar-associated chromatin loci (NADs; Fig. 4). Fisher's exact test. **f,** O-MAP-Seq identifies novel nucleolar-localized transcripts. Note co-localization between ENSG00000-286147.1 RNA-FISH and NPM1 Immunofluorescence (arrows; highlighted in zoom insets), quantified on right. **g,** Nucleolar transcripts are enriched in Transposable Element (TE) domains. Left: Z-scores of variance stabilizing transformed (VST) data, corresponding to major TE families. Right: volcano plot of individual TE classes in 47S O-MAP-Seq data.

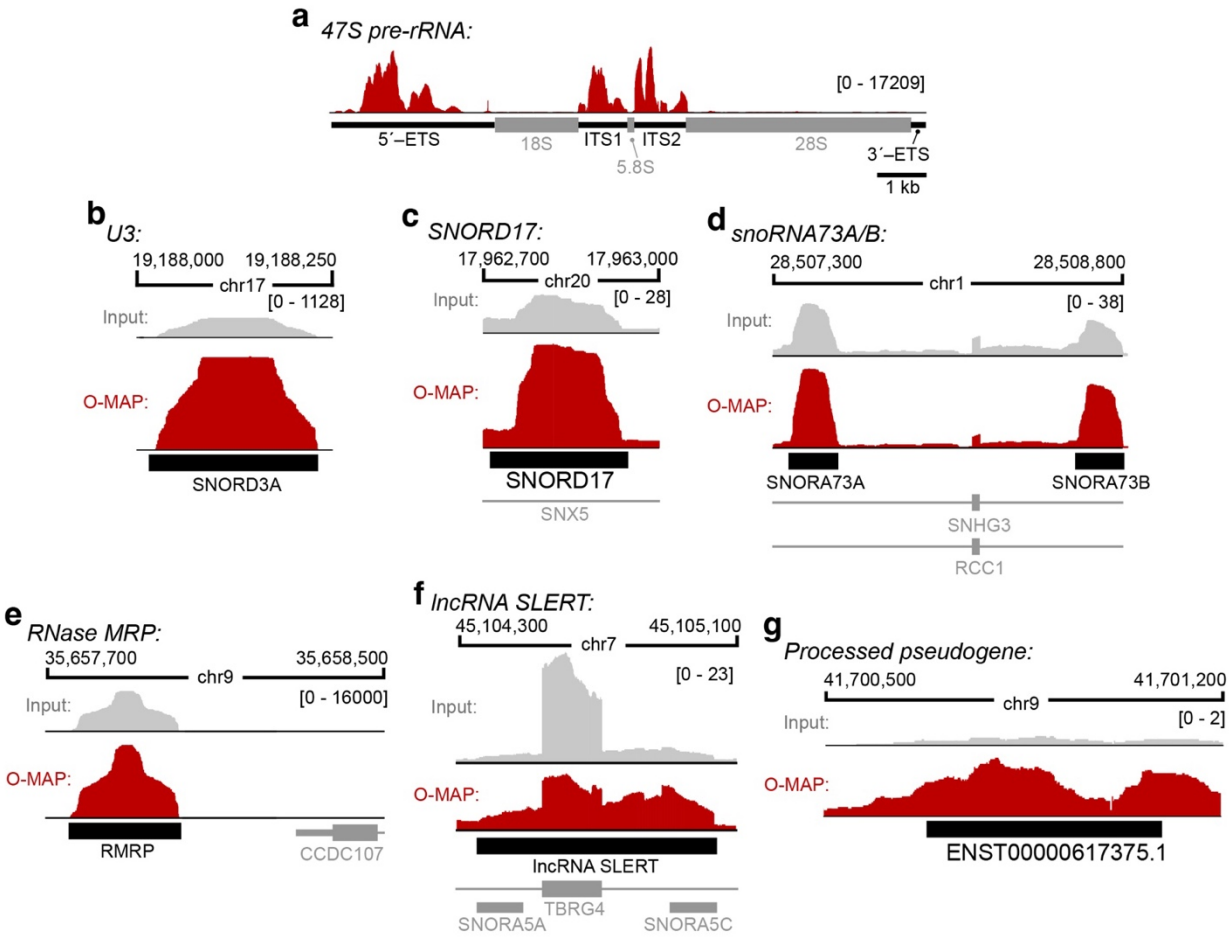



Figure 1.3. **47S O-MAP-Seq enriches known and novel nucleolar transcripts**. **a,** the 47S pre-rRNA. Note prominent enrichment for the 5′–ETS, IT1, and ITS2 "transcribed spacer" domains. Sequences corresponding to the mature 18S, 5.8S, and 28S rRNAs are removed during sequencing library preparation. Reads are aligned to a custom genome assembly containing a single copy of the rDNA consensus sequence (courtesy of T. Moss, U. Laval) annotated as a unique chromosome. **b,** the U3 noncoding RNA, which directs key cleavage events during ribosomal biogenesis, **c–d,** exemplar Box C/D (c) and H/ACA (d) small nucleolar RNAs (snoRNAs). SnoRNAs are often expressed within the introns of protein-coding genes (gray). **e,** RNase MRP (enriched) upstream of the CCDC107 gene (not enriched). **f,** lncRNA SLERT, which is processed from the two H/ACA snoRNAs embedded in the TBRG4 gene. **g,** Example of a novel nucleolar transcript—a processed pseudogene—discovered by O-MAP-Seq.

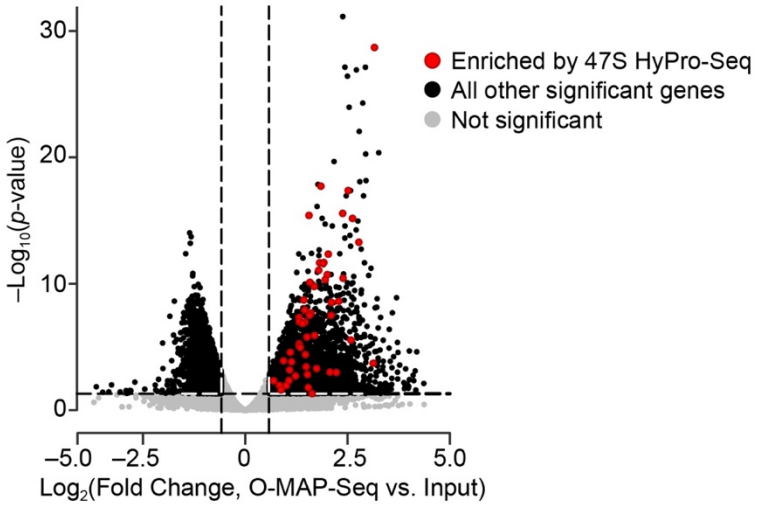



Figure 1.4. **47S O-MAP-Seq and HyPro-Seq enrich common transcripts.** Volcano plot of 47S O-MAP-Seq data, merging both coding and noncoding genes. Transcripts that were reportedly enriched by 47S HyPro-Seq (PMID: 34741808) are indicated (red). Note that differences in transcript annotation complicated systematic comparison to HyPro-Seq data. In all, only 53 of the significantly enriched transcripts reported by HyPro-Seq were detected in our data, all of which were also enriched (L2FC > 0.58, $p_{adj} < 0.05$, p-values calculated by DESeq2 (Love et al., 2014)) by 47S O-MAP-Seq.

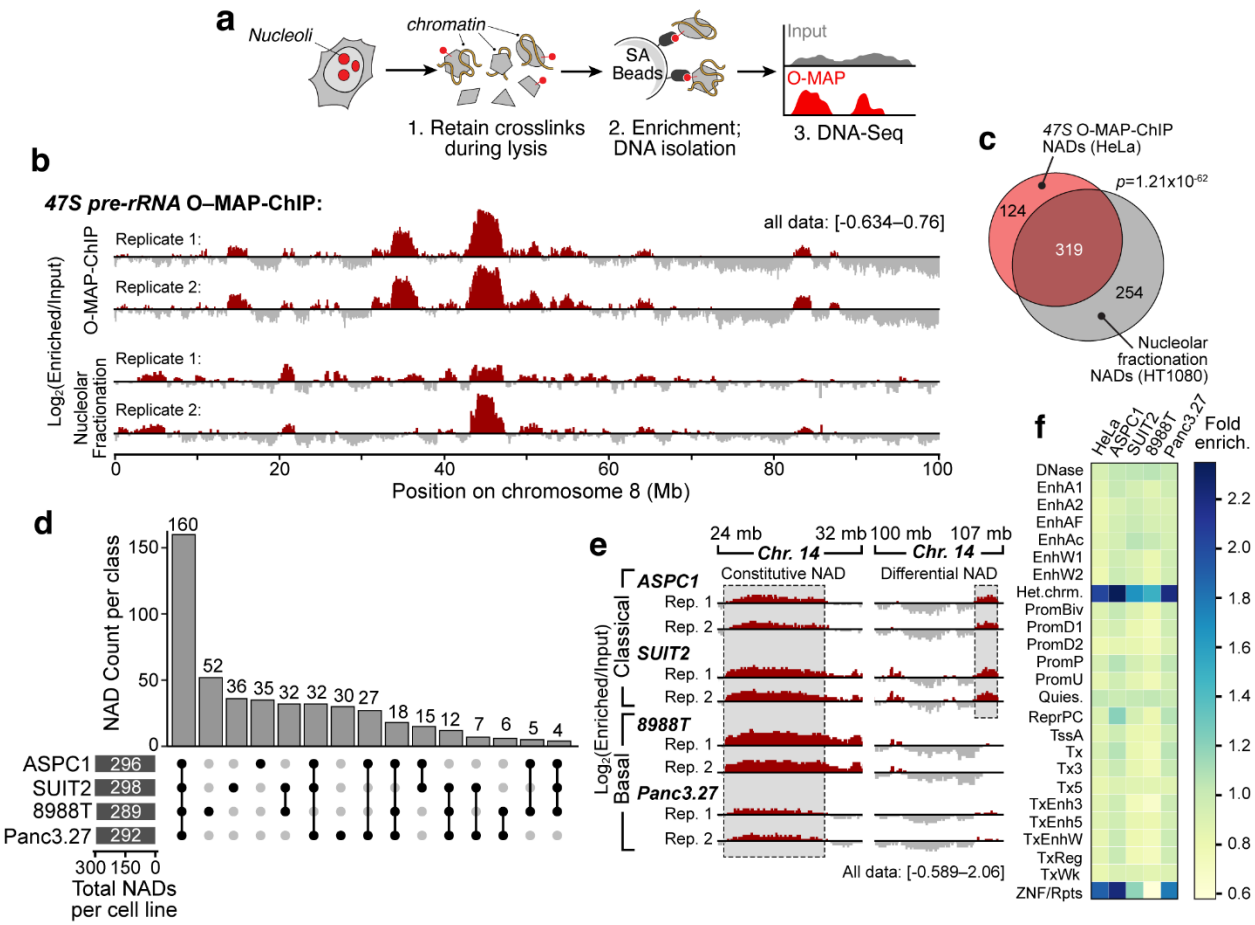



Figure 1.5. **O-MAP-ChIP for probing RNA-proximal genomic loci. a**, Approach. In situ biotinylated proteins are used to enrich nearby chromatin loci, which are then quantified by DNA-sequencing. **b-f**, O-MAP-ChIP characterization of Nucleolar Associated Domains (NADs). **b**, 47S-targeted O-MAP in HeLa cells recapitulates the known human NAD architecture. Most of chromosome 8 is shown. Note higher reproducibility of O-MAP data. Nucleolar fractionation data taken from (van Koningsbruggen et al., 2010). **c**, Conservation of NAD architecture between HeLa and HT1080 cells. **d**, Parallelized analysis of NAD architecture across four Pancreatic Ductal Adenocarcinoma (PDA) cell lines. Upset Plot summarizing NAD conservation, or lack thereof, between lines. The total number of NADs in each line appears nearly invariant. **e**, O-MAP-ChIP identifies NADs that are differentially regulated between Classical and Basal PDA subtypes. Examples of constitutive (left) and Differential (right) NADs on Chromosome 14 are shown. **f,** ChromHMM (Ernst & Kellis, 2012) analysis reveals differential enrichment of chromatin signatures among HeLa and PDA cell line NADs.

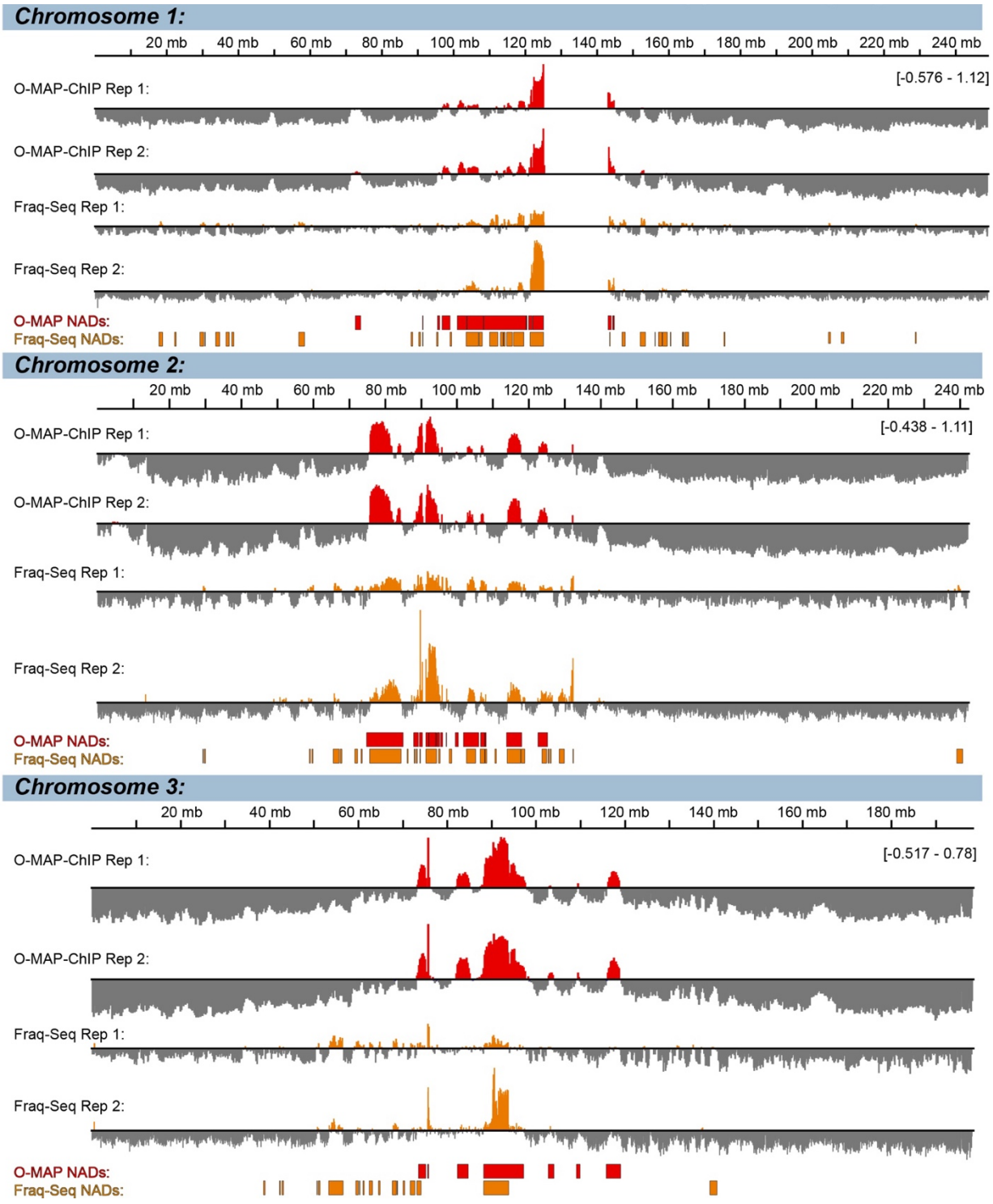
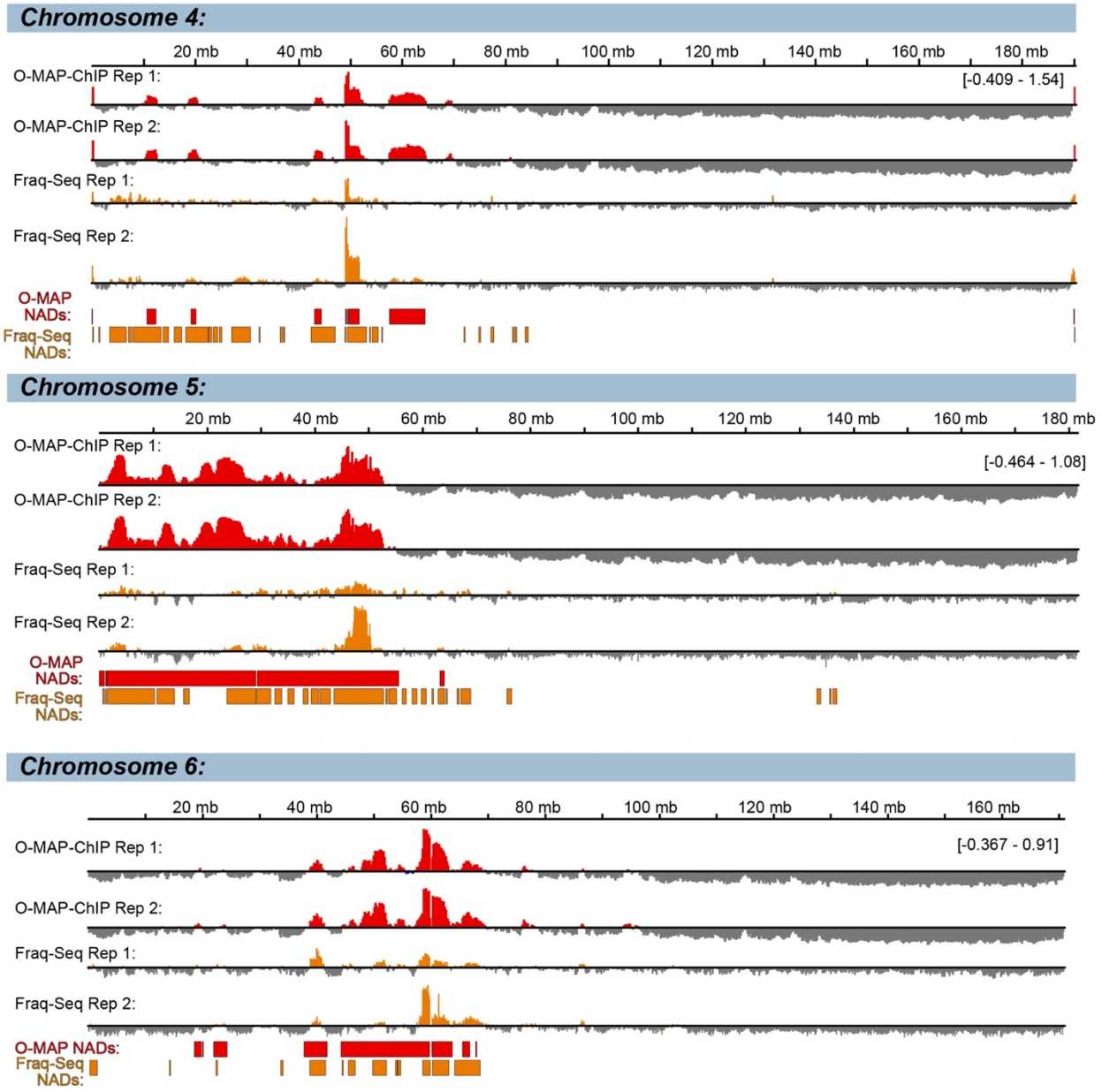
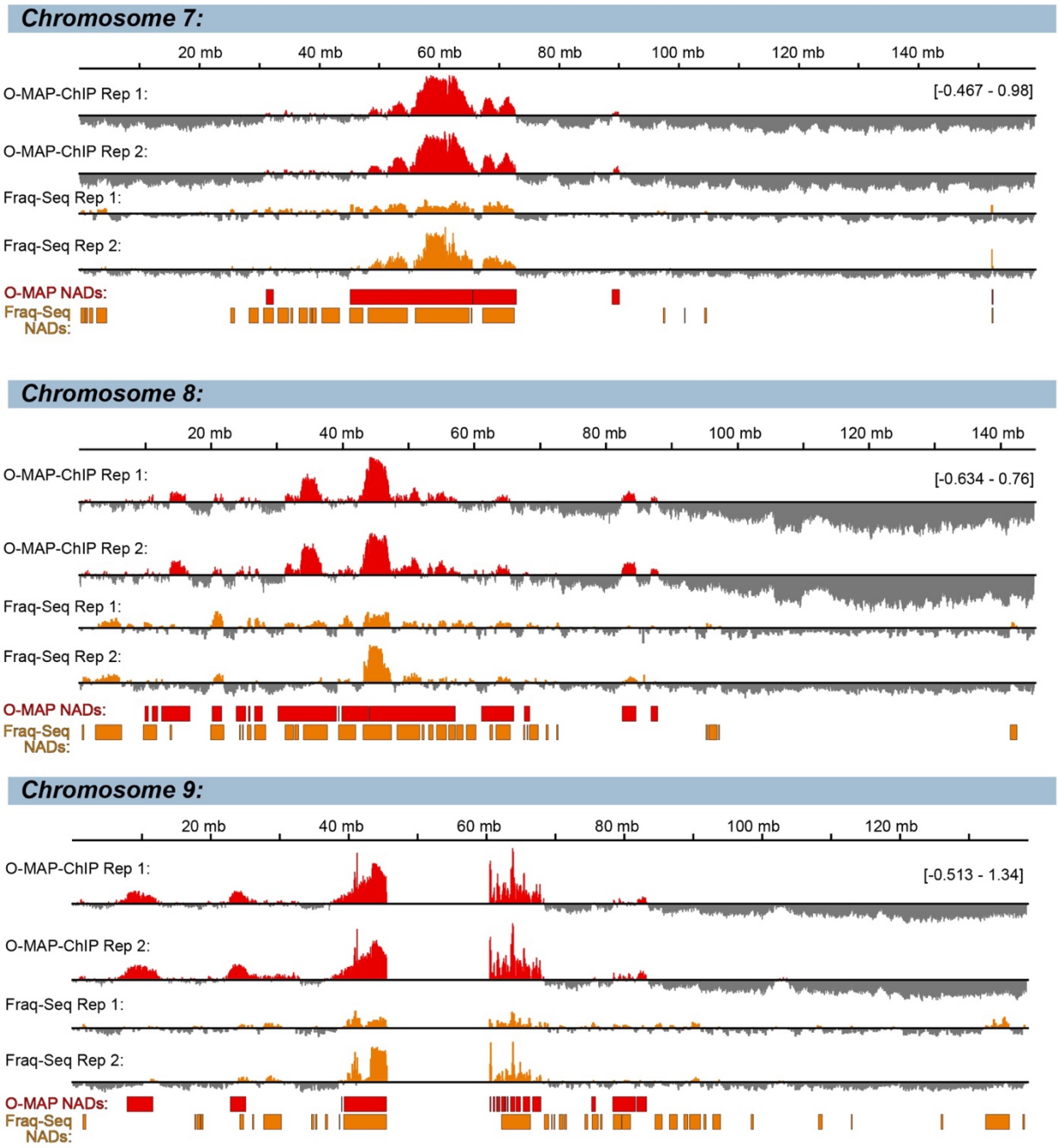
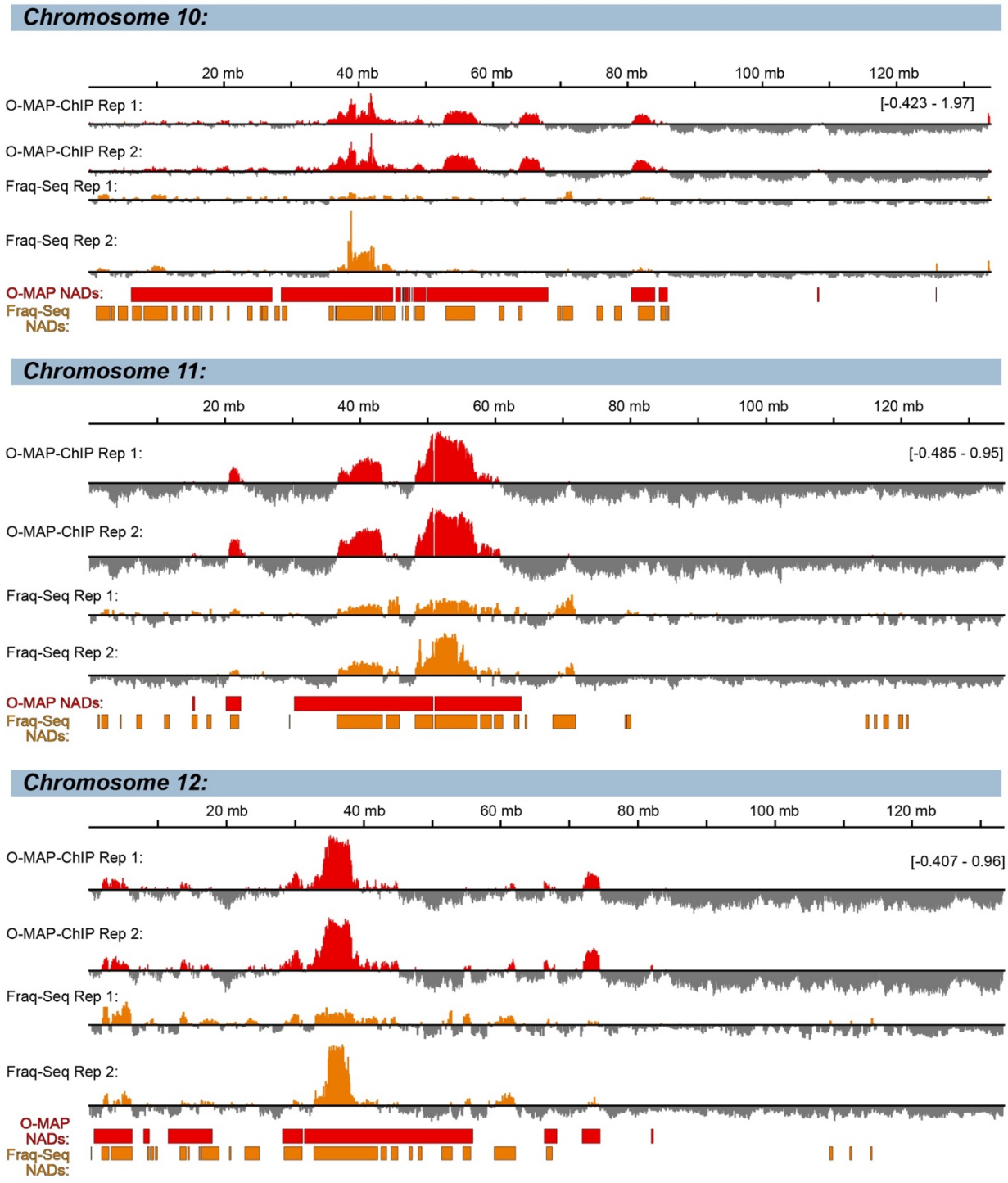


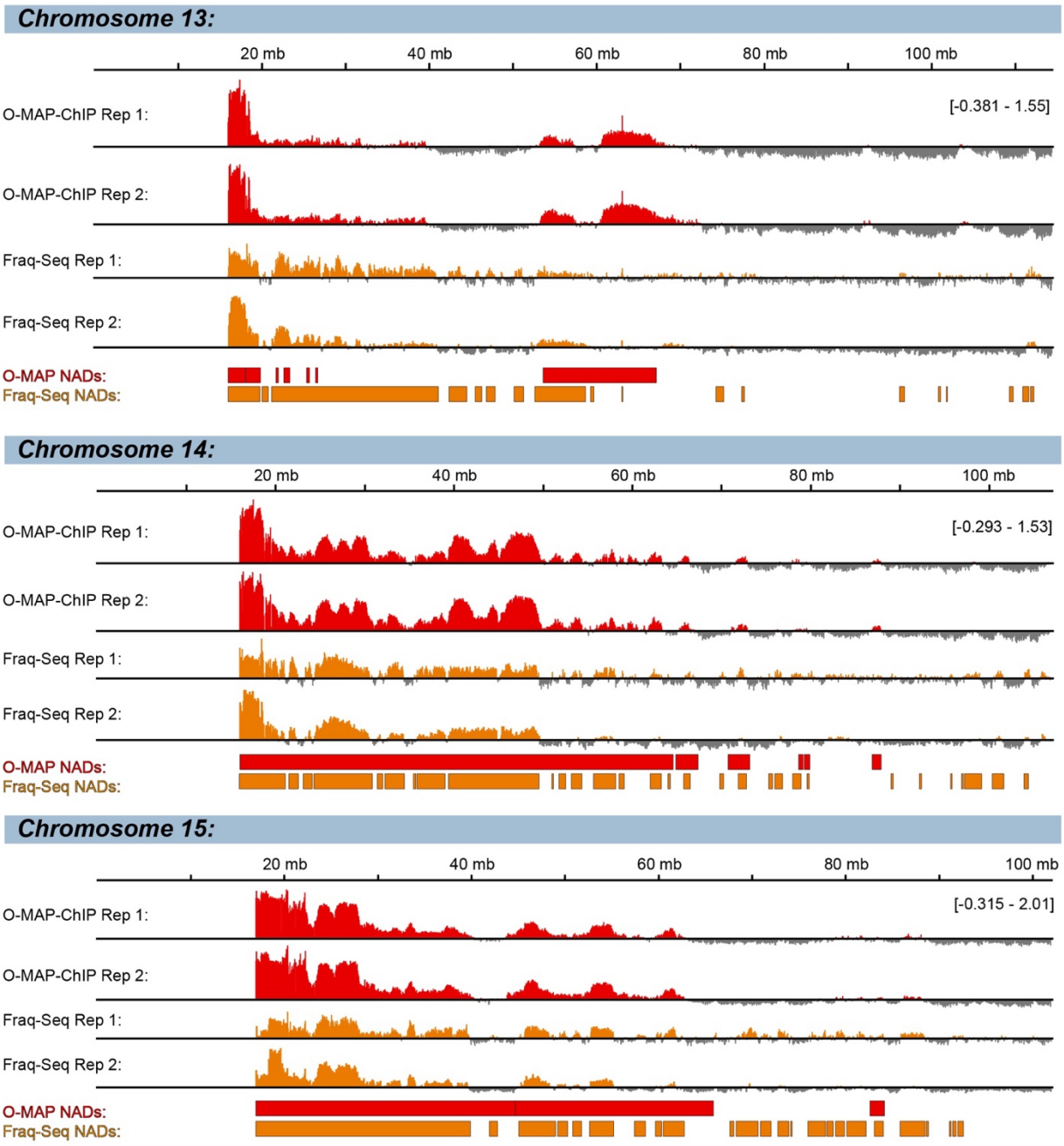

Figure 1.6. **Genomic maps of HeLa Nucleolar-Associated Domains (NADs).** Data are Log2(O-MAP-ChIP/Input). NADs were called by merging peaks from EDD and EPIC2 (red blocks). The same pipeline was also applied to HT1080 NADs, probed by fractionation-sequencing (Fraq-Seq; PMID 20826608). Although all replicates are shown, high noise/low information in Fraq-Seq

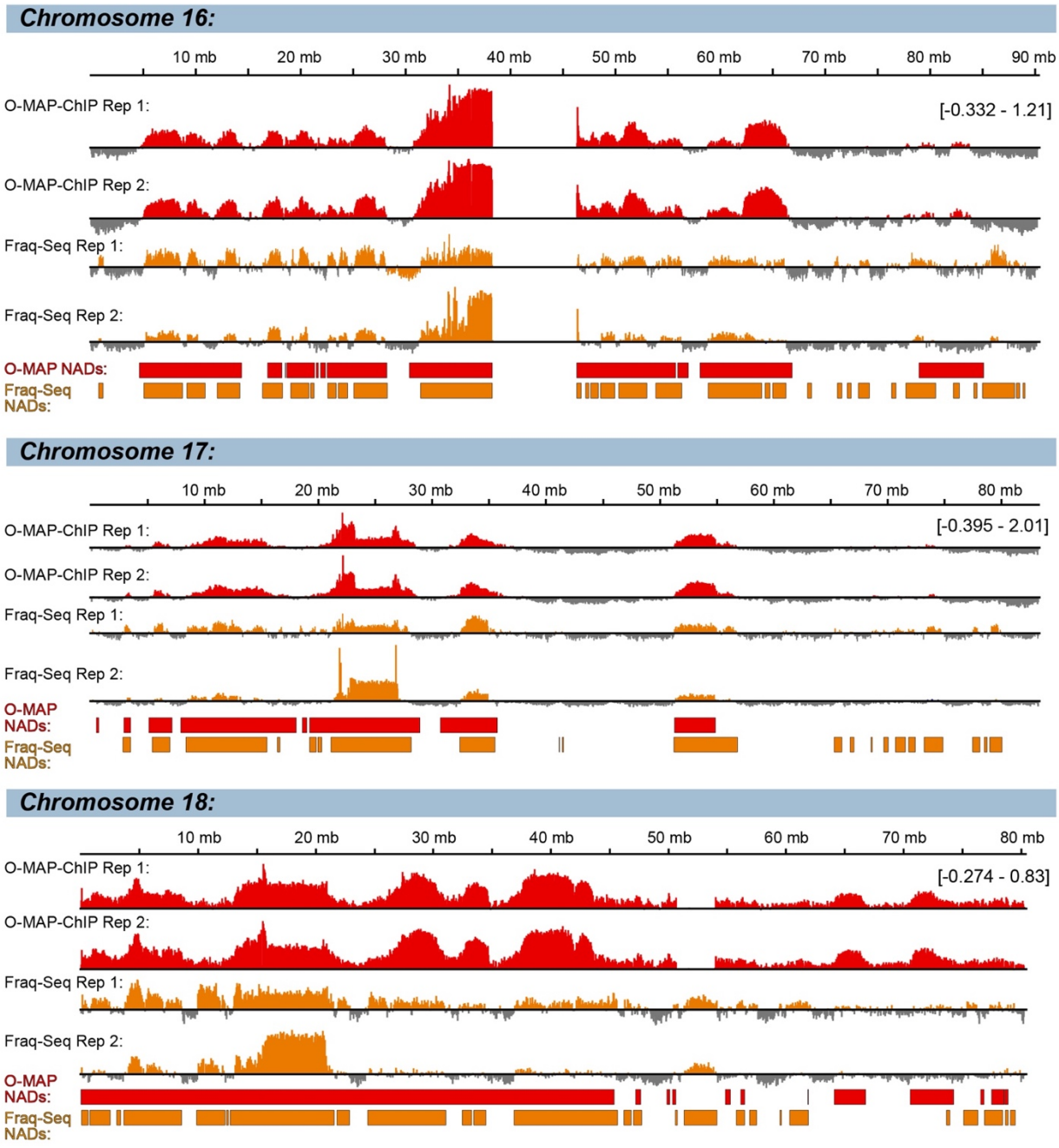
replicate 2 resulted in EDD failure after 10,000 Monte Carlo simulations; only replicate 1 was used for NAD calls (orange blocks). Figure continued on the next seven pages.

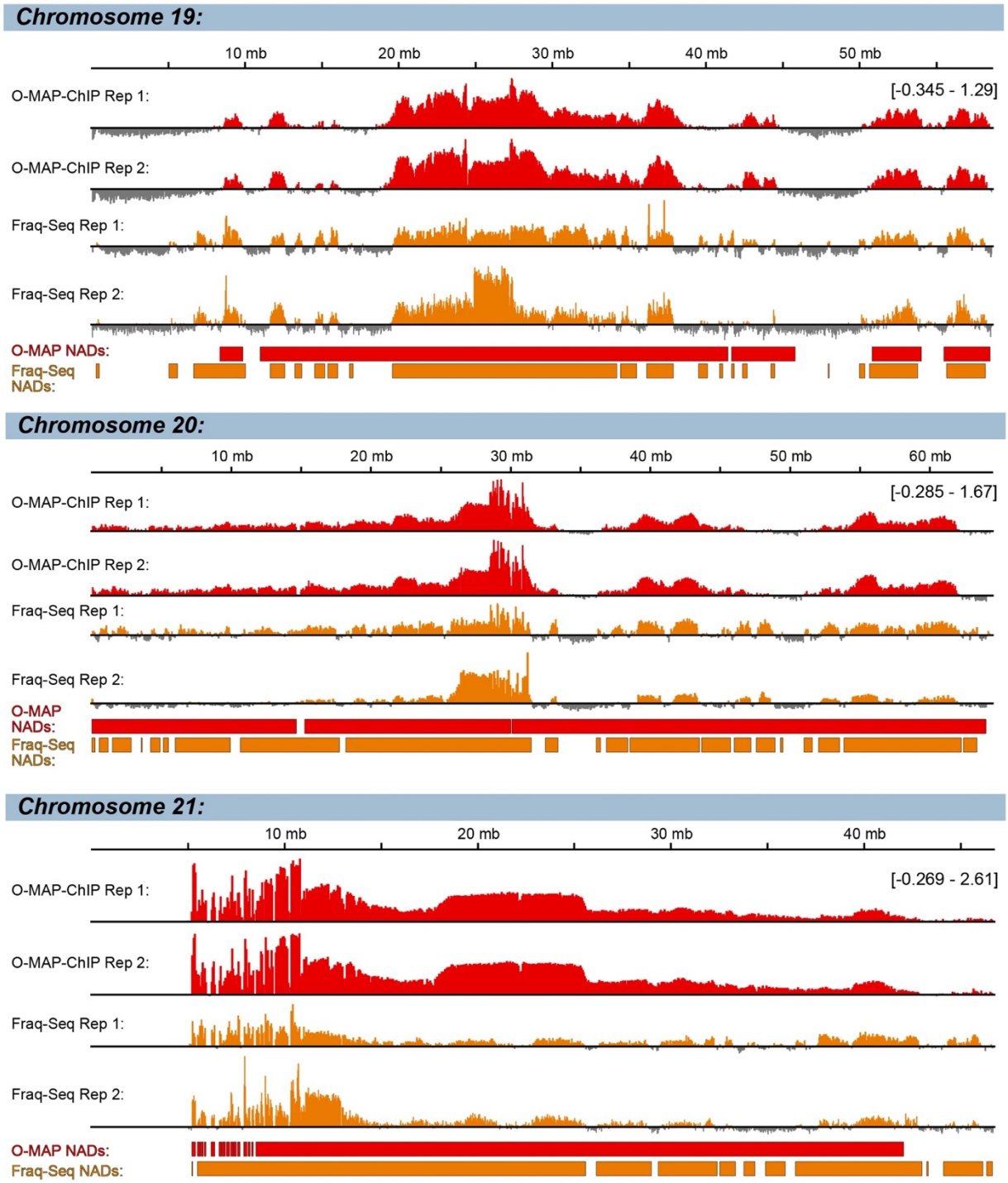



Chromosome 7:
20 mb
40 mb
60 mb
80 mb
100 mb
120 mb
140 mb
O-MAP-ChIP Rep 1:
[-0.467 - 0.98]
O-MAP-ChIP Rep 2:
Fraq-Seq Rep 1:
Fraq-Seq Rep 2:
O-MAP NADs:
Fraq-Seq NADs:
Chromosome 8:
20 mb
40 mb
60 mb
80 mb
100 mb
120 mb
140 mb
O-MAP-ChIP Rep 1:
[-0.634 - 0.76]
O-MAP-ChIP Rep 2:
Fraq-Seq Rep 1:
Fraq-Seq Rep 2:
O-MAP NADs:
Fraq-Seq NADs:
Chromosome 9:
20 mb
40 mb
60 mb
80 mb
100 mb
120 mb
O-MAP-ChIP Rep 1:
[-0.513 - 1.34]
O-MAP-ChIP Rep 2:
Fraq-Seq Rep 1:
Fraq-Seq Rep 2:
O-MAP NADs:
Fraq-Seq NADs:


Chromosome 10:
20 mb
40 mb
60 mb
80 mb
100 mb
120 mb
O-MAP-ChIP Rep 1:
[-0.423 - 1.97]
O-MAP-ChIP Rep 2:
Fraq-Seq Rep 1:
Fraq-Seq Rep 2:
O-MAP NADs:
Fraq-Seq NADs:
Chromosome 11:
20 mb
40 mb
60 mb
80 mb
100 mb
120 mb
O-MAP-ChIP Rep 1:
[-0.485 - 0.95]
O-MAP-ChIP Rep 2:
Fraq-Seq Rep 1:
Fraq-Seq Rep 2:
O-MAP NADs:
Fraq-Seq NADs:
Chromosome 12:
20 mb
40 mb
60 mb
80 mb
100 mb
120 mb
O-MAP-ChIP Rep 1:
[-0.407 - 0.96]
O-MAP-ChIP Rep 2:
Fraq-Seq Rep 1:
Fraq-Seq Rep 2:
O-MAP NADs:
Fraq-Seq NADs:


Chromosome 13:
20 mb
40 mb
60 mb
80 mb
100 mb
O-MAP-ChIP Rep 1:
[-0.381 - 1.55]
O-MAP-ChIP Rep 2:
Fraq-Seq Rep 1:
Fraq-Seq Rep 2:
O-MAP NADs:
Fraq-Seq NADs:
Chromosome 14:
20 mb
40 mb
60 mb
80 mb
100 mb
O-MAP-ChIP Rep 1:
[-0.293 - 1.53]
O-MAP-ChIP Rep 2:
Fraq-Seq Rep 1:
Fraq-Seq Rep 2:
O-MAP NADs:
Fraq-Seq NADs:
Chromosome 15:
20 mb
40 mb
60 mb
80 mb
100 mb
O-MAP-ChIP Rep 1:
[-0.315 - 2.01]
O-MAP-ChIP Rep 2:
Fraq-Seq Rep 1:
Fraq-Seq Rep 2:
O-MAP NADs:
Fraq-Seq NADs:


Chromosome 16:
10 mb
20 mb
30 mb
40 mb
50 mb
60 mb
70 mb
80 mb
90 mb
O-MAP-ChIP Rep 1:
[-0.332 - 1.21]
O-MAP-ChIP Rep 2:
Fraq-Seq Rep 1:
Fraq-Seq Rep 2:
O-MAP NADs:
Fraq-Seq NADs:
Chromosome 17:
10 mb
20 mb
30 mb
40 mb
50 mb
60 mb
70 mb
80 mb
O-MAP-ChIP Rep 1:
[-0.395 - 2.01]
O-MAP-ChIP Rep 2:
Fraq-Seq Rep 1:
Fraq-Seq Rep 2:
O-MAP NADs:
Fraq-Seq NADs:
Chromosome 18:
10 mb
20 mb
30 mb
40 mb
50 mb
60 mb
70 mb
80 mb
O-MAP-ChIP Rep 1:
[-0.274 - 0.83]
O-MAP-ChIP Rep 2:
Fraq-Seq Rep 1:
Fraq-Seq Rep 2:
O-MAP NADs:
Fraq-Seq NADs:


Chromosome 19:
10 mb
20 mb
30 mb
40 mb
50 mb
O-MAP-ChIP Rep 1:
[-0.345 - 1.29]
O-MAP-ChIP Rep 2:
Fraq-Seq Rep 1:
Fraq-Seq Rep 2:
O-MAP NADs:
Fraq-Seq NADs:
Chromosome 20:
10 mb
20 mb
30 mb
40 mb
50 mb
60 mb
O-MAP-ChIP Rep 1:
[-0.285 - 1.67]
O-MAP-ChIP Rep 2:
Fraq-Seq Rep 1:
Fraq-Seq Rep 2:
O-MAP NADs:
Fraq-Seq NADs:
Chromosome 21:
10 mb
20 mb
30 mb
40 mb
O-MAP-ChIP Rep 1:
[-0.269 - 2.61]
O-MAP-ChIP Rep 2:
Fraq-Seq Rep 1:
Fraq-Seq Rep 2:
O-MAP NADs:
Fraq-Seq NADs:

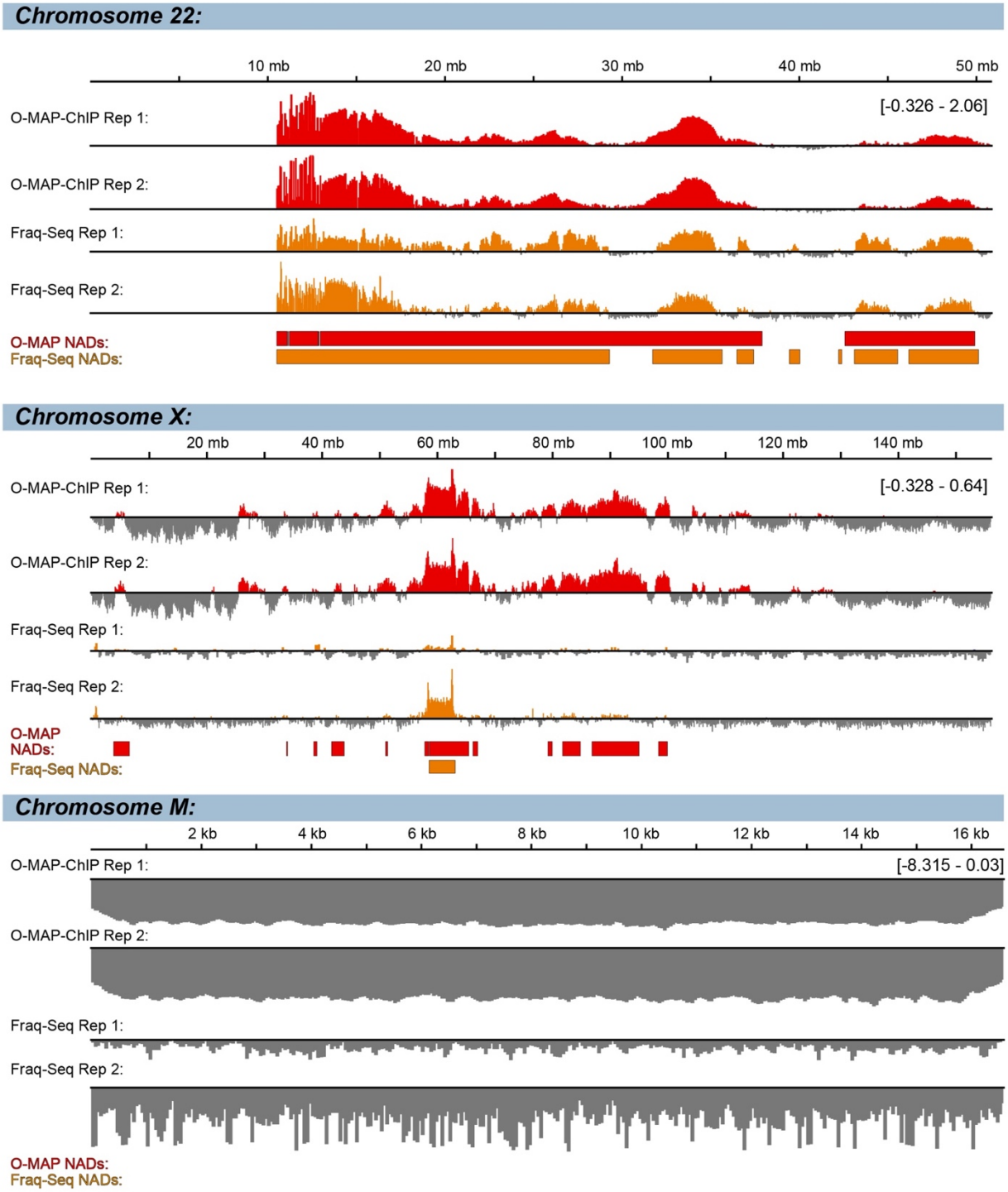



Figure 1.6, continued. The mitochondrial genome exhibits a complete absence of nucleolar interactions, as expected. Note that HeLa (O-MAP-ChIP) and HT1080 (Fraq-Seq) cells were derived from patients of different genders, which may explain the divergence in NAD architecture on the X-chromosome.

# Chapter 2. LNCRNA TARGETED PROXIMITY OMICS ENABLES NEW INSIGHTS ON THE INTERACTOME OF GENE-SLIENCING LNCRNAS.

## 2.1 ABSTRACT

Long non-coding RNAs (lncRNAs) are a diverse class of regulatory molecules that play critical roles in gene expression control. Among them, lncRNAs involved in transcriptional silencing have garnered significant interest due to their impact on genome integrity and epigenetic regulation. Here, we employed Oligonucleotide-directed proximity interactome MAPping (O-MAP) to comprehensively map the nearby biomolecules of two prominent gene silencers, X-inactive-specific transcript (Xist) and Kcnq1 Opposite Strand/Antisense Transcript 1 (Kcnq1ot1), along with the Xist nascent transcript. Here, O-MAP allowed for high-resolution identification of the chromatin loci, RNAs, and genomic loci at each lncRNA. Our results reveal that Xist, a key player in X-chromosome inactivation, specifically interacts with loci on the inactive X-chromosome (Xi) and exhibits distinct epigenetic signatures associated with polycomb repressive complexes. In contrast, Kcnq1ot1, which governs imprinted gene regulation, interacts with genomic loci on chromosome 7 and other autosomes, and its epigenetic signatures are indicative of heterochromatin formation. The nascent Xist transcript exhibited discrete regions of enrichment on the Xi before splicing into the mature isoform that spreads along the entire chromosome. Remarkably, we identify a shared pool of interacting transcripts among Xist, Kcnq1ot1, and nascent Xist, suggesting common mechanisms underlying their gene silencing functions. Notably, the lncRNAs Malat1 and Neat1 are among the shared interactors, with Malat1 showing a potential regulatory role in Kcnq1ot1-mediated gene silencing. Moreover, proteomic analysis reveals a considerable overlap in protein interactors among the three lncRNAs, with unique signatures suggesting distinct functions. Overall, our study provides unprecedented insight into the molecular

intricacies of gene silencing lncRNAs and their interactions with chromatin loci, RNAs, and proteins. O-MAP offers a powerful tool to comprehensively map lncRNA-scaffolded structures, shedding light on their functional roles in gene expression control and opening new avenues for understanding their broader impact in various diseases characterized by gene dysregulation.

## 2.2 INTRODUCTION

Long non-coding RNAs (lncRNAs) are an emerging class of regulatory molecules (Wang & Chang, 2011). lncRNAs are longer than 200 nucleotides and share a similar biogenesis pathway to canonical protein-coding mRNAs, but differ in that lncRNAs are not translated into functional proteins (Fang et al., 2018; Ponting et al., 2009). Once considered transcriptional noise, the pervasive presence and diverse functions of lncRNAs in the complex landscape of the genome have captivated the attention of researchers worldwide. It is estimated there are 16,000 to 100,000 lncRNAs expressed in the human genome, however most lack evidence for any function (Fang et al., 2018; Uszczynska-Ratajczak et al., 2018). Despite this, studies are increasingly showing that a handful of lncRNAs act as functional molecules. These lncRNAs function across areas that include regulation of transcription, post-transcriptional regulation, structural/scaffolding functions, and genome integrity (Statello et al., 2021; Wang & Chang, 2011). The majority of these functions primarily relate to the nucleus, as most lncRNAs are retained in the nucleus due to sequence motifs that interact with nuclear proteins and Watson-Crick chromatin association via R-loop formation (Derrien et al., 2012; C.-J. Guo, Ma, et al., 2020; C.-J. Guo, Xu, et al., 2020).

Among the multifaceted roles played by nuclear localized lncRNAs, their ability to transcriptionally silence DNA has garnered significant interest. This gene silencing phenomenon can occur through multiple mechanisms. Some *cis*-acting lncRNAs, such as Anrassf1, recruit chromatin modifying proteins, such as polycomb repressive complexes (PRCs), to silence nearby

neighboring genes (Beckedorff et al., 2013). Other nuclear lncRNAs, such as Airn and Chaserr, can also silence genes by blocking the recruitment of RNA polymerase II or causing RNA polymerase II to disassociate from the genomic region (Latos et al., 2012; Rom et al., 2019; Sleutels et al., 2002). Considering all nuclear lncRNAs, the X-inactive-specific transcript (Xist) and the Kcnq1 overlapping transcript 1 (Kcnq1ot1) represent two prominent lncRNAs that have been extensively studied as exemplars of transcriptional silencers.

Xist, discovered nearly three decades ago, plays a pivotal role in X-chromosome inactivation (XCI), a process that equalizes gene expression in genetic females by silencing one of the two X chromosomes (Heard et al., 1997; Wutz, 2011). During XCI onset near the start of embryogenesis, the maternal or paternal X-chromosome begins to transcribe the lncRNA Xist (Avner & Heard, 2001). The mechanism governing the decision to silence the maternal or paternal allele is still under active investigation. Upon transcription, Xist lncRNA molecules rapidly coat their chromosome of origin in *cis*, recruiting chromatin modifying proteins such as Spen, hnRNPK, and polycomb complexes to silence most of the chromosome (Chu et al., 2015; Colognori et al., 2019; Jeon & Lee, 2011; McHugh et al., 2015; Pintacuda et al., 2017). During this process, the chromosome is condensed in three-dimensional space into a spheroid region commonly called the inactive X-chromosome (Xi) or the Barr body (Heard et al., 1997). Thereafter, the epigenetic state of the Xi and localization of Xist are maintained (Loda et al., 2022). Although much is known about the protein interactors that impart Xist's silencing functions, how Xist remains tethered to the Xi is an open question. Especially when considering that ectopic Xist expression on other chromosomes is sufficient to drive condensation and silencing of that chromosome (Chiang et al., 2018). Moreover, the exact mechanism of how Xist spreads across the Xi is unknown.

Kcnq1ot1 is a key player in imprinted gene regulation, governing the monoallelic expression of a cluster of imprinted genes within the Kcnq1 locus that span over 400 kb (Edwards & Ferguson-Smith, 2007; Pandey et al., 2008). Here, the paternal allele harboring the Kcnq1 locus always expresses Kcnq1ot1; unlike Xist the decision of allelic expression is not random (Barlow & Bartolomei, 2014; Pandey et al., 2008). This is because the Kcnq1ot1 promoter is methylated on the maternal chromosome and un-methylated on the paternal chromosome. Interestingly, Kcnq1ot1 transcripts are only localized to a region of genes near the Kcnq1 locus and condense that region into a spheroid nuclear domain, whereas Xist traverses along its entire chromosome (Redrup et al., 2009). How Kcnq1ot1 localizes to a singular area on its chromosome is unknown. Despite distinct differences in chromosome selection for expression and localization on the chromosome, Kcnq1ot1 functions similarly to Xist in that they both interact with chromatin modifiers, recruit epigenetic factors, and establish repressive chromatin domains to mediate gene silencing. For example, it has been found that Kcnq1ot1 recruits the chromatin modifying proteins of the PRC2 complex and G9a to impart its silencing function (Pandey et al., 2008). Moreover, Kcnq1ot1's region of silencing is epigenetically maintained (Andresini et al., 2019; Chiesa et al., 2012). Through their physical association with the genome, Xist and Kcnq1ot1 influence the spatial organization of chromatin, ultimately leading to the establishment of transcriptionally silent nuclear compartments.

Understanding the molecular intricacies of lncRNAs involved in transcriptional silencing will unveil an unprecedented layer of complexity in gene regulation. Here, we use O-MAP to comprehensibly map the genomic loci and interacting transcriptomes of Xist and Kcnq1ot1. We also target nascent Xist transcripts to interrogate what biomolecules the nascent transcript interacts with as it shuttles along the Xi. Overall, O-MAP allowed us to identify key interactors shared by

all three (Xist, Kcnq1ot1, nascent Xist) interactomes that may help impart their ability to transcriptionally silence DNA and/or localize these lncRNAs to their specific genomic loci. Moreover, we demonstrate the localization of each lncRNA on its chromosome with base-pair resolution. Elucidating the biomolecular interactions of Xist and Kcnq1ot1 will provide insight on the fundamental principles of epigenetic regulation and open new avenues for deciphering the broader impact of lncRNAs in gene expression control. Furthermore, unraveling the mechanisms employed by these lncRNAs may pave the way for potential therapeutic interventions in various diseases where gene dysregulation is a hallmark.

## 2.3 RESULTS

**O-MAP precisely targets gene silencing lncRNAs and precursor lncRNAs**

The Kcnq1ot1 domain and Xi both use long noncoding RNAs to silence large chromatin domains in *cis*. The Kcnq1ot1 domain is scaffolded by the Kcnq1ot1 transcript and constitutively inactivates a ~1 Mb domain on the paternal allele of chromosome 7. The Xi is scaffolded by the Xist transcript and epigenetically inactivates almost all genes on one copy of the X chromosome (Fig. 2.1, a). To study these nuclear RNA-scaffolded structures, O-MAP was targeted to Kcnq1ot1 and Xist in Patski mouse embryonic fibroblasts (Lingenfelter et al., 1998). As predicted, both Xist and Kcnq1ot1 O-MAP induced strong biotinylation in a single prominent nuclear punctum, and which exclusively co-localized with RNA-FISH signal for the respective lncRNA (Fig. 2.1, b). We were also curious to examine the X-inactivation center (XIC) where the nascent Xist transcript localizes to within the Xi. To study this RNA-scaffolded structure, we targeted O-MAP to Xist introns. Excitingly, we observed strong O-MAP signal for nascent Xist within a cloud of RNA-FISH signal for the mature Xist transcript (Fig. 2.1, c).

**Xist localizes to the X chromosome while the imprinting RNA Kcnq1ot1 touches multiple genomic loci**

We next examined if O-MAP could map the chromatin loci for the targeted lncRNAs. Our strategy relied on formaldehyde crosslinking and the capture of *in situ* biotinylated proteins to enrich nearby DNA, similar to Chromatin Immunoprecipitation (ChIP; Fig. 2.2, a). When we applied O-MAP-ChIP to Xist and mapped our data genome-wide, we observed robust enrichment along the entirety of the X-Chromosome (Fig. 2.2, b). Patski cells are a hybrid line derived from a cross between *M. musculus* and *M. spretus*, and they exhibit skewed XCI in which the musculus X-chromosome is constitutively inactivated (Lingenfelter et al., 1998). This allowed us to quantify the allelic segregation of Xist O-MAP-ChIP reads, by measuring distinct SNPs between maternal and paternal alleles. Strikingly, the average allelic proportions across Xist O-MAP X-chromosomal peaks were highly specific to the inactive (*musculus*) allele (94%; Fig. 2.2, c). Considerable evidence indicates that the Xist RNA makes direct chromatin contacts exclusively with loci on the inactive X-Chromosome (Engreitz et al., 2013; Simon et al., 2013). This corresponds to our data that demonstrate both mature and nascent Xist O-MAP-ChIP capture regions specific to the X-Chromosome (Fig. 2.2, g).

When we mapped our Kcnq1ot1 O-MAP-ChIP data genome-wide, we observed robust enrichment along the majority of chromosome 7. Signal on chromosome 7 was strongest at the Kcnq1ot1 locus, located at the distal end of chromosome 7. Interestingly, as the data on chromosome 7 progressed, signal located further away from the Kcnq1ot1 locus monotonically decreased in intensity (Fig. 2.2, e). This corresponds to evidence suggesting that Kcnq1ot1 localizes to a domain of genes around its own locus on chromosome 7 (Pandey et al., 2008). Comparing nascent Xist O-MAP-ChIP to mature Xist, we observed very discrete regions of

enrichment just on the Xi (Fig. 2.2, f and h). This is in line with evidence suggesting that the nascent Xist transcript anchors onto discrete regions on the Xi before being spliced into the mature isoform that spreads along the length of the Xi (Markaki et al., 2021; Pandya-Jones et al., 2020). Of note, regions of enrichment for nascent Xist on the Xi appear to largely de-enrich for the mature Xist isoform (Fig. 2.2 h), suggesting the two transcripts compartmentalize into separate but neighboring regions. Intriguingly, the imprinting RNA Kcnq1ot1 differs from Xist in that it interacts with many genomic loci on other chromosomes, while Xist and nascent Xist are exclusively tethered to the Xi (Fig. 2.3, g). Comparing the epigenetic signatures of each lncRNA's chromatin domains, Xist was heavily enriched for weak polycomb repressive signatures. It is well established that Xist recruits polycomb repressive complexes to achieve heterochromatinization of the Xi (Colognori et al., 2019). Intriguingly, we observe the nascent Xist transcript does not enrich for polycomb signatures, but strikingly enriches for the quiescent 2 signature. This signature is unique amongst all quiescent signatures in that it is very low in CpG dinucleotide composition and low for DNA methylation (van der Velde et al., 2021). The most prominent epigenetic signature for Kcnq1ot1 was a weak enrichment for general heterochromatin. This finding was expected given that Kcnq1ot1's role is to promote the heterochromatinization the genomic loci to which it localizes (Pandey et al., 2008).

**Gene silencing RNAs interact with shared and discrete transcripts**

Differentiated cells typically express only 100–200 copies of the Xist transcript (Brockdorff, 2019); these maintain epigenetic silencing by physically coating the inactive X-chromosome (Xi) and compacting it into a discrete subnuclear compartment (Strehle & Guttman, 2020). The protein- and chromatin-interactions that enable Xist to drive this process have been extensively studied (Disteche, 2012; Loda & Heard, 2019), but to our knowledge the RNA

composition of the Xist-coated Xi compartment remains uncharacterized. We thus sought to elucidate the Xi transcriptome by O-MAP-Seq, using our validated Xist-targeting probe set and optimized O-MAP-Seq workflow (Fig. 2.3, a), in mouse Patski cells (Lingenfelter et al., 1998). This revealed striking enrichment of the Xist transcript itself (29-fold enriched over input), and of lncRNAs Jpx and Ftx (average 12.1–fold enrichment), which are located from the same genomic neighborhood and which also contribute to XCI (Fig. 2.3, b,left). Furthermore, we observed prominent enrichment of several X-linked genes that are known to escape XCI (e.g. Mid1; Shroom4) (Disteche & Berletch, 2015), likely due to the capture of nascent transcripts expressed from the Xi (Fig. 2.4). This was paralleled by the enrichment of the recently discovered noncoding RNA GM14636 (Covarrubias et al., 2017), suggesting that it too might escape XCI (Fig. 2.3, b, left). RNA-FISH analysis supported this hypothesis: though mono- and biallelic expression of GM14636 was variable in Patski cells, approximately 41.5% of observed foci co-localized with Xist, indicating penetrant expression from the Xi (Fig. 2.3, c, right). Finally, we noted that Xist O-MAP transcripts were highly enriched for several classes of X-linked transposable elements—most notably, the L1MCb and L1MD1 subfamilies of LINE1 elements (Fig. 2.3, b, right). These elements are overrepresented on the X-chromosome, and have been hypothesized to play a role in XCI (Lyon, 1998). Intriguingly, Xist O-MAP-Seq also enriched for Kcnq1ot1 (Figs. 2.3, b and d, left). Like Xist, Kcnq1ot1 physically coats a Mb-scale domain near the site of its transcription, ultimately driving that domain's heterochromatinization and silencing (Pandey et al., 2008). RNA FISH confirmed that Kcnq1ot1 co-localizes with Xist at higher-than-expected frequency (14% of puncta, n = 896; Fig. 2.3, d, right). This might indicate that Xist and Kcnq1ot1, although bound to their respective chromatin loci in *cis*, share a common pool of resources or visit a shared subnuclear structure that supports their epigenetic silencing functions.

O-MAP-Seq for Kcnq1ot1 revealed it most predominately enriches for nearby transcripts located on chromosome 7 (Fig. 2.3, e, red). Strikingly, O-MAP-Seq for nascent Xist revealed that it interacts with a large number of autosomal transcripts (312, Fig. 2.3, f, black). Suggesting the nascent Xist transcript has a discrete domain separate from that of the mature Xist isoform. Interestingly, all three lncRNAs targeted were in close proximity to Neat1 and Malat1 lncRNAs (Fig. 2.3, b, e, and f). Providing further evidence that these gene silencing lncRNAs share common interactors to achieve their epigenetic functions. In total, Kcnq1ot1, Xist, and nascent Xist all interact with a set of 42 RNAs (Fig. 2.3, g) that are primarily composed of mRNAs and to-be-experimentally-confirmed (TEC) transcripts (Fig. 2.3, h) that span 17 chromosomes (Fig. 2.3, i). Notably, nascent Xist and Kcnq1ot1 share more interacting transcripts than nascent Xist and its mature isoform (Fig. 2.3, g).

**Xist localizes to nucleoli and the lncRNAs Malat1 and Neat1**

Neat1 and Malat1 were key RNA interactors shared by Kcnq1ot1, nascent Xist, and mature Xist lncRNAs (Fig. 2.3). We next sought to validate if Malat1 and Neat1 localize to the Xi and if they affect the expression of Xist and Kcnq1ot1 or the formation or localization of the Xi. We performed high-resolution confocal imaging of Patski cells, visualizing Xist by O-MAP, Malat1 or Neat1 by RNA-FISH, and because Xist is known to localize near nucleoli (Loda & Heard, 2019), immunofluorescence for NPM1. Intriguingly, we found that Xist colocalizes with Neat1, Malat1, and nucleoli in a manner where one side of the Xist cloud touches nucleoli while the other interacts with Neat1 and Malat1 (Fig. 2.5, a and b). This is consistent with our hypothesis that the Xi (as demarked by Xist) may form two discrete compartments—a repressive compartment that co-localizes with perinucleolar heterochromatin, and a transcriptional activating compartment that localizes near a speckle or paraspeckle.

Notably, there is emerging evidence that Malat1 and Neat1 interact with each other, wherein Malat1 localizes to the 5′-end of Neat1 (Cai et al., 2020). To assess the degree Neat1 and Malat1 affect the localization or formation of the Xi, antisense oligonucleotides (ASOs) were designed against each RNA and transfected into Patski cells. Cells were then imaged for Xist O-MAP with RNA-FISH for Neat1 or Malat1 and immunofluorescence for NPM1. Quantitative RT–PCR was also utilized to demonstrate the success of Neat1 or Malat1 knockdown compared to control ASO, as well as the specificity of each ASO for its target RNA (Fig. 2.5, e). Compared to a control ASO, knockdown of Neat1 or Malat1 was successful as gauged by RNA-FISH, but did not affect the formation of the Xi or its localization to nucleoli (Fig 2.5, c and d). Interestingly, while neither Neat1 nor Malat1 knockdown affected Xist expression, Malat1 knockdown resulted in a slight but significant ($p < 0.05$; paired two-tailed Students' t-test) increase in Kcnq1ot1 levels (Fig. 2.5, e).

**O-MAP unifies known Xist proteomes and demonstrates a shared pool of proteins amongst gene-silencing lncRNAs**

Having identified the interacting genomic loci and transcriptome of Xist, Kcnq1ot1, and nascent Xist, we also sought to identify interacting proteins. In our approach, we performed a nuclear enrichment before lysing cells to enhance the signal-to-noise ratio of biotinylated proteins before performing label-free quantification mass spectrometry (Fig. 2.6, a). To validate our method, we examined the degree to which Xist O-MAP enriched proteins corroborated with prior Xist interacting proteomic approaches (Jachowicz et al., 2022; Minajigi et al., 2015). Excitingly, all but one previously identified Xist-interacting protein was enriched in Xist O-MAP-MS (Fig. 2.6, b). Intriguingly, most of these prior identified proteins were also enriched in Kcnq1ot1 and nascent Xist O-MAP-MS (Fig. 2.6, c-d); again, indicating these gene-silencing RNAs interact with

a pool of common biomolecules. Surprisingly, Spen, thought to play a role in the maintenance of the Xi and the initiation of XCI (Robert-Finestra et al., 2021) was not enriched in nascent Xist O-MAP-MS, indicating that nascent Xist functions spreads across the Xi and functions independently of Spen. When comparing the three enriched proteomes, we observed that most proteins were shared amongst Xist, Kcnq1ot1, and nascent Xist (Fig. 2.6, e), perhaps indicating these proteins may be used for similar functions between the RNAs or are indicative of a common nearby subnuclear structure. Moreover, Xist harbored the highest degree of unique interacting proteins, which may indicate proteins with unique functions in the process of XCI (Fig. 2.6, e). Indeed, many proteins already established as being Xist-interacting, such as Spen and Rbm15, were more heavily enriched at mature Xist compared to nascent Xist and mock samples (Fig. 2.6, f). Nascent Xist also had its own unique proteome, enriching heavily for the splicing factors Rnh1 and Lsm8, among others (Fig. 2.6, f & g).

Comparing gene ontology signatures amongst the lncRNAs, we found that most protein functions are shared, such as chromatin organization and lncRNA processing (Fig. 2.7). Unique gene ontology signatures did exist for each lncRNA. For example, Xist solely enriched the COP9 signalsome complex which was found to be important for genome integrity (Meir et al., 2015). Whereas Kcnq1ot1 and nascent Xist were enriched for proteins involved in non-homologous end joining. This may be understood by the common phenomenon that the process of lncRNA transcription often recruits DNA repair proteins (Statello et al., 2021). Similarly, when comparing protein-protein interaction networks between the XCI related transcripts Xist and nascent Xist, we observed that most proteins within the identified networks are shared (Fig. 2.8, a-d). For example, the majority of the minichromosome maintenance complex, important for DNA replication and repair (Bochman & Schwacha, 2009), is represented by both Xist and nascent Xist enriched

proteins (a). In other complexes, such as the SWI/SNF complex that that plays a role in chromatin remodeling (Tang et al., 2010), some proteins were uniquely enriched by mature Xist while others were shared by mature Xist and nascent Xist.

Collectively, these findings provide insight on the protein composition of gene silencing RNAs and demonstrate that their proteomes consist of unique and shared proteins. Gene ontology offers a glimpse into the myriad of distinct and overlapping functions of the targeted gene silencing lncRNA proteomes. Moreover, our investigations into the protein-protein interaction networks of the targeted XCI-related lncRNAs offer insight on where Xist and nascent Xist connect functionally at the level of higher-order protein complexes.

## 2.4 DISCUSSION

In this study, we employed O-MAP to precisely target gene silencing lncRNAs and a precursor lncRNA. Our results provide valuable insights into the localization, interactions, and epigenetic signatures of the RNA-scaffolded structures these lncRNAs nucleate.

We first investigated the nuclear RNA-scaffolded structures of the Kcnq1ot1 domain and the Xi, using O-MAP targeting Kcnq1ot1 and Xist, respectively, in Patski mouse embryonic fibroblasts. As expected, we observed strong O-MAP signals that specifically co-localized with the respective lncRNA transcripts. The Xist O-MAP-ChIP data revealed robust enrichment along the entire X chromosome, confirming the known localization of Xist to the Xi. Allelic segregation analysis demonstrated that Xist O-MAP-ChIP reads were highly specific to the inactive allele, further supporting the exclusive interaction of Xist with loci on the Xi.

Genome-wide mapping of Kcnq1ot1 O-MAP-ChIP data showed robust enrichment along the majority of chromosome 7, with the strongest signal at the Kcnq1ot1 locus. The signal intensity decreased as the distance from the Kcnq1ot1 locus increased, consistent with the known

localization of Kcnq1ot1 to a cluster of genes surrounding its transcription site (Pandey et al., 2008). Additionally, nascent Xist O-MAP-ChIP data revealed discrete regions of enrichment on the Xi, indicating specific interaction sites before the splicing of nascent Xist into its mature isoform. Notably, the epigenetic signatures of each targeted lncRNA's chromatin domains differed. Xist enriched for weak polycomb repressive signatures commonly associated with the Xi (Loda & Heard, 2019). Kcnq1ot1 exhibited a general heterochromatin signature, consistent with its role in promoting the heterochromatinization of the genomic loci it interacts with (Pandey et al., 2008). Nascent Xist was strongly enriched in a unique quiescent signature lacking CpG dinucleotides and methylation.

Furthermore, our O-MAP-Seq analysis of the Xist-coated Xi provided insights into the RNA composition of the Xi structure. We observed enrichment of Xist, as well as the lncRNAs Jpx and Ftx, which are located within the same genomic neighborhood and contribute to X-chromosome inactivation (Loda & Heard, 2019). Additionally, several X-linked genes known to escape X chromosome inactivation, such as Mid1 and Shroom4, were also enriched, likely due to the capture of nascent transcripts that escape XCI (Disteche & Berletch, 2015). Interestingly, Xist O-MAP-Seq also enriched several chromatin-regulatory lncRNAs, including the imprinting regulator Kcnq1ot1, suggesting a potential shared pool of resources or a common subnuclear structure supporting their epigenetic silencing functions.

O-MAP-Seq data for Kcnq1ot1 and Xist revealed they predominantly enriched for nearby transcripts in *cis*, while nascent Xist exhibited many interactions with autosomal transcripts. This suggests that the nascent Xist transcript has a discrete subnuclear domain separate from the mature Xist isoform. Intriguingly, the shared interactors among Kcnq1ot1, nascent Xist, and mature Xist included the RNAs Malat1 and Neat1. Additionally, we identified a set of 42

interacting RNAs, primarily composed of mRNAs and to-be-experimentally-confirmed (TEC) transcripts, spanning 17 chromosomes. These findings provide evidence of common interactors that may contribute to the epigenetic functions of gene-silencing lncRNAs. Notably, nascent Xist and Kcnq1ot1 shared more interacting transcripts than nascent Xist and its mature isoform. A possible explanation for this could with how nascent Xist and the initially transcribed Kcnq1ot1 transcript localize to chromatin. Kcnq1ot1 expression is imprinted due to lack of methylation of the paternal Kcnq1ot1 locus (Pandey et al., 2008) and nascent Xist localizes to areas on the Xi devoid of methylation. Lack of methylation could therefore be a key determinant of how cis-acting gene-silencing lncRNAs initially localize on their chromosome in *cis*, and the common Kcnq1ot1/nascent Xist interacting RNAs could help regulate that process.

We also explored the localization and potential functional roles of Malat1 and Neat1 in relation to Xist. Our imaging analysis revealed co-localization of Xist with Neat1, Malat1, and nucleoli, suggesting a spatial relationship and potential interactions between these RNAs and nucleoli. However, knockdown of Neat1 or Malat1 did not impact the formation or localization of the Xi, indicating that they are not essential for these processes. Interestingly, while Malat1 knockdown did not affect Xist expression, it led to a slight increase in Kcnq1ot1 levels, suggesting a potential regulatory role of Malat1 in Kcnq1ot1-mediated gene silencing. Future studies will examine how this increased expression of Kcnq1ot1 influences its target transcripts and other genes on chromosome 7.

To identify interacting proteins, we employed an additional nuclear enrichment step in the O-MAP-MS protocol followed by label-free quantification mass spectrometry (Fig. 2.6, a). Remarkably, our Xist O-MAP-MS approach successfully enriched almost all previously identified Xist interacting proteins, confirming the reliability and validity of our method (Fig. 2.6, b). Equally

intriguing, these known Xist-interacting proteins were also enriched in Kcnq1ot1 and nascent Xist O-MAP-MS (Fig. 2.6, c-d), suggesting a shared pool of proteins among these gene-silencing RNAs. Notably, Spen, previously associated with Xi maintenance and XCI initiation, was not enriched in nascent Xist O-MAP-MS, indicating independent functions of nascent Xist across the Xi (Fig. 2.6, c). When comparing the enriched proteomes of Xist, Kcnq1ot1, and nascent Xist, we observed a considerable overlap in proteins (Fig. 2.6, e), suggesting potential similarities in their functional roles or associations with a common subnuclear structure. Interestingly, Xist exhibited the highest number of unique interacting proteins, potentially related to its distinct involvement in XCI (Fig. 2.6, f). Additionally, gene ontology analysis revealed shared protein functions, including chromatin organization and ncRNA processing, among the three lncRNAs (Fig. 2.7). Nevertheless, each lncRNA displayed unique gene ontology signatures; for instance, Xist exclusively enriched the COP9 signalsome complex known for its role in genome integrity (Meir et al., 2015), while Kcnq1ot1 and nascent Xist were associated with proteins involved in non-homologous end joining, possibly due to the recruitment of DNA repair proteins during lncRNA transcription (Statello et al., 2021). These findings demonstrate that our O-MAP approach provides valuable insight into the shared and distinct protein interactions of gene-silencing lncRNAs, shedding light on their functional mechanisms and potential implications in cellular processes.

Despite the valuable insights gained from this study, there are several limitations that should be considered. First, the cells used in this study were in the maintenance phase of gene silencing, which might limit the identification of dynamic interactors that occur during the process of inactivation. Further investigations using cells at different stages of gene silencing, including the initiation and establishment phases, could provide a more comprehensive understanding of the RNA-scaffolded structures and their interactors. In terms of future directions, it would be

intriguing to explore the combined knockdown of Malat1 and Neat1 to investigate their potential synergistic effects on the formation and localization of the Xi, as well as their impact on the expression of Xist and Kcnq1ot1. Furthermore, expanding the study to other cell types and species would enhance our understanding of the conservation and variability of gene silencing lncRNA-scaffolded structures in different biological contexts. Investigating the dynamics of these structures over time, and in response to environmental cues or developmental processes could provide a deeper understanding of their functional roles and regulatory mechanisms.

In conclusion, O-MAP provides precise targeting and characterization of gene silencing lncRNAs and a precursor lncRNA. Utilizing this, we identified the interactions with chromatin loci and epigenetic signatures of Xist, Kcnq1ot1, and nascent Xist. Our findings highlight the specific interactions of Xist with the Xi and the surprising diverse interactions of Kcnq1ot1 with many genomic loci. Furthermore, we identified shared interactors among Xist, Kcnq1ot1, and nascent Xist, including the lncRNAs Malat1 and Neat1, suggesting potential co-regulation and functional associations. Overall, this study expands our understanding of lncRNA-scaffolded structures involved in gene silencing processes and provides valuable insights into their localization, interactions, and epigenetic signatures.

## 2.5 METHODS

**Cell culture**

Patski cells (a gift from Dr. C. Disteche, UW were cultured in High Glucose DMEM with Pyruvate (Thermo Fisher; 11995073), supplemented with 10% (v/v) Fetal Bovine Serum (FBS, Thermo Fisher; 26140079), 100 units/mL penicillin and 100 µg/mL streptomycin (Thermo Fisher;

15140122), and 1x GlutaMAX (Thermo Fisher; 35050061). Cells were maintained at 37°C, under 5% CO2.

For imaging experiments, cells were cultured in two-well Lab-Tek borosilicate glass #1.0 chambers (Thermo Fisher; 155380). For biochemistry and high-throughput sequencing experiments, cells were cultured in six-well plates. When necessary, material from multiple wells was harvested and merged into a single lysate, as described below.

**lncRNA O-MAP main protocol**

Patski cells were used for both O-MAP omics and O-MAP imaging. For omics, cells were plated in six-well dishes ($3.5\times10^5$ cells/well; plated one day prior to harvest). For imaging-only experiments, cells were plated at $7\times10^4$ cells per well, in two-chamber LabTeks. In all cases, RNase-free reagents and manipulations were used throughout. The core O-MAP protocol follows the same two-day protocol as previously described (Tsue, Kania et al., 2023).

**lncRNA O-MAP day 1**

Cells were quickly washed with 1x DPBS (Thermo Fisher; 14190250) and fixed in 2% (v/v) formaldehyde (Electron Microscopy Sciences; 15710) in 1x PBS (Sigma; 6506) for 10 minutes. The fixation solution was aspirated and quenched by two five-minute washes with 250mM glycine in 1x PBS, with rocking (3 rpm on a platform rocker (Everlast Rocker 247)). Cells were then briefly washed with DPBS and permeabilized with 0.5% (v/v) triton-X 100 in PBS for 10 min with rocking. Three five-minute DPBS washes were then performed to prevent over-permeabilization. Samples were then incubated with a 0.5% (v/v) $H_2O_2$ solution for 10 minutes with rocking. Two five-minute washes with DPBS followed to prevent $H_2O_2$ damage to cells. Cells were then equilibrated in Formamide Wash Buffer (40% (v/v) deionized formamide (VWR, 0314); 2X SSC (Thermo Fisher; AM9765); 0.1% Tween-20) for five minutes with rocking. Wash buffer was then

aspirated, and to each well 50 µL of Primary Probe Mix (Supplementary File 1) (0.1 µM primary oligo probe pool, in 1x Hybridization Buffer: 40% deionized formamide; 2x SSC; 0.1% (v/v) Tween-20; 10% (w/v) dextran sulfate (SIGMA; D8906); in nuclease-free water) was added. A clean 12-well glass coverslip was then applied to spread the Primary Probe Mix across the samples. A kimwipe soaked in RNase-free water was placed between the wells, and plates were sealed with Parafilm to create a hybridization chamber. Plates were then incubated for 8 hours at 42°C without rocking.

**lncRNA O-MAP day 2**

After primary hybridization, 30% Formamide Wash Buffer was added to the samples. Coverslips were then removed, and cells were allowed to wash for five-minutes rocking. Two additional five-minute washes with fresh 30% Formamide Wash Buffer followed. Then 50 µL of O-MAP Secondary Probe Mix (100 nM SABER1-HRP oligo in 30% Hybridization Buffer) was added to each well, and a clean coverslip was added. Samples then incubated for one hour at 37°C in the dark. For lncRNA O-MAP imaging, endogenous biotin blocking was performed (see lncRNA O-MAP Imaging below). Otherwise, to each well 1-mL of PBST (0.1% (v/v) Tween-20 in 1X PBS) was added. Coverslips were then removed, and samples washed for 15 minutes in PBST rocking at room temperature. Three more 15-minute washes with PBST followed. Following aspiration, 1-mL of Labeling Solution was added (0.8 µM biotinyl-tyramide (Sigma; SML2135), 1 mM $H_2O_2$, 1x PBST) and in situ biotinylation was carried out for 10-60 minutes depending on the target RNA (Supplementary File 1). The biotinylation reaction was then quenched by the addition of 1-mL of PBST-q (10 nM sodium azide, 10 nM sodium ascorbate in 1X PBST), for three five-minute washes rocking.

**lncRNA O-MAP Imaging**

After hybridization with SABER1-HRP, samples were washed three times for five minutes in PBST. To block endogenous biotin signal, samples were then incubated for 30 minutes with 1-mL of 1% nuclease-free BSA (VWR; 97061-420) in PBST with rocking, followed by a 15-minute incubation with 1-mL of Neutravidin solution (10 µg/mL neutravidin (Thermo Fisher; 31000), 1% (w/v) nuclease-free BSA, in 1x PBST) rocking. After three five-minute washes in PBST, samples were incubated with 1-mL of biotin solution (10 µg/mL D-biotin (Thermo Fisher; B20656) in 1x PBST) to occupy free neutravidin binding sites. Samples were then washed three times in PBST before proceeding with in situ biotinylation and quenching as described above. After biotinylation samples were incubated with 1-mL of 1 µg/mL neutravidin-Dy550 (ThermoFisher; 84606) in a 1% BSA PBST solution for one hour at room temperature with rocking. Samples were then washed three times with PBST, stained with DAPI, and either imaged immediately or mounted in Vectashield (Vector Labs; H-1900-10).

**Combined lncRNA O-MAP, RNA-FISH , and immunofluorescence**

To perform O-MAP with RNA-FISH and immunofluorescence, cells were first fixed in PFA, quenched with glycine, permeabilized in Triton X-100, underwent peroxidase inactivation, equilibrated in formamide, and hybridized to primary probes as described above (see lncRNA O-MAP Day 1). The primary probes were appended with a landing pad for recruitment of either SABER1 or SABER2 sequences (Supplementary File 1). O-MAP was then performed for imaging which includes endogenous biotin block (lncRNA O-MAP Imaging). Following quenching, samples were equilibrated in formamide by three five-minute washes in 1-mL 30% Formamide Wash Buffer. This buffer was aspirated and 50 µL of FISH secondary probe mix (100 nM SABER2–AF647 oligo in 30% Formamide Hybridization Buffer) was added, followed by the placement of a clean coverslip over the sample. Samples were incubated at 37°C in the dark, then

washed three times with PBST. Before proceeding to immunofluorescence, samples were next blocked with 5% (w/v) nuclease-free BSA in 1X PBST for 30 minutes. Samples were then incubated with rabbit-NPM1 (ThermoFisher; PA517742, used at 1:100 dilution) in 5% BSA in 1X PBST for 1 hour 45 minutes. Samples were then washed 3x5 minutes in 1X PBST. Then samples were incubated with anti-rabbit-Atto488 (ThermoFisher; R37116, used at 1:300 dilution) in 1% BSA in 1XPBST along with 1 µg/mL neutravidin-Dy550. Four fifteen-minute washes in 1X PBST ensued, followed by DAPI staining. Samples were imaged immediately in 1X PBST.

When performing pairwise fluorescent imaging, either O-MAP with RNA-FISH, O-MAP with immunofluorescence, or RNA-FISH with immunofluorescence, the O-MAP protocol preceded the RNA-FISH or immunofluorescence protocol, and RNA-FISH preceded the immunofluorescence protocol.

**lncRNA O-MAP ChIP**

lncRNA O-MAP ChIP was performed as previously described (Tsue, Kania et al., 2023). Briefly, O-MAP labeled cells ($5.0x10^6$ per replicate; two biological replicates per experimental condition) were harvested in PBSTq using cell scrapers, and pelleted at 3,000 x g for 10 minutes at 4°C. Then PBSTq was aspirated and cells were flash-frozen in liquid $N_2$ before storing at -80°C until use. Samples were thawed on ice and resuspended by gentle pipetting in 1mL CLB (20 mM Tris pH 8.0, 85 mM KCl, 0.5% (v/v) NP-40) supplemented with 1x Halt EDTA-Free protease inhibitor cocktail and 10 mM sodium azide, for 10 minutes. Lysates were clarified by centrifugation at 3,000 x g for five minutes at 4 °C followed by aspiration of supernatant. Two more rounds of resuspension with CLB, clarification, and aspiration ensued. Then pellets were lysed by gentle pipetting in 1-mL of NLB (10 mM Tris-HCl pH 7.5, 1% (v/v) NP-40, 0.5% (w/v) sodium Deoxycholate, 0.1% (w/v) SDS) supplemented with 1x Halt EDTA-Free protease inhibitor

cocktail and 10 mM sodium azide, for 10 minutes. Lysates were then placed into an ice-cold thermal block and sheared with a Branson Digital Sonifier outfitted with a double stepped microtip, at 10-12 Watts over 30 s intervals (0.7 s on; 1.3 s off), with 30 s resting steps between intervals. A total of 18 intervals were performed per sample, and after every 5 intervals the thermal block was placed on ice for 2 minutes to recalibrate to ice-cold temperature. This resulted in an average shearing size of approximately 200-400 bp.

Samples were then clarified by centrifugation at 15,000xg for 10-minutes at 4°C, and supernatants were moved to fresh tubes. Each sample then had 10 µL removed as 'input'. To the remaining 990 µL, 100 µL of streptavidin-coated magnetic beads equilibrated by two NLB washes was added. Samples rotated end-over-end for 2 hours at room-temperature. The beads were then washed with the following series of buffers (1 mL each, 5-minutes per wash, at room-temperature, with end-over-end rotation): (1) NLB, supplemented with 5 mM EDTA, 10 mM sodium azide and protease inhibitors (1x Halt EDTA-free Protease Inhibitor Cocktail), 150 mM NaCl; (2) NLB, supplemented with 5 mM EDTA, 10mM sodium azide and protease inhibitors, (3–4) two washes in 1 M KCl, 10 mM Tris-HCl pH 7.5, 5 mM EDTA, (5–6) two washes in 2 M Urea, 10 mM Tris-HCl pH 7.5, 5 mM EDTA, (7) 10 mM Tris-HCl pH 7.5, 1% (w/v) SDS, (8–9) 10 mM Tris-HCl pH 7.5, 1 mM EDTA.

Following the final aspiration, beads were resuspended in 100 µL of Elution Buffer (2% (v/v) N-lauryl sarcosine, 10 mM EDTA, 5 mM DTT, in 1x PBS, supplemented with 200 µg proteinase K). Samples were then incubated at 65°C while shaking at 700 rpm in a Mixer HC. Following this, samples were transferred to 0.2-mL tubes and incubated overnight in a thermocycler at 65 °C. Oligonucleotides were then extracted by addition of an equal volume of phenol pH 6.6, followed by two equal volume absolute chloroform extractions. To each sample 1

µg GlycoBlue and NaCl to 300 mM final were added, followed by ethanol precipitation overnight at -20 °C. Samples were then pelleted by centrifugation at 15,000 x g for 30 minutes at 4 °C. Pellets were washed twice with 80% ethanol and resuspended in 20 µL nuclease-free water. Residual RNA was removed by supplementation of 10 µg RNase Cocktail Enzyme Mix (Thermo Fisher; AM2286) and incubation at 37 °C for 1 hour. Finally, DNA was purified by phenol extraction and ethanol precipitation as described above, and resuspended in 20 µL of nuclease-free water.

**lncRNA O-MAP ChIP library preparation and sequencing**

DNA samples were processed for sequencing and sequenced as previously described (Tsue, Kania et al., 2023). Briefly, samples were quantified using a NanoDrop One and 300 ng of DNA from each sample was used for library preparation, using the NEBNext Ultra II DNA Library Prep Kit and NEBNext Multiplex Oligos for Illumina (NEB;E7645 and E7335), according to the manufacturer's instructions. Two biological replicates were used per condition, and each library was given a unique index. Library concentrations were quantified using the NEBNext Library Quantification Kit for Illumina (NEB; E7360). Library quality was confirmed using an Agilent 4200 TapeStation with a "DNA High Sensitivity" kit (Fred Hutch Genomics Core). Libraries were pooled in equimolar concentrations to 20 nM aggregate in nuclease-free water, with no more than eight libraries per pool. Each pool was subjected to 150 cycles of paired-end sequencing on an Illumina HiSeq 4000, run in high output mode (Azenta Life Sciences).

**lncRNA O-MAP Seq**

O-MAP labeled cells ($9.0x10^6$ cells; four biological replicates per experimental condition) were harvested by scrapping in PBSTq. Samples were then pelleted by centrifugation at 3,000xg for ten-minutes at 4°C, aspirated, flash-frozen in liquid $N_2$, and stored at -80°C until use. Samples were thawed on ice and resuspended by gentle pipetting in 1000 µL ice cold RIPA Buffer (50 mM

Tris-HCl pH 7.5, 150 mM NaCl, 0.1% (w/v) SDS, 0.5% (w/v) sodium Deoxycholate, 1% (v/v) Triton X-100, 5 mM EDTA, 0.5 mM DTT), supplemented with 1x EDTA-Free Halt Protease Inhibitor Cocktail, 0.1 U/μL RNase-OUT (Thermo Fisher; 10777019)), and 10mM sodium azide, and rocked end-over-end at 4°C for five minutes. Cells were placed in an ice-cold thermal block and then sheared using a Branson Digital Sonifier 250 outfitted with a double stepped microtip, at 10–12 Watts for 30 s intervals (0.7 s on; 1.3 s off), with 30 s resting steps between intervals, seven intervals total. After every five intervals the thermal block was re-equilibrated to an ice-cold temperature by incubation on ice for two-minutes. After sonication, lysates were clarified by centrifugation at 15,000xg for ten-minutes at 4°C. Supernatant was placed in fresh tubes and diluted with 1-mL Nuclear Lysis Buffer (NLB: 25 mM Tris-HCl pH 7.5, 150 mM KCl, 0.5% (v/v) NP-40, 5 mM EDTA, 0.5 mM DTT), supplemented with 1X Halt Protease Inhibitor Cocktail, 0.1 U/μL RNase-OUT and 10 mM sodium azide. Each sample then had 5% of its volume taken as "input," and 100 μL Pierce streptavidin magnetic bead slurry that had been equilibrated by two washes in 1:1 RIPA:NLB supplemented with 10mM sodium azide, 0.1 U/μL RNase-OUT, and 1X Halt Protease Inhibitor Cocktail was added to the remainder. Samples were incubated for 2 hr at room temperature with end-over-end agitation. Beads were then washed with the following series of buffers (1 mL each, 5 min per wash at room-temperature with end-over-end agitation). All buffers were supplemented with 1x EDTA-Free Halt protease inhibitor cocktail, 0.1 U/μL RNase-OUT, and 0.5 mM DTT, unless otherwise noted: (1) RIPA, supplemented with 10 mM sodium azide; (2) RIPA alone (3) High Salt Buffer (1 M KCl, 50 mM Tris-HCl pH 7.5, 5 mM EDTA) (4) Urea Buffer (2 M Urea, 50mM Tris-HCl pH 7.5, 5 mM EDTA) (5) 1:1 RIPA:NLB, without protease inhibitors (6) NLB, without protease inhibitors, (7–8) two washes in TE buffer (10 mM Tris-HCl pH 7.5, 1 mM EDTA), without protease inhibitors.

RNA was isolated from both input and O-MAP-enriched samples by proteolysis in 100 µL Elution Buffer (2% (v/v) N-lauryl sarcosine, 10 mM EDTA, 5 mM DTT, 200 µg proteinase K (Thermo Fisher; AM2548), in 1xPBS). Reactions were shaken at 700 rpm in a Mixer HC (USA Scientific) for 1 hour at 42°C, followed by 1 hour at 60°C. RNA was then extracted once with 1 volume of phenol pH 4.3, and twice thereafter with an equal volume of absolute chloroform. Reactions were supplemented with 1µL Glycoblue (Thermo Fisher; AM9515) and NaCl to 300 mM, and ethanol precipitated at –20°C overnight. RNA was harvested by centrifugation at 15,000 x g for 30 minutes at 4°C, washed twice with 70% ethanol, and resuspended in 84.75 µL nuclease free water. Contaminating DNA was removed by digestion with 5 U RQ1 RNase-free DNase I (Promega; M6101) in 100 µL of the manufacturer's supplied buffer (1x final concentration) at 37°C for 30 min, and this reaction was terminated by addition of EDTA to 15 mM, final. RNA was purified by phenol extraction and ethanol precipitation, as described above, and resuspended in 15 µL nuclease free water. Sample concentration was measured using a Nanodrop One (Thermo Fisher).

**lncRNA O-MAP Seq library preparation and sequencing**

RNA samples were processed for sequencing as previously described (Tsue, Kania et al., 2023) and sequenced with no more than 8 libraries per lane. Briefly, ribosomal RNA was first depleted by RNase-H digestion, using a pool of DNA oligonucleotides antisense to rRNA, as previously described (Adiconis et al., 2013). Murine rRNA and mitochondrial-rRNA probes were designed using Oligo-ASST (Phelps et al., 2021) using FASTA files for the 5.8S, 5S, 12S, 16S, 18S, and 28S rRNA species from the *Mus musculus* complete genome (NCBI:txid10090). Probes were

pooled and 1 µg of the antisense probe pool was added to 1 µg of RNA in 200 mM NaCl, 100 mM Tris-HCl, pH 7.4, at a final volume of 10 µL. Samples were heated at 95°C for 2 minutes then slowly cooled to 45°C at a rate of -0.1°C/s, using a ProFlex PCR system. Then, reactions were supplemented with 10 µL of RNase H mix (10 U Hybridase Thermostable RNase H (Lucigen; H39500), 20 mM MgCl2) that was pre-heated to 45°C. Samples were incubated at 45°C for 30 minutes then placed on ice. RNA was purified by acidic phenol-chloroform extraction and ethanol precipitation. Residual DNA was removed using DNase I, and RNA was purified again by the RNA Clean and Concentrator-5 Kit.

Samples were quantified on a Nanodrop One. Sequencing libraries were made from 300 ng RNA using the NEBNext Ultra II Directional RNA Library Prep Kit and NEBNext Multiplex Oligos for Illumina (NEB; E7760 and E7735), according to manufacturer's instructions. Four biological replicates were used for each condition and each library was given a unique index. Libraries were quantified using the NEBNext Library Quantification Kit for Illumina, following manufacturer's instructions, and library quality was confirmed using an Agilent 4200 TapeStation with a "DNA High Sensitivity" kit (Fred Hutch Genomics Core). Libraries were pooled in equimolar concentrations to 20 nM aggregate concentration in nuclease-free water, with no more than 8 libraries per pool. The pool underwent 150 cycles of paired-end sequencing, on one lane of an Illumina HiSeq 4000 per pool, run in high output mode (Azenta Life Sciences).

**preTTN O-MAP-ChIP data analysis**

Deep sequencing reads were trimmed using TrimGalore! (*Babraham Bioinformatics - Trim Galore!*) with parameters -q 30 --phred33, and mapped to the mm10 genome using Bowtie2 version 2.4 (Langmead & Salzberg, 2012). Duplicate reads were removed with the Picard MarkDuplicates function (*Picard Tools - By Broad Institute*). Coverage bigWig files were

generated using deepTools version 3.5 (Ramírez et al., 2014) with a binsize of 1 and normalizing with RPKM. Log2FC bigWig files were generated by comparing O-MAP ChIP data to matched input samples. Regions of enrichment were called using Epic2 (Stovner & Sætrom, 2019) using a bin size of 20,000. Epigenomic analysis of enriched DNA loci was performed using ChromHMM version 1.22 (Ernst & Kellis, 2012). The OverlapEnrichment function was called using a 15-state chromatin state file for mice (ENCFF099OAB). Fold enrichment of each epigenomic signature was plotted as a heatmap using seaborn version 0.10.1 (Waskom, 2021).

SNP analysis of the allelic segregation of XIST-interacting chromatin regions (Fig. 2.2, c) was performed as previously described, using a method optimized for the Patski cell line (Bonora et al., 2018). Briefly, analysis was restricted to MACS2-enriched regions along chromosome X with a coverage of at least five reads. Each region's allelic proportion was then calculated by computing (Xi/(Xa+Xi)) for each read within the region. Regions with an allelic proportion >0.7 were designated as Xi-specific; those with an allelic proportion <0.3 were designated as Xa-specific; those between 0.3 and 0.7 were classified as common to both alleles.

**lncRNA O-MAP-Seq data analysis**

Raw RNA-seq FASTA files were aligned to mm10 using HISAT2 version 2.2.1 (Kim et al., 2019), in paired-end setting with default parameters. The resulting SAM files were converted to BAM format and sorted using Samtools (H. Li et al., 2009) version 1.15.1. Bigwig files for visualizing strand specific information were created using deepTools version 3.5.19 with parameters: --filterRNAstrand forward/reverse --binSize 1 --normalizeusingBPM. Mapped reads were quantified using StringTie (Pertea et al., 2015) version 2.2.1 and the StringTie output was prepared

for differential expression analysis using the prepDE.py function. The resulting count matrices were used for differential expression analysis using DESeq2 (Love et al., 2014) with a FDR cutoff of 0.05. Volcano plots were generated using EnhancedVolcano version 1.12.0 (Blighe, 2018/2023). All statistical analyses (Fisher's exact test, hypergeometric distribution test, or Student's t-test, where appropriate) were performed in R or in python using the ggplot2 (Wickham, 2009), seaborn (Waskom, 2021), or matplotlib (Hunter, 2007) modules.

**lncRNA O-MAP proteomics sample preparation**

Briefly, O-MAP labeled cells ($9.0x10^6$ per replicate; four biological replicates per experimental condition) were harvested in PBSTq without sodium ascorbate using cell scrapers, and pelleted at 3,000 x g for 10 minutes at 4°C. Then PBSTq was aspirated and cells were flash-frozen in liquid $N_2$ before storing at -80°C until use. Samples were thawed on ice and resuspended by gentle pipetting in 1mL CLB (20 mM Tris pH 8.0, 85 mM KCl, 0.5% (v/v) NP-40) supplemented with 1x Halt EDTA-Free protease inhibitor cocktail and 10 mM sodium azide, for 10 minutes. Lysates were clarified by centrifugation at 3,000 x g for five minutes at 4 °C followed by aspiration of supernatant. After two more rounds of resuspension with CLB, clarification, and aspiration, pellets were lysed by gentle pipetting in 1-mL of NLB (10 mM Tris-HCl pH 7.5, 1% (v/v) NP-40, 0.5% (w/v) sodium Deoxycholate, 0.1% (w/v) SDS) supplemented with 1x Halt EDTA-Free protease inhibitor cocktail and 10 mM sodium azide, for 10 minutes. Lysates were then placed into an ice-cold thermal block and sheared with a Branson Digital Sonifier outfitted with a double stepped microtip, at 10-12 Watts over 30 s intervals (0.7 s on; 1.3 s off), with 30 s resting steps between intervals. A total of 4 intervals were performed per sample, and after each sample was sonicated the thermal block was placed on ice for 2 minutes to recalibrate to ice-cold temperature.

Samples were then supplemented with 47.47 µL of 20% SDS to bring the final SDS concentration to 1%. The samples were then boiled at 95°C for one hour to ensure complete reverse-crosslinking. Samples were allowed to cool to 50 °C before placing on ice. Then each sample was sonicated for one more cycle. Samples were then clarified by centrifugation at 15,000xg for 10-minutes at 4°C, and supernatants were moved to fresh tubes. To each tube 75 µL of streptavidin-coated magnetic beads equilibrated by two NLB washes was added. Samples rotated end-over-end for 2 hours at room-temperature. The beads were then washed with 250 µl of NLB, supplemented with 5 mM EDTA, 10 mM sodium azide, protease inhibitors (1x Halt EDTA-free Protease Inhibitor Cocktail), and 150 mM NaCl. Beads were then moved to fresh PCR-strip tubes (Simport Scientific; T320-2N) to minimize bead loss. The following series of washes were then performed (200 µl each, 1 minute per wash, at room-temperature, with end-over-end rotation): (1-4) NLB, supplemented with 5 mM EDTA, 10mM sodium azide and protease inhibitors (5-8) 1X PBS in nuclease-free water (9-16) 1 M KCl, 10 mM Tris-HCl pH 7.5, 5 mM EDTA, (17-24) 2 M Urea, 10 mM Tris-HCl pH 7.5, 5 mM EDTA (25-28) 200 mM EPPS pH 8.5 in nuclease-free water. After the final wash the beads were moved to fresh PCR strip tubes, aspirated, and resuspended in 50 µl 200 mM EPPS pH 8.5.

To place the PCR strip tubes in a Mixer HC (USA Scientific), aluminum foil was wrapped around the end of a 1-mL pipet tip. These hollow aluminum cones were then placed in the 1.5-mL wells of the Mixer HC, and to each a single sample in a PCR tube was added. TCEP was then added to each sample to a final concentration of 10 mM and incubated at RT for 1-hour shaking at 700 rpm. Then iodoacetamide was added to each sample to a final concentration of 20 mM and incubated in the dark for one-hour at RT shaking at 700 rpm. Alkylation was quenched by the addition of DTT (5 mM final) and incubation at RT for 15 minutes shaking at 700 rpm. To each sample 300

ng of mass spectrometry grade LysC (FujiFilm; 121-05063) was added and samples incubated at 37C for three-hours shaking at 700 rpm. This was followed by the addition of 300 ng of sequencing grade modified trypsin (Promega; V5113) and incubation for exactly 16-hours at 37C shaking at 700 rpm.

The following day beads were magnetically removed, and the clarified samples were placed in fresh PCR strip tubes. Samples were then frozen at -80 °C to inactivate LysC/trypsin and were stored no longer than three-days. Samples were then thawed to room temperature and to each 0.5 µL of 1% HPLC-grade formic acid was added to achieve pH < 3 and desalted using C18 StageTips according to the published protocol with minor adjustments (Tsue, Kania et al., 2023). Briefly, the StageTip was activated with 50µL of HPLC methanol, followed by addition of 50uL of Stage Tip B (80% acetonitrile in 0.1% trifluoroacetic acid (TFA)) and 50µL of Stage Tip A (0.1% TFA). After sample loading, peptides were washed with 50uL of Stage Tip A.

**O-MAP-MS data acquisition and analysis**

Peptides from StageTips were eluted using 50uL of 45% acetonitrile, 0.1% TFA into a 96-well plate. The samples were dried down and resuspended in Stage Tip A. Peptides were separated on an EASY-nLC 1200 System (Thermo Fisher Scientific) using 20 cm long fused silica capillary columns (100 µm ID, laser pulled in-house with Sutter P-2000, Novato CA) packed with 3 µm 120 Å reversed phase C18 beads (Dr. Maisch, Ammerbuch, DE). The liquid chromatography (LC) gradient was 90 min long with 5−35% B at 300 nL/min. LC solvent A was 0.5% (v/v) aq. acetic acid and LC solvent B was 0.5% (v/v) acetic acid, 80% acetonitrile. MS data was collected with a Thermo Fisher Scientific Orbitrap Fusion Lumos using a data-independent acquisition (DIA) method with a 120K resolution Orbitrap MS1 scan and 12 m/z isolation window, 15K resolution Orbitrap MS2 scans for precursors from 400-1000 m/z.

Data .raw files were converted to .mzML using MSConvert 3.0.21251-d2724a5 and spectral libraries built using MSFragger-DIA ((Kong et al., 2017)) (w/ FragPipe version 19.1) with quantification through DIA-NN version 1.8.232. The database search was against the UniProt murine database (updated September 17th, 2021) containing 20420 sequences and 20420 reverse-sequence decoys. For the MSFragger analysis, both precursor and (initial) fragment mass tolerances were set to 20 ppm. Spectrum deisotoping, mass calibration, and parameter optimization were enabled. Enzyme specificity was set to "stricttrypsin" and up to two missed trypsin cleavages were allowed. Oxidation of methionine, acetylation of protein N-termini, −18.0106 Da on N-terminal Glutamic acid, and −17.0265 Da on N-terminal Glutamine and Cysteine were set as variable modifications. Carbamidomethylation of Cysteine was set as a fixed modification. Maximum number of variable modifications per peptide was set to 3.

FragPipe/DIA-NN output files were processed and analyzed using the Perseus software package v1.5.6.0. Expression columns (protein MS intensities) were log2 transformed and normalized by subtracting the median log2 expression value from each expression value within each MS run. For statistical testing of significant differences in expression, a two-sample Student's t test was applied.

**Antisense Oligonucleotide (ASO) Knockdown and Quantitative PCR**

Antisense LNA GapmeR ASOs were designed against Neat1 and Malat1 using Qiagen's proprietary GeneGlobe platform. Cells were transfected with 50 nM (final concentration) of ASO using RNAi Max (ThermoFisher; 13778075) following manufacturer's instructions. Cells then incubated for 48 hours. For quantitative PCR (qPCR), each well of cells had media aspirated followed immediately by addition of 1-mL TRIzol reagent (Invitrogen; 15596026). RNA extraction proceeded according to manufacturer's protocol. Contaminating DNA was removed by

incubation with DNase I and the use of the RNA Clean and Concentrator-5 Kit as described above (see preTTN O-MAP seq). RNA was eluted in 30 µL nuclease-free water. RNA samples were reverse transcribed using SuperScript IV Reverse Transcriptase (Thermo Fisher; 18090010), priming with random hexamers according to the manufacturer's protocol. cDNA was diluted with nuclease-free water, mixed with gene specific primers, supplemented with Rox-normalized PowerUp SYBR Green Master Mix (Thermo Fisher; A25777), and aliquoted into 384-well plates. qPCR was performed on a QuantStudio 5 Real-Time PCR System (Thermo Fisher; A34322) in both biological and technical quadruplicate. Gene expression was calculated using the 2-ΔΔCt method using the 12S rRNA.

**Image acquisition**

Fluorescence widefield microscopy was performed on a Leica DM IL, equipped with a HC Fluotar 100x oil immersion objective with a 1.32 numerical aperture and planar correction (Leica; 11506527), a white LED light source (Leica; EL6000) and a DFC365 FX digital camera (Leica; 11547004). The following filter cubes were used: Texas Red (Leica TX2 ET; 11504180; used with Dylight-550 conjugates), Cy5 (Leica Y5 ET; 11504181, used for Alexafluor-647), GFP (Leica GFP ET; 11504174, used for Atto-488 and AF488), and DAPI (Leica DAPI ET; 11504204). Illumination intensity was adjusted using the light source manual control; acquisition times ranged from 40–1000 ms, as controlled by the Leica LASX software. Fluorescence confocal microscopy was performed on a Leica SP8X microscope (UW Keck Imaging Center), outfitted with a HC CS2 63x oil immersion objective, with 1.40 numerical aperture with both planar and apochromatic correction. The pixel size was 0.18 µm for fixed cell distance quantification, 0.06 µm for fixed cell representative images, and 0.045 for FFPE distance quantification and representative images. Samples were illuminated using a 470–670nm tunable White Light Laser system, with a typical

laser power of 0.1% for DAPI, 3% for 550 nm, and 10% for 647 nm. Gain and offset settings were adjusted to avoid pixel saturation. Images were line-averaged twice, with an average pixel dwell time of 1.58 µs. A bit-depth of 16 was used and zoom factor of 1-3 was used for all images.

**Image processing**

Images were processed using Fiji (Schindelin et al., 2012) and ImageJ (Schneider et al., 2012), and multicolor overlays were made using the screen setting in Adobe Photoshop. All confocal images are maximum projections of z-stacks. Brightness and contrast were adjusted for display purposes using Fiji and ImageJ or Adobe Photoshop. In all cases, contrast adjustment was applied to improve signal visibility, by changing the minimum (black) and maximum (white) values only. Automated despeckling was applied when necessary to reduce residual background signal.

## 2.6 ACKNOWLEDGMENTS

We thank N. Peters for imaging assistance; A. Mishra, R.D. Hawkins, and G. Bonara for sequencing assistance; and C. Disteche and X. Deng for the kind donation of materials. This work was supported by National Institutes of Health Grants 1R01GM138799-01 and T32GM007750, a Safeway Albertsons Early Career Award in Cancer Research, a Brotman Baty Institute Catalytic Collaborations Award, the UW Royalty Research Fund (RRF), and the UW Student Technology Fund (STF). E.E.K. was supported by the NSF DEB2016186 and the AHA 902616. Imaging at the UW Keck Center was supported by NIH S10 OD016240 and the UW Student Technology Fee. This work used an EASYnLC1200 UHPLC and Thermo Scientific Orbitrap Fusion Lumos Tribrid mass spectrometer purchased with funding from a National Institutes of Health SIG grant S10OD021502. NGS data analysis was facilitated through the use of advanced computational, storage, and networking infrastructure provided by the Hyak supercomputer system and funded by the STF at the University of Washington.

## 2.7 FIGURES

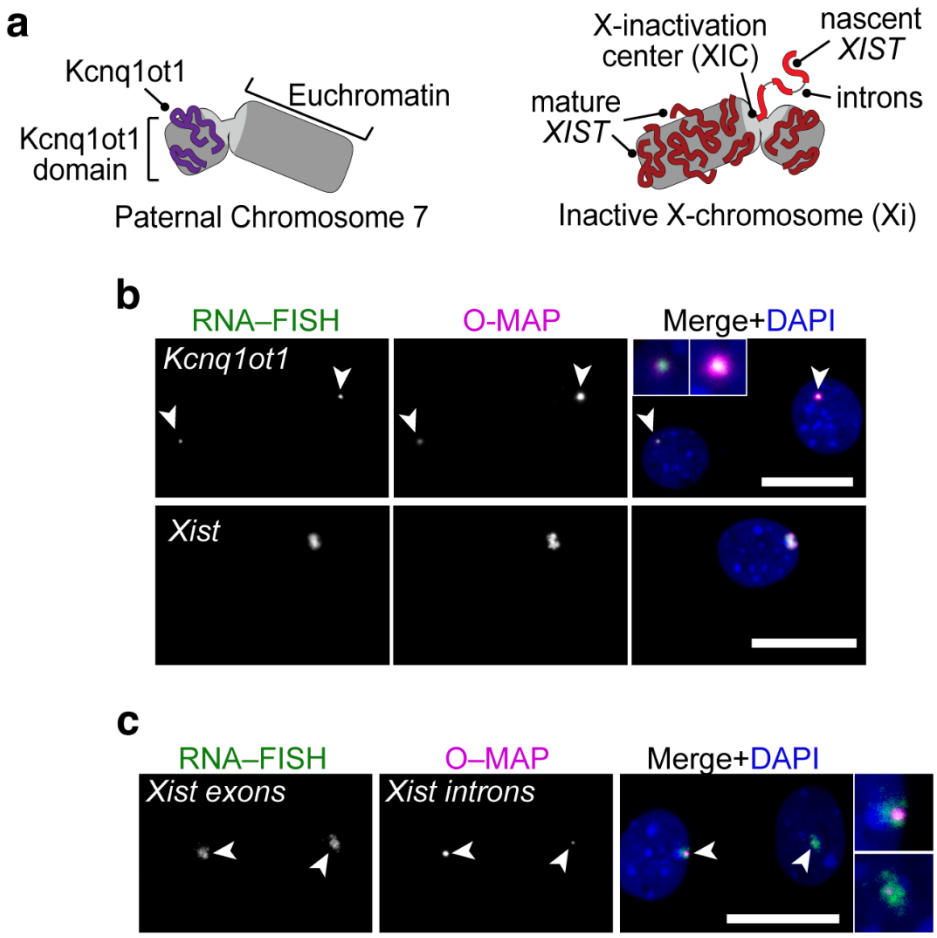
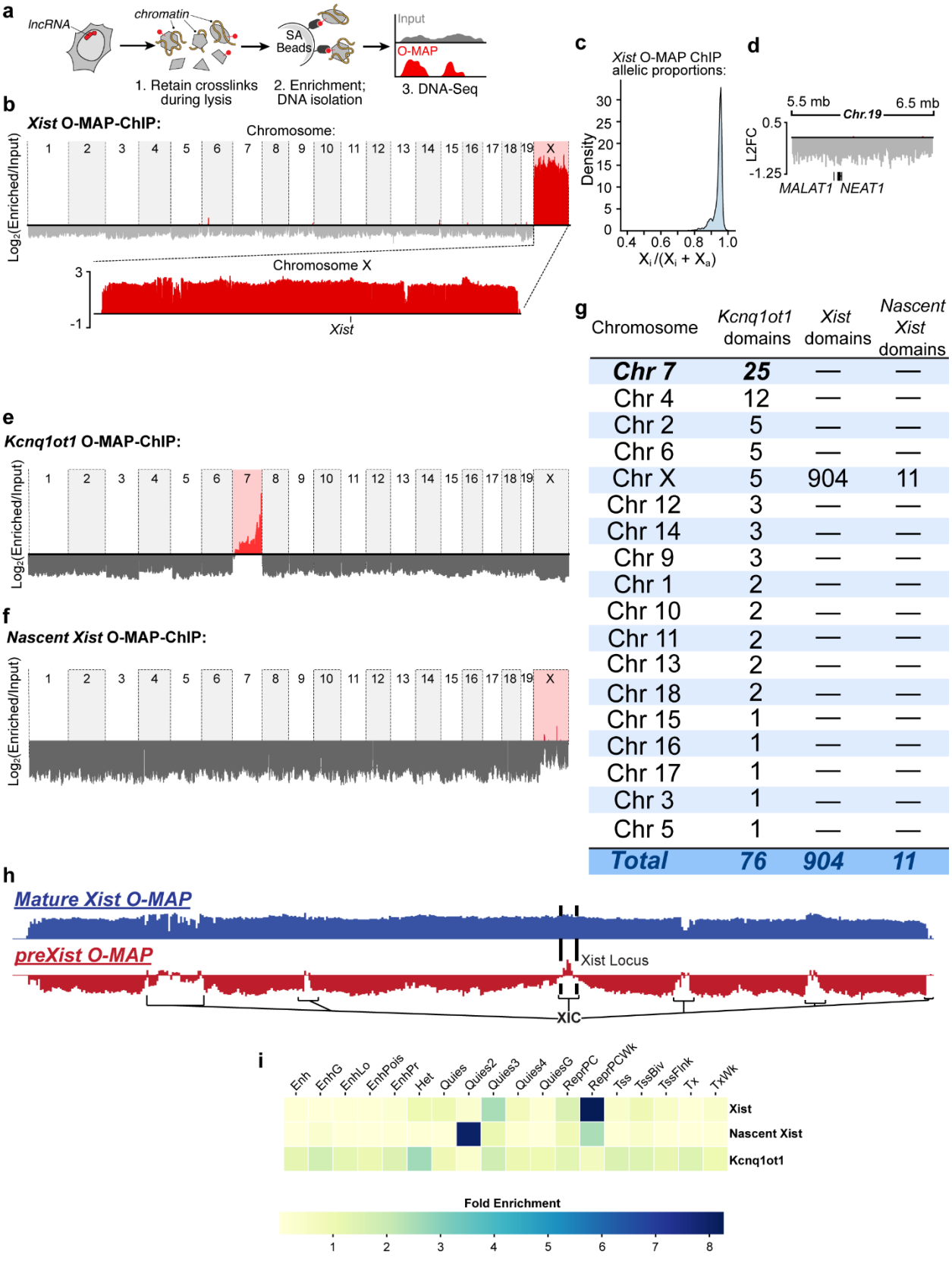



Figure 2.1. **O-MAP precisely targets gene silencing lncRNAs and precursor lncRNAs. a,** lncRNAs targeted. Kcnq1ot1 coats a cluster of genes on the terminal end of paternal chromosome 7, condensing it into the Kcnq1ot1 heterochromatin domain (left). Xist coats the entire inactive X chromosome (Xi), condensing it into the inactive Barr body. From the Xi, nascent Xist is transcribed and spreads across the Xi. O-MAP was targeted to nascent Xist by via Xist's introns. **b,** O-MAP Probe Validation Assay applied to Kcnq1ot1 and Xist. Note conspicuous overlap between O-MAP (magenta) and RNA-FISH (green) signals (left). **c,** O-MAP at nascent transcripts probes subnuclear neighborhoods. Although mature Xist coats the entire inactive X-chromosome

(Xi), nascent Xist transcripts uniquely denote the X-inactivation center (XCI). O-MAP targeting Xist introns induces biotinylation at confined foci within the Xi “cloud.” All scale bars: 20 μm.




| Chromosome | *Kcnq1ot1* domains | *Xist* domains | *Nascent Xist* domains |
|---|---|---|---|
| ***Chr 7*** | ***25*** | — | — |
| Chr 4 | 12 | — | — |
| Chr 2 | 5 | — | — |
| Chr 6 | 5 | — | — |
| Chr X | 5 | 904 | 11 |
| Chr 12 | 3 | — | — |
| Chr 14 | 3 | — | — |
| Chr 9 | 3 | — | — |
| Chr 1 | 2 | — | — |
| Chr 10 | 2 | — | — |
| Chr 11 | 2 | — | — |
| Chr 13 | 2 | — | — |
| Chr 18 | 2 | — | — |
| Chr 15 | 1 | — | — |
| Chr 16 | 1 | — | — |
| Chr 17 | 1 | — | — |
| Chr 3 | 1 | — | — |
| Chr 5 | 1 | — | — |
| ***Total*** | ***76*** | ***904*** | ***11*** |

Figure 2.2. **Xist localizes to the X chromosome while the imprinting RNA Kcnq1ot1 touches multiple genomic loci. a,** Approach. O-MAP was targeted to a lncRNA. Biotinylated material was natively pulled down from crude lysate with streptavidin beads. DNA was then isolated and sequenced. **b,** Xist O-MAP-ChIP predominantly labels the X-chromosome (right, and inset). Data for the entire mouse genome are shown; Xist genomic locus is noted below. **c,** Xist O-MAP-ChIP is specific to the inactive X-chromosome (Xi). Histogram of Allelic proportions for ChIP data, quantified using SNPs specific to the Xi and Xa77. **d,** Putative interactions between autosomal loci and the Xi. The kcnq1ot1 locus—but not MALAT1 and NEAT1 loci—appear enriched in Xist O-MAP-ChIP data. L2FC: log2(Fold change, Enriched/Input). **e,** Kcnq1ot1 most strongly labels chromosome 7. Signal is strongest at the Kcnq1ot1 locus on chromosome 7 and decreases distally along chromosome 7. **f,** Nascent Xist localizes to specific regions on chromosome X. Data for the entire mouse genome are shown. **g,** Table summarizing all observed lncRNA chromatin domains. **h,** Mature Xist is spread across the Xi while nascent Xist localizes to discrete regions on the Xi. Data are log2(Fold change, Enriched/Input). **i,** ChromHMM analysis reveals differential chromatin signatures amongst the targeted lncRNAs. Note Xist primarily enriches for weak polycomb repressive signatures, nascent Xist primarily enriches for quiescent 2 signatures, and Kcnq1ot1 primairly enriches for heterochromatin signatures.

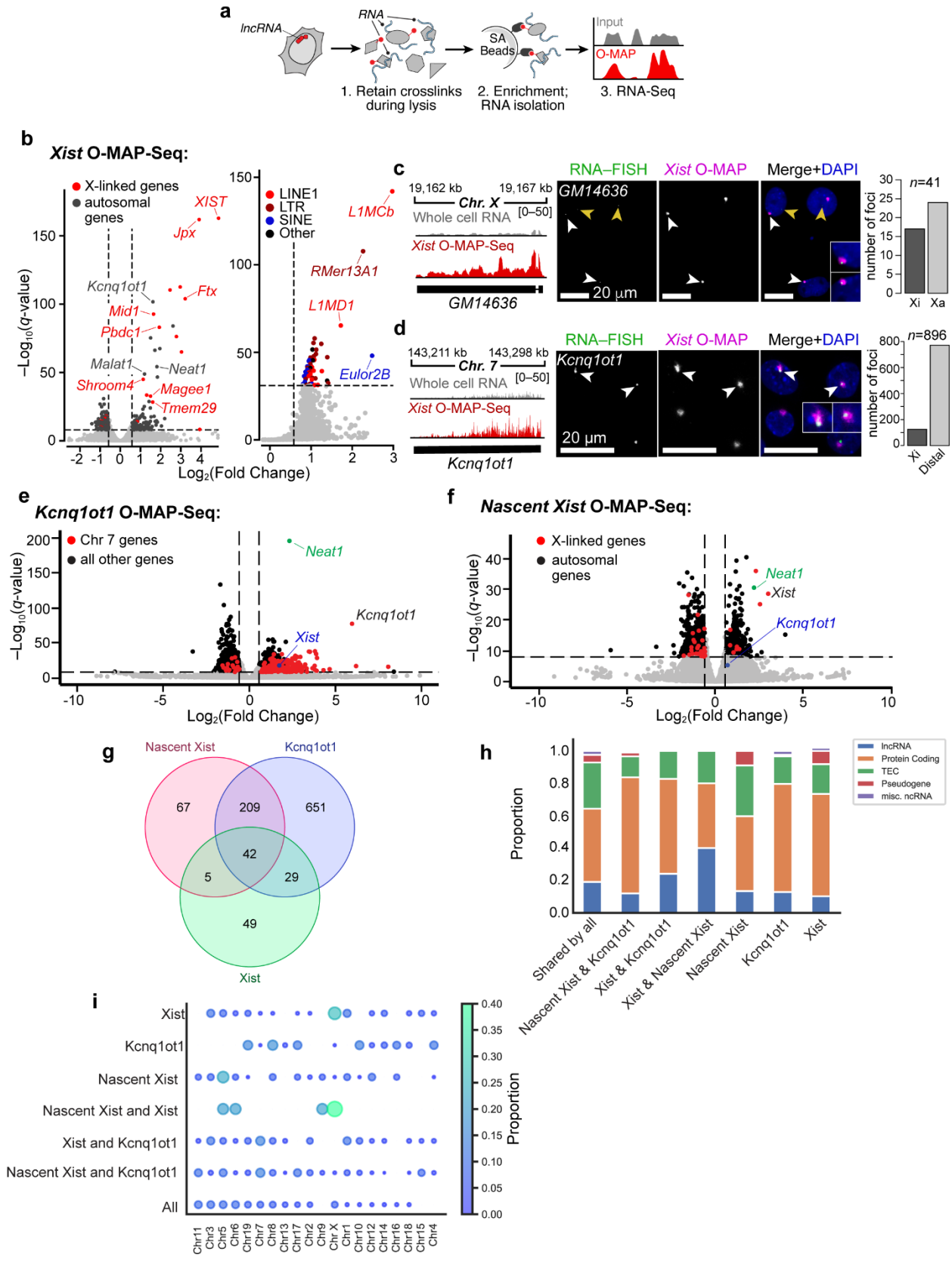

a
lncRNA
RNA
SA Beads
Input
O-MAP
1. Retain crosslinks during lysis
2. Enrichment; RNA isolation
3. RNA-Seq
b
Xist O-MAP-Seq:
X-linked genes
autosomal genes
XIST
Jpx
Ftx
Kcnq1ot1
Mid1
Pbdc1
Malat1
Neat1
Shroom4
Magee1
Tmem29
LINE1
LTR
SINE
Other
L1MCb
RMer13A1
L1MD1
Eulor2B
−Log10(q-value)
Log2(Fold Change)
c
19,162 kb
19,167 kb
Chr. X
Whole cell RNA
[0–50]
Xist O-MAP-Seq
GM14636
RNA–FISH
Xist O-MAP
Merge+DAPI
20 μm
number of foci
n=41
Xi
Xa
d
143,211 kb
143,298 kb
Chr. 7
Whole cell RNA
[0–50]
Xist O-MAP-Seq
Kcnq1ot1
RNA–FISH
Xist O-MAP
Merge+DAPI
Kcnq1ot1
20 μm
number of foci
n=896
Xi
Distal
e
Kcnq1ot1 O-MAP-Seq:
Chr 7 genes
all other genes
Neat1
Kcnq1ot1
Xist
−Log10(q-value)
Log2(Fold Change)
f
Nascent Xist O-MAP-Seq:
X-linked genes
autosomal genes
Neat1
Xist
Kcnq1ot1
−Log10(q-value)
Log2(Fold Change)
g
Nascent Xist
Kcnq1ot1
Xist
67
209
651
42
5
29
49
h
Proportion
lncRNA
Protein Coding
TEC
Pseudogene
misc. ncRNA
Shared by all
Nascent Xist & Kcnq1ot1
Xist & Kcnq1ot1
Xist & Nascent Xist
Nascent Xist
Kcnq1ot1
Xist
i
Xist
Kcnq1ot1
Nascent Xist
Nascent Xist and Xist
Xist and Kcnq1ot1
Nascent Xist and Kcnq1ot1
All
Proportion
Chr11 Chr3 Chr5 Chr6 Chr19 Chr7 Chr8 Chr13 Chr17 Chr2 Chr9 Chr X Chr1 Chr10 Chr12 Chr14 Chr16 Chr18 Chr15 Chr4

Figure 2.3. **Gene silencing RNAs interact with shared and discrete transcripts. a,** Approach. O-MAP was targeted to a lncRNA. Biotinylated material was natively pulled down from crude lysate with streptavidin beads. DNA was then isolated and sequenced. **b,** Volcano plots of Xist O-MAP-Seq data. Left: enrichment for X-linked transcripts that escape X-chromosome inactivation (XCI, red) and autosomal transcripts, including several chromatin-regulatory lncRNAs (indicated). Right: enrichment for several classes of X-linked transposable elements. (enriched data are L2FC > 0.58, padj < $10^{-9}$). **c,** Gm14636 is a novel XCI-escape gene. Left: Gm14636 enrichment by O-MAP-Seq. Middle: co-localization of Gm14636 RNA-FISH and Xist O-MAP Note penetrant mono- or bi-allelic expression from the Xa (gold arrows) and/or Xi (white arrowheads), quantified on the right. **d,** Kcnq1ot1 localizes near the Xi. Panel arrangement parallels that of **c**. **e,** Volcano plot of Kcnq1ot1 O-MAP-Seq data. Enrichment for chromosome 7 transcripts (red) and autosomal transcripts, including several chromatin-regulatory lncRNAs (indicated). (enriched data are L2FC > 0.58, padj < $10^{-9}$). **f,** Volcano plot of nascent Xist O-MAP-Seq data. Enrichment for X-linked transcripts (red) and autosomal transcripts, including several chromatin-regulatory lncRNAs (indicated). (enriched data are L2FC > 0.58, padj < $10^{-9}$). **g,** Venn diagram indicating shared and discrete interacting transcripts amongst the targeted lncRNAs. **h,** Stacked bar plot indicating the proportion of RNA biotypes constituting the shared and discrete interacting transcripts amongst the targeted lncRNAs. **i,** Interacting RNA origin for shared and discrete interacting transcripts amongst the targeted lncRNAs.

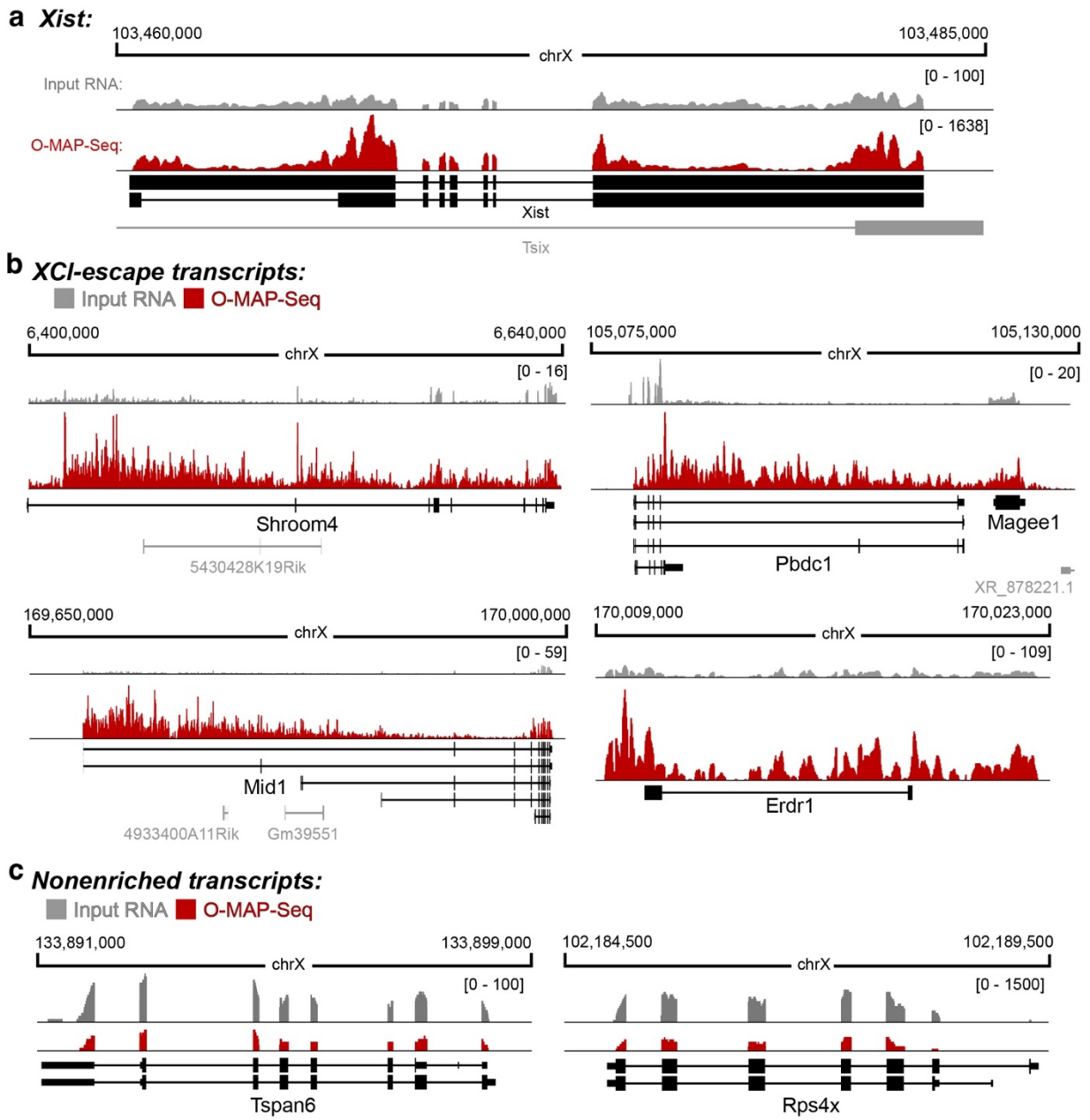



Figure 2.4. **Xist O-MAP-Seq enriches nascent transcripts of XCI-escape genes. a,** Enrichment of the Xist gene itself. Note different scales for input RNA and O-MAP-Seq tracks. The lack of intronic reads suggests that O-MAP-Seq has predominantly targeted and captured the mature Xist transcript. The absence of reads mapping to the antisense noncoding RNA TsiX, a lncRNA that is monoallelically expressed from the Xa (gray), confirms that O-MAP-Seq is precisely labeling the Xi. **b,** Enriched XCI-escape genes appear to be nascent transcripts. In all cases, read densities for both Input RNA (gray) and O-MAP-Seq (red) are shown, using matched scales. Note prominent intronic read density for all XCI-escape genes (Shroom4, Pbdc1, Magee1, Mid1, Erdr1). Transcript structures denoting the most prominent isoforms are displayed; other nearby genes not known to escape XCI are denoted in gray. **c,** Xist O-MAP-Seq does not appear to preferentially capture nascent transcripts of non-XCI-escape genes. Two examples (Tspan6 and Rps4x) are shown, neither of which was enriched by Xist O-MAP-Seq. Note the absence of prominent intronic read

density, suggesting that the intronic signatures observed in (b) are not general artifacts of the O-MAP-Seq pipeline.

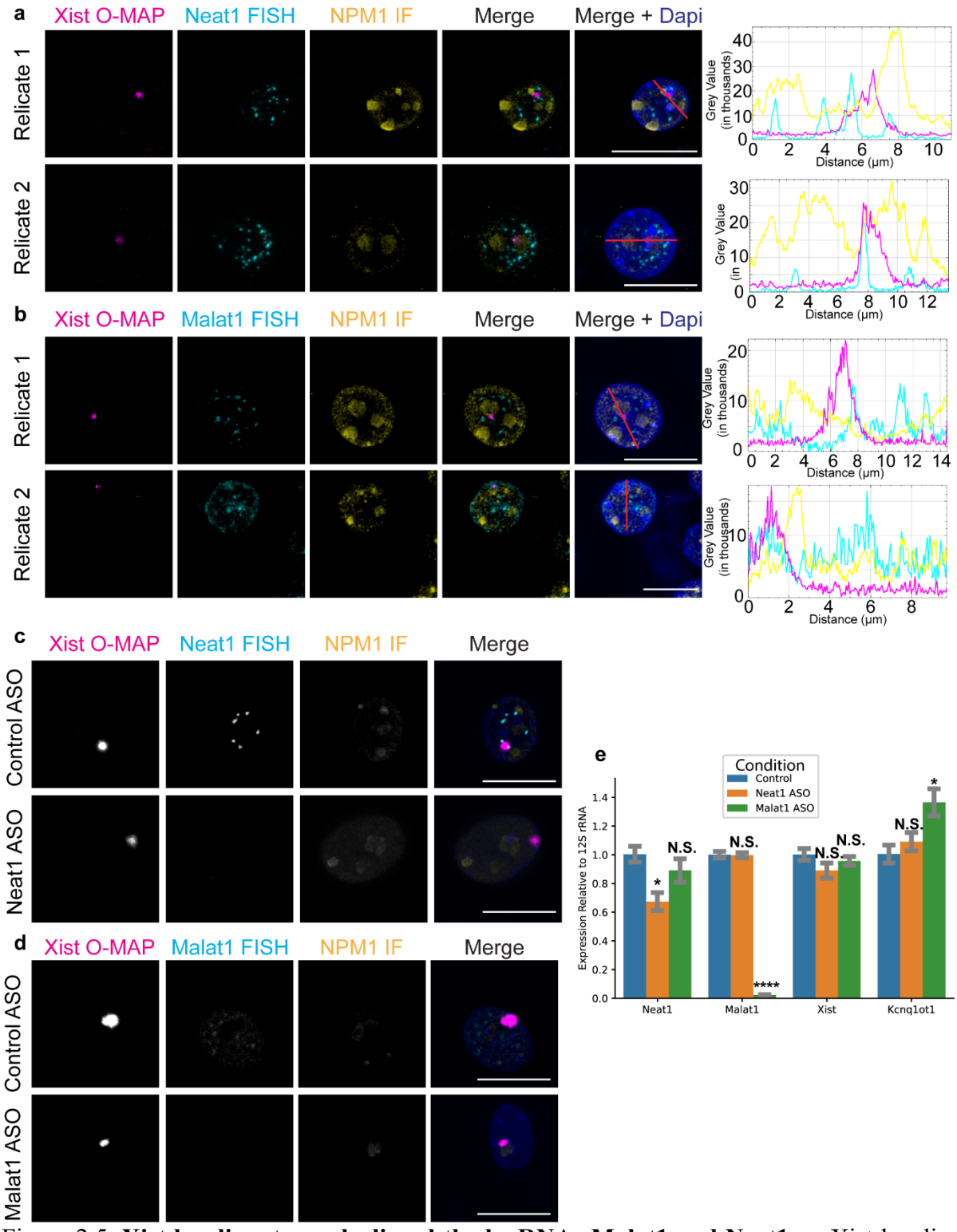



Figure 2.5. **Xist localizes to nucleoli and the lncRNAs Malat1 and Neat1. a,** Xist localizes to nucleoli and Neat1. Left: Confocal imaging for Xist O-MAP, Neat1 RNA FISH, and NPM1 immunofluorescence. A composite image with and without Dapi signal is also shown. Right: Line plot indicating grey value intensity for Xist O-MAP (magenta), Neat1 RNA FISH (cyan), and NPM1 immunofluorescence (yellow). Note that Xist and Neat1 localize to the periphery of

nucleoli. **b,** Xist localizes to nucleoli and Malat1. Left: Confocal imaging for Xist O-MAP, Malat1 RNA FISH, and NPM1 immunofluorescence. A composite image with and without Dapi signal is also shown. Right: Line plot indicating grey value intensity for Xist O-MAP (magenta), Malat1 RNA FISH (cyan), and NPM1 immunofluorescence (yellow). Note that Xist and Malat1 localize to the periphery of nucleoli. **c,** Antisense oligonucleotide (ASO) directed against of Neat1 effectively knocks down Neat1 levels via RNA-FISH (bottom) compared to control ASO (top). **d,** Antisense oligonucleotide (ASO) directed against of Malat1 effectively knocks down Malat1 levels via RNA-FISH (bottom) compared to control ASO (top). **e,** qPCR validation that Neat1 and Malat1 ASOs effectively knockdown their target RNA without affecting expression of the other lncRNA or Xist. Malat1 ASO knockdown results in a slight increase in Kcnq1ot1 expression. Data are biological quadruplicates run in technical quadruplicates. Statistical significance by paired 2-tailed student's t-test. All scale bars are 20 μm, and red lines indicate where line plots originate.

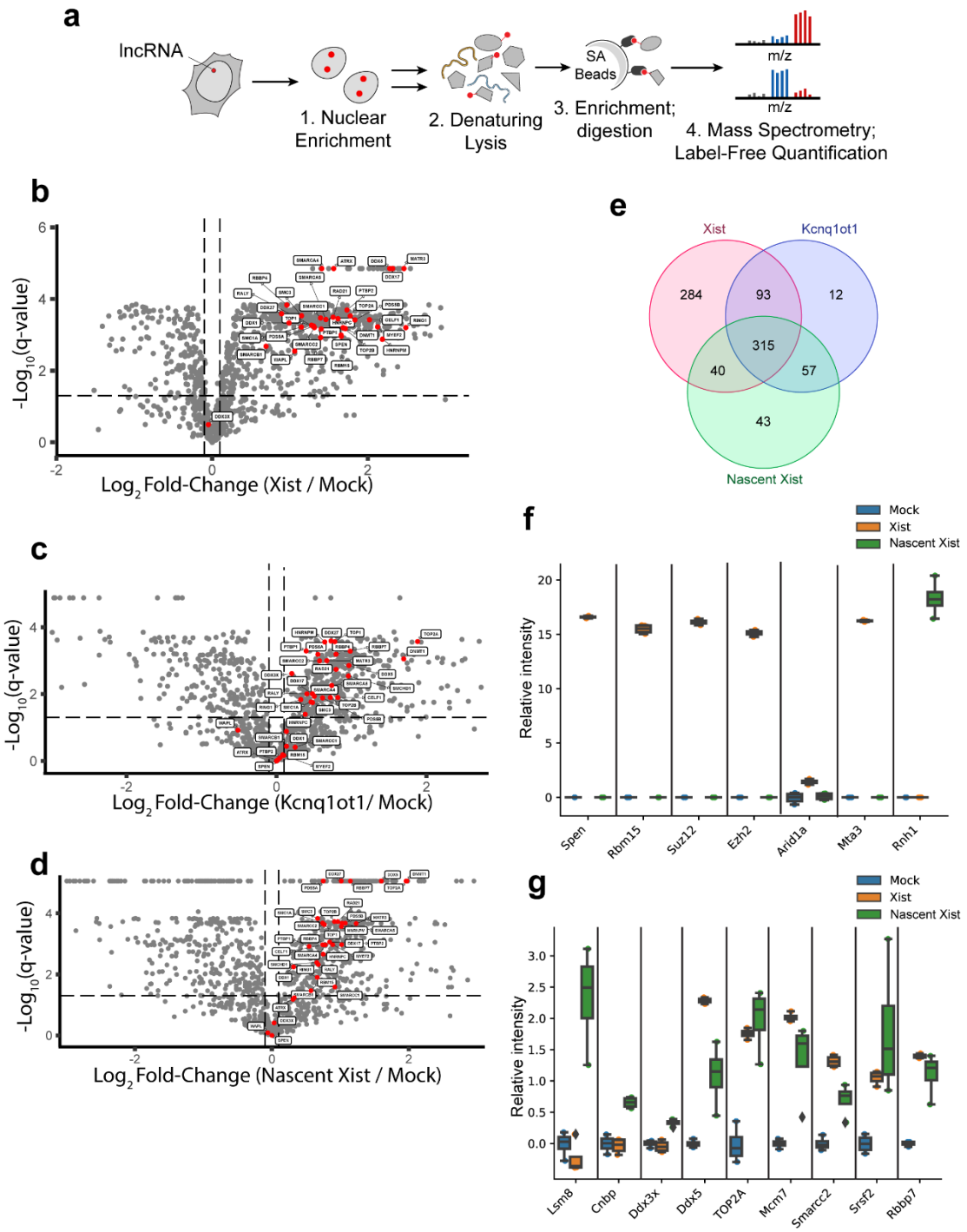



Figure 2.6. **O-MAP unifies known Xist proteomes and demonstrates a shared pool of proteins amongst gene-silencing lncRNAs. a,** Approach. O-MAP was targeted to a lncRNA. Nuclei were isolated and underwent a denaturing lysis procedure. Biotinylated material was enriched and proteins were analyzed via label-free quantification mass spectrometry. **b-d,** Volcano plot summarizing lncRNA O-MAP-MS hits specific for Xist **(b, right)**, Kcnq1ot1 **(c, right)**, and nascent Xist **(d, right)**. In each plot, previously identified Xist protein interactors are highlighted in red (Markaki et al., 2021; Minajigi et al., 2015). **e,** Three-way Venn diagram depicting O-MAP

enriched protein intersections amongst all targeted lncRNAs. **f-g,** Relative intensity plots of Xist-specific, nascent Xist specific, and shared Xist/nascent Xist enriched proteins. Relative intensity was calculated for each individual replicate (n = 4) by normalizing each replicate to the mean of the mock replicates for each respective protein.

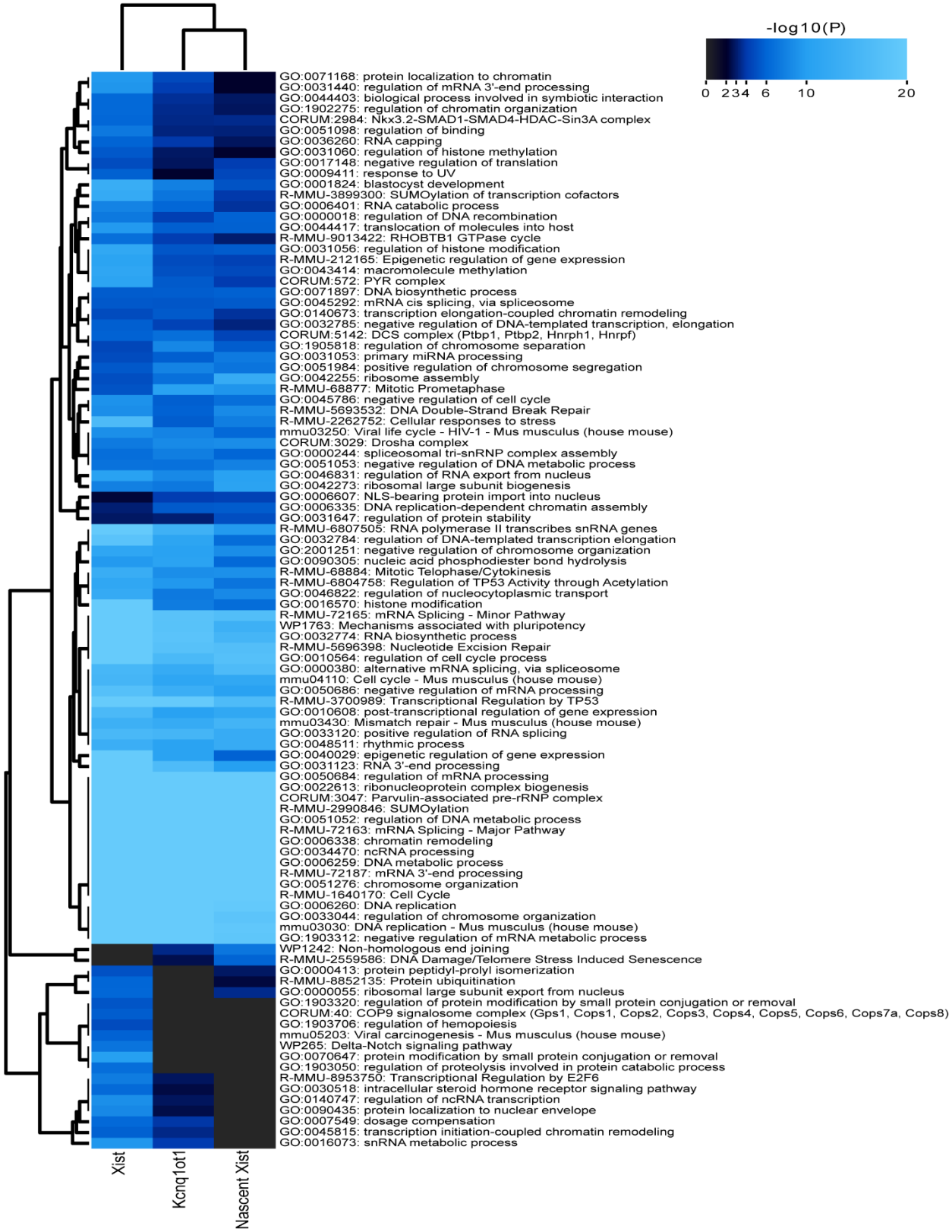



Figure 2.7. **Gene silencing lncRNA proteomes demonstrate shared and unique gene**

**ontologies.** Enriched terms across enriched proteomes for O-MAP targeted Xist, Kcnq1ot1, and nascent Xist. Heatmap is labeled according to 0.05 FDR-corrected p-value.

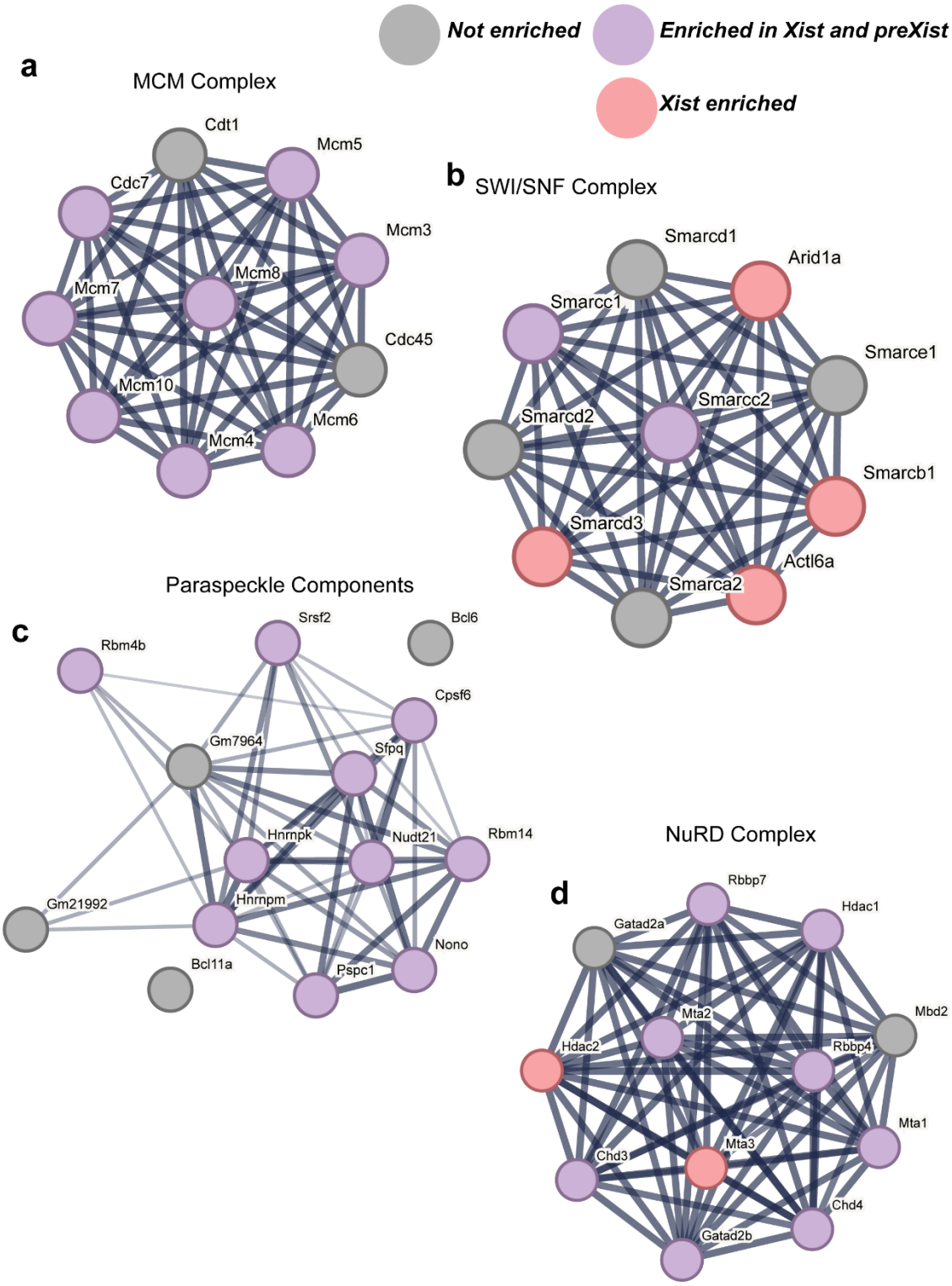



Figure 2.8. **X-chromosome inactivation lncRNA proteomes demonstrate shared and unique protein-protein interaction networks. a-f,** STRING network depictions of selected complexes.

Grey circles represent complex proteins that are not enriched in Xist nor nascent Xist proteomes, red circles represent complex proteins enriched uniquely in the Xist proteome compared to nascent Xist, and purple circles represent complex proteins enriched in both Xist and nascent Xist proteomes.

# Chapter 3. NASCENT TRANSCRIPT-TARGETED PROXIMITY OMICS ENABLES HIGH-RESOLUTION MICRODISSECTION OF A PRE-MRNA-SCAFFOLDED NUCLEAR REGULATORY COMPARTMENT AND REVEALS NEW MOLECULAR PLAYERS IN DILATED CARDIOMYOPATHY

## 3.1 ABSTRACT

Dilated cardiomyopathy (DCM) is a leading cause of global heart failure. A particularly aggressive form of DCM — characterized by conduction system disorders, ventricular arrhythmia, and rapid progression to heart failure — is caused by mutations in RNA Binding Motif protein 20 (RBM20). It has been shown that these RBM20 mutations lead to aberrant alternative splicing of key cardiac genes. However, the mechanisms by which mutant RBM20 dysregulates alternative splicing and drives DCM pathogenesis remain unclear. Recently, it has been discovered in cardiomyocytes (CMs) that RBM20 forms discrete subnuclear foci that are scaffolded on titin pre-mRNA (pre-TTN). These RBM20 foci organize chromatin architecture, co-localizing many essential cardiac genes and coordinating their expression and splicing. Disruption of RBM20 foci remodels this architecture and dysregulates cardiac gene splicing, similar to what is observed in RBM20 DCM. This suggests that RBM20 foci may be critical to both CM development and DCM etiology. But, many of the factors that comprise RBM20 foci, and the genomic loci they regulate, are unknown. Elucidating these components will be crucial for deciphering the mechanism of RBM20 DCM. However, revealing these factors is challenging since RNA-scaffolded structures like RBM20 foci are too dynamic and fragile to be purified and analyzed by current methods. Here, we overcome this challenge using Oligonucleotide-directed proximity interactome MAPping (O-MAP) to functionally dissect RBM20 foci in a cellular model of RBM20 DCM. O-MAP identified the proteins, RNAs, and genomic loci near pre-TTN RNA in WT and ΔRBM20 cardiomyocytes.

Further analyses of differentially enriched material revealed key differences in epigenetic signatures, alternative splicing, and RNA-binding protein composition. Moreover, we reveal the downregulation of a key protein involved in mitochondrial reactive oxygen species (mROS) regulation and functionally demonstrate impaired mROS handling in the RBM20 DCM phenotype. Overall, this work provides transformative insight into the mechanisms by which RNA-scaffolded structures drive cardiac development, enables detailed functional studies, and reveals new biomarkers and therapeutic targets for RBM20 DCM.

## 3.2 INTRODUCTION

Nuclear organization is intricately linked to cell differentiation, development, and function. The three-dimensional arrangement of chromatin, the positioning of nuclear bodies, and the distribution of transcription factors and epigenetic marks all contribute to the regulation of gene expression and ultimately determine cell fate (Cremer et al., 2006; Hnisz et al., 2016; Misteli, 2004, 2007; van Steensel & Belmont, 2017). Given the importance of nuclear organization in regulating gene expression and determining cell fate, subnuclear bodies such as nuclear speckles have attracted significant attention due to their involvement in a wide range of nuclear processes (Dundr, 2012; Spector & Lamond, 2011). Nuclear speckles are subnuclear bodies located in the interchromatin space of the nucleus and are involved in diverse nuclear functions, including transcriptional regulation, RNA processing, and DNA repair (Galganski et al., 2017; Mao et al., 2011; Spector & Lamond, 2011). The composition of nuclear speckles varies depending on cellular context and can dynamically change in response to environmental stimuli or cellular signaling (Galganski et al., 2017; Spector & Lamond, 2011). Nuclear speckles also have an important role in splicing regulation (Misteli, 2000; Spector & Lamond, 2011). It has been demonstrated that nuclear speckles provide a platform for the concentration and assembly of splicing factors and pre-

mRNA, thereby facilitating the splicing process (Y. Chen & Belmont, 2019; Shopland et al., 2003). This is supported by the observation that many splicing factors are enriched in nuclear speckles, including members of the serine/arginine-rich protein family and heterogeneous nuclear ribonucleoproteins (hnRNPs) (Geuens et al., 2016; Hall et al., 2006; Howard & Sanford, 2015).

Recent studies have revealed the existence of RNA-binding motif protein 20 (RBM20)-containing nuclear speckles in cardiomyocytes, which are involved in regulating the splicing of specific cardiac genes (Bertero et al., 2019; Brauch et al., 2009; Fenix et al., 2021; W. Guo et al., 2012; Watanabe et al., 2018). RBM20 is a cardiac-specific splicing factor that plays a role in the regulation of alternative splicing of pre-mRNA (W. Guo et al., 2012; Maatz et al., 2014). The unique speckles RBM20 localizes to are scaffolded by nascent titin (TTN) pre-mRNA, the second largest human gene, and which is essential for the proper assembly and function of the sarcomere in striated muscle (Granzier & Labeit, 2004; W. Guo et al., 2012). The Murry group has previously demonstrated that RBM20 nuclear speckles recruit genomic loci encoding cardiac genes that exhibit alternative splicing events resulting in the production of their appropriate cardiac isoforms (Bertero et al., 2019). The identification of these specialized RBM20 nuclear speckles highlights the complex regulation of gene expression in the heart and suggests that cell-type specific splicing speckles may be a common feature of gene regulation and transcriptional modification in other tissues (Galganski et al., 2017; Shopland et al., 2003).

It is firmly established that RBM20 mutations lead to a loss of RBM20 speckles and the development of dilated cardiomyopathy (DCM) (Beraldi et al., 2014; Brauch et al., 2009; W. Guo et al., 2012). DCM is a complex cardiac disorder characterized by progressive ventricular dilation and impaired systolic function (Hershberger et al., 2010; McNally & Mestroni, 2017). The genetic etiology of DCM is multifactorial, with mutations in the RBM20 gene being implicated in up to

3% of familial and sporadic cases (Brauch et al., 2009; Haas et al., 2015; Refaat et al., 2012). Loss of RBM20 function has been shown to result in altered splicing patterns of TTN and other sarcomeric genes, leading to the disruption of the structural integrity of the cardiac sarcomere and subsequent development of DCM (W. Guo et al., 2012, 2018). Recent studies have highlighted the potential of RBM20 as a therapeutic target for DCM, with emerging evidence suggesting that restoration of RBM20 function through upregulation or base-editing could improve cardiac function and prevent disease progression (Fernández-Ruiz, 2020; Nishiyama et al., 2022). Further research of RBM20 speckles is needed to fully understand the molecular mechanisms underlying RBM20-associated DCM and to develop effective therapeutic strategies for this debilitating disorder. Despite the crucial role of cardiac RBM20 speckles in the regulation of gene expression and alternative splicing, the characterization of these specific nuclear speckles remains challenging due to general dynamic and heterogeneous nature of nuclear speckles, and the difficulty to purify unique nuclear speckles through canonical biochemical approaches (Dopie et al., 2020; Galganski et al., 2017; Mintz et al., 1999; Saitoh et al., 2004; Spector & Lamond, 2011).

Here, using oligonucleotide-directed proximity-interactome mapping (O-MAP) (Tsue et al., 2023), we characterized the DNA-loci, RNAs, and proteins present at cardiac-specific RBM20 speckles in wild-type (WT) human cardiomyocytes, as well as at the nascent TTN transcript in mutant cells lacking RBM20 (ΔRBM20). In O-MAP, cells are chemically fixed and pools of antisense DNA probes are hybridized to a target RNA, similar to high-resolution RNA-FISH (Kishi et al., 2019). These probes then recruit a secondary oligo that is linked to Horseradish Peroxidase (HRP). Upon addition of its biotin-tyramide substrate, HRP generates freely diffusing biotin radicals that affinity-tag nearby proteins, RNAs, and DNA loci, enabling their isolation by simple streptavidin pulldown (Y. Chen et al., 2018; Rhee et al., 2013). In WT cells, O-MAP

revealed that RBM20 speckles localized near bivalent DNA domains, and encompassed transcripts encoding for calcium regulatory proteins, proteins involved in chromosome organization and a specific cohort of RNA-binding and splicing factors. In ΔRBM20 cells, we observed a complete loss of all sites of *trans* DNA interactions, enrichment for transcripts encoding for proteins involved in DCM etiology, and enrichment for core speckle proteins. Alternative splicing analyses revealed that RBM20 speckles and ΔRBM20 nascent TTN speckles are abundant in cardiac specific transcripts that undergo alternative splicing. In addition, leveraging the power of O-MAP, we uncovered a significant impairment in the mediation of mitochondrial reactive oxygen species (mROS) in the absence of RBM20. These findings suggest a previously unknown role for RBM20 in regulating the redox balance of cardiac cells through the modulation of mROS levels. Overall, our study not only sheds light on the molecular mechanisms underlying RBM20-DCM but also provides a streamlined approach for characterizing the complete interactome of cell-type-specific subnuclear compartments.

## 3.3 RESULTS

**Nascent TTN O-MAP Design and Validation**

Canonical protein directed approaches to capture nuclear RBM20-interactions (e.g. immunoprecipitation, live-cell proximity-biotinylation) are ill equipped to probe RBM20-foci. Of note, a sizable fraction of RBM20 is localized in a diffuse nucleoplasmic compartment outside of the foci, and presumably form distinct molecular interactions that would confound analysis (Bertero et al., 2019; Fenix et al., 2021). Moreover, RBM20 foci are too dynamic and fragile to be isolated and analyzed by current biochemical approaches, requiring new tools that can identify these critical regulatory factors within their native context. To overcome this challenge, we employed O-MAP targeted to the introns of the nascent TTN pre-mRNA, since these intronic

sequences are exclusively localized at the site of TTN transcription upon which RBM20 foci are organized (S. Li et al., 2013) (Fig 3.1, a). Using OligoMiner, 110 probes targeting fourteen introns near RBM20 repressed exons were designed (Fig 3.1, b) (S. Li et al., 2013). Using these probes, we applied O-MAP in RUES2 ES-derived human cardiomyocytes (CMs). Encouragingly, O-MAP induced precisely defined biotinylation in WT cells (Fig 3.1, c, top). In CMs bearing a deletion in the TTN promoter (Bertero et al., 2019) there was no detectable O-MAP biotinylation nor immunofluorescence for TTN protein itself, as expected (Fig 3.1, c, bottom). To test the specificity of the probe pool for TTN pre-mRNA, the pool was split in half to allow simultaneous O-MAP and RNA-FISH imaging (Fig. 1.1, b) (Tsue, Kania et al., 2023). Excitingly, O-MAP in WT cardiomyocytes induced defined biotinylation exclusively near nascent TTN RNA-FISH, with no detectable spatial off-targeting (Fig 3.1, d). This strongly suggests that our O-MAP probe pool is precisely and specifically targeting nascent *TTN* transcripts. Having assessed the specificity of O-MAP for nascent TTN RNA, we sought to determine the colocalization of O-MAP for RBM20 foci. In WT cardiomyocytes, O-MAP signal precisely biotinylated at RBM20 foci, as assessed by RBM20 immunofluorescence (Fig 3.1, e, top). Object based colocalization found the overlap of RBM20 and nascent TTN O-MAP to be statistically significant ($p < 0.01$ (Fig 3.1, f)). In ΔRBM20 cells, O-MAP resulted in signal since TTN is still transcribed, but no RBM20 protein was detected by immunofluorescence (Fig 3.1, e, bottom). Lastly, as proof-of-principle, combined O-MAP, RNA-FISH, and immunofluorescence demonstrate the co-localization of nascent TTN O-MAP, RNA-FISH, and RBM20, demonstrating that pre-*TTN* O-MAP specifically biotinylated RBM20 factories, while avoiding the remaining nucleoplasmic pool of RBM20 (Fig 3.1, g).

**Nascent TTN O-MAP-ChIP identifies *cis*- and *trans*- chromatin interactions with pre-*TTN* foci**

Having validated O-MAP's ability to precisely biotinylate RBM20 foci, we applied O-MAP ChIP to systematically identify chromatin contacts made by the pre-*TTN* factory (Fig 3.2, a). Chromatin domains associating with RBM20 foci in WT CMs or with nascent TTN in ΔRBM20 CMs were sequenced and mapped (average of ~54 M reads; 96.88% alignment). We observed substantial enrichment of megabase-scale domains along chromosome 2 (which harbors the TTN gene), largely mirroring the "A/B" subnuclear compartmentalization previously established for this cell line (Fig 3.2, b) (Bertero, 2019). Pearson correlation of signal along chromosome 2 demonstrated a moderate correlation (Average r = 0.41) between TTN pre-mRNA O-MAP *cis*-interacting domains (CIDs) and DNase Hi-C; mock O-MAP showed no significant correlation (Average r = -0.24). We also observed 19 *trans*-interacting domains (TIDs) across nine chromosomes (Fig 3.2, c and d). Strikingly, all *cis* and *trans* interactions are ablated upon RBM20 loss. In ΔRBM20 cells, the only remaining enriched DNA locus in contains the TTN locus itself (Fig 3.3, e). This strongly suggests that the architecture of pre-TTN factory depends not only on nascent *TTN* (Bertero et al., 2019), but on its interactions with the RBM20 protein.

We performed DNA-FISH to validate putative TID 3-dimensional localization near the TTN locus. *Cis* and *trans* enriched DNA loci were validated by performing DNA-FISH for putative TIDs and *cis* interactions in both WT and ΔRBM20 CMs. In total, eight *trans* loci and one cis loci were examined by DNA-FISH (Fig. 3.2 & Fig. 3.9). A *cis* negative control, a *trans* negative control, and DNA-FISH for the TTN locus were also utilized (Fig. 3.2 & Fig. 3.9). Resulting DNA-FISH for TTN colocalized with RBM20, demonstrating the specificity of the probe pool (Fig. 3.9, b). Strikingly, of the loci tested, each *cis* and all but one trans loci examined

were significantly closer to the TTN locus than were negative controls, highlighting the sensitivity of TTN pre-mRNA O-MAP to identify nearby chromatin (Fig 3.2, f-i). Of note, the *cis* negative control is also located farther away from the TTN locus in ΔRBM20 cells (Fig. 3.9, a). This change was not captured by O-MAP likely because the WT interaction occurred beyond the O-MAP biotinylation radius, meaning the increased distance observed in ΔRBM20 cells would also not be captured by O-MAP.

**pre-*TTN* Trans-Interacting chromatin Domains are enriched for transcriptionally silent, bivalent chromatin**

Having validated TIDs at RBM20 foci, we next sought to characterize them. To examine the epigenomic signatures of *cis* and *trans* interactions, we employed the use of ChromHMM using an epigenome file for RUES2 CMs. *Cis* interactions enriched for genic enhancers, while TIDs enriched heavily for bivalent enhancers and mildly for repressed polycomb and bivalent transcription start sites (Fig 3.3, a). Bivalent enhancers are known to contain transcription factors crucial for embryonic stem cell differentiation (Blanco et al., 2020). Strikingly, loss of RBM20 led to an increase in transcription of TID-enclosed genes (Fig. 3.3, b). Bivalent domains are denoted as having both H3K27me3 and H3K4me3 marks. Immunofluorescence for H3K4me3 combined with nascent TTN O-MAP demonstrated a similar degree of H3K4me3 proximity to TTN O-MAP between WT and ΔRBM20 CMs (Fig. 3.3, c). Intriguingly, both phenotypes demonstrated bright H3K4me3 puncta directly overlapping with nascent TTN O-MAP in roughly half their populations, indicating intense transcriptional activity near the TTN locus (Fig 3.3, c). H3K27me3 immunofluorescence combined with nascent TTN O-MAP also demonstrated a similar degree of proximity, in both WT and ΔRBM20 CMs (Fig 3.3, d). Hence, although RBM20 is required to organize CID and TID architecture around the actively transcribed *TTN* locus, it appears to have

no influence on the local compartmentalization of modified histone domains near that locus. Of note, the center of pre-*TTN* O-MAP signal appears to preside in a zone of depletion between peaks in the H3K27me3 signal (Fig 3.3, d). Taken together, this indicates the nascent TTN transcript heavily colocalizes with H3K4me3 at its core, with sites of H3K27me3 surrounding in close proximity.

Bivalent enhancers frequently regulate core transcription factors that oversee important developmental or tissue differentiation events (Sha & Boyer, 2008). We therefore examined the transcription factor genes harbored in our pre-*TTN* TIDs, by GO-term analysis, to assess if they were related to cardiomyocyte differentiation. Strikingly, the top ten biological process GO-terms were all related to cell differentiation and cardiomyogenesis (q-value < 0.05) (Fig 3.3, e). Intriguingly, cardiac conduction system development and ventricle development terms were also enriched; these are frequently dysregulated in RBM20-DCM (Koelemen et al., 2021; Wyles et al., 2016). When parsing enriched GO-terms so that each term contains no more than 30% similarity to another term, cardiac chamber morphogenesis, DNA-templated transcription, neural crest differentiation, and lung development were the top significant terms (q-value < 0.05). Ontology with DisGeNET highlighted hypoplastic heart defect and ventricular septal defect as being in the top ten enriched disease pathways (q-value < 0.05) (Fig. 3.9, c). We next examined if the TID-enclosed transcription factors implicated in these analyses are dysregulated upon RBM20 loss. Strikingly, we observed by qPCR that the heart-related transcription factors demonstrated a significant upregulation of MSX2, FOXF1, FOXL1, and FOXC2 upon RBM20 deletion (Fig 3.3, f). Using a male model of RBM20-DCM (Fenix et al., 2021) we found these same transcription factors were also upregulated in a mutation dose-dependent manner. No-template and no-reverse-transcriptase controls resulted in no detectable signal past 60 amplification cycles.

Correspondingly, an increase in these transcription factors in ΔRBM20 cells leads to upregulation of the transcripts they regulate, as exemplified by FOXF1- and FOXL1-regulated genes (Fig. 3.3, g and h).

Collectively, these data suggest that RBM20 interacts with pre-TTN mRNA to transcriptionally repress chromatin domains harboring genes essential for cardiomyogenesis. Since these cells are terminally differentiated cardiomyocytes, we speculate this regulatory mechanism enables the downregulation of cardiomyocyte-specific differentiation pathways that would be redundant at this mature stage of differentiation.

**pre-*TTN* factories are hubs of pre-mRNA splicing**

Having gauged O-MAP's ability to dissect chromatin loci at RBM20 foci, we moved to examine the transcriptome (Fig 3.4, a). General comparisons of streptavidin-enriched material normalized to whole-cell RNA for each phenotype demonstrates that WT and ΔRBM20 enriched transcripts are both shared and unique. However, in both cell types the cohort of pre-*TTN*-proximal transcripts appeared to localize there post-transcriptionally. For example, of the 2447 transcripts enriched in WT cells, only 177 (7.2%) and 16 (0.65%) were encoded in CID- or TID-localized genes, respectively (Fig. 3.4, b). Since these transcripts' localization could not be explained simply by chromatin localization, we next sought to identify other genomic features that might explain these phenomena, by directly comparing pre-*TTN*-proximal RNAs isolated from WT and ΔRBM20 CMs, normalized to total RNA from each respective phenotype. This allowed us to identify pre-TTN-proximal RNAs specific to each cell type, yielding a catalog of 447 and 473 WT- and ΔRBM20-specific transcripts, respectively (Fig 3.4, c-e). In both WT and ΔRBM20 cells, pre-*TTN* O-MAP enriched heavily for protein-coding transcripts, although the ΔRBM20 phenotype enriched slightly more non-coding transcripts compared to WT (Fig 3.4, c). Of note,

the TTN transcript itself was not differentially enriched in either condition, indicating that it is transcribed at similar levels in both phenotypes (Fig 3.4, e). Moreover, TID-encoded genes are not preferentially enriched in WT or ΔRBM20 cells, further demonstrating that cell type-specificity of the pre-*TTN*-proximal transcriptome is not determined by local chromatin architecture (Fig 3.4, f). However, transcript structure itself appears to be a major driver of this localization. On average, the transcripts enriched in WT cells were ~34 Kb longer (Fig 3.9, d) and ~10% less GC-rich than those enriched in ΔRBM20 cells (Fig 3.4, g). Suggesting that the ΔRBM20 pre-TTN interactome may be driven in part by sequence composition and length.

To validate the three-dimensional distances of several interacting transcripts to pre-TTN, we performed pre-TTN O-MAP with nascent RNA-FISH. A transcript not enriched in either phenotype, SDHA, did not change its distance from TTN O-MAP between conditions (1.8 μm and 1.9 μm, WT *vs* KO respectively) (Fig 3.4, h, bottom). Two transcripts encoding proteins related to DCM—OBSCN and MYOM1—were in closer proximity to nascent TTN O-MAP in ΔRBM20 CMs (1.8 μm *vs* 1.2 μm for OBSCN, and 2.0 μm *vs* 1.4 μm for MYOM1, WT *vs* KO respectively) (Fig 3.4, h). While a transcript encoding a protein acting within mitochondrial calcium ion homeostasis, IMMT, was heavily proximal to TTN O-MAP signal in WT CMs compared to ΔRBM20 CMs (average distance: 0.5 μm and 1.7 μm, WT *vs* KO respectively) (Fig 3.4, h). These differences in distance indicate these transcripts localize to distinct subnuclear locations in the nucleus when RBM20 is lost, which could lead to their downstream alternative splicing (Fig. 3.5, f-i). Strikingly, these findings are recapitulated in fetal human heart tissue sections (Fig 3.4, i), suggesting that interactions discovered in our O-MAP model system are also formed *in vivo*.

**Alternative splicing events are enriched in pre-TTN-proximal transcripts in ΔRBM20 cardiomyocytes**

RBM20 is known to play a role in alternative splicing by promoting exon exclusion (Dauksaite & Gotthardt, 2018). To assess if there were differences in alternative splicing upon ΔRBM20 we analyzed whole-cell RNA-sequencing from WT and ΔRBM20 cells for alternative splicing events using rMATs (Shen et al., 2014). Principal component analysis demonstrates that the differences observed are indeed due to the ΔRBM20 phenotype, as WT and ΔRBM20 replicates cluster tightly together (Fig 3.5, a). In total there were 3,477 differential splicing events observed between WT and ΔRBM20 cells (FDR<0.05; ΔPercent Spliced-in (ΔPSI) 0.1), and largest difference in global splicing events by category between phenotypes was skipped-exons (3.15% difference) (Fig 3.5, b). Intriguingly, alternative splicing events were significantly higher among transcripts associated with nascent TTN RNA, compared to the transcriptome overall (Fig 3.5, c, left). Skipped-exon, mutual exclusive splicing, and retained intron events appeared to be the primary drivers of these splicing differences (Fig 3.5, c, right). We next sought to characterize the differences in splicing observed in pre-TTN-proximal transcripts. Indeed, nascent TTN RNA interacts with more differentially spliced transcripts in ΔRBM20 cells (Fig 3.5, d, left). Retained intron and alternative 3′ splice-site events accounted for this difference (Fig 3.5, d, right). Skipped exons accounted for the largest contribution in differential splicing, and plotting these events on a volcano plot showed clear differences in RBM20 targets previously validated by eCLIP, such as TTN and NADSYN1 (fig 3.5, e, top). Of note, the global differences in splicing could not be attributed to loss of RBM20 alone, as eCLIP-validated RBM20 target RNAs comprised only a small fraction of these events (Fig. 3.5, e, top).

Gene ontology of global alternative splicing events revealed enrichment for alternative splicing of transcripts encoding for proteins involved in DNA damage response, chromatin organization, and heart development. Excitingly, gene ontology of splicing events occurring exclusively near nascent TTN RNA encode for proteins related to actin cytoskeleton organization, muscle cell differentiation, and striated muscle cell development (Fig. 1.9, e). Sashimi plots for all significant alternative splicing events were generated (available upon request). We chose to validate IMMT as it is a transcript enriched at RBM20 foci, involved in calcium regulation, and alternatively spliced (Fig 3.5, f). Indeed, qPCR demonstrated higher levels of exon 6 retention in ΔRBM20 CMs compared to WT CMs (Fig 3.5, g, left). This was also recapitulated in male WTC11 iPSC-CMs with diseases-associated RBM20-DCM point mutations (Fenix et al., 2021), confirming that this event was not an artifact of the RUES2-CM system employed, and that its induction didn't require total loss of RMB20 function (Fig 3.5, g, right). Similarly, OBSCN was chosen for validation as it is a transcript enriched at nascent TTN RNA in ΔRBM20 CMs, encodes for a protein causal in DCM when mutated (Marston et al., 2015) is aberrantly spliced under RBM20-DCM causing mutations (Fenix et al., 2021; G. Wu et al., 2021), and is alternatively spliced in our data (Fig 3.5, h). As with *IMMT*, qPCR demonstrated higher levels of the shorter OBSCN isoform in WT compared to ΔRBM20 CMs (Fig 3.5 I, left). This also recapitulated in the WTC11 RBM20-DCM CMs (Fig 3.5 I, right).

Collectively, these data demonstrate interactions with pre-TTN mRNA and RBM20 are crucial for regulating splicing in cardiomyocytes and that loss of RBM20 leads to aberrant splicing similar to what is observed in lines harboring RBM20-DCM mutations. Moreover, pre-TTN-proximal splicing events primarily consist of retained-intron events, meaning that RBM20 is not directly splicing the majority of pre-TTN-proximal transcripts. Intriguingly, these pre-TTN

proximal transcripts undergoing differential splicing between WT and ΔRBM20 cells collectively enrich for gene ontology signatures related to muscle cell and cardiac function, as well as actin organization which is well established as being disrupted in DCM (Luk et al., 2009; Olson et al., 1998, 2002).

**The pre-TTN-factory proteome**

Having assessed the DNA-loci and transcriptome around nascent TTN RNA in WT and ΔRBM20 CMs, we moved to identify interacting proteins. Nascent TTN O-MAP yields very little directly biotinylated material, relative to more abundant structures like nucleoli, for which the original O-MAP-MS pipeline was developed (Tsue, Kania et al., 2023). To increase signal-noise, we employed a modified protocol that enriches for nuclear material prior to pulldown, and minimizes protein loss during the enrichment and post-processing steps (Fig 3.6, a, methods). In total, we enriched for 483 proteins at RBM20 foci in WT CMs, and 640 proteins at nascent TTN RNA in ΔRBM20 CMs (Fig 3.6, b). Due to inherent noise coming from pre-biotinylated material and potential residual mitochondrial peroxidase activity, we further refined this list for subsequent analysis by focusing on proteins with known nuclear localization. This resulted in 254 proteins enriched at RBM20 foci and 180 proteins enriched at nascent TTN in ΔRBM20 CMs. Gene ontology demonstrated that RBM20 foci localize proteins important for chromosome organization, RNA splicing, and RNP biogenesis (Fig 3.6, b). For example, chromatin-regulatory factors included SAFB2, SAFB and their receptor, SLTM, as well as most components of the Cohesin Complex, and High Mobility Group proteins HMGB2 and 3 (Fig 3.6, b, right). We further observed the striking enrichment of many proteins involved in DNA replication and repair, including all members of the MCM complex, FEN1, PCNA, and APEX1. This suggests that

expression of the *TTN* transcript may induce DNA damage, as is observed for other long, repeat-laden transcripts (Belotserkovskii et al., 2018). Observed RNA regulatory proteins included factors with a broad role in general RNA metabolism (e.g., hnRNPAB, hnRNPQ), those that regulate specific cohorts of RNA targets (e.g., QKI, SRSF9, PUM1), and nucleolar factors (e.g., RPL23A, PES1).

In ΔRBM20 CMs, the nascent TTN transcript proteome appeared markedly less driven by RNA-protein interactions as none of the RNA-binding motifs enriched in ΔRBM20 pre-TTN-proximal transcripts correspond to proteins enriched in the ΔRBM20 pre-TTN-proximal proteome. The ΔRBM20 pre-TTN-proximal proteome comprised a varied assortment of splicing factors and nuclear speckle components (Fig 3.6, b). For example, two core nuclear speckle proteins SON and SRRM2, and RBM25 were enriched relative to WT (Fig 3.6, b). This suggests that in the absence of the influence RBM20 exerts on pre-TTN's local compartment, the transcript explores or assembles a largely generic splicing speckle. Intriguingly, ALPK3—a transcription factor kinase thought to be involved in all forms of cardiomyopathy—was one of the strongest enriched proteins in ΔRBM20 CMs.

The novel WT hits QKI and SAFB were found to overlap with pre-TTN RNA FISH in fetal heart tissue compared to the negative control RBM25 (Fig 3.6, c and d). Strikingly, both QKI and SAFB formed prominent nuclear puncta that almost exclusively overlapped with pre-TTN nascent transcripts. SAFB is a cryptic nuclear matrix-associated scaffolding factor with both chromatin- and RNA-binding functions (Norman et al., 2016), and has recently been implicated in playing roles in the organization of other RNA-scaffolded subnuclear compartments (Hong et al., 2015; Norman et al., 2016). That this factor is almost exclusively compartmentalized in the pre-TTN factory suggests that, in cardiomyocytes it may play a similar role in coupling TTN mRNA

biogenesis to chromatin organization. Likewise, QKI is a developmentally regulated splicing factor with putative roles in mRNA nuclear export (Pilotte et al., 2001), and which has recently been implicated in contributing to DCM etiology (X. Chen, Yin, et al., 2021). Its co-compartmentalization with pre-TTN suggests that the pre-TTN factory may globally coordinate and regulate QKI targets—and those recognized by the other RNA-binding proteins observed in our data—transcriptome wide.

To test this hypothesis, we examined if known RNA-binding protein recognition motifs were enriched in WT or ΔRBM20 pre-TTN-proximal transcripts, using Transite (Krismer et al., 2020) (Fig 3.6, e). This analysis revealed marked enrichment of binding motifs for several WT pre-TTN proximal enriched proteins represented in our proteomics data, including QKI and ELAVL1 (HuR). Indeed, consensus motifs for each of these factors are observed in the majority of WT pre-TTN-proximal transcripts, but are almost completely depleted from ΔRBM20 pre-TTN-proximal transcripts (Fig 3.6, f-g). Strikingly, the binding motifs enriched in ΔRBM20 pre-TTN-proximal transcripts did not have their respective proteins enriched in ΔRBM20 pre-TTN-proximal proteins. This strongly suggests that, in the presence of RBM20, the pre-TTN factory recruits a specific cohort of RNA-binding proteins, along with their bound target transcripts. These interactions seem to be guided by specific RNA-protein interactions. Conversely, when RBM20 is absent, the proteins localizing to pre-TTN mRNA lack such specific RNA-protein interactions. Earlier we noted a higher GC percentage in ΔRBM20 pre-TTN-proximal transcripts compared to the wild-type (Fig. 3.4, g). Moreover, we observed an enrichment of two proteins in the ΔRBM20 pre-TTN-proximal proteome, SON and SRRM2, known to be involved in forming nuclear speckles (Ilik et al., 2020). Emerging evidence suggests that SRRM2 prefers interacting with transcripts that have a high GC content. This, in addition to our differential alternative splicing

analysis above (Fig. 3.5), suggests that upon loss of RBM20 the pre-TTN factory might undergo a shift in localization from its unique and specialized subnuclear splicing compartment to a more general nuclear speckle. Consequently, in ΔRBM20 cells, the presence of pre-TTN could potentially interfere with the natural splicing processes of other transcripts that often occur at nuclear speckles.

**Loss of RBM20 downregulates TXNRD2 expression and impairs mROS-handling.**

We were curious to examine the physiological impact RBM20 loss might have on cells, beyond its immediate effects on pre-mRNA splicing. In our O-MAP seq data we noted the prominent enrichment of TXNRD2 at WT RBM20 foci (Fig 3.7, a). The protein encoded by this transcript plays a key role in mitochondrial ROS regulation. Of note, we also observe oxygen bubbles in the WT CMs upon O-MAP $H_2O_2$ treatment that are absent in the ΔRBM20 condition, suggesting impaired ROS regulation in ΔRBM20 cells. Moreover, it is known that mutation or KO of TXNRD2 leads to severe forms of DCM (Sibbing et al., 2011). Raw sequencing of whole cell RNA between WT and ΔRBM20 indicated that the overall expression of TXNRD2 is nearly tenfold lower in ΔRBM20 cells. Given the implications of DCM etiology and the observed alterations in gene expression, we sought to characterize the protein expression of TXNRD2 and mitochondrial ROS mediation in our CMs. Using qPCR, we validated the substantial decrease in TXNRD2 expression observed in our sequencing data (Fig 3.7, b). Moreover, in our RBM20-DCM mutation lines, we observed a decrease in TXNRD2 expression in a mutation dependent manner (Fig 3.7, c). To see if the phenomenon of TXNRD2 expression is related to RBM20 functionality, we pulled RNA-sequencing data from another group with CMs harboring a different RBM20 mutation, R634Q. Strikingly, TXNRD2 expression was undetectable at the RNA level in R634Q heterozygous or homozygous mutant lines (Fig 3.7, d). Furthermore, terminally differentiated

human iPSC-cardiomyocytes in which the point mutant was corrected by CRISPR editing *in vivo* (Nishiyama, 2022) failed to rescue *TXNRD2* expression. Using immunofluorescence for TXNRD2 and the mitochondrial protein ATP5A, we observed decreased levels of TXNRD2 protein in ΔRBM20 cells compared to WT (Fig 3.7, e, left). Fluorescence quantification normalized to ATP5A demonstrated there is roughly five times less TXNRD2 protein in ΔRBM20 cells (Fig 3.7, e, right). To examine the functional impact this had on mitochondrial ROS production, we used a MitoSOX red assay. Basal levels of mitochondrial ROS were not significantly different between cell lines. However, upon inducing an oxidative challenge (100 μM $H_2O_2$, 30 minutes), WT CMs demonstrated a swift stabilization of mitochondrial ROS at approximately 4 minutes, while ΔRBM20 CMs exhibited notably elevated levels of mitochondrial ROS and required significantly longer times (~20 minutes) to stabilize (Fig 3.7, f, g). These data strongly suggest that loss of RBM20, via the dysregulation of the pre-*TTN*-proximal transcript *TXNRD2*, exhibits gross defects in mitochondrial ROS-handling. A similar dysregulation of mitochondrial ROS-metabolism has been observed in patient tissue (Ramaccini et al., 2021).

## 3.4 DISCUSSION

In this study, we leveraged the O-MAP toolkit to investigate the DNA loci, RNAs, and proteins associated with TTN pre-mRNA speckles in a Dilated Cardiomyopathy (DCM) model system, consisting of wild-type (WT) and ΔRBM20 hESC-derived cardiomyocytes (Fig. 3.8). To the best of our knowledge, this represents the first comprehensive multi-omic analysis of a specific locus-associated nuclear speckle, as previous studies have primarily conducted speckle analyses *en masse* (Dopie et al., 2020; Hall et al., 2006; Mintz et al., 1999; Saitoh et al., 2004). By characterizing the constituents of TTN pre-mRNA speckles, we gained insights into their crucial biological functions within their native context and explored their potential role in the etiology of

DCM (Fig 3.8). Furthermore, we expect that the optimized O-MAP protocols, tailored for the investigation of spatial biology surrounding low-abundance transcripts with limited enriched material, will facilitate the examination of other delicate RNA-scaffolded structures.

Chromatin organization plays a crucial role in regulating gene expression and maintaining cellular function. By compacting and organizing DNA, chromatin structures determine which genes are accessible for transcription and influence various cellular processes, such as DNA replication, repair, and cell differentiation (G. Li & Reinberg, 2011; Sexton & Cavalli, 2015; Zheng & Xie, 2019). Dysregulation of chromatin organization can lead to cardiovascular diseases, including cardiac hypertrophy, heart failure, and congenital heart defects, where alterations in chromatin structure contribute to dysregulated gene expression patterns and impaired cardiac function (P. Han et al., 2011; Hang et al., 2010; Papait et al., 2013). Therefore, the use O-MAP ChIP represents a valuable approach to clarify the molecular basis of DCM. O-MAP ChIP analysis revealed a total of 19 trans-interacting domains (TIDs) and 30 cis domains on chromosome 2 in WT cells. Strikingly, the loss of RBM20 resulted in the absence of all chromatin interacting domains, except for a domain containing the TTN locus. This finding underscores the crucial role of RBM20 in preserving the structural integrity of pre-TTN speckles in terms of chromatin localization. The *cis* loci identified exhibited a correlation with chromatin A compartment signal in the ΔRBM20 cell line, suggesting that RBM20 assists in localizing TTN pre-mRNA around actively transcribing genes located on chromosome 2. Furthermore, the TIDs demonstrated enrichment for polycomb-repressed and bivalent enhancer chromatin states, indicating their epigenetic repression. Notably, genes located within TIDs exhibited significantly lower transcription levels in WT cells compared to ΔRBM20 cells. Within the TID loci, genes encoding transcription factors crucial for cardiomyogenesis were identified, and their

transcription, along with the genes they regulate, was found to increase in the absence of RBM20. Collectively, these findings establish a novel framework demonstrating the coordinated action of pre-TTN mRNA and RBM20 in orchestrating the transcriptional repression of chromatin domains harboring genes essential for cardiomyogenesis. We speculate this regulatory mechanism enables the downregulation of redundant differentiation pathways over the course of development, thereby contributing to the fine-tuning of cardiac development.

The primary established role of RBM20 is to facilitate exon exclusion in alternative splicing (Dauksaite & Gotthardt, 2018). Previous studies have identified a small number of transcripts presumed to be present at pre-TTN speckles in WT cardiomyocytes, undergoing exon-exclusion alternative splicing (Bertero et al., 2019; Fenix et al., 2021). In this study, we employed O-MAP Seq to comprehensively capture all transcripts associated with pre-TTN speckles. Our analysis revealed an enrichment of transcripts encoding calcium regulatory proteins in WT cardiomyocytes, while transcripts encoding proteins linked to DCM etiology were enriched in ΔRBM20 cells. Furthermore, we observed a significantly higher occurrence of alternative splicing events at pre-TTN speckles compared to the rest of the transcriptome. Specifically focusing on transcripts localized at pre-TTN speckles in WT or ΔRBM20 cells, we observed substantially elevated levels of alternative splicing in transcripts associated with ΔRBM20 pre-TTN speckles. Gene ontology analysis of these transcripts revealed an enrichment in processes related to actin cytoskeletal organization and muscle cell development. Given that ΔRBM20 pre-TTN speckles also exhibited an enrichment of transcripts associated with DCM, the differential alternative splicing observed in transcripts at pre-TTN speckles may contribute to the etiology of DCM. These findings suggest the regulation of alternative splicing at pre-TTN speckles plays a significant role in the pathogenesis of DCM, further implicating RBM20 in the fine-tuning of gene expression

during cardiac development and disease progression (Bertero et al., 2019; Fenix et al., 2021; Koelemen et al., 2021).

Differential analysis of RNA-binding motifs within transcripts enriched at pre-TTN speckles revealed a significant enrichment of QKI motifs in WT pre-TTN speckles. QKI is a splicing factor that exhibits high expression in cardiomyocytes and is essential for the differentiation of human embryonic stem cells (hESCs) into functional cardiomyocytes (X. Chen, Liu, et al., 2021). Notably, we observed colocalization of QKI with pre-TTN speckles in fetal heart tissue, further supporting the notion that pre-TTN speckles serve as central hubs for cardiac-specific functions. Proteomic analysis using O-MAP identified an enrichment of QKI and chromatin organizational proteins associated with DNA damage in WT cardiomyocytes. Immunofluorescence experiments also confirmed the colocalization of SAFB, a protein known to localize to nuclear stress bodies in areas of DNA damage (Collins et al., 2020), with pre-TTN speckles. The importance of DNA repair for cardiac function and the permanent exit of proliferating cardiomyocytes from the cell cycle is well established (de Boer et al., 2023; Puente et al., 2014). Therefore, we propose that pre-TTN speckles may serve as specialized nuclear stress bodies that function as sites of DNA damage and repair, potentially triggered by the presence of R-loops formed by the long pre-TTN transcript.

In ΔRBM20 cells, pre-TTN foci were enriched for spliceosome proteins and the essential nuclear speckle proteins SRRM2 and SON. This finding is consistent with the higher degree of alternative splicing events coordinated by ΔRBM20 pre-TTN foci. Moreover, the transcripts present at ΔRBM20 pre-TTN foci exhibited higher GC content, and emerging evidence suggests that SRRM2 preferentially interacts with GC-rich transcripts (Roca, 2023). Therefore, we

speculate that the nascent TTN transcript in ΔRBM20 cells sequesters itself into other spliceosomes, depleting splicing factors and contributing to the etiology of DCM.

Collectively, our findings suggest that pre-TTN speckles play a multifaceted role in cardiac biology. They serve as important sites for the recruitment of splicing factors, including QKI, while also being implicated in DNA damage and repair processes. The dysregulation of these functions in the absence of RBM20 underscores its crucial role in maintaining the integrity and functionality of pre-TTN speckles, providing new insights into the pathogenesis of DCM.

Finally, we demonstrate the functional insights provided by O-MAP analyses. We observed a substantial enrichment of the transcript TXNRD2 at pre-TTN foci in WT cells by O-MAP-Seq, whereas its presence was generally reduced in ΔRBM20 cells as observed by whole-cell RNA-seq. Mutations or knockout of TXNRD2 have been identified as causal factors in severe dilated cardiomyopathy (DCM) (Conrad et al., 2004; Sibbing et al., 2011). In our study, we observed a significant downregulation of both transcript and protein levels of TXNRD2 in ΔRBM20 cells. This downregulation was further supported by publicly available data from two different RBM20-DCM mutation models, and we confirmed this downregulation through quantitative PCR analysis in R636S heterozygous and homozygous lines. The role of TXNRD2 is to modulate levels of mitochondrial reactive oxygen species (mROS), and in ΔRBM20 cells, we observed an impaired ability to swiftly stabilize mROS compared to WT cells. This finding provides evidence for the functional implications of dysregulating pre-TTN foci, as elevated levels of mROS have been implicated in cardiovascular disease (Kornfeld et al., 2015). However, the exact mechanism underlying the downregulation of TXNRD2 in the context of RBM20 deficiency remains to be fully understood. These results highlight the functional consequences associated with the dysregulation of pre-TTN foci. The significant downregulation of TXNRD2, a critical

mediator of mROS levels, in ΔRBM20 cells supports the link between disrupted pre-TTN foci and DCM pathogenesis. Further investigations are required to elucidate the precise factors and mechanisms contributing to the observed TXNRD2 downregulation. Nonetheless, these findings emphasize the importance of maintaining the integrity and proper functioning of pre-TTN foci in preserving cardiovascular health.

The optimized O-MAP protocols for low-abundance transcripts and recovered material described in this study provide significant advancements by enhancing the applicability of O-MAP to considerably more RNA species and biological contexts. These improvements lay the foundation for future investigations aimed at unraveling the biological significance of enriched DNA loci, RNAs, and proteins associated with RBM20 speckle assembly, chromatin organization, and splicing regulation. To achieve this, targeted gene knockout using CRISPR/Cas9, RNA interference, or antisense oligonucleotide knockdown techniques will be employed. For example, RNAi against SAFB followed by measuring the formation of RBM20 speckles or their localization with validated hits via fluorescence-based assays. These experimental approaches will enable a comprehensive understanding of the elaborate landscape of genome regulation by RBM20 speckles during cardiomyogenesis and dilated cardiomyopathy (DCM), potentially leading to the discovery of novel RNA-based therapeutic targets.

It is important to note that our study is limited by the utilization of hESC and hiPSC-derived cardiomyocytes, which more closely resemble fetal human heart than adult tissue (Y. Jiang et al., 2018). However, future investigations are already underway to address this limitation by conducting O-MAP analyses in adult human heart tissues. By comparing the enriched material from stem cell-derived cardiomyocytes to that obtained from adult heart tissues, we further aim to assess the correlation and bridge the gap between developmental stages. Despite this limitation,

the data presented in this study hold significant importance and relevance. Our comparison between wild-type and disease state samples provides valuable insights that are unattainable using human heart tissue due to the scarcity of RBM20-DCM samples. Furthermore, the exceptional sensitivity of the O-MAP technique in dissecting specific nuclear speckles opens new avenues for investigating other distinct nuclear speckles in the context of disease and fundamental biology.

## 3.5 METHODS

**Cell culture and tissue sections**

RUES2 hESCs (RUESe002-A; WiCell) and WTC-11 hiPSCs (a gift of Bruce R. Conklin; also available from Coriell #GM25256) were maintained on Matrigel (Corning; 354277) in Essential 8 media (ThermoFisher; A1517001). Media was exchanged every other day. Cells were passaged with Versene (ThermoFisher; 15040066) into low confluent wells with 10 μM Y-27632 (Stem Cell Technologies; 72302).

hiPSC and hESC lines were differentiated into hiPSC- or hESC-cardiomyocytes (CMs) using a method previously described (Bertero et al., 2019; Fenix et al., 2021) Briefly, 4.5-9 x $10^4$ cells were plated onto 12-well plates pre-coated with 80 μg/μl growth factor-reduced Matrigel in mTeSR1 (StemCell Technologies; 85850) and 10 μM Y-27632 for 24 hours. Cells were then treated with 1 μM CHIR-99021 (Cayman Chemical; 13122). On day 0, media was exchanged for RPMI-1640 with glutamine supplemented with 500 μg/ml BSA, 213 μg/ml ascorbic acid (RBA media; Millipore Sigma), and 5 μM CHIR-99021, and maintained for 48 h. On day 2, culture media was replaced with RBA media supplemented with 2 μM WNT-C59 (Selleck Chemicals; S7037) and maintained for 48 h. On day 4, media was replaced with RBA. Then on day 6, cells were cultured in RPMI supplemented with B27 (Thermo Fisher; 17504044), with media exchanged every two days. To increase cell viability after cryopreservation and thawing, on day 13 cells

underwent a 45-minute heat shock at 42 °C using pre-warmed media. Then on day 14, cells were dissociated with TrypLe (Thermo Fisher; 12563011) with gentle trituration and were strained with a 100 μM filter (Fisher Scientific; 22363549). Dissociated cells were frozen in CryoStor media (1 million cells/100 μL CryoStor; Millipore Sigma; C2874) and frozen slowly at -80 °C using a Mr. Frosty Freezing Container (Thermo Fisher; 5100-0001). The next day cells were stored in liquid $N_2$.

PFA-fixed and paraffin embedded fetal heart tissue sections were acquired from the National Disease Research exchange. The University of Washington Pathology Research Services Laboratory was used to cut blocks into 5 μm sections.

**TTN pre-mRNA O-MAP main protocol**

RUES2-CMs were used for both O-MAP omics and O-MAP imaging. For omics, CM stocks were thawed onto 12-well plates coated in Matrigel at 1.2 million cells/well in RPMI/B27 supplemented with 10 μM Y-27632 and 10% FBS. For imaging, $2.5x10^5$ cells were plated onto 10mm glass bottom plates (Cellvis; D35-10-1.5-N) . The next day media was exchanged for RPMI/B27 and exchanged every two days thereafter until cells reached day 21 post-differentiation. The core O-MAP protocol follows the same two-day protocol as previously described (Tsue, Kania et. al. 2023), but with slight optimizations for cardiomyocytes. These optimizations are: fixation with 4% PFA and inactivation of endogenous peroxidases with $H_2O_2$ solution for 20 minutes.

**TTN pre-mRNA O-MAP day 1**

Cells were quickly washed with 1x DPBS (Thermo Fisher; 14190250) and fixed in 4% (v/v) formaldehyde (Electron Microscopy Sciences; 15710) in 1x PBS (Sigma; 6506) for 10 minutes. The fixation solution was aspirated and quenched by two five-minute washes with 250 mM glycine in 1x PBS, with rocking (3 rpm on a platform rocker (Everlast Rocker 247)). Cells were then

briefly washed with DPBS and permeabilized with 0.5% (v/v) triton-X 100 in PBS for 10 min with rocking. Three five-minute DPBS washes were then performed to prevent over-permeabilization. Samples were then incubated with a 0.5% (v/v) $H_2O_2$ solution for 20 minutes with rocking. Two five-minute washes with DPBS followed to prevent $H_2O_2$ damage to cells. Cells were then equilibrated in Formamide Wash Buffer (40% (v/v) deionized formamide (VWR, 0314); 2X SSC (Thermo Fisher; AM9765); 0.1% Tween-20) for five minutes with rocking. Wash buffer was then aspirated, and to each well 50 µL of Primary Probe Mix (Supplementary File 1) (0.1 µM primary oligo probe pool, in 1x Hybridization Buffer: 40% deionized formamide; 2x SSC; 0.1% (v/v) Tween-20; 10% (w/v) dextran sulfate (SIGMA; D8906); in nuclease-free water) was added. A clean 12-well glass coverslip was then applied to spread the Primary Probe Mix across the samples. A kimwipe soaked in RNase-free water was placed between the wells, and plates were sealed with Parafilm to create a hybridization chamber. Plates were then incubated for 8 hours at 42°C without rocking.

**preTTN O-MAP day 2**

After primary hybridization, 30% Formamide Wash Buffer was added to the samples. Coverslips were then removed, and cells were allowed to wash for five-minutes rocking. Two additional five-minute washes with fresh 30% Formamide Wash Buffer followed. Then 50 µL of O-MAP Secondary Probe Mix (100 nM SABER1-HRP oligo in 30% Hybridization Buffer) was added to each well, and a clean coverslip was added. Samples then incubated for one hour at 37°C in the dark. For preTTN O-MAP imaging, endogenous biotin blocking was performed (see preTTN O-MAP Imaging below). Otherwise, to each well 1-mL of PBST (0.1% (v/v) Tween-20 in 1X PBS) was added. Coverslips were then removed, and samples washed for 15 minutes in PBST rocking at room temperature. Three more 15-minute washes with PBST followed. Following aspiration, 1-

mL of Labeling Solution was added (0.8 µM biotinyl-tyramide (Sigma; SML2135), 1 mM $H_2O_2$, 1x PBST) and in situ biotinylation was carried out for 45 minutes. The biotinylation reaction was then quenched by the addition of 1-mL of PBST-q (10 nM sodium azide, 10 nM sodium ascorbate in 1X PBST), for three five-minute washes rocking.

**preTTN O-MAP Imaging**

After hybridization with SABER1-HRP, samples were washed three times for five minutes in PBST. To block endogenous biotin signal, samples were then incubated for 30 minutes with 1-mL of 1% nuclease-free BSA (VWR; 97061-420) in PBST with rocking, followed by a 15-minute incubation with 1-mL of Neutravidin solution (10 µg/mL neutravidin (Thermo Fisher; 31000), 1% (w/v) nuclease-free BSA, in 1x PBST) rocking. After three five-minute washes in PBST, samples were incubated with 1-mL of biotin solution (10 µg/mL D-biotin (Thermo Fisher; B20656) in 1x PBST) to occupy free neutravidin binding sites. Samples were then washed three times in PBST before proceeding with in situ biotinylation and quenching as described above. After biotinylation samples were incubated with 1-mL of 1 µg/mL neutravidin-Dy550 (ThermoFisher; 84606) in a 1% BSA PBST solution for one hour at room temperature with rocking. Samples were then washed three times with PBST, stained with DAPI, and either imaged immediately or mounted in Vectashield (Vector Labs; H-1900-10).

**Combined preTTN O-MAP, RNA-FISH , and immunofluorescence**

To perform O-MAP with RNA-FISH and immunofluorescence, cells were first fixed in PFA, quenched with glycine, permeabilized in Triton X-100, underwent peroxidase inactivation, equilibrated in formamide, and hybridized to primary probes as described above (see preTTN O-MAP day 1). The primary probes were appended with a landing pad for recruitment of either

SABER1 or SABER2 sequences. O-MAP was then performed for imaging which includes endogenous biotin block (see preTTN O-MAP Imaging). Following quenching, samples were equilibrated in formamide by three five-minute washes in 1-mL 30% Formamide Wash Buffer. This buffer was aspirated and 50 µL of FISH secondary probe mix (100 nM SABER2–AF647 oligo in 30% Formamide Hybridization Buffer) was added, followed by the placement of a clean coverslip over the sample. Samples were incubated at 37°C in the dark, then washed three times with PBST. Before proceeding to immunofluorescence, samples were next blocked with 5% (w/v) nuclease-free BSA in 1X PBST for 30 minutes. Samples were then incubated with rabbit-RBM20 (ThermoFisher; PA5-57404, used at 1:800 dilution) in 5% BSA in 1X PBST for 1 hour 45 minutes. Samples were then washed 3x5 minutes in 1X PBST. Then samples were incubated with anti-rabbit-Atto488 (ThermoFisher; R37116, used at 1:300 dilution) in 1% BSA in 1XPBST along with 1 µg/mL neutravidin-Dy550. Four fifteen-minute washes in 1X PBST ensued, followed by DAPI staining. Samples were imaged immediately in 1X PBST.

When performing pairwise fluorescent imaging, either O-MAP with RNA-FISH, O-MAP with immunofluorescence, or RNA-FISH with immunofluorescence, the O-MAP protocol preceded the RNA-FISH or immunofluorescence protocol and RNA-FISH preceded the immunofluorescence protocol. For these experiments, rabbit-TTN (Myomedix; #Z1Z2, used at 1:250 dilution), rabbit-QKI (Sigma; HPA019123, used at 1:125 dilution), rabbit-SAFB (Sigma; HPA020076, used at 1:50 dilution), rabbit-H3K27Me3 (CST; 9733, used at 1:800 dilution), rabbit-H3k4Me3 (Active Motif; 39159, used at 1:500 dilution), rabbit-TrxR2 (ThermoFisher; PA5-29458, used at 1:1000 dilution), or mouse-ATP5A (Abcam; ab14748, used at 1:500) were used.

**preTTN O-MAP ChIP**

preTTN O-MAP ChIP was performed as previously described (Tsue, Kania, et al (2023), but with 24 sonication cycles. Briefly, O-MAP labeled cells ($3.6x10^6$ per replicate; two biological replicates per experimental condition) were harvested in PBSTq using cell scrapers, and pelleted at 3,000 x g for 10 minutes at 4°C. Then PBSTq was aspirated and cells were flash-frozen in liquid $N_2$ before storing at -80°C until use. Samples were thawed on ice and resuspended by gentle pipetting in 1mL CLB (20 mM Tris pH 8.0, 85 mM KCl, 0.5% (v/v) NP-40) supplemented with 1x Halt EDTA-Free protease inhibitor cocktail and 10 mM sodium azide, for 10 minutes. Lysates were clarified by centrifugation at 3,000 x g for five minutes at 4 °C followed by aspiration of supernatant. Two more rounds of resuspension with CLB, clarification, and aspiration ensued. Then pellets were lysed by gentle pipetting in 1-mL of NLB (10 mM Tris-HCl pH 7.5, 1% (v/v) NP-40, 0.5% (w/v) sodium Deoxycholate, 0.1% (w/v) SDS) supplemented with 1x Halt EDTA-Free protease inhibitor cocktail and 10 mM sodium azide, for 10 minutes. Lysates were then placed into an ice-cold thermal block and sheared with a Branson Digital Sonifier outfitted with a double stepped microtip, at 10-12 Watts over 30 s intervals (0.7 s on; 1.3 s off), with 30 s resting steps between intervals. A total of 24 intervals were performed per sample, and after every 5 intervals the thermal block was placed on ice for 2 minutes to recalibrate to ice-cold temperature. This resulted in an average shearing size of approximately 200-400 bp, as confirmed by agarose gel electrophoresis.

Samples were then clarified by centrifugation at 15,000xg for 10-minutes at 4°C, and supernatants were moved to fresh tubes. Each sample then had 10 µL removed as 'input'. To the remaining 990 µL, 100 µL of streptavidin-coated magnetic beads equilibrated by two NLB washes was added. Samples rotated end-over-end for 2 hours at room-temperature. The beads were then washed with the following series of buffers (1 mL each, 5-minutes per wash, at room-temperature, with end-

over-end rotation): (1) NLB, supplemented with 5 mM EDTA, 10 mM sodium azide and protease inhibitors (1x Halt EDTA-free Protease Inhibitor Cocktail), 150 mM NaCl; (2) NLB, supplemented with 5 mM EDTA, 10mM sodium azide and protease inhibitors, (3–4) two washes in 1 M KCl, 10 mM Tris-HCl pH 7.5, 5 mM EDTA, (5–6) two washes in 2 M Urea, 10 mM Tris-HCl pH 7.5, 5 mM EDTA, (7) 10 mM Tris-HCl pH 7.5, 1% (w/v) SDS, (8–9) 10 mM Tris-HCl pH 7.5, 1 mM EDTA.

Following the final aspiration, beads were resuspended in 100 µL of Elution Buffer (2% (v/v) N-lauryl sarcosine, 10 mM EDTA, 5 mM DTT, in 1x PBS, supplemented with 200 µg proteinase K). Samples were then incubated at 65°C while shaking at 700 rpm in a Mixer HC. Following this, samples were transferred to 0.2-mL tubes and incubated overnight in a thermocycler at 65 °C. Oligonucleotides were then extracted by addition of an equal volume of phenol pH 6.6, followed by two equal volume absolute chloroform extractions. To each sample 1 µg GlycoBlue and NaCl to 300 mM final was added, followed by ethanol precipitation overnight at -20 °C. Samples were then pelleted by centrifugation at 15,000 x g for 30 minutes at 4 °C. Pellets were washed twice with 80% ethanol and resuspended in 20 µL nuclease-free water. Residual RNA was removed by supplementation of 10 µg RNase Cocktail Enzyme Mix (Thermo Fisher; AM2286) and incubation at 37 °C for 1 hour. Finally, DNA was purified by phenol extraction and ethanol precipitation as described above, and resuspended in 20 µL of nuclease-free water.

**preTTN O-MAP ChIP library preparation and sequencing**

DNA samples were processed for sequencing and sequenced as previously described (Tsue, Kania, et al. (2023). Briefly, samples were quantified using a NanoDrop One and 300 ng of DNA from each sample was used for library preparation, using the NEBNext Ultra II DNA Library Prep Kit and NEBNext Multiplex Oligos for Illumina (NEB;E7645 and E7335), according to the

manufacturer's instructions. Two biological replicates were used per condition, and each library was given a unique index. Library concentrations were quantified using the NEBNext Library Quantification Kit for Illumina (NEB; E7360). Library quality was confirmed using an Agilent 4200 TapeStation with a "DNA High Sensitivity" kit (Fred Hutch Genomics Core). Libraries were pooled in equimolar concentrations to 20 nM aggregate in nuclease-free water, with no more than eight libraries per pool. Each pool was subjected to 150 cycles of paired-end sequencing on an Illumina HiSeq 4000, run in high output mode (Azenta Life Sciences).

**DNA-FISH Probe Design**

One-thousand DNA-FISH probes were designed for 12 different DNA loci using the online PaintSHOP application (Hershberg et al., 2021). Briefly, segments of DNA from the center of experimentally determined enriched regions (see preTTN O-MAP ChIP data analysis below) were uploaded to PaintSHOP as a BED file. Once loaded, parameters for DNA Probe Design were set as follows: Repeat -allowed, Max Off Target Score -100, Max K-mer count -5, Min Prob Value -0, Optimize Set -show, Number of probes per target -1000, Enable -off. For Append Sequences the following parameters were used: Design Scheme -DNA Probe Design, (5′ Outer Primer Sequence -Append, Orientation -Forward, Format -Same for all probes, Select Sequence Set -Custom, ATCCTAGCCCATACGGCAATG), (5′ Bridge Sequence -Append, Orientation -Forward, Format -Unique for each target, Select Sequence Set -Kishi et al. 2019 Bridges), (5′ Inner Primer Sequence -Append, Orientation -Forward, Format -Same for all probes, Select Sequence Set -Custom TGAATAGCAGCGGTGGCAAAC), (3′ Appending Scheme -Primer/Bridge/Universal), (3′ Outer Primer Sequence -Append, Orientation -Reverse Complement, Format -Same for all probes, Select Sequence Set -Custom GTATCGTGCAAGGGTGAATGC).

**DNA-FISH Probe Amplification**

Probes were ordered from Twist Bioscience (Oligo pool, Tier 3L ssDNA, >6,001 up to 12,000 oligos (151-200nt)). Oligos were resuspended in Tris-EDTA to 6.0 ng/µL. The probes were then PCR amplified using Phusion HF (ThermoFisher, F531S) and the primers Forward 5′-ATCCTAGCCCATACGGCAATG-3′ and Reverse 5′-TAATACGACTCACTATAGGGGTATCGTGCAAGGGTGAATGC-3′ to add a T7 sequence to the probes. This reaction ran on a ProFlex PCR System (ThermoFisher) using the following parameters – Initial denaturation: 98°C, 5s; Denaturation: 98°C, 5s; Annealing: 58°C, 30s; Extension: 72°C, 15s; Final extension: 72°C, 5 min – where denaturation, annealing, and extension were performed for 10 cycles. DNA products were then purified using the Monarch PCR and DNA Cleanup Kit (NEB; T1030S). RNA was then in vitro transcribed from the DNA products using the HiScribe T7 High Yield RNA Synthesis Kit (NEB; E2040S). Briefly, 10 µL of PCR product, 10 µL of NTP buffer mix (from HiScribe Kit), 2 µL T7 Pol Mix (from HiScribe kit), 1 µL RNase-OUT, and 7 µL nuclease-free water were combined in a PCR strip tube. This mixture was gently mixed and allowed to incubate 5-16 hours at 37°C. ssDNA was then transcribed from RNA using the Maxima H Minus RT enzyme (ThermoFisher; EP0751). Briefly, 7 µL dNTP (10 mM each), 14 µL Maxima H Minus RT buffer, 10 µL of Forward primer (100 µM) and of 4 µL RNase-OUT were added to a PCR strip tube. To this, the entire 30 µL T7 reaction was added, followed by 5 µL of Maxima H Minus RT enzyme. The reaction was gently mixed and then incubated at 50°C for two-hours. To digest remaining RNA, 50 µL of 1:1 1 M NaOH and 0.5M EDTA was added to the samples followed by a 12-minute incubation at 95 °C. The final ssDNA product was purified using the Oligo Clean and Concentrator Kit (Zymo Research; D4060) according to manufacturer's instructions.

**DNA-FISH Bridge Polymerase Exchange Reaction (PER)**

ssDNA oligonucleotides antisense to the 5′ Bridge Sequence (see above; DNA-FISH Probe Design) were ordered from IDT (Supplementary File 1). To facilitate recruitment of more fluorophore-oligo conjugates, the bridge oligos underwent PER. Briefly, to first remove any contaminating dGTP, 17 µL nuclease-free water, 5 µL 10X PBS, 5 µL 100mM MgSO4, 2.5 µL 6 mM dHTPs, 5 µL Clean.G (1 µM, Table X), 5 µL of respective hairpin (10 µM, Supplementary File 1), and 0.5 µL Bst LF polymerase (NEB; M0275S) were added to a PCR strip tube and incubated 37°C for 15 minutes. Then 10 µL of the respective bridge oligo was added to the reaction. Samples were incubated at 37°C for 2-hours followed by a 20-minute incubation at 80°C to inactivate the polymerase. PER bridge oligos were then purified using the Oligo Clean and Concentrator Kit according to manufacturer's instructions, and eluted in 10 uL nuclease-free water.

**DNA-FISH**

To perform DNA FISH, 1.2 µM of the ssDNA probe pool and 1.2 µM of the target PER bridge oligo were added to a 1.5-mL tube. The oligos were then dried using a Speed-Vac (Savant; DNA Speed Vac DNA120). The dried oligos were then supplemented with 50% Formamide (v/v), 2X SSC, 0.1% Tween-20, and 0.4 µg/mL RNase-A in nuclease-free water. The oligos were then placed on ice while cells were processed for DNA-FISH. Volumes used for this protocol are all 1-mL unless specified differently. Cells were first briefly washed in 1X DPBS then fixed in 4% PFA in 1X PBS for 10 minutes. PFA was aspirated and cells were washed with DPBS for 1 minute, then permeabilized with 0.5% Triton X-100 in 1X PBS for 10 minutes. Cells were then washed with 1X DPBS for two-minutes then incubated in 0.1N HCl in 1X PBS for 5 minutes at room temperature. The solution was aspirated followed by two one-minute washes with 2X SSCT, and

one two-minute wash with 50% Formamide Wash Buffer at room temperature. Following aspiration, cells were incubated with pre-warmed 60°C 50% Formamide Wash Buffer and incubated at 60°C for 20-minutes. After incubation, the wash buffer was aspirated and 50 µL of the hybridization solution (made above) was added. Then samples were denatured at 78°C for 4 minutes on a thermal block in a water bath. Samples were then incubated at 37°C overnight.

The next day, samples were rinsed twice with pre-warmed 60°C 2x SSCT. This was followed by four 5-minute washes with pre-warmed 60°C 2x SSCT and two 5-minute washes with room-temperature 2x SSCT. Samples were then rinsed with 1X PBS and 50 µL of Fluorophore solution (1 µM p30-AF647, 1µM p31-Atto488, in 1X PBS) was added. Samples were protected from light and incubated for one-hour at 37°C. Following incubation, samples were rinsed with 1X PBST followed by two 5-minute washes in 1X PBST, and two 5-minute 1X PBS washes. Samples were imaged the same day in 1X PBS.

**FFPE Tissue Section Preparation for RNA-FISH and Immunofluorescence**

Tissue sections were prepared using the FFPE Sample Preparation and Pretreatment protocol from the RNAscope 2.5 assay, using Coplin jars. Briefly, to deparaffinize sections slides were placed in Xylenes (Sigma; 534056) for 5-minutes at RT, followed by another 5-minute incubation in a separate jar of Xylenes. The slides were then immediately placed in fresh 100% EtOH for 1-minute at RT with agitation, followed by another 1-minute incubation in a separate jar of fresh 100% EtOH. Slides were then dried for 3 minutes on a Thermostat slide warmer (No. 26020) with its thermostat set to 4.5. Then ~5-8 drops of RNAscope Hydrogen Peroxide was added to cover the entire section, and incubated for 10 min at RT. Slides were then washed twice for 1-minute in nuclease-free water followed by submerging in boiling RNAscope 1X Target Retreival Reagents solution for 10-minutes.Slides were removed from boiling solution and placed immediately in RT

nuclease-free water to wash for 1-minute, followed by another 1-minute water wash, and a 1-minute EtOH wash. Then slides were dried for 3-minutes on a Thermostat slide warmer. An Immedge hydrophobic barrier pen (Vector labs; H-4000) was then used to create a hydrophobic barrier around each section, and dried for 3-minutes at RT. Five drops of RNAscope Protease Plus was then added to entirely cover each section. Sections then incubated in hybridization chambers at 40°C for 40 minutes. Slides were then washed twice for one-minute in RT nuclease-free water, then twice for two-minutes with 150 µL 30% Formamide Wash Buffer. 150 µL of primary RNA-FISH probes were then added, and ensuin combined RNA-FISH and immunofluorescence was performed as described above (see Combined O-MAP, RNA-FISH , and immunofluorescence).

**preTTN O-MAP Seq**

O-MAP labeled cells ($4.8x10^6$ cells; four biological replicates per experimental condition) were harvested by scrapping in PBSTq. Samples were then pelleted by centrifugation at 3,000xg for ten-minutes at 4°C, aspirated, flash-frozen in liquid $N_2$, and stored at -80°C until use. Samples were thawed on ice and resuspended by gentle pipetting in 1000 µL ice cold RIPA Buffer (50 mM Tris-HCl pH 7.5, 150 mM NaCl, 0.1% (w/v) SDS, 0.5% (w/v) sodium Deoxycholate, 1% (v/v) Triton X-100, 5 mM EDTA, 0.5 mM DTT), supplemented with 1x EDTA-Free Halt Protease Inhibitor Cocktail, 0.1 U/µL RNase-OUT (Thermo Fisher; 10777019)), and 10mM sodium azide, and rocked end-over-end at 4°C for five minutes. Cells were placed in an ice-cold thermal block and then sheared using a Branson Digital Sonifier 250 outfitted with a double stepped microtip, at 10–12 Watts for 30 s intervals (0.7 s on; 1.3 s off), with 30 s resting steps between intervals, five intervals total. After every five intervals the thermal block was re-equilibrated to an ice-cold temperature by incubation on ice for two-minutes. After sonication, lysates were clarified by centrifugation at 15,000xg for ten-minutes at 4°C. Supernatant was placed in fresh tubes and

diluted with 1-mL Nuclear Lysis Buffer (NLB: 25 mM Tris-HCl pH 7.5, 150 mM KCl, 0.5% (v/v) NP-40, 5 mM EDTA, 0.5 mM DTT), supplemented with 1X Halt Protease Inhibitor Cocktail, 0.1 U/μL RNase-OUT and 10 mM sodium azide. Each sample then had 5% of its volume taken as "input," and 100 μL Pierce streptavidin magnetic bead slurry that had been equilibriated by two washes in 1:1 RIPA:NLB supplemented with 10mM sodium azide, 0.1 U/μL RNase-OUT, and 1X Halt Protease Inhibitor Cocktail was added to the remainder. Samples were incubated for 2 hr at room temperature with end-over-end agitation. Beads were then washed with the following series of buffers (1 mL each, 5 min per wash at room-temperature with end-over-end agitation). All buffers were supplemented with 1x EDTA-Free Halt protease inhibitor cocktail, 0.1 U/μL RNase-OUT, and 0.5 mM DTT, unless otherwise noted: (1) RIPA, supplemented with 10 mM sodium azide; (2) RIPA alone (3) High Salt Buffer (1 M KCl, 50 mM Tris-HCl pH 7.5, 5 mM EDTA) (4) Urea Buffer (2 M Urea, 50mM Tris-HCl pH 7.5, 5 mM EDTA) (5) 1:1 RIPA:NLB, without protease inhibitors (6) NLB, without protease inhibitors, (7–8) two washes in TE buffer (10 mM Tris-HCl pH 7.5, 1 mM EDTA), without protease inhibitors.

RNA was isolated from both input and O-MAP-enriched samples by proteolysis in 100 μL Elution Buffer (2% (v/v) N-lauryl sarcoside, 10 mM EDTA, 5 mM DTT, 200 μg proteinase K (Thermo Fisher; AM2548), in 1xPBS). Reactions were shaken at 700 rpm in a Mixer HC (USA Scientific) for 1 hour at 42°C, followed by 1 hour at 60°C. RNA was then extracted once with 1 volume of phenol pH 4.3, and twice thereafter with an equal volume of absolute chloroform. Reactions were supplemented with 1μL Glycoblue (Thermo Fisher; AM9515) and NaCl to 300 mM, and ethanol precipitated at –20°C overnight. RNA was harvested by centrifugation at 15,000 x g for 30 minutes at 4°C, washed twice with 70% ethanol, and resuspended in 40 μL nuclease-free water. To each sample, 5 μL DNA Digestion Buffer and 5 μL DNase I (Zymo Research; R1013) was added,

gently mixed by inversion, and then incubated for 15-minutes at room temperature. RNA was then purified using an RNA Clean and Concentrator-5 Kit (Zymo Research; R1013), according to manufacturer's instructions. RNA was eluted in 15 µL nuclease-free water, and its concentration was measured using a Nanodrop One (Thermo Fisher).

**preTTN O-MAP Seq library preparation and sequencing**

RNA samples were processed for sequencing as previously described (Tsue, Kania et al., 2023) and sequenced with no more than 8 libraries per lane. Briefly, ribosomal RNA was first depleted by RNase-H digestion, using a pool of DNA oligonucleotides antisense to rRNA, as previously described5. 1 µg of the antisense probe pool was added to 1 µg of RNA in 200 mM NaCl, 100 mM Tris-HCl, pH 7.4, at a final volume of 10 µL. Samples were heated at 95°C for 2 minutes then slowly cooled to 45°C at a rate of -0.1°C/s, using a ProFlex PCR system. Then, reactions were supplemented with 10 µL of RNase H mix (10 U Hybridase Thermostable RNase H (Lucigen; H39500), 20 mM MgCl2) that was pre-heated to 45°C. Samples were incubated at 45°C for 30 minutes then placed on ice. RNA was purified by acidic phenol-chloroform extraction and ethanol precipitation. Residual DNA was removed using DNase I, and RNA was purified again by the RNA Clean and Concentrator-5 Kit.

Samples were quantified on a Nanodrop One. Sequencing libraries were made from 300 ng RNA using the NEBNext Ultra II Directional RNA Library Prep Kit and NEBNext Multiplex Oligos for Illumina (NEB; E7760 and E7735), according to manufacturer's instructions. Four biological replicates were used for each condition and each library was given a unique index. Libraries were quantified using the NEBNext Library Quantification Kit for Illumina, following manufacturer's instructions, and library quality was confirmed using an Agilent 4200 TapeStation with a "DNA High Sensitivity" kit (Fred Hutch Genomics Core). Libraries were pooled in equimolar

concentrations to 20 nM aggregate concentration in nuclease-free water, with no more than 8 libraries per pool. The pool underwent 150 cycles of paired-end sequencing, on one lane of an Illumina HiSeq 4000 per pool, run in high output mode (Azenta Life Sciences).

**preTTN O-MAP-ChIP data analysis**

Deep sequencing reads were trimmed using TrimGalore!6 with parameters -q 30 --phred33, and mapped to the GRCH38 genome using Bowtie2 version 2.4.47. Duplicate reads were removed with the Picard MarkDuplicates function (*Picard Tools - By Broad Institute*). Coverage bigWig files were generated using deepTools version 3.5.19 (Ramírez et al., 2014) with a binsize of 1 and normalizing with RPKM. Log2FC bigWig files were generated by comparing O-MAP ChIP data to matched input samples. Regions of enrichment were called using Enriched Domain Detector (EDD) (Lund et al., 2014). EDD was performed using default settings and blacklisting regions from the ENCODE Blacklist (Amemiya et al., 2019). Epigenomic analysis of enriched DNA loci was performed using ChromHMM version 1.22 (Ernst & Kellis, 2012). The OverlapEnrichment function was called using a chromatin state file for RUES2 cardiomyocytes (Boix et al., 2021) (Accelerating Medicines in Partnerships; accession DSR313BZF). Fold enrichment of each epigenomic signature for cis, trans, and all DNA loci was plotted as a heatmap using seaborn version 0.10.11 (Waskom, 2021). Transcription factors were pulled from the Human Transcription Factor database version 1.011 (Lambert et al., 2018), and those encoded in preTTN O-MAP TIDs underwent gene ontology using Metascape (Zhou et al., 2019).

**preTTN O-MAP-Seq data analysis**

Raw RNA-seq FASTA files were aligned to GRCH38 using HISAT2 version 2.2.11 (Kim et al., 2019), in paired-end setting with default parameters. The resulting SAM files were converted to BAM format and sorted using Samtools (H. Li et al., 2009) version 1.15.1. Bigwig files for

visualizing strand specific information were created using deepTools version 3.5.1 (Ramírez et al., 2014) with parameters: --filterRNAstrand forward/reverse --binSize 1 --normalizeusingBPM. Mapped reads were quantified using StringTie (Pertea et al., 2015) version 2.2.1 and the StringTie output was prepared for differential expression analysis using the prepDE.py function. The resulting count matrices were used for differential expression analysis using DESeq2 (Love et al., 2014) with a FDR cutoff of 0.05. For TTN foci specific differential analysis between WT and ΔRBM20 phenotypes, the DESeq2 likelihood ratio test was used to calculate (streptavidin-enriched transcripts for WT / streptavidin-enriched transcripts for ΔRBM20) / (whole-cell transcriptome for WT / whole-cell transcriptome for ΔRBM20). For computing differentially enriched transcripts between WT and ΔRBM20 cells, DESeq2 was performed for WT whole-cell transcriptome compared to ΔRBM20 whole-cell transcriptome using standard settings. The rMATS (Shen et al., 2014) package was used for alternative splicing analysis utilizing the BAM files generated by HISAT2 with the parameters: -t paired –readLength 150. Low coverage splicing events were filtered to include an average of at least 5 reads between replicates of each group. Differential events were further filtered for an FDR of 0.05 and a change in PSI between the two groups > 10% ($\delta$ PSI > 0.1). Sashimi plots for each rMATS event type (skipped exon, retained intron, mutually exclusive exons, alternative 5' splice site, and alternative 3' splice site) were generated using rmats2sashimiplot (*Rmats2sashimiplot*) using the BAM files generated by HISAT2 for coordinate location and the rMATS events file for splicing event information. Transite (Krismer et al., 2020) was used for differential RNA-binding protein profiling of transcripts enriched at TTN foci in WT versus ΔRBM20 cells using the Transcript Set Motif Analysis with the parameters: --matrix-based --Homo sapiens --svg --Transite motif database --maximum binding sites 5. The oRNAment database24 was used to identify transcripts with QKI binding

motifs. Volcano plots were generated using EnhancedVolcano version 1.12.0 (Blighe, 2018/2023). All statistical analyses (Fisher's exact test, hypergeometric distribution test, or Student's t-test, where appropriate) were performed in R or in python using the ggplot2 (Wickham, 2009), seaborn (Waskom, 2021), or matplotlib (Hunter, 2007) modules. Gene ontology analysis was performed using Metascape (Zhou et al., 2019), Enrichr (E. Y. Chen et al., 2013), and GSEA (Subramanian et al., 2005).

**Quantitative PCR**

For quantitative PCR (qPCR), each well of cells in a 12-well plate had media aspirated followed immediately by addition of 1-mL TRIzol reagent (Invitrogen; 15596026). RNA extraction proceeded according to manufacturer's protocol. Contaminating DNA was removed by incubation with DNase I and the use of the RNA Clean and Concentrator-5 Kit as described above (see preTTN O-MAP seq). RNA was eluted in 30 µL nuclease-free water. RNA samples were reverse transcribed using SuperScript IV Reverse Transcriptase (Thermo Fisher; 18090010), priming with random hexamers according to the manufacturer's protocol. cDNA was diluted with nuclease-free water, mixed with gene specific or isoform specific primers, supplemented with Rox-normalized PowerUp SYBR Green Master Mix (Thermo Fisher; A25777), and aliquoted into 384-well plates. qPCR was performed on a QuantStudio 5 Real-Time PCR System (Thermo Fisher; A34322) in both biological and technical quadruplicate. Gene expression was calculated using the 2-ΔΔCt method using the housekeeping gene RPLP0. Splicing isoform gene expression was calculated using the 2-ΔCt method as previously described2; 2^-(Ct spliced isoform – Ct constitutive isoform). Primer sequences for qPCR are listed in (Supplementary File 1) (Al-Maqtari et al., 2017; J. Chen et al., 2020; P. Chen et al., 2020; Kirschneck et al., 2017; Krstic et al., 2016; Kukharsky et al., 2020; Laptev et al., 2020; Miyashita et al., 2020; Y. Wu et al., 2021; B. Zhang et al., 2012).

**preTTN O-MAP proteomics sample preparation**

Briefly, O-MAP labeled cells ($2.4x10^6$ per replicate; five biological replicates per experimental condition) were harvested in PBSTq without sodium ascorbate using cell scrapers, and pelleted at 3,000 x g for 10 minutes at 4°C. Then PBSTq was aspirated and cells were flash-frozen in liquid $N_2$ before storing at -80°C until use. Samples were thawed on ice and resuspended by gentle pipetting in 1mL CLB (20 mM Tris pH 8.0, 85 mM KCl, 0.5% (v/v) NP-40) supplemented with 1x Halt EDTA-Free protease inhibitor cocktail and 10 mM sodium azide, for 10 minutes. Lysates were clarified by centrifugation at 3,000 x g for five minutes at 4 °C followed by aspiration of supernatant. Two more rounds of resuspension with CLB, clarification, and aspiration ensued. Then pellets were lysed by gentle pipetting in 1-mL of NLB (10 mM Tris-HCl pH 7.5, 1% (v/v) NP-40, 0.5% (w/v) sodium Deoxycholate, 0.1% (w/v) SDS) supplemented with 1x Halt EDTA-Free protease inhibitor cocktail and 10 mM sodium azide, for 10 minutes. Lysates were then placed into an ice-cold thermal block and sheared with a Branson Digital Sonifier outfitted with a double stepped microtip, at 10-12 Watts over 30 s intervals (0.7 s on; 1.3 s off), with 30 s resting steps between intervals. A total of 4 intervals were performed per sample, and after each sample was sonicated the thermal block was placed on ice for 2 minutes to recalibrate to ice-cold temperature. Samples were then supplemented with 47.47 µL of 20% SDS to bring the final SDS concentration up to 1%. The samples were then boiled at 95°C for one hour to ensure complete reverse-crosslinking. Samples were allowed to cool to 50 °C before placing on ice. Then each sample was sonicated for one more cycle. Samples were then clarified by centrifugation at 15,000xg for 10-minutes at 4°C, and supernatants were moved to fresh tubes. To each tube 75 µL of streptavidin-coated magnetic beads equilibrated by two NLB washes was added. Samples rotated end-over-end

for 2 hours at room-temperature. The beads were then washed with 250 µl of NLB, supplemented with 5 mM EDTA, 10 mM sodium azide, protease inhibitors (1x Halt EDTA-free Protease Inhibitor Cocktail), and 150 mM NaCl. Beads were then moved to fresh PCR-strip tubes (Simport Scientific; T320-2N) to minimize bead loss. The following series of washes were then performed (200 µl each, 1 minute per wash, at room-temperature, with end-over-end rotation): (1-4) NLB, supplemented with 5 mM EDTA, 10mM sodium azide and protease inhibitors (5-8) 1X PBS in nuclease-free water (9-16) 1 M KCl, 10 mM Tris-HCl pH 7.5, 5 mM EDTA, (17-24) 2 M Urea, 10 mM Tris-HCl pH 7.5, 5 mM EDTA (25-28) 200 mM EPPS pH 8.5 in nuclease-free water. After the final wash the beads were moved to fresh PCR strip tubes, aspirated, and resuspended in 50 µl 200 mM EPPS pH 8.5.

To place the PCR strip tubes in a Mixer HC (USA Scientific), aluminum foil was wrapped around the end of a 1-mL pipet tip. These hollow aluminum cones were then placed in the 1.5-mL wells of the Mixer HC, and to each a single sample in a PCR tube was added. TCEP was then added to each sample to a final concentration of 10 mM and incubated at RT for 1-hour shaking at 700 rpm. Then iodoacetamide was added to each sample to a final concentration of 20 mM and incubated in the dark for one-hour at RT shaking at 700 rpm. Alkylation was quenched by the addition of DTT (5 mM final) and incubation at RT for 15 minutes shaking at 700 rpm. To each sample 300 ng of mass spectrometry grade LysC (FujiFilm; 121-05063) was added and samples incubated at 37C for three-hours shaking at 700 rpm. This was followed by the addition of 300 ng of sequencing grade modified trypsin (Promega; V5113) and incubation for exactly 16-hours at 37C shaking at 700 rpm.

The following day beads were magnetically removed, and the clarified samples were placed in fresh PCR strip tubes. Samples were then frozen at -80 °C to inactivate LysC/trypsin and

were stored no longer than three-days. Samples were then thawed to room temperature and to each 0.5 µL of 1% HPLC-grade formic acid was added to achieve pH < 3 and desalted using C18 StageTips according to the published protocol with minor adjustments (Tsue, Kania et al., 2023). Briefly, the StageTip was activated with 50µL of HPLC methanol, followed by addition of 50uL of Stage Tip B (80% acetonitrile in 0.1% trifluoroacetic acid (TFA)) and 50µL of Stage Tip A (0.1% TFA). After sample loading, peptides were washed with 50uL of Stage Tip A.

**O-MAP-MS data acquisition and analysis**

Peptides from StageTips were eluted using 50uL of 45% acetonitrile, 0.1% TFA into a 96-well plate. The samples were dried down and resuspended in Stage Tip A. Peptides were separated on an EASY-nLC 1200 System (Thermo Fisher Scientific) using 20 cm long fused silica capillary columns (100 µm ID, laser pulled in-house with Sutter P-2000, Novato CA) packed with 3 µm 120 Å reversed phase C18 beads (Dr. Maisch, Ammerbuch, DE). The liquid chromatography (LC) gradient was 90 min long with 5−35% B at 300 nL/min. LC solvent A was 0.5% (v/v) aq. acetic acid and LC solvent B was 0.5% (v/v) acetic acid, 80% acetonitrile. MS data was collected with a Thermo Fisher Scientific Orbitrap Fusion Lumos using a data-independent acquisition (DIA) method with a 120K resolution Orbitrap MS1 scan and 12 m/z isolation window, 15K resolution Orbitrap MS2 scans for precursors from 400-1000 m/z.

Data .raw files were converted to. mzML using MSConvert 3.0.21251-d2724a5 and spectral libraries built using MSFragger-DIA (Kong et al., 2017) (w/ FragPipe version 19.1) with quantification through DIA-NN version 1.8.232. The database search was against the UniProt human database (updated September 17th, 2021) containing 20420 sequences and 20420 reverse-sequence decoys. For the MSFragger analysis, both precursor and (initial) fragment mass tolerances were set to 20 ppm. Spectrum deisotoping, mass calibration, and parameter optimization

were enabled. Enzyme specificity was set to "stricttrypsin" and up to two missed trypsin cleavages were allowed. Oxidation of methionine, acetylation of protein N-termini, −18.0106 Da on N-terminal Glutamic acid, and −17.0265 Da on N-terminal Glutamine and Cysteine were set as variable modifications. Carbamidomethylation of Cysteine was set as a fixed modification. Maximum number of variable modifications per peptide was set to 3.

FragPipe/DIA-NN output files were processed and analyzed using the Perseus software package v1.5.6.0. Expression columns (protein MS intensities) were log2 transformed and normalized by subtracting the median log2 expression value from each expression value within each MS run. For statistical testing of significant differences in expression, a two-sample Student's t test was applied.

**Image acquisition**

Fluorescence widefield microscopy was performed on a Leica DM IL, equipped with a HC Fluotar 100x oil immersion objective with a 1.32 numerical aperture and planar correction (Leica; 11506527), a white LED light source (Leica; EL6000) and a DFC365 FX digital camera (Leica; 11547004). The following filter cubes were used: Texas Red (Leica TX2 ET; 11504180; used with Dylight-550 conjugates), Cy5 (Leica Y5 ET; 11504181, used for Alexafluor-647), GFP (Leica GFP ET; 11504174, used for Atto-488 and AF488), and DAPI (Leica DAPI ET; 11504204). Illumination intensity was adjusted using the light source manual control; acquisition times ranged from 40–1000 ms, as controlled by the Leica LASX software. Fluorescence confocal microscopy was performed on a Leica SP8X microscope (UW Keck Imaging Center), outfitted with a HC CS2 63x oil immersion objective, with 1.40 numerical aperture with both planar and apochromatic correction. The pixel size was 0.18 μm for fixed cell distance quantification, 0.06 μm for fixed cell representative images, and 0.045 for FFPE distance quantification and representative images. Samples were illuminated using a 470–670nm tunable White Light Laser system, with a typical

laser power of 0.1% for DAPI, 3% for 550 nm, and 10% for 647 nm. Gain and offset settings were adjusted to avoid pixel saturation. Images were line-averaged twice, with an average pixel dwell time of 1.58 µs. A bit-depth of 16 was used and zoom factor of 1-3 was used for all images.

**Mitochondrial reactive oxygen species (mROS) quantification**

mROS levels were measured using the MitoSox assay. Briefly, 75,000 cells were plated per well in a 96-well plate (ThermoFisher; 167425) that was pre-coated in Matrigel. Cells were allowed to grow until day 21 post-differentiation. A fresh 5mM stock solution of MitoSox Red (ThermoFisher; M36008) was created by dissolving the contents of the vial in 13 µL anhydrous DMSO. The MitoSox solution was then diluted in 1X DPBS (Thermo Fisher; 14190250) to a working concentration of 5 µM and protected from light. Cells were briefly washed once in 1X DPBS before adding 100 µL of the MitoSox red (5 µM) solution. Cells were then incubated at 37C for 30 minutes. MitoSox red solution was aspirated, and cells were washed three times with 1X DPBS. Then 100 µL of DPBS was added and the cells were imaged using a BioTek Synergy H1 with the Red Filter Set and an excitation of 530 +/-25 and emission of 590 +/-35. To measure kinetics of mROS over time, $H_2O_2$ (100 µM final) was added to the MitoSox red solution, and fluorescent measurements were made at intervals of 2 minutes for 30 minutes (Andersson et al., 2011; Chung & Duchen, 2022).

**Image processing**

Images were processed using Fiji (Schindelin et al., 2012) and ImageJ (Schneider et al., 2012), and multicolor overlays were made using the screen setting in Adobe Photoshop. All confocal images are maximum projections of z-stacks. Brightness and contrast were adjusted for display purposes using Fiji and ImageJ or Adobe Photoshop. In all cases, contrast adjustment was applied to improve signal visibility, by changing the minimum (black) and maximum (white) values only.

Automated despeckling was applied when necessary (e.g. in DNA-FISH images with weak, diffuse speckling between cells) to reduce residual background signal. To measure the 3D distances between DNA-FISH puncta for the TTN locus and DNA-FISH puncta for an experimental TID, nuclei were first segmented in Fiji. Each nucleus was then duplicated in 3D retaining all Z-slices for DAPI and for FISH signal. Then, for each 3D composite image, the minimum edge-edge distances between TTN FISH loci and TID FISH loci were measured using the object-based ImageJ Distance Analysis (DiAna) plug-in (Gilles et al., 2017), using the Gaussian-filtering setting to remove noise. At least 50 cells per condition were analyzed; the exact number is listed in each figure. The same pipeline was used to measure the distance between preTTN O-MAP signal and interacting RNA-FISH signal. DiAna was also used for object-based colocalization, and the shuffle function was used to calculate the statistical significance of object-based colocalization and generate the cumulative distribution plots.

## 3.6 ACKNOWLEDGMENTS

We thank N. Peters and A. Fenix for imaging assistance, and C. Murry, A. Bertero, and A. Fenix for the kind donation of materials. This work was supported by National Institutes of Health Grants 1R01GM138799-01 and T32GM007750, a Safeway Albertsons Early Career Award in Cancer Research, a Brotman Baty Institute Catalytic Collaborations Award, the UW Royalty Research Fund (RRF), and the UW Student Technology Fund (STF). E.E.K. was supported by the NSF DEB2016186 and the AHA 902616. Imaging at the UW Keck Center was supported by NIH S10 OD016240 and the UW Student Technology Fee. This work used an EASY-nLC1200 UHPLC and Thermo Scientific Orbitrap Fusion Lumos Tribrid mass spectrometer purchased with funding from a National Institutes of Health SIG grant S10OD021502. NGS data analysis was facilitated

through the use of advanced computational, storage, and networking infrastructure provided by the Hyak supercomputer system and funded by the STF at the University of Washington.

## 3.7 FIGURES

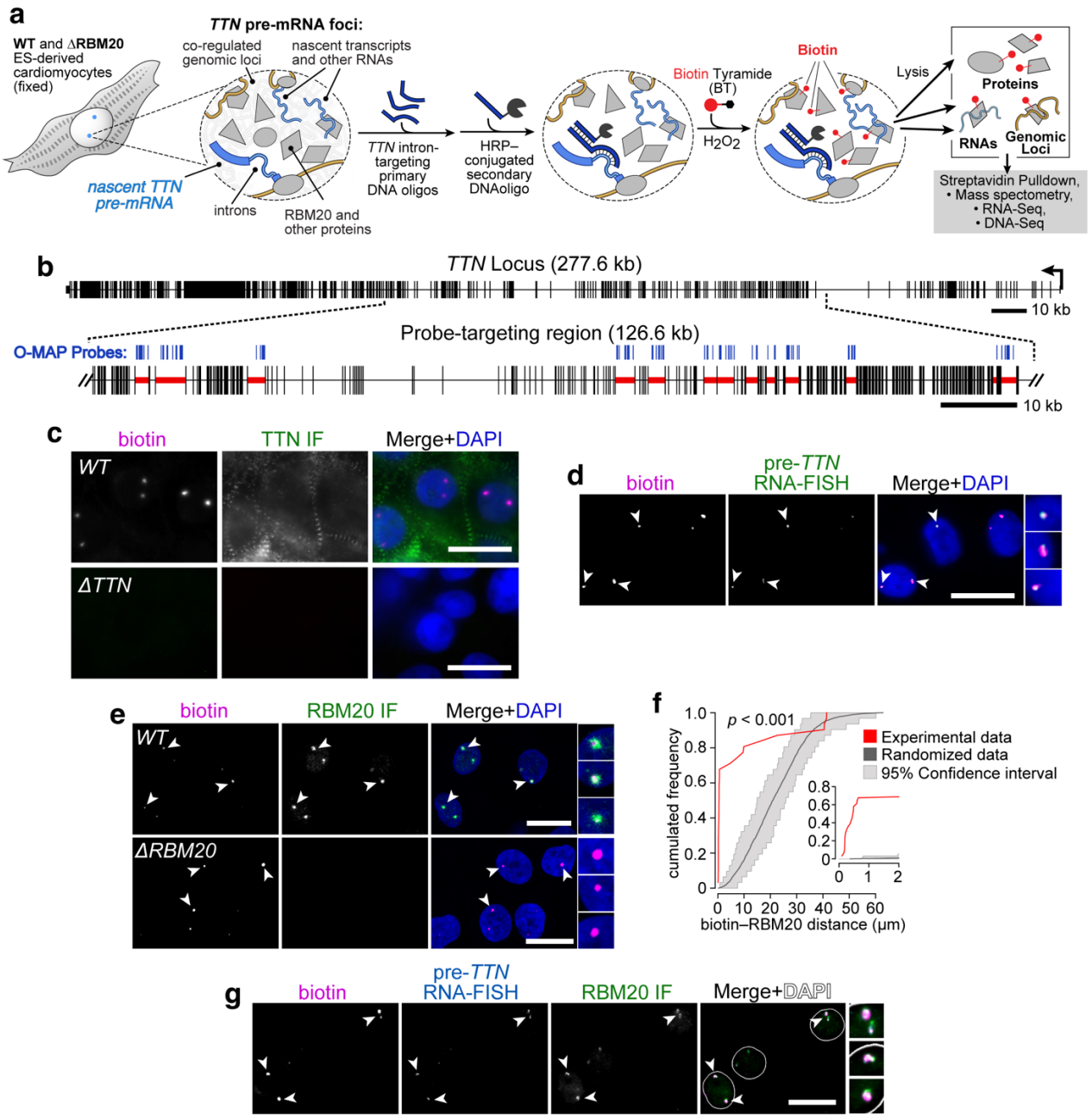



Figure 3.1. **Nascent TTN O-MAP Design and Validation. a,** Overview of TTN pre-mRNA O-MAP. ES-derived cardiomyocytes are chemically fixed, and pools of antisense DNA probes are hybridized to the introns of nascent TTN mRNA. These probes recruit a common HRP-conjugated probe that catalyzes in situ proximity-biotinylation. Biotinylated material is enriched with streptavidin beads and biomolecules of interest are isolated for downstream omic analysis. **b,** 110 probes were designed to target 14 introns within nascent TTN mRNA that have been previously

identified as binding sites for RBM20. **c,** pre-TTN O-MAP induces biotinylation in WT ES-derived cardiomyocytes (top). Deletion of the TTN promoter ablates biotinylation due to loss of the target RNA (bottom). In situ biotinylation imaged by neutravidin staining; TTN by immunofluorescence. **d,** Primary probes were split into sub-pools that enable O-MAP and RNA-FISH to be performed simultaneously. Colocalization of in situ biotinylation and pre-TTN RNA-FISH indicate specific targeting of probes. **e,** Colocalization of in situ biotinylation and RBM20 immunofluorescence in WT ES-derived cardiomyocytes (top). ΔRBM20 ES-derived cardiomyocytes permit pre-TTN O-MAP in situ biotinylation and lack RBM20 immunofluorescence (bottom). **f,** Object based colocalization of RBM20 immunofluorescence to pre-TTN O-MAP. The distribution of the distances from experimental images (red) falls outside the confidence interval of the distance distribution obtained for 100 shuffled images in which object locations are random (grey). **g,** Combined pre-TTN O-MAP, pre-TTN RNA-FISH, and RBM20 immunofluorescence demonstrate the co-localization of in situ biotinylation, pre-TTN mRNA, and RBM20 protein.

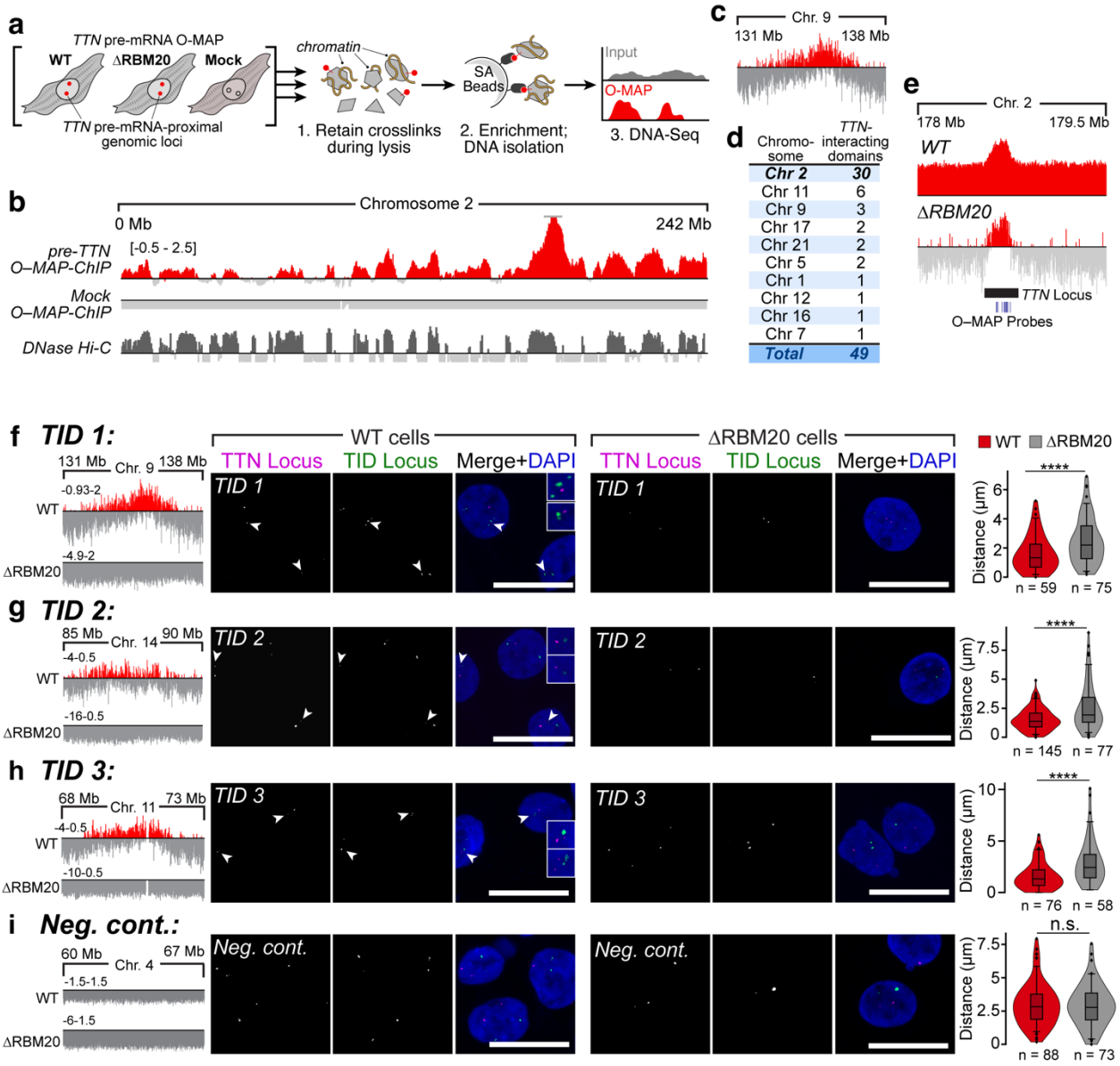



Figure 3.2. **Nascent TTN O-MAP-ChIP for chromatin interactions. a,** Approach. pre-TTN O-MAP was applied to WT and ΔRBM20 cardiomyocytes. A mock control where the HRP-recruiting probes were omitted was also used. In situ biotinylated proteins were used to enrich nearby chromatin loci, which are then quantified by DNA-sequencing. **b,** pre-TTN O-MAP ChIP enriched cis DNA loci on chromosome 2 (top) recapitulate DNase Hi-C on chromosome 2 (bottom) (Bertero et al., 2019). Mock O-MAP ChIP displays no enrichment (middle). Data are L2FC: log2(Fold change, Enriched/Input). **c,** Putative example of a trans-interacting domain (TID) on chromosome 9. Data are L2FC (Enriched/Input). **d,** A total of 72 pre-TTN O-MAP enriched chromatin domains spanned 12 chromosomes. **e,** The TTN locus (black bar) is enriched from pre-TTN O-MAP in both WT and ΔRBM20 cardiomyocytes. Enriched signal exists in both outside of O-MAP probes (blue), indicating signal is not from sequencing of probes. **f-h,** DNA-FISH for putative TIDs and the TTN locus. TIDs are located closer to the TTN locus in WT cells (left) compared to KO cells (right). Distances are measured in microns, number of cells measured per condition located below each violin. **i,** DNA-FISH for a negative control locus and the TTN locus. The negative control locus is

equidistant from the TTN locus in both WT and KO cells. Distances are measured in microns, number of cells measured per condition located below each violin. Violin plots represent aggregated data and indicate the number of cells. Internal box and whisker plots indicate median, 25th, and 75th percentile, and the 5-95 percentile range. p-values by Kruskal-Wallis test followed by Dunn's multiple comparisons vs WT unless otherwise noted. n.s. > 0.05, **** < 0.001.

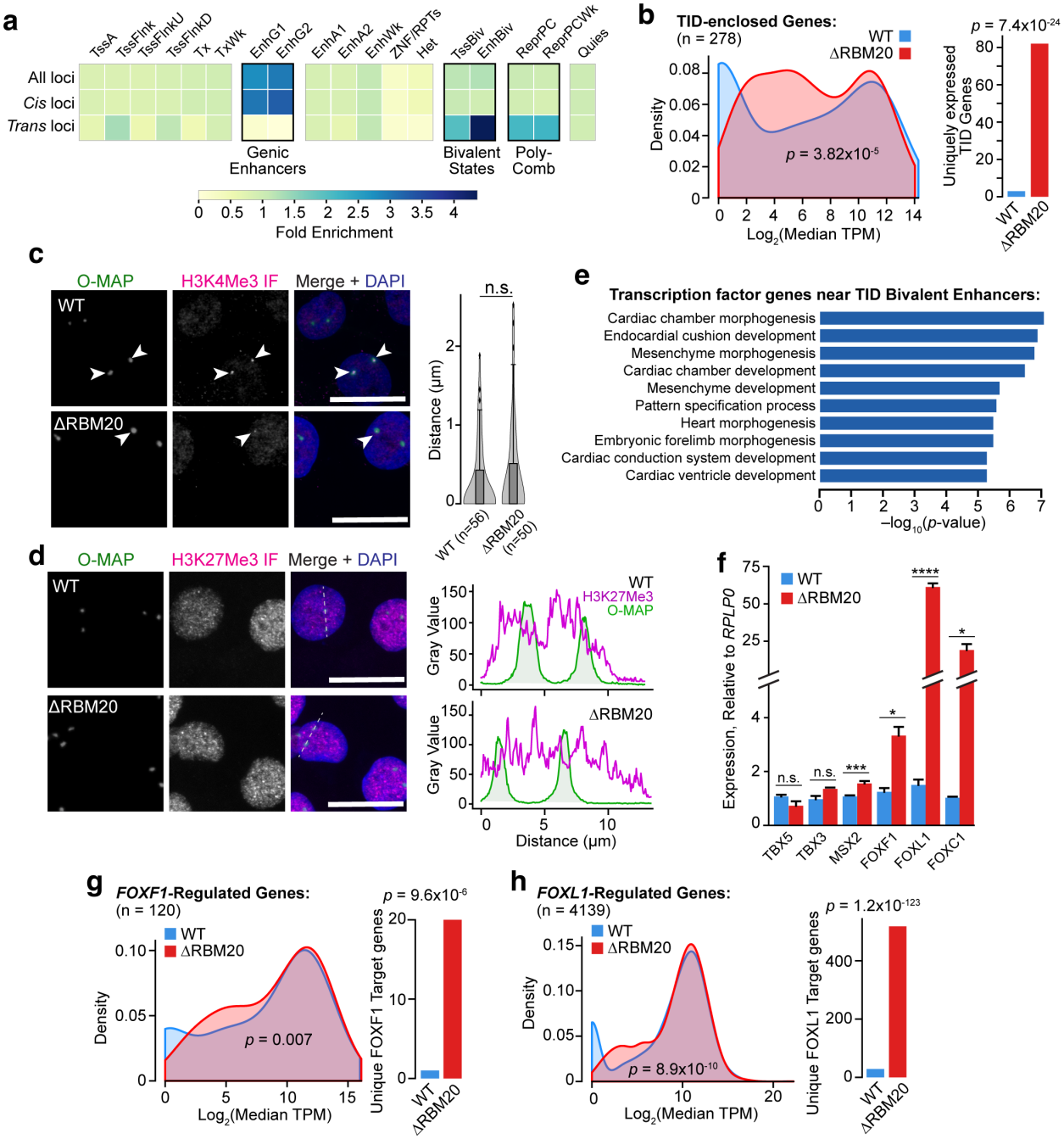



Figure 3.3. **Nascent TTN O-MAP TID characterization. a,** ChromHMM analysis reveals differential enrichment of chromatin signatures among pre-TTN cis- and trans-interacting domains (TIDs). **b,** TID-enclosed genes are higher expressed in ΔRBM20 compared to WT cells. log2 median TPM of TID-enclosed genes as a density plot for WT (blue) and ΔRBM20 (red) cells; p-value student's t-test (left). Total count of uniquely expressed TID-enclosed genes; p-value Fisher's Exact Test (right). **c,** pre-TTN O-MAP with H3K4Me3 immunofluorescence. Distance between O-MAP and H3K4Me3 puncta were not statistically different between WT (top) and ΔRBM20 cells (bottom). Distance is measured in microns and the number of cells per condition is indicated underneath each violin. **d,** pre-TTN O-MAP with H3K27Me3 immunofluorescence in WT (top) and ΔRBM20 cells (bottom). In both conditions line plots indicate O-MAP signal exists

within troughs of H3K27Me3 signal (right). Dashed yellow lines indicate the location of origin for the line plots. Distance measured in microns, signal intensity measured as grey values. **e,** Gene Ontology (GO) analysis for transcription factors within enriched TIDs. The top ten enriched terms are shown. **f,** Quantitative PCR (qPCR) for TID-encoded transcription factors in WT (blue) and ΔRBM20 (red) cells. Expression relative to RPLP0. **g,** FOXF1 regulated genes are higher expressed in ΔRBM20 cells. log2 median TPM of FOXF1 regulated genes as a density plot for WT (blue) and ΔRBM20 (red) cells; p-value student's t-test (left). Total count of uniquely expressed FOXF1 regulated genes; p-value Fisher's Exact Test (right). **h,** FOXL1 regulated genes are higher expressed in ΔRBM20 cells. log2 median TPM of FOXL1 regulated genes as a density plot for WT (blue) and ΔRBM20 (red) cells; p-value student's t-test (left). Total count of uniquely expressed FOXF1 regulated genes; p-value Fisher's Exact Test (right).

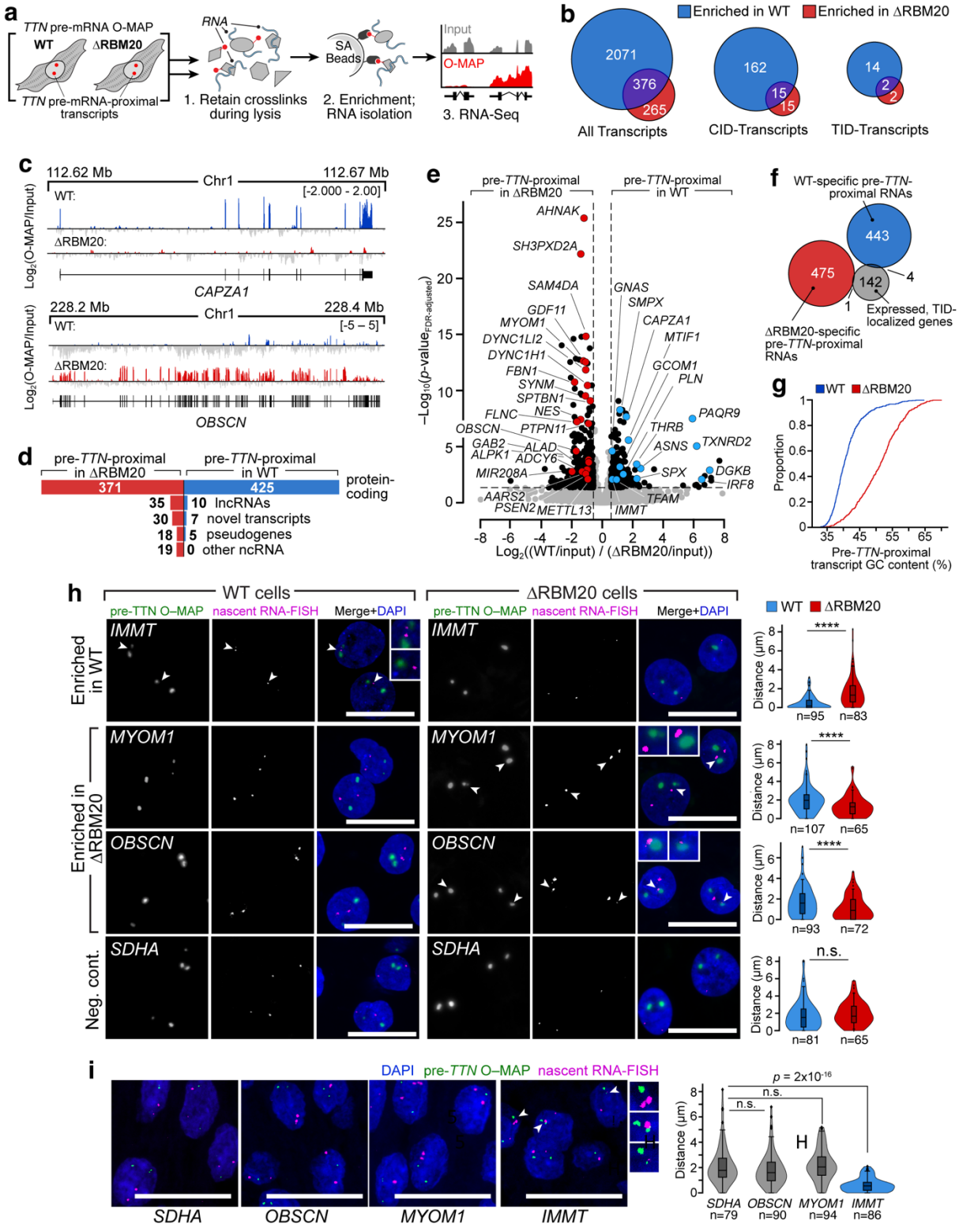



Figure 3.4. **Nascent TTN O-MAP Seq for RNA interactions. a,** Approach. pre-TTN O-MAP was applied to WT and ΔRBM20 cardiomyocytes. In situ biotinylated proteins were used to enrich nearby RNAs, which are then quantified by RNA-sequencing. **b,** Nearly half of O-MAP enriched

transcripts in ΔRBM20 (red) cells are represented in WT (blue) O-MAP enriched transcripts. This finding is consistent when considering all enriched transcripts (left), enriched transcripts originating from CIDs (middle), and transcripts originating from TIDs (right). **c,** Examples of differentially enriched pre-TTN O-MAP transcripts. CAPZA1 (top) is enriched in WT (blue) pre-TTN O-MAP-Seq. OBSCN (bottom) is enriched in ΔRBM20 (red) pre-TTN O-MAP-Seq. **d,** Summary of pre-TTN-proximal enriched RNA classes in WT and ΔRBM20 cells. **e,** Volcano plot demonstrating enrichment of O-MAP-Seq transcripts in WT cells (right) and enrichment of O-MAP-Seq transcripts in ΔRBM20 cells (left) (n=4 biological replicates). **f,** Differentially enriched pre-TTN O-MAP transcripts demonstrate unbiased enrichment for transcripts expressed from TIDs. **g,** Pre-TTN-proximal transcripts enriched in ΔRBM20 cells (red) have a higher GC content (mean Δ10%) compared to pre-TTN-proximal transcripts enriched in WT cells (blue). **h-i,** Pairwise distance analysis between pre-TTN O-MAP and nascent RNA-FISH for differentially enriched pre-TTN-proximal transcripts and a negative control. **h**, Left: Representative images in WT hESC-cardiomyocytes. Middle: Representative images in ΔRBM20 hESC-cardiomyocytes. Right: Minimal distance per cell of each pre-TTN O-MAP foci to nascent RNA-FISH signal. **i**, Representative images in Fetal heart tissue. Violin plots represent aggregated data and indicate the number of cells. Internal box and whisker plots indicate median, 25th, and 75th percentile, and the 5-95 percentile range. p-values by Kruskal-Wallis test followed by Dunn's multiple comparisons vs WT unless otherwise noted. n.s. > 0.05, **** < 0.001.

Kania *et. al.* **Figure 5.**

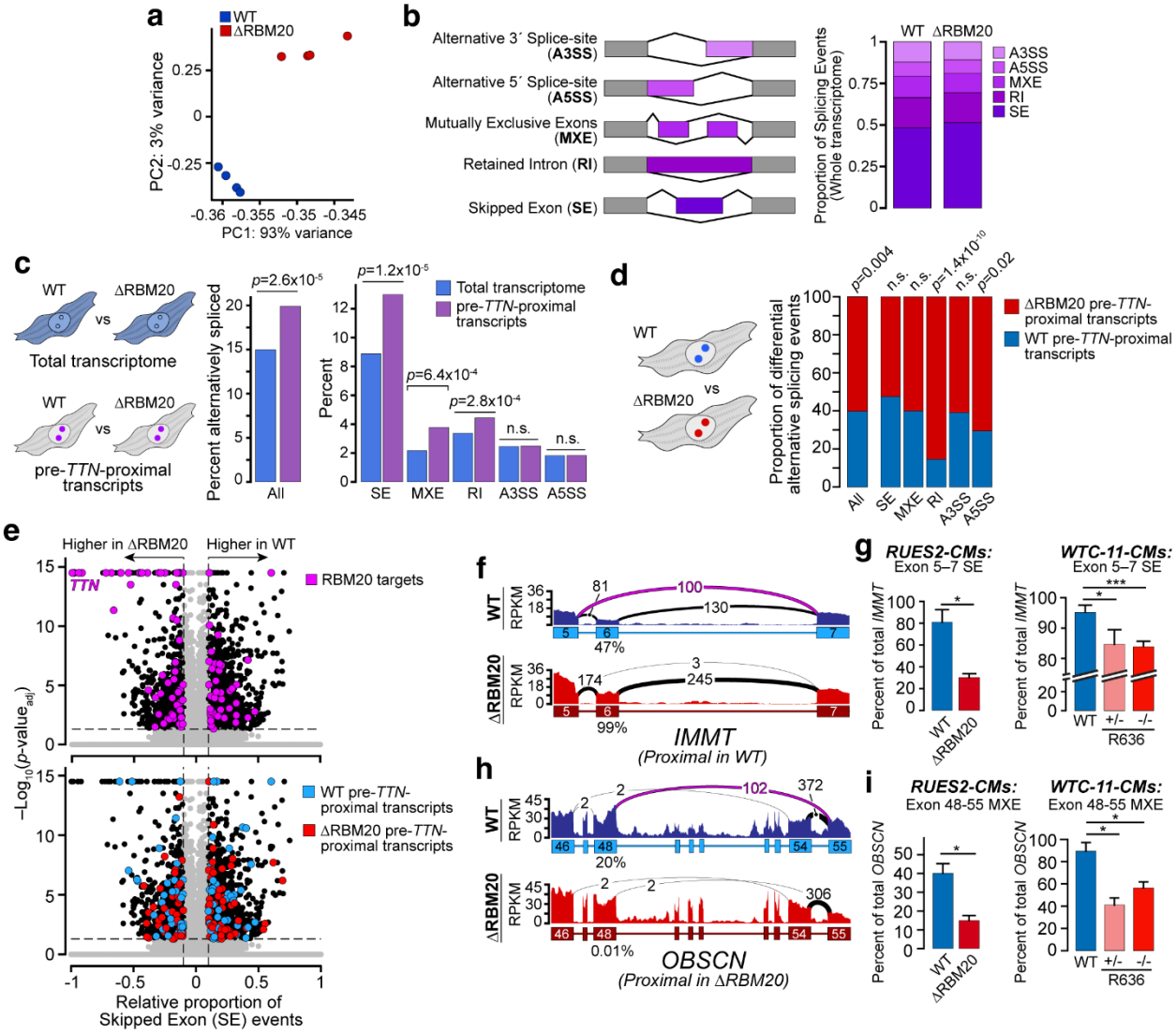



Figure 3.5. **Alternative splicing events are enriched at pre-TTN transcripts in ΔRBM20 cardiomyocytes. a,** Principal component analysis (PCA) of alternative splicing between WT (blue) and ΔRBM20 (red) replicates. **b,** Schematic of alternative splicing events considered for analysis (left). Percentage of splicing events considered differential between whole-cell WT and ΔRBM20 cardiomyocytes (FDR< 0.05, δPSI > 0.1) (right). **c,** Percentage of global differential alternative splice events (FDR< 0.05, δPSI > 0.1) between WT and ΔRBM20 cells in the whole transcriptome (blue) vs transcripts significantly enriched by O-MAP-seq (purple). **d,** Percentage of differential alternative splice events (FDR< 0.05, δPSI > 0.1) between transcripts enriched by preTTN-O-MAP-Seq in WT (blue) and ΔRBM20 (red) cardiomyocytes. **e,** Overview of transcripts undergoing differential SE alternative splicing between WT (positive δPSI) and ΔRBM20 (negative δPSI) conditions. In black are significant events (FDR< 0.05, δPSI > 0.1). Left: In red are transcripts confirmed to bind to RBM20 through ECLIP. Right: In blue are WT-specific pre-TTN-proximal transcripts and in red are ΔRBM20 pre-TTN-proximal transcripts. **f-i,** Sashimi-plot visualization (left) and qPCR validation in WT and ΔRBM20 female ESC-derived cardiomyocytes

(middle) and male WT, heterozygous R636S, and homozygous R636S iPSC-derived cardiomyocytes (right) of differentially spliced transcripts enriched at pre-TTN mRNA.

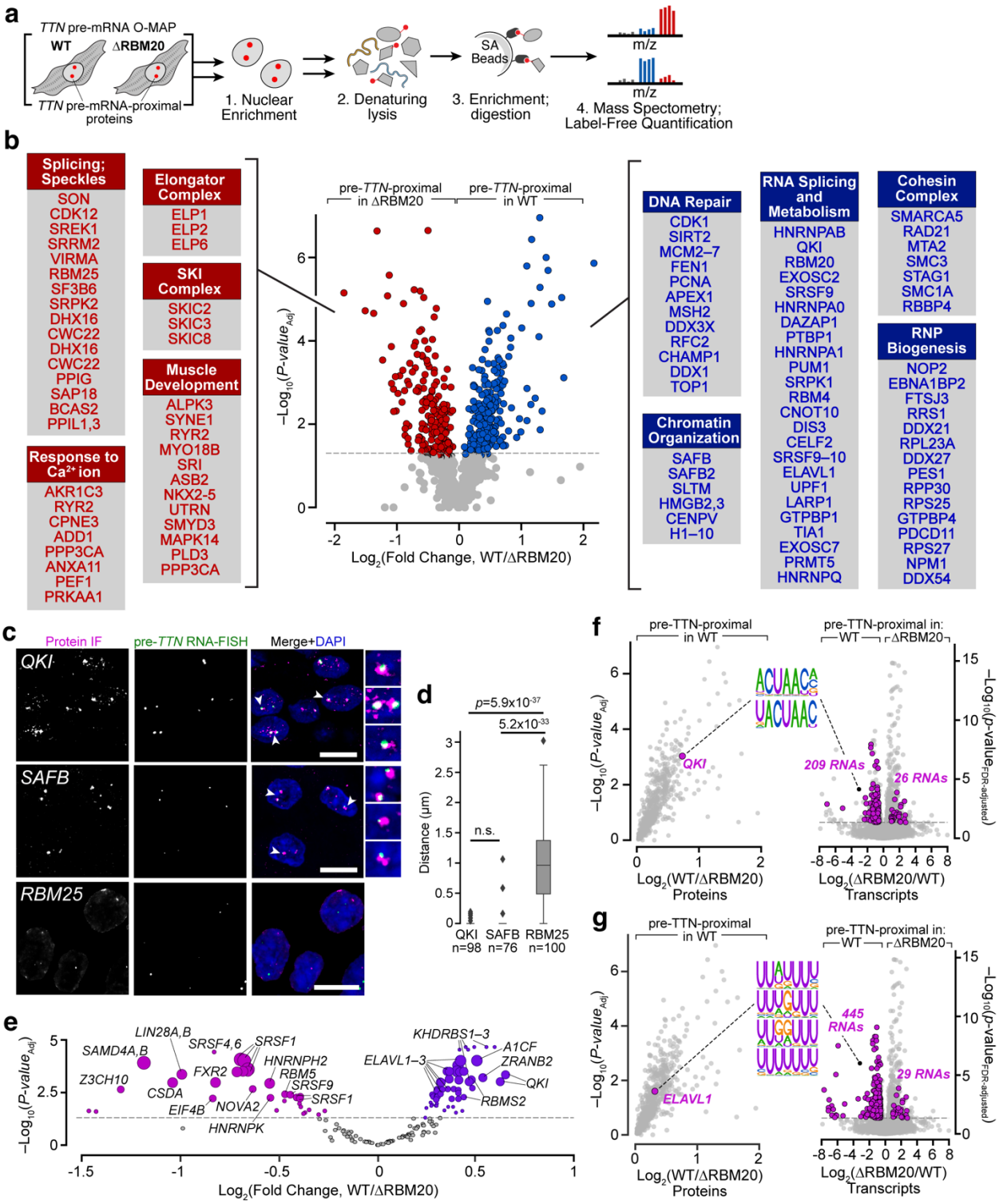



Figure 3.6. Figure 6. **The pre-TTN-factory proteome**. **a,** O-MAP-MS experimental scheme. pre-TTN O-MAP was performed in parallel in WT and ΔRBM20 CMs. Following nuclear isolation, biotinylated proteins were enriched under denaturing conditions and analyzed by mass spectrometry with label-free quantification (LFQ). **b,** Volcano plot summarizing pre-TTN O-MAP-MS hits specific for WT (blue; right) and ΔRBM20 (red; left) CMs. Prominent functional

classes and representative protein hits are shown for each. **c,** novel proteomic hits compartmentalize within pre-TTN factories. Immunofluorescence for two WT-specific hits, QKI and SAFB, and negative control, RBM25, in WT human fetal heart tissue sections. pre-TTN factories were visualized by RNA-FISH. Arrowheads denote conspicuous colocalization, highlighted in zoomed images (right). All scale bars, 20 μm. **d,** quantification of 3D-distances between pre-TTN and each protein hit. Box and whisker plots indicate median, 25th, and 75th percentile, and the 5–95 percentile range. p-values by Kruskal-Wallis test followed by Dunn's multiple comparisons, vs RBM25. **e,** Differential enrichment of RNA-binding protein recognition motifs in WT- and ΔRBM20-specific pre-TTN-proximal transcripts. The size of each motif's circle is proportional to the total number of transcripts harboring that motif. Significant differential motifs are highlighted in purple (Log2(Fold Change) >0.58; FDR < 0.05). **f–g,** pre-TTN factories co-compartmentalize RNA-binding proteins (RBPs) along with their target transcripts. Left: significant enrichment of each protein (highlighted in magenta) in O-MAP-MS. Data are the same as in b. Middle: Logos for the major RNA recognition motifs for each RBP. Right: Incidence of these motifs within the WT- and ΔRBM20-pre-TTN-proximal transcriptomes, observed by O-MAP-Seq (highlighted in magenta). Data are the same as in (Fig. 4e), but with inverted x-axes. The number of transcripts bearing each binding motif is indicated. **f,** QKI; g, ELAVL1.

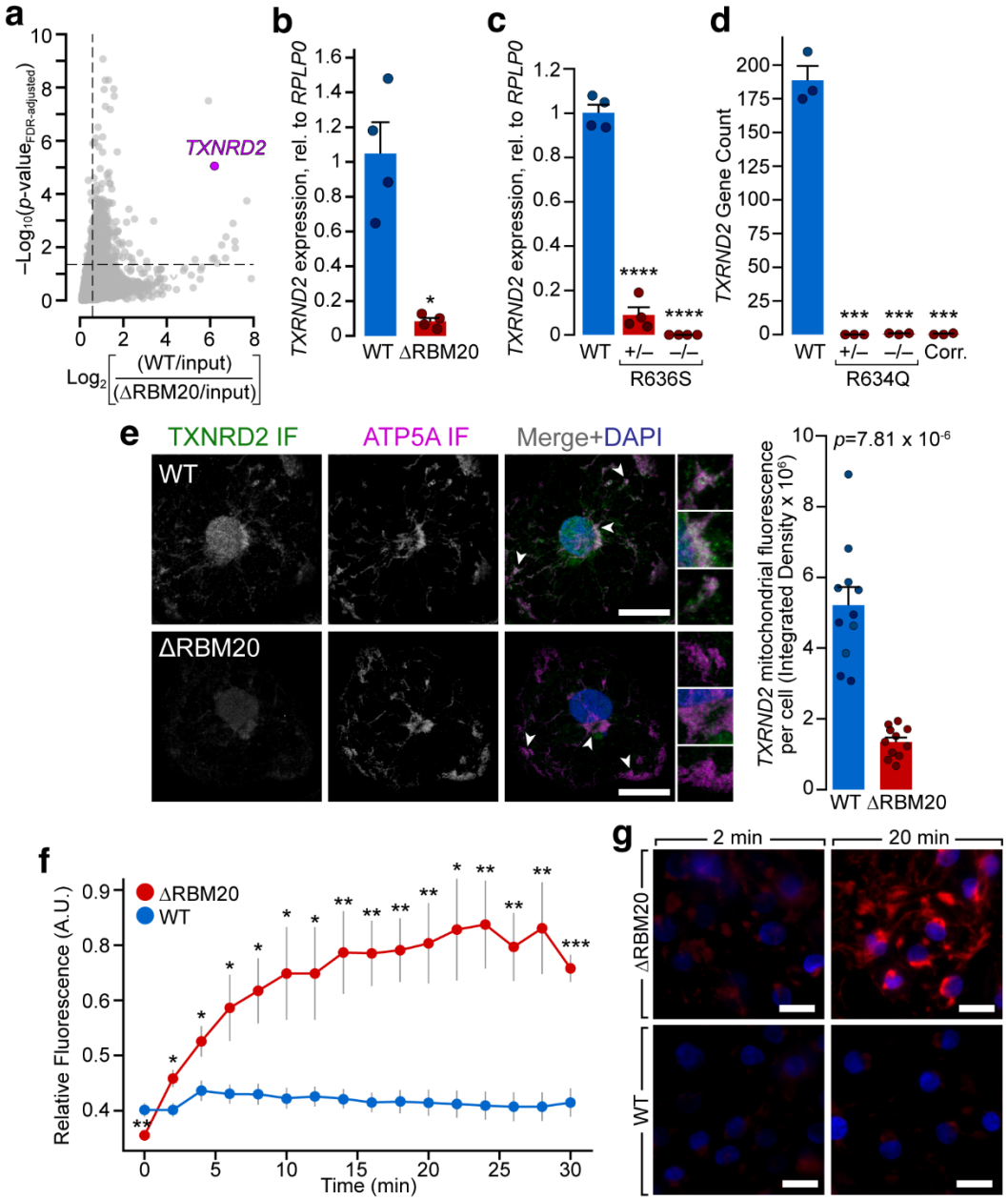



Figure 3.7. **Loss of RBM20 downregulates TXNRD2 expression and impairs mROS-handling. a,** The TXNRD2 mRNA is strongly enriched in WT CMs. Data are the same as (Fig. 4e). b–c, qRT-PCR analysis demonstrating loss of TXNRD2 expression in: **b,** RUES2-derived ΔRBM20 CMs and **c,** WTC-11-derived CMs bearing DCM-associated point mutations. Expression normalized to RPLP0. ( 4 biological replicates run in technical quadruplicate, Significance: paired 2-tailed student's T-test, * < 0.05; *** < 0.005). **d,** TXNRD2 expression is ablated in mouse models of RBM20-DCM, and is not restored upon genetic correction of the RBM20 allele. Data from (Nishiyama, 2022). **e,** Left: Representative confocal images of TXNRD2 and the mitochondrial marker ATP5A in WT and ΔRBM20 cells. Overlap (or lack thereof) denoted by arrowheads and highlighted in zoomed images. Right: TXNRD2 protein levels, measured by relative fluorescence of TXNRD2 to ATP5A. Significance: paired 2-tailed student's T-test, *** < 0.005. **f,** ΔRBM20 cells are deficient in mROS-handling. MitoSox timecourse assays performed

with 100 μM $H_2O_2$ challenge. Significance: paired 2-tailed student's t-test, * < 0.05, **<0.01, ***<0.005, relative to the matched WT time point (n=4). **g,** Representative images of MitoSOX fluorescence at 2-min and 20 min post $H_2O_2$ challenge. All scale bars, 20 μm.

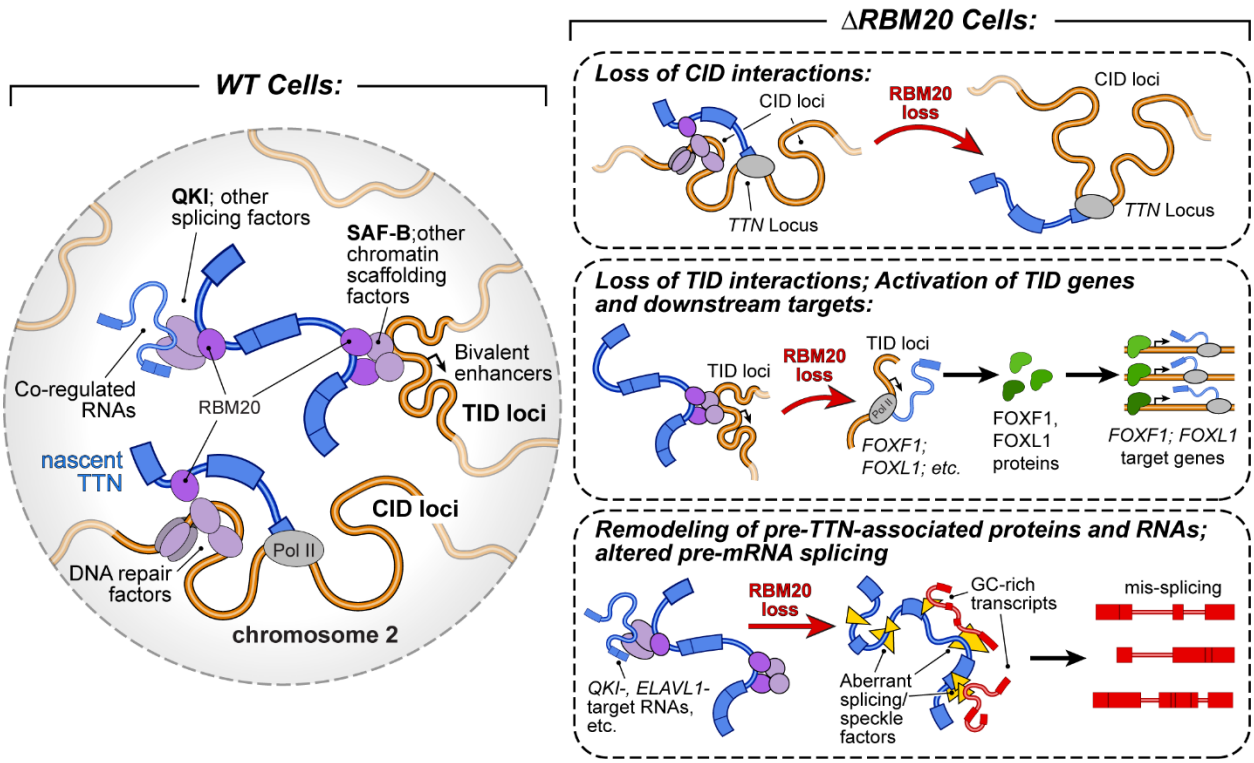



Figure 3.8**. Architectural model of the pre-TTN factory**. Left: overview of the compartment-wide interactions discovered by O-MAP. For simplicity, factory components are depicted as directly binding to pre-TTN and RBM20, though these may interact through higher-order assemblies or biomolecular condensates. Our data cannot discriminate between interactions with nascent transcripts and incompletely-spliced intermediate species, both of which are depicted. "Co-regulated RNAs:" QKI- and ELAVL1-bound transcripts, among others (Fig. 3.4). Right: broad summary of architectural alterations induced by loss of RBM20.

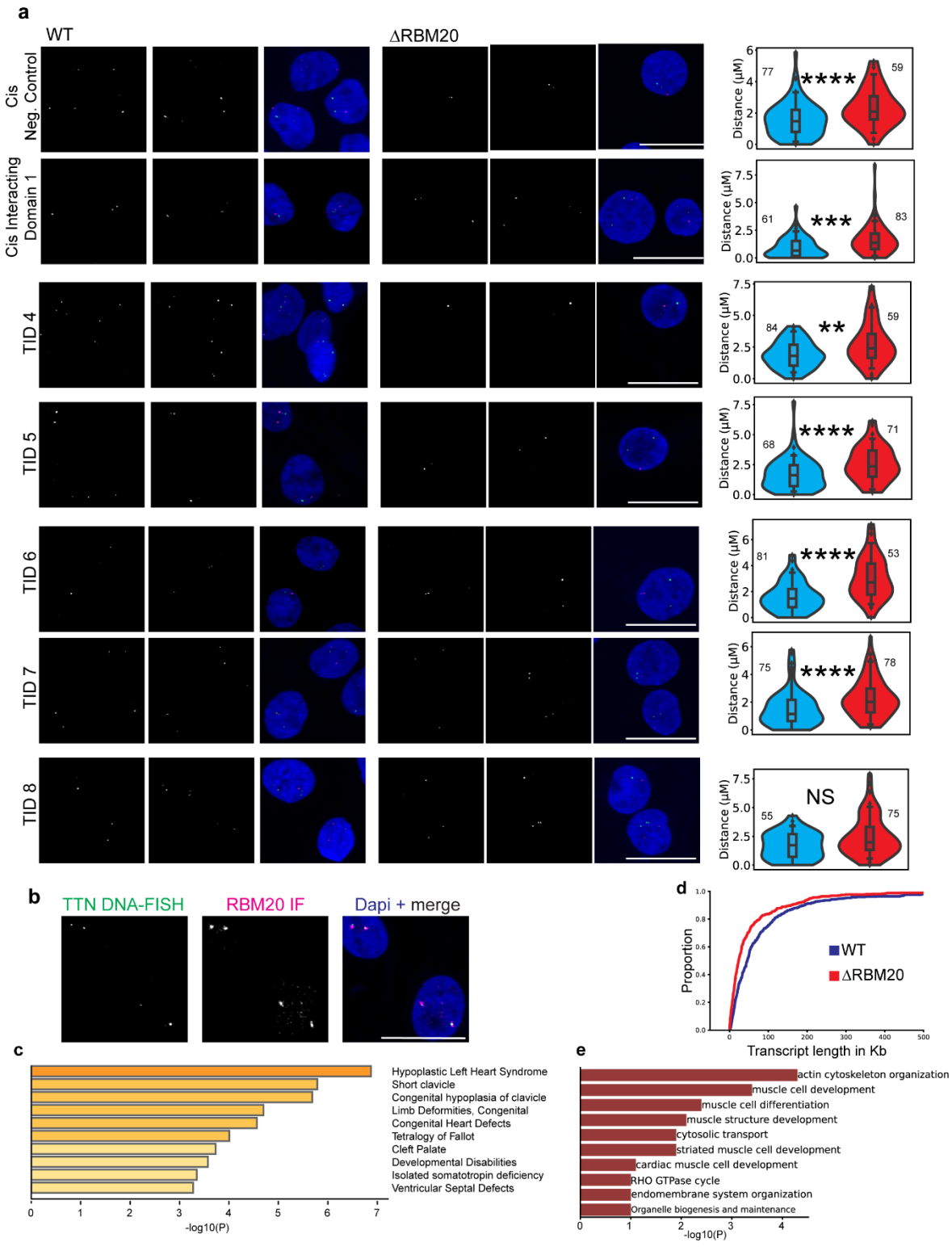

a
WT
ΔRBM20
Cis Neg. Control
Cis Interacting Domain 1
TID 4
TID 5
TID 6
TID 7
TID 8
Distance (μM)
77
59
****
61
83
***
84
59
**
68
71
****
81
53
****
75
78
****
55
75
NS
b
TTN DNA-FISH
RBM20 IF
Dapi + merge
c
Hypoplastic Left Heart Syndrome
Short clavicle
Congenital hypoplasia of clavicle
Limb Deformities, Congenital
Congenital Heart Defects
Tetralogy of Fallot
Cleft Palate
Developmental Disabilities
Isolated somatotropin deficiency
Ventricular Septal Defects
-log10(P)
d
Proportion
WT
ΔRBM20
Transcript length in Kb
e
actin cytoskeleton organization
muscle cell development
muscle cell differentiation
muscle structure development
cytosolic transport
striated muscle cell development
cardiac muscle cell development
RHO GTPase cycle
endomembrane system organization
Organelle biogenesis and maintenance
-log10(P)

Figure 3.9**. Supplemental**. **a,** DNA-FISH for one cis-interacting negative control (top), one cis-interacting domain (second), and 5 putative TIDs. Representative DNA-FISH images for each target are shown in WT (left) and ΔRBM20 cells (middle). Minimal 3-dimensional distances between TTN locus DNA-FISH and the putative cis-interacting domain or TID were measured per cell in WT and ΔRBM20 cells (right). Distances are measured in microns. Violin plots represent aggregated data and indicate the number of cells. Internal box and whisker plots indicate median, 25th, and 75th percentile, and the 5-95 percentile range. p-values by Kruskal-Wallis test followed by Dunn's multiple comparisons vs WT unless otherwise noted. n.s. > 0.05, **** < 0.001, *** < 0.005, ** < 0.01. **b,** Colocalization of TTN locus DNA-FISH with RBM20 immunofluorescence demonstrates the specifity of the designed DNA-FISH probe pool. **c,** Top ten enriched DisGeNET disease pathway ontologies for transcription factors encoded in WT enriched bivalent enhancer regions demonstrate their roles in cardiac diseases, including hypoplasic left heart syndrome (top) and ventricular septal defects (bottom). **d,** WT pre-TTN-proximal transcripts are longer than ΔRBM20 pre-TTN-proximal transcripts. **e,** Top ten enriched gene ontology terms for pre-TTN-proximal transcripts that are differentially alternatively spliced.

## Chapter 4. CLOSING SUMMARY

In this thesis, I first expanded the O-MAP suite of technologies to comprehensively capture RNA interactions with genomic loci and other RNAs. I then optimized the O-MAP toolkit to visualize low-abundant RNAs and to capture complete interactomes (RNA-DNA, RNA-RNA, and RNA-protein interactions), using a minimum of 1.2 million cells. Applying O-MAP, I uncovered critical insights into gene silencing lncRNAs, revealing their interactions with genomic loci, RNAs, and proteins at specific sites such as the Kcnq1ot1 silencing body and the Xi. By focusing on the nascent Xist transcript, I differentiated key components of the X-chromosome inactivation center, enhancing our understanding of how gene silencing lncRNAs function and localize on chromosomes.

Furthermore, I extended the O-MAP technology to study the nascent TTN factory in a model of RBM20-DCM, shedding light on its biomolecular composition in both normal and diseased conditions. These investigations identified aberrant alternative splicing and the downregulation of a critical mROS handling component in the DCM state, and functional validation showcased impaired mROS handling in cells with the RBM20-DCM phenotype.

The primary data output from the suite of O-MAP tools is the proximity and localization of biomolecules around RNA-scaffolded subnuclear structures, necessitating functional inferences that require testing. Leveraging O-MAP, I demonstrated functional insights through alternative splicing, ASO knock-down, and mROS handling assays. Future directions are delineated in the respective chapters, ranging from genetic manipulations of genomic loci and RNAs to the perturbation of proteins via knockdown or knockout approaches. These assays are tailored to the specific biomolecules under study and the experimental context.

In summary, O-MAP offers a potent toolkit for characterizing RNA-scaffolded structures across diverse biological and disease settings. Its readout stimulates hypothesis generation, propelling investigations into how these structures exert their function. Validated components emerge as biomarkers, drug targets, or regulators. Overall, this thesis provides accessible, universal tools to unravel the composition of challenging subnuclear RNA-scaffolded bodies and exemplifies strategies for functionally dissecting their roles.

# REFERENCES


Adiconis, X., Borges-Rivera, D., Satija, R., DeLuca, D. S., Busby, M. A., Berlin, A. M., Sivachenko, A., Thompson, D. A., Wysoker, A., Fennell, T., Gnirke, A., Pochet, N., Regev, A., & Levin, J. Z. (2013). Comparative analysis of RNA sequencing methods for degraded or low-input samples. *Nature Methods*, *10*(7), 623–629. https://doi.org/10.1038/nmeth.2483

Al-Maqtari, T., Hong, K. U., Vajravelu, B. N., Moktar, A., Cao, P., Moore, J. B., & Bolli, R. (2017). Transcription factor-induced activation of cardiac gene expression in human c-kit+ cardiac progenitor cells. *PloS One*, *12*(3), e0174242. https://doi.org/10.1371/journal.pone.0174242

Amemiya, H. M., Kundaje, A., & Boyle, A. P. (2019). The ENCODE Blacklist: Identification of Problematic Regions of the Genome. *Scientific Reports*, *9*(1), 9354. https://doi.org/10.1038/s41598-019-45839-z

Andersson, D. C., Fauconnier, J., Yamada, T., Lacampagne, A., Zhang, S.-J., Katz, A., & Westerblad, H. (2011). Mitochondrial production of reactive oxygen species contributes to the β-adrenergic stimulation of mouse cardiomycytes. *The Journal of Physiology*, *589*(Pt 7), 1791–1801. https://doi.org/10.1113/jphysiol.2010.202838

Andresini, O., Rossi, M. N., Matteini, F., Petrai, S., Santini, T., & Maione, R. (2019). The long non-coding RNA Kcnq1ot1 controls maternal p57 expression in muscle cells by promoting H3K27me3 accumulation to an intragenic MyoD-binding region. *Epigenetics & Chromatin*, *12*(1), 8. https://doi.org/10.1186/s13072-019-0253-1

Avner, P., & Heard, E. (2001). X-chromosome inactivation: Counting, choice and initiation. *Nature Reviews. Genetics*, *2*(1), 59–67. https://doi.org/10.1038/35047580

*Babraham Bioinformatics—Trim Galore!* (n.d.). Retrieved May 1, 2023, from https://www.bioinformatics.babraham.ac.uk/projects/trim_galore/

Bachellerie, J. P., Cavaillé, J., & Hüttenhofer, A. (2002). The expanding snoRNA world. *Biochimie*, *84*(8), 775–790. https://doi.org/10.1016/s0300-9084(02)01402-5

Bai, B., Liu, H., & Laiho, M. (2014). Small RNA expression and deep sequencing analyses of the nucleolus reveal the presence of nucleolus-associated microRNAs. *FEBS Open Bio*, *4*, 441–449. https://doi.org/10.1016/j.fob.2014.04.010

Balas, M. M., & Johnson, A. M. (2018). Exploring the mechanisms behind long noncoding RNAs and cancer. *Non-Coding RNA Research*, *3*(3), 108–117. https://doi.org/10.1016/j.ncrna.2018.03.001

Bar, D. Z., Atkatsh, K., Tavarez, U., Erdos, M. R., Gruenbaum, Y., & Collins, F. S. (2018). Biotinylation by antibody recognition—A method for proximity labeling. *Nature Methods*, *15*(2), 127–133. https://doi.org/10.1038/nmeth.4533

Barlow, D. P., & Bartolomei, M. S. (2014). Genomic Imprinting in Mammals. *Cold Spring Harbor Perspectives in Biology*, *6*(2), a018382. https://doi.org/10.1101/cshperspect.a018382

Barutcu, A. R., Wu, M., Braunschweig, U., Dyakov, B. J. A., Luo, Z., Turner, K. M., Durbic, T., Lin, Z.-Y., Weatheritt, R. J., Maass, P. G., Gingras, A.-C., & Blencowe, B. J. (2022). Systematic mapping of nuclear domain-associated transcripts reveals speckles and lamina as hubs of functionally distinct retained introns. *Molecular Cell*, *82*(5), 1035-1052.e9. https://doi.org/10.1016/j.molcel.2021.12.010

Bassett, A. R., Akhtar, A., Barlow, D. P., Bird, A. P., Brockdorff, N., Duboule, D., Ephrussi, A., Ferguson-Smith, A. C., Gingeras, T. R., Haerty, W., Higgs, D. R., Miska, E. A., &

Ponting, C. P. (2014). Considerations when investigating lncRNA function in vivo. *ELife*, *3*, e03058. https://doi.org/10.7554/eLife.03058

Beckedorff, F. C., Ayupe, A. C., Crocci-Souza, R., Amaral, M. S., Nakaya, H. I., Soltys, D. T., Menck, C. F. M., Reis, E. M., & Verjovski-Almeida, S. (2013). The Intronic Long Noncoding RNA ANRASSF1 Recruits PRC2 to the RASSF1A Promoter, Reducing the Expression of RASSF1A and Increasing Cell Proliferation. *PLoS Genetics*, *9*(8), e1003705. https://doi.org/10.1371/journal.pgen.1003705

Beliveau, B. J., Kishi, J. Y., Nir, G., Sasaki, H. M., Saka, S. K., Nguyen, S. C., Wu, C., & Yin, P. (2018). OligoMiner provides a rapid, flexible environment for the design of genome-scale oligonucleotide in situ hybridization probes. *Proceedings of the National Academy of Sciences*, *115*(10), E2183–E2192. https://doi.org/10.1073/pnas.1714530115

Belotserkovskii, B. P., Tornaletti, S., D'Souza, A. D., & Hanawalt, P. C. (2018). R-loop generation during transcription: Formation, processing and cellular outcomes. *DNA Repair*, *71*, 69–81. https://doi.org/10.1016/j.dnarep.2018.08.009

Beraldi, R., Li, X., Martinez Fernandez, A., Reyes, S., Secreto, F., Terzic, A., Olson, T. M., & Nelson, T. J. (2014). Rbm20-deficient cardiogenesis reveals early disruption of RNA processing and sarcomere remodeling establishing a developmental etiology for dilated cardiomyopathy. *Human Molecular Genetics*, *23*(14), 3779–3791. https://doi.org/10.1093/hmg/ddu091

Bernat, V., & Disney, M. D. (2015). RNA structures as mediators of neurological diseases and as drug targets. *Neuron*, *87*(1), 28–46. https://doi.org/10.1016/j.neuron.2015.06.012

Bertero, A., Fields, P. A., Ramani, V., Bonora, G., Yardimci, G. G., Reinecke, H., Pabon, L., Noble, W. S., Shendure, J., & Murry, C. E. (2019). Dynamics of genome reorganization

during human cardiogenesis reveal an RBM20-dependent splicing factory. *Nature Communications*, *10*(1), Article 1. https://doi.org/10.1038/s41467-019-09483-5

Binder, J. X., Pletscher-Frankild, S., Tsafou, K., Stolte, C., O'Donoghue, S. I., Schneider, R., & Jensen, L. J. (2014). COMPARTMENTS: Unification and visualization of protein subcellular localization evidence. *Database: The Journal of Biological Databases and Curation*, *2014*, bau012. https://doi.org/10.1093/database/bau012

Blanco, E., González-Ramírez, M., Alcaine-Colet, A., Aranda, S., & Croce, L. D. (2020). The Bivalent Genome: Characterization, Structure, and Regulation. *Trends in Genetics*, *36*(2), 118–131. https://doi.org/10.1016/j.tig.2019.11.004

Blighe, K. (2023). *EnhancedVolcano: Publication-ready volcano plots with enhanced colouring and labeling* [R]. https://github.com/kevinblighe/EnhancedVolcano (Original work published 2018)

Bochman, M. L., & Schwacha, A. (2009). The Mcm Complex: Unwinding the Mechanism of a Replicative Helicase. *Microbiology and Molecular Biology Reviews : MMBR*, *73*(4), 652–683. https://doi.org/10.1128/MMBR.00019-09

Boix, C. A., James, B. T., Park, Y. P., Meuleman, W., & Kellis, M. (2021). Regulatory genomic circuitry of human disease loci by integrative epigenomics. *Nature*, *590*(7845), 300–307. https://doi.org/10.1038/s41586-020-03145-z

Brauch, K. M., Karst, M. L., Herron, K. J., de Andrade, M., Pellikka, P. A., Rodeheffer, R. J., Michels, V. V., & Olson, T. M. (2009). Mutations in ribonucleic acid binding protein gene cause familial dilated cardiomyopathy. *Journal of the American College of Cardiology*, *54*(10), 930–941. https://doi.org/10.1016/j.jacc.2009.05.038

Brockdorff, N. (2019). Localized accumulation of Xist RNA in X chromosome inactivation. *Open Biology*, *9*(12), 190213. https://doi.org/10.1098/rsob.190213

Cai, Z., Cao, C., Ji, L., Ye, R., Wang, D., Xia, C., Wang, S., Du, Z., Hu, N., Yu, X., Chen, J., Wang, L., Yang, X., He, S., & Xue, Y. (2020). RIC-seq for global in situ profiling of RNA-RNA spatial interactions. *Nature*, *582*(7812), 432–437. https://doi.org/10.1038/s41586-020-2249-1

Cao, M., Zhao, J., & Hu, G. (2019). Genome-wide methods for investigating long noncoding RNAs. *Biomedicine & Pharmacotherapy*, *111*, 395–401. https://doi.org/10.1016/j.biopha.2018.12.078

Carter, D., Chakalova, L., Osborne, C. S., Dai, Y., & Fraser, P. (2002). Long-range chromatin regulatory interactions in vivo. *Nature Genetics*, *32*(4), 623–626. https://doi.org/10.1038/ng1051

Caudron-Herger, M., Pankert, T., & Rippe, K. (2016). Regulation of nucleolus assembly by non-coding RNA polymerase II transcripts. *Nucleus (Austin, Tex.)*, *7*(3), 308–318. https://doi.org/10.1080/19491034.2016.1190890

Caudron-Herger, M., Pankert, T., Seiler, J., Németh, A., Voit, R., Grummt, I., & Rippe, K. (2015). Alu element-containing RNAs maintain nucleolar structure and function. *The EMBO Journal*, *34*(22), 2758–2774. https://doi.org/10.15252/embj.201591458

Cech, T. R., & Steitz, J. A. (2014). The noncoding RNA revolution-trashing old rules to forge new ones. *Cell*, *157*(1), 77–94. https://doi.org/10.1016/j.cell.2014.03.008

Chen, E. Y., Tan, C. M., Kou, Y., Duan, Q., Wang, Z., Meirelles, G. V., Clark, N. R., & Ma'ayan, A. (2013). Enrichr: Interactive and collaborative HTML5 gene list enrichment analysis tool. *BMC Bioinformatics*, *14*(1), 128. https://doi.org/10.1186/1471-2105-14-128

Chen, J., Rong, X., Liu, X., Zheng, D., Rong, X., Chen, F., Zhao, P., Liu, F., & Ruan, J. (2020). FOXC2 is a prognostic biomarker and contributes to the growth and invasion of human hepatocellular carcinoma. *Cancer Cell International*, *20*, 196. https://doi.org/10.1186/s12935-020-01265-0

Chen, P., Xiao, Y., Wang, Y., Zheng, Z., Chen, L., Yang, X., Li, J., Wu, W., & Zhang, S. (2020). Intracellular calcium current disorder and disease phenotype in OBSCN mutant iPSC-based cardiomyocytes in arrhythmogenic right ventricular cardiomyopathy. *Theranostics*, *10*(24), 11215–11229. https://doi.org/10.7150/thno.45172

Chen, X., Liu, Y., Xu, C., Ba, L., Liu, Z., Li, X., Huang, J., Simpson, E., Gao, H., Cao, D., Sheng, W., Qi, H., Ji, H., Sanderson, M., Cai, C.-L., Li, X., Yang, L., Na, J., Yamamura, K., … Sun, N. (2021). QKI is a critical pre-mRNA alternative splicing regulator of cardiac myofibrillogenesis and contractile function. *Nature Communications*, *12*(1), Article 1. https://doi.org/10.1038/s41467-020-20327-5

Chen, X., Yin, J., Cao, D., Xiao, D., Zhou, Z., Liu, Y., & Shou, W. (2021). The Emerging Roles of the RNA Binding Protein QKI in Cardiovascular Development and Function. *Frontiers in Cell and Developmental Biology*, *9*, 668659. https://doi.org/10.3389/fcell.2021.668659

Chen, Y., & Belmont, A. S. (2019). Genome organization around nuclear speckles. *Current Opinion in Genetics & Development*, *55*, 91–99. https://doi.org/10.1016/j.gde.2019.06.008

Chen, Y., Zhang, Y., Wang, Y., Zhang, L., Brinkman, E. K., Adam, S. A., Goldman, R., van Steensel, B., Ma, J., & Belmont, A. S. (2018). Mapping 3D genome organization relative

to nuclear compartments using TSA-Seq as a cytological ruler. *The Journal of Cell Biology*, *217*(11), 4025–4048. https://doi.org/10.1083/jcb.201807108

Chiang, J.-C., Jiang, J., Newburger, P. E., & Lawrence, J. B. (2018). Trisomy silencing by XIST normalizes Down syndrome cell pathogenesis demonstrated for hematopoietic defects in vitro. *Nature Communications*, *9*, 5180. https://doi.org/10.1038/s41467-018-07630-y

Chiesa, N., De Crescenzo, A., Mishra, K., Perone, L., Carella, M., Palumbo, O., Mussa, A., Sparago, A., Cerrato, F., Russo, S., Lapi, E., Cubellis, M. V., Kanduri, C., Cirillo Silengo, M., Riccio, A., & Ferrero, G. B. (2012). The KCNQ1OT1 imprinting control region and non-coding RNA: New properties derived from the study of Beckwith–Wiedemann syndrome and Silver–Russell syndrome cases. *Human Molecular Genetics*, *21*(1), 10–25. https://doi.org/10.1093/hmg/ddr419

Chu, C., & Chang, H. Y. (2018). ChIRP-MS: RNA-Directed Proteomic Discovery. *Methods in Molecular Biology (Clifton, N.J.)*, *1861*, 37–45. https://doi.org/10.1007/978-1-4939-8766-5_3

Chu, C., Quinn, J., & Chang, H. Y. (2012). Chromatin isolation by RNA purification (ChIRP). *Journal of Visualized Experiments: JoVE*, *61*, 3912. https://doi.org/10.3791/3912

Chu, C., Zhang, Q. C., da Rocha, S. T., Flynn, R. A., Bharadwaj, M., Calabrese, J. M., Magnuson, T., Heard, E., & Chang, H. Y. (2015). Systematic discovery of Xist RNA binding proteins. *Cell*, *161*(2), 404–416. https://doi.org/10.1016/j.cell.2015.03.025

Chung, C.-Y., & Duchen, M. R. (2022). A Plate Reader-Based Measurement of the Cellular ROS Production Using Dihydroethidium and MitoSOX. In N. Tomar (Ed.), *Mitochondria: Methods and Protocols* (pp. 333–337). Springer US. https://doi.org/10.1007/978-1-0716-2309-1_24

Collins, M., Li, Y., & Bowser, R. (2020). RBM45 associates with nuclear stress bodies and forms nuclear inclusions during chronic cellular stress and in neurodegenerative diseases. *Acta Neuropathologica Communications*, *8*(1), 91. https://doi.org/10.1186/s40478-020-00965-y

Colognori, D., Sunwoo, H., Kriz, A. J., Wang, C.-Y., & Lee, J. T. (2019). Xist Deletional Analysis Reveals an Interdependency between Xist RNA and Polycomb Complexes for Spreading along the Inactive X. *Molecular Cell*, *74*(1), 101-117.e10. https://doi.org/10.1016/j.molcel.2019.01.015

Conrad, M., Jakupoglu, C., Moreno, S. G., Lippl, S., Banjac, A., Schneider, M., Beck, H., Hatzopoulos, A. K., Just, U., Sinowatz, F., Schmahl, W., Chien, K. R., Wurst, W., Bornkamm, G. W., & Brielmeier, M. (2004). Essential Role for Mitochondrial Thioredoxin Reductase in Hematopoiesis, Heart Development, and Heart Function. *Molecular and Cellular Biology*, *24*(21), 9414–9423. https://doi.org/10.1128/MCB.24.21.9414-9423.2004

Course, M. M., Gudsnuk, K., Smukowski, S. N., Winston, K., Desai, N., Ross, J. P., Sulovari, A., Bourassa, C. V., Spiegelman, D., Couthouis, J., Yu, C.-E., Tsuang, D. W., Jayadev, S., Kay, M. A., Gitler, A. D., Dupre, N., Eichler, E. E., Dion, P. A., Rouleau, G. A., & Valdmanis, P. N. (2020). Evolution of a Human-Specific Tandem Repeat Associated with ALS. *American Journal of Human Genetics*, *107*(3), 445–460. https://doi.org/10.1016/j.ajhg.2020.07.004

Covarrubias, S., Robinson, E. K., Shapleigh, B., Vollmers, A., Katzman, S., Hanley, N., Fong, N., McManus, M. T., & Carpenter, S. (2017). CRISPR/Cas-based screening of long non-

coding RNAs (lncRNAs) in macrophages with an NF-κB reporter. *The Journal of Biological Chemistry*, *292*(51), 20911–20920. https://doi.org/10.1074/jbc.M117.799155

Cremer, T., Cremer, M., Dietzel, S., Müller, S., Solovei, I., & Fakan, S. (2006). Chromosome territories—A functional nuclear landscape. *Current Opinion in Cell Biology*, *18*(3), 307–316. https://doi.org/10.1016/j.ceb.2006.04.007

Dauksaite, V., & Gotthardt, M. (2018). Molecular basis of titin exon exclusion by RBM20 and the novel titin splice regulator PTB4. *Nucleic Acids Research*, *46*(10), 5227–5238. https://doi.org/10.1093/nar/gky165

de Boer, M., te Lintel Hekkert, M., Chang, J., van Thiel, B. S., Martens, L., Bos, M. M., de Kleijnen, M. G. J., Ridwan, Y., Octavia, Y., van Deel, E. D., Blonden, L. A., Brandt, R. M. C., Barnhoorn, S., Bautista-Niño, P. K., Krabbendam-Peters, I., Wolswinkel, R., Arshi, B., Ghanbari, M., Kupatt, C., … Duncker, D. J. (2023). DNA repair in cardiomyocytes is critical for maintaining cardiac function in mice. *Aging Cell*, *22*(3), e13768. https://doi.org/10.1111/acel.13768

Derenzini, M., Trerè, D., Pession, A., Montanaro, L., Sirri, V., & Ochs, R. L. (1998). Nucleolar function and size in cancer cells. *The American Journal of Pathology*, *152*(5), 1291–1297.

Derrien, T., Johnson, R., Bussotti, G., Tanzer, A., Djebali, S., Tilgner, H., Guernec, G., Martin, D., Merkel, A., Knowles, D. G., Lagarde, J., Veeravalli, L., Ruan, X., Ruan, Y., Lassmann, T., Carninci, P., Brown, J. B., Lipovich, L., Gonzalez, J. M., … Guigó, R. (2012). The GENCODE v7 catalog of human long noncoding RNAs: Analysis of their gene structure, evolution, and expression. *Genome Research*, *22*(9), 1775–1789. https://doi.org/10.1101/gr.132159.111

Disteche, C. M. (2012). Dosage Compensation of the Sex Chromosomes. *Annual Review of Genetics*, *46*, 537–560. https://doi.org/10.1146/annurev-genet-110711-155454

DISTECHE, C. M., & BERLETCH, J. B. (2015). X-chromosome inactivation and escape. *Journal of Genetics*, *94*(4), 591–599.

Diwakarla, C., Hannan, K., Hein, N., & Yip, D. (2017). Advanced pancreatic ductal adenocarcinoma—Complexities of treatment and emerging therapeutic options. *World Journal of Gastroenterology*, *23*(13), 2276–2285. https://doi.org/10.3748/wjg.v23.i13.2276

Dobin, A., Davis, C. A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., & Gingeras, T. R. (2013). STAR: Ultrafast universal RNA-seq aligner. *Bioinformatics*, *29*(1), 15. https://doi.org/10.1093/bioinformatics/bts635

Dopie, J., Sweredoski, M. J., Moradian, A., & Belmont, A. S. (2020). Tyramide signal amplification mass spectrometry (TSA-MS) ratio identifies nuclear speckle proteins. *The Journal of Cell Biology*, *219*(9), e201910207. https://doi.org/10.1083/jcb.201910207

Dundr, M. (2012). Nuclear bodies: Multifunctional companions of the genome. *Current Opinion in Cell Biology*, *24*(3), 415–422. https://doi.org/10.1016/j.ceb.2012.03.010

Edwards, C. A., & Ferguson-Smith, A. C. (2007). Mechanisms regulating imprinted genes in clusters. *Current Opinion in Cell Biology*, *19*(3), 281–289. https://doi.org/10.1016/j.ceb.2007.04.013

Engreitz, J. M., Pandya-Jones, A., McDonel, P., Shishkin, A., Sirokman, K., Surka, C., Kadri, S., Xing, J., Goren, A., Lander, E. S., Plath, K., & Guttman, M. (2013). The Xist lncRNA exploits three-dimensional genome architecture to spread across the X chromosome. *Science (New York, N.Y.)*, *341*(6147), 1237973. https://doi.org/10.1126/science.1237973

Engreitz, J. M., Sirokman, K., McDonel, P., Shishkin, A., Surka, C., Russell, P., Grossman, S. R., Chow, A. Y., Guttman, M., & Lander, E. S. (2014). RNA-RNA Interactions Enable Specific Targeting of Noncoding RNAs to Nascent Pre-mRNAs and Chromatin Sites. *Cell*, *159*(1), 188–199. https://doi.org/10.1016/j.cell.2014.08.018

Ernst, J., & Kellis, M. (2012). ChromHMM: Automating chromatin state discovery and characterization. *Nature Methods*, *9*(3), 215–216. https://doi.org/10.1038/nmeth.1906

Fang, S., Zhang, L., Guo, J., Niu, Y., Wu, Y., Li, H., Zhao, L., Li, X., Teng, X., Sun, X., Sun, L., Zhang, M. Q., Chen, R., & Zhao, Y. (2018). NONCODEV5: A comprehensive annotation database for long non-coding RNAs. *Nucleic Acids Research*, *46*(Database issue), D308–D314. https://doi.org/10.1093/nar/gkx1107

Farley, K. I., Surovtseva, Y., Merkel, J., & Baserga, S. J. (2015). Determinants of mammalian nucleolar architecture. *Chromosoma*, *124*(3), 323–331. https://doi.org/10.1007/s00412-015-0507-z

Fazal, F. M., Han, S., Parker, K. R., Kaewsapsak, P., Xu, J., Boettiger, A. N., Chang, H. Y., & Ting, A. Y. (2019). Atlas of Subcellular RNA Localization Revealed by APEX-Seq. *Cell*, *178*(2), 473-490.e26. https://doi.org/10.1016/j.cell.2019.05.027

Fenix, A. M., Miyaoka, Y., Bertero, A., Blue, S., Spindler, M. J., Tan, K. K. B., Perez-Bermejo, J., Chan, A. H., Mayer, S. J., Nguyen, T., Russell, C. R., Lizarraga, P., Truong, A., So, P.-L., Kulkarni, A., Chetal, K., Sathe, S., Sniadecki, N. J., Yeo, G. W., … Salomonis, N. (2021). *Gain-of-function cardiomyopathic mutations in RBM20 rewire splicing regulation and re-distribute ribonucleoprotein granules within processing bodies* (p. 2021.06.02.446820). https://doi.org/10.1101/2021.06.02.446820

Fernández-Ruiz, I. (2020). Upregulating RBM20 as a therapy for DCM. *Nature Reviews Cardiology*, *17*(11), Article 11. https://doi.org/10.1038/s41569-020-00458-4

Galganski, L., Urbanek, M. O., & Krzyzosiak, W. J. (2017). Nuclear speckles: Molecular organization, biological function and role in disease. *Nucleic Acids Research*, *45*(18), 10350–10368. https://doi.org/10.1093/nar/gkx759

Geuens, T., Bouhy, D., & Timmerman, V. (2016). The hnRNP family: Insights into their role in health and disease. *Human Genetics*, *135*, 851–867. https://doi.org/10.1007/s00439-016-1683-5

Gilles, J.-F., Dos Santos, M., Boudier, T., Bolte, S., & Heck, N. (2017). DiAna, an ImageJ tool for object-based 3D co-localization and distance analysis. *Methods (San Diego, Calif.)*, *115*, 55–64. https://doi.org/10.1016/j.ymeth.2016.11.016

Granzier, H. L., & Labeit, S. (2004). The Giant Protein Titin. *Circulation Research*, *94*(3), 284–295. https://doi.org/10.1161/01.RES.0000117769.88862.F8

Guo, C.-J., Ma, X.-K., Xing, Y.-H., Zheng, C.-C., Xu, Y.-F., Shan, L., Zhang, J., Wang, S., Wang, Y., Carmichael, G. G., Yang, L., & Chen, L.-L. (2020). Distinct Processing of lncRNAs Contributes to Non-conserved Functions in Stem Cells. *Cell*, *181*(3), 621-636.e22. https://doi.org/10.1016/j.cell.2020.03.006

Guo, C.-J., Xu, G., & Chen, L.-L. (2020). Mechanisms of Long Noncoding RNA Nuclear Retention. *Trends in Biochemical Sciences*, *45*(11), 947–960. https://doi.org/10.1016/j.tibs.2020.07.001

Guo, W., Schafer, S., Greaser, M. L., Radke, M. H., Liss, M., Govindarajan, T., Maatz, H., Schulz, H., Li, S., Parrish, A. M., Dauksaite, V., Vakeel, P., Klaassen, S., Gerull, B., Thierfelder, L., Regitz-Zagrosek, V., Hacker, T. A., Saupe, K. W., Dec, G. W., …

Gotthardt, M. (2012). RBM20, a gene for hereditary cardiomyopathy, regulates titin splicing. *Nature Medicine*, *18*(5), 766–773. https://doi.org/10.1038/nm.2693

Guo, W., Zhu, C., Yin, Z., Wang, Q., Sun, M., Cao, H., & Greaser, M. L. (2018). Splicing Factor RBM20 Regulates Transcriptional Network of Titin Associated and Calcium Handling Genes in The Heart. *International Journal of Biological Sciences*, *14*(4), 369–380. https://doi.org/10.7150/ijbs.24117

Haas, J., Frese, K. S., Peil, B., Kloos, W., Keller, A., Nietsch, R., Feng, Z., Müller, S., Kayvanpour, E., Vogel, B., Sedaghat-Hamedani, F., Lim, W.-K., Zhao, X., Fradkin, D., Köhler, D., Fischer, S., Franke, J., Marquart, S., Barb, I., … Meder, B. (2015). Atlas of the clinical genetics of human dilated cardiomyopathy. *European Heart Journal*, *36*(18), 1123–1135a. https://doi.org/10.1093/eurheartj/ehu301

Hall, L. L., Smith, K. P., Byron, M., & Lawrence, J. B. (2006). THE MOLECULAR ANATOMY OF A SPECKLE. *The Anatomical Record. Part A, Discoveries in Molecular, Cellular, and Evolutionary Biology*, *288*(7), 664–675. https://doi.org/10.1002/ar.a.20336

Han, P., Hang, C. T., Yang, J., & Chang, C.-P. (2011). Chromatin remodeling in cardiovascular development and physiology. *Circulation Research*, *108*(3), 378–396. https://doi.org/10.1161/CIRCRESAHA.110.224287

Han, S., Zhao, B. S., Myers, S. A., Carr, S. A., He, C., & Ting, A. Y. (2020). RNA-protein interaction mapping via MS2- or Cas13-based APEX targeting. *Proceedings of the National Academy of Sciences of the United States of America*, *117*(36), 22068–22079. https://doi.org/10.1073/pnas.2006617117

Han, Y., Branon, T. C., Martell, J. D., Boassa, D., Shechner, D., Ellisman, M. H., & Ting, A. (2019). Directed Evolution of Split APEX2 Peroxidase. *ACS Chemical Biology*, *14*(4), 619–635. https://doi.org/10.1021/acschembio.8b00919

Hang, C. T., Yang, J., Han, P., Cheng, H.-L., Shang, C., Ashley, E., Zhou, B., & Chang, C.-P. (2010). Chromatin regulation by Brg1 underlies heart muscle development and disease. *Nature*, *466*(7302), 62–67. https://doi.org/10.1038/nature09130

Heard, E., Clerc, P., & Avner, P. (1997). X-Chromosome Inactivation in Mammals. *Annual Review of Genetics*, *31*(1), 571–610. https://doi.org/10.1146/annurev.genet.31.1.571

Hershberg, E. A., Camplisson, C. K., Close, J. L., Attar, S., Chern, R., Liu, Y., Akilesh, S., Nicovich, P. R., & Beliveau, B. J. (2021). PaintSHOP enables the interactive design of transcriptome- and genome-scale oligonucleotide FISH experiments. *Nature Methods*, *18*(8), 937–944. https://doi.org/10.1038/s41592-021-01187-3

Hershberger, R. E., Morales, A., & Siegfried, J. D. (2010). Clinical and genetic issues in dilated cardiomyopathy: A review for genetics professionals. *Genetics in Medicine: Official Journal of the American College of Medical Genetics*, *12*(11), 655–667. https://doi.org/10.1097/GIM.0b013e3181f2481f

Hnisz, D., Day, D. S., & Young, R. A. (2016). Insulated neighborhoods: Structural and functional units of mammalian gene control. *Cell*, *167*(5), 1188–1200. https://doi.org/10.1016/j.cell.2016.10.024

Hong, E., Best, A., Gautrey, H., Chin, J., Razdan, A., Curk, T., Elliott, D. J., & Tyson-Capper, A. J. (2015). Unravelling the RNA-Binding Properties of SAFB Proteins in Breast Cancer Cells. *BioMed Research International*, *2015*, 395816. https://doi.org/10.1155/2015/395816

Howard, J. M., & Sanford, J. R. (2015). THE RNAissance Family: SR proteins as multifaceted regulators of gene expression. *Wiley Interdisciplinary Reviews. RNA*, *6*(1), 93–110. https://doi.org/10.1002/wrna.1260

Huber, L. A., Pfaller, K., & Vietor, I. (2003). Organelle proteomics: Implications for subcellular fractionation in proteomics. *Circulation Research*, *92*(9), 962–968. https://doi.org/10.1161/01.RES.0000071748.48338.25

Hunter, J. D. (2007). Matplotlib: A 2D Graphics Environment. *Computing in Science & Engineering*, *9*(3), 90–95. https://doi.org/10.1109/MCSE.2007.55

Jachowicz, J. W., Strehle, M., Banerjee, A. K., Blanco, M. R., Thai, J., & Guttman, M. (2022). Xist spatially amplifies SHARP/SPEN recruitment to balance chromosome-wide silencing and specificity to the X chromosome. *Nature Structural & Molecular Biology*, *29*(3), 239–249. https://doi.org/10.1038/s41594-022-00739-1

Jain, A., & Vale, R. D. (2017). RNA phase transitions in repeat expansion disorders. *Nature*, *546*(7657), 243–247. https://doi.org/10.1038/nature22386

Jeon, Y., & Lee, J. T. (2011). YY1 tethers Xist RNA to the inactive X nucleation center. *Cell*, *146*(1), 119–133. https://doi.org/10.1016/j.cell.2011.06.026

Jiang, S., Kotani, N., Ohnishi, T., Miyagawa-Yamguchi, A., Tsuda, M., Yamashita, R., Ishiura, Y., & Honke, K. (2012). A proteomics approach to the cell-surface interactome using the enzyme-mediated activation of radical sources reaction. *Proteomics*, *12*(1), 54–62. https://doi.org/10.1002/pmic.201100551

Jiang, Y., Park, P., Hong, S.-M., & Ban, K. (2018). Maturation of Cardiomyocytes Derived from Human Pluripotent Stem Cells: Current Strategies and Limitations. *Molecules and Cells*, *41*(7), 613–621. https://doi.org/10.14348/molcells.2018.0143

Jin, Y., Tam, O. H., Paniagua, E., & Hammell, M. (2015). TEtranscripts: A package for including transposable elements in differential expression analysis of RNA-seq datasets. *Bioinformatics (Oxford, England)*, *31*(22), 3593–3599. https://doi.org/10.1093/bioinformatics/btv422

Kaewsapsak, P., Shechner, D. M., Mallard, W., Rinn, J. L., & Ting, A. Y. (2017). Live-cell mapping of organelle-associated RNAs via proximity biotinylation combined with protein-RNA crosslinking. *ELife*, *6*, e29224. https://doi.org/10.7554/eLife.29224

Kim, D., Paggi, J. M., Park, C., Bennett, C., & Salzberg, S. L. (2019). Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nature Biotechnology*, *37*(8), 907–915. https://doi.org/10.1038/s41587-019-0201-4

Kirschneck, C., Batschkus, S., Proff, P., Köstler, J., Spanier, G., & Schröder, A. (2017). Valid gene expression normalization by RT-qPCR in studies on hPDL fibroblasts with focus on orthodontic tooth movement and periodontitis. *Scientific Reports*, *7*(1), 14751. https://doi.org/10.1038/s41598-017-15281-0

Kishi, J. Y., Lapan, S. W., Beliveau, B. J., West, E. R., Zhu, A., Sasaki, H. M., Saka, S. K., Wang, Y., Cepko, C. L., & Yin, P. (2019). SABER enables amplified and multiplexed imaging of RNA and DNA in cells and tissues. *Nature Methods*, *16*(6), 533–544. https://doi.org/10.1038/s41592-019-0404-0

Koelemen, J., Gotthardt, M., Steinmetz, L. M., & Meder, B. (2021). RBM20-Related Cardiomyopathy: Current Understanding and Future Options. *Journal of Clinical Medicine*, *10*(18), 4101. https://doi.org/10.3390/jcm10184101

Kong, A. T., Leprevost, F. V., Avtonomov, D. M., Mellacheruvu, D., & Nesvizhskii, A. I. (2017). MSFragger: Ultrafast and comprehensive peptide identification in mass

spectrometry-based proteomics. *Nature Methods*, *14*(5), 513–520. https://doi.org/10.1038/nmeth.4256

Kornfeld, O. S., Hwang, S., Disatnik, M.-H., Chen, C.-H., Qvit, N., & Mochly-Rosen, D. (2015). Mitochondrial Reactive Oxygen Species at the Heart of the Matter: New Therapeutic Approaches for Cardiovascular Diseases. *Circulation Research*, *116*(11), 1783–1799. https://doi.org/10.1161/CIRCRESAHA.116.305432

Krismer, K., Bird, M. A., Varmeh, S., Handly, E. D., Gattinger, A., Bernwinkler, T., Anderson, D. A., Heinzel, A., Joughin, B. A., Kong, Y. W., Cannell, I. G., & Yaffe, M. B. (2020). Transite: A Computational Motif-Based Analysis Platform That Identifies RNA-Binding Proteins Modulating Changes in Gene Expression. *Cell Reports*, *32*(8), 108064. https://doi.org/10.1016/j.celrep.2020.108064

Krstic, M., Macmillan, C. D., Leong, H. S., Clifford, A. G., Souter, L. H., Dales, D. W., Postenka, C. O., Chambers, A. F., & Tuck, A. B. (2016). The transcriptional regulator TBX3 promotes progression from non-invasive to invasive breast cancer. *BMC Cancer*, *16*(1), 671. https://doi.org/10.1186/s12885-016-2697-z

Kukharsky, M. S., Ninkina, N. N., An, H., Telezhkin, V., Wei, W., Meritens, C. R. de, Cooper-Knock, J., Nakagawa, S., Hirose, T., Buchman, V. L., & Shelkovnikova, T. A. (2020). Long non-coding RNA Neat1 regulates adaptive behavioural response to stress in mice. *Translational Psychiatry*, *10*(1), 171. https://doi.org/10.1038/s41398-020-0854-2

Lambert, S. A., Jolma, A., Campitelli, L. F., Das, P. K., Yin, Y., Albu, M., Chen, X., Taipale, J., Hughes, T. R., & Weirauch, M. T. (2018). The Human Transcription Factors. *Cell*, *172*(4), 650–665. https://doi.org/10.1016/j.cell.2018.01.029

Langmead, B., & Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nature Methods*, *9*(4), 357–359. https://doi.org/10.1038/nmeth.1923

Laptev, I., Shvetsova, E., Levitskii, S., Serebryakova, M., Rubtsova, M., Zgoda, V., Bogdanov, A., Kamenski, P., Sergiev, P., & Dontsova, O. (2020). METTL15 interacts with the assembly intermediate of murine mitochondrial small ribosomal subunit to form m4C840 12S rRNA residue. *Nucleic Acids Research*, *48*(14), 8022–8034. https://doi.org/10.1093/nar/gkaa522

Latos, P. A., Pauler, F. M., Koerner, M. V., Şenergin, H. B., Hudson, Q. J., Stocsits, R. R., Allhoff, W., Stricker, S. H., Klement, R. M., Warczok, K. E., Aumayr, K., Pasierbek, P., & Barlow, D. P. (2012). Airn transcriptional overlap, but not its lncRNA products, induces imprinted Igf2r silencing. *Science (New York, N.Y.)*, *338*(6113), 1469–1472. https://doi.org/10.1126/science.1228110

Li, G., & Reinberg, D. (2011). Chromatin higher-order structures and gene regulation. *Current Opinion in Genetics & Development*, *21*(2), 175–186. https://doi.org/10.1016/j.gde.2011.01.022

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., & 1000 Genome Project Data Processing Subgroup. (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics (Oxford, England)*, *25*(16), 2078–2079. https://doi.org/10.1093/bioinformatics/btp352

Li, S., Guo, W., Dewey, C. N., & Greaser, M. L. (2013). Rbm20 regulates titin alternative splicing as a splicing repressor. *Nucleic Acids Research*, *41*(4), 2659–2672. https://doi.org/10.1093/nar/gks1362

Lingenfelter, P. A., Adler, D. A., Poslinski, D., Thomas, S., Elliott, R. W., Chapman, V. M., & Disteche, C. M. (1998). Escape from X inactivation of Smcx is preceded by silencing during mouse development. *Nature Genetics*, *18*(3), 212–213. https://doi.org/10.1038/ng0398-212

Loda, A., Collombet, S., & Heard, E. (2022). Gene regulation in time and space during X-chromosome inactivation. *Nature Reviews Molecular Cell Biology*, *23*(4), Article 4. https://doi.org/10.1038/s41580-021-00438-7

Loda, A., & Heard, E. (2019). Xist RNA in action: Past, present, and future. *PLoS Genetics*, *15*(9), e1008333. https://doi.org/10.1371/journal.pgen.1008333

Lomberk, G., Blum, Y., Nicolle, R., Nair, A., Gaonkar, K. S., Marisa, L., Mathison, A., Sun, Z., Yan, H., Elarouci, N., Armenoult, L., Ayadi, M., Ordog, T., Lee, J.-H., Oliver, G., Klee, E., Moutardier, V., Gayet, O., Bian, B., … Urrutia, R. (2018). Distinct epigenetic landscapes underlie the pathobiology of pancreatic cancer subtypes. *Nature Communications*, *9*(1), Article 1. https://doi.org/10.1038/s41467-018-04383-6

Love, M. I., Huber, W., & Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology*, *15*(12), 550. https://doi.org/10.1186/s13059-014-0550-8

Luk, A., Ahn, E., Soor, G. S., & Butany, J. (2009). Dilated cardiomyopathy: A review. *Journal of Clinical Pathology*, *62*(3), 219–225. https://doi.org/10.1136/jcp.2008.060731

Lukong, K. E., Chang, K., Khandjian, E. W., & Richard, S. (2008). RNA-binding proteins in human genetic disease. *Trends in Genetics: TIG*, *24*(8), 416–425. https://doi.org/10.1016/j.tig.2008.05.004

Lund, E., Oldenburg, A. R., & Collas, P. (2014). Enriched domain detector: A program for detection of wide genomic enrichment domains robust against local variations. *Nucleic Acids Research*, *42*(11), e92. https://doi.org/10.1093/nar/gku324

Lyon, M. F. (1998). X-chromosome inactivation: A repeat hypothesis. *Cytogenetics and Cell Genetics*, *80*(1–4), 133–137. https://doi.org/10.1159/000014969

Maatz, H., Jens, M., Liss, M., Schafer, S., Heinig, M., Kirchner, M., Adami, E., Rintisch, C., Dauksaite, V., Radke, M. H., Selbach, M., Barton, P. J. R., Cook, S. A., Rajewsky, N., Gotthardt, M., Landthaler, M., & Hubner, N. (2014). RNA-binding protein RBM20 represses splicing to orchestrate cardiac pre-mRNA processing. *The Journal of Clinical Investigation*, *124*(8), 3419–3430. https://doi.org/10.1172/JCI74523

Machyna, M., & Simon, M. D. (2017). Catching RNAs on chromatin using hybridization capture methods. *Briefings in Functional Genomics*, *17*(2), 96–103. https://doi.org/10.1093/bfgp/elx038

Mao, Y. S., Zhang, B., & Spector, D. L. (2011). Biogenesis and function of nuclear bodies. *Trends in Genetics: TIG*, *27*(8), 295–306. https://doi.org/10.1016/j.tig.2011.05.006

Markaki, Y., Chong, J. G., Wang, Y., Jacobson, E. C., Luong, C., Tan, S. Y. X., Jachowicz, J. W., Strehle, M., Maestrini, D., Banerjee, A. K., Mistry, B. A., Dror, I., Dossin, F., Schöneberg, J., Heard, E., Guttman, M., Chou, T., & Plath, K. (2021). Xist nucleates local protein gradients to propagate silencing across the X chromosome. *Cell*, *184*(25), 6212. https://doi.org/10.1016/j.cell.2021.11.028

Mars, J.-C., Sabourin-Felix, M., Tremblay, M. G., & Moss, T. (2018). A Deconvolution Protocol for ChIP-Seq Reveals Analogous Enhancer Structures on the Mouse and Human

Ribosomal RNA Genes. *G3 (Bethesda, Md.)*, *8*(1), 303–314. https://doi.org/10.1534/g3.117.300225

Marston, S., Montgiraud, C., Munster, A. B., Copeland, O., Choi, O., dos Remedios, C., Messer, A. E., Ehler, E., & Knöll, R. (2015). OBSCN Mutations Associated with Dilated Cardiomyopathy and Haploinsufficiency. *PLoS ONE*, *10*(9), e0138568. https://doi.org/10.1371/journal.pone.0138568

Matheson, T. D., & Kaufman, P. D. (2016). Grabbing the genome by the NADs. *Chromosoma*, *125*(3), 361–371. https://doi.org/10.1007/s00412-015-0527-8

McDonel, P., & Guttman, M. (2019). Approaches for Understanding the Mechanisms of Long Noncoding RNA Regulation of Gene Expression. *Cold Spring Harbor Perspectives in Biology*, *11*(12), a032151. https://doi.org/10.1101/cshperspect.a032151

McHugh, C. A., Chen, C.-K., Chow, A., Surka, C. F., Tran, C., McDonel, P., Pandya-Jones, A., Blanco, M., Burghard, C., Moradian, A., Sweredoski, M. J., Shishkin, A. A., Su, J., Lander, E. S., Hess, S., Plath, K., & Guttman, M. (2015). The Xist lncRNA interacts directly with SHARP to silence transcription through HDAC3. *Nature*, *521*(7551), 232–236. https://doi.org/10.1038/nature14443

McHugh, C. A., & Guttman, M. (2018). RAP-MS: A Method to Identify Proteins that Interact Directly with a Specific RNA Molecule in Cells. *Methods in Molecular Biology (Clifton, N.J.)*, *1649*, 473–488. https://doi.org/10.1007/978-1-4939-7213-5_31

McNally, E. M., & Mestroni, L. (2017). Dilated Cardiomyopathy: Genetic Determinants and Mechanisms. *Circulation Research*, *121*(7), 731–748. https://doi.org/10.1161/CIRCRESAHA.116.309396

Meir, M., Galanty, Y., Kashani, L., Blank, M., Khosravi, R., Fernández-Ávila, M. J., Cruz-García, A., Star, A., Shochot, L., Thomas, Y., Garrett, L. J., Chamovitz, D. A., Bodine, D. M., Kurz, T., Huertas, P., Ziv, Y., & Shiloh, Y. (2015). The COP9 signalosome is vital for timely repair of DNA double-strand breaks. *Nucleic Acids Research*, *43*(9), 4517–4530. https://doi.org/10.1093/nar/gkv270

MILI, S., & STEITZ, J. A. (2004). Evidence for reassociation of RNA-binding proteins after cell lysis: Implications for the interpretation of immunoprecipitation analyses. *RNA*, *10*(11), 1692–1694. https://doi.org/10.1261/rna.7151404

Minajigi, A., Froberg, J., Wei, C., Sunwoo, H., Kesner, B., Colognori, D., Lessing, D., Payer, B., Boukhali, M., Haas, W., & Lee, J. T. (2015). A comprehensive Xist interactome reveals cohesin repulsion and an RNA-directed chromosome conformation. *Science (New York, N.Y.)*, *349*(6245), 10.1126/science.aab2276 aab2276. https://doi.org/10.1126/science.aab2276

Mintz, P. J., Patterson, S. D., Neuwald, A. F., Spahr, C. S., & Spector, D. L. (1999). Purification and biochemical characterization of interchromatin granule clusters. *The EMBO Journal*, *18*(15), 4308–4320. https://doi.org/10.1093/emboj/18.15.4308

Misteli, T. (2000). Cell biology of transcription and pre-mRNA splicing: Nuclear architecture meets nuclear function. *Journal of Cell Science*, *113 ( Pt 11)*, 1841–1849. https://doi.org/10.1242/jcs.113.11.1841

Misteli, T. (2004). Spatial positioning; a new dimension in genome function. *Cell*, *119*(2), 153–156. https://doi.org/10.1016/j.cell.2004.09.035

Misteli, T. (2007). Beyond the sequence: Cellular organization of genome function. *Cell*, *128*(4), 787–800. https://doi.org/10.1016/j.cell.2007.01.028

Mitrea, D. M., Mittasch, M., Gomes, B. F., Klein, I. A., & Murcko, M. A. (2022). Modulating biomolecular condensates: A novel approach to drug discovery. *Nature Reviews Drug Discovery*, *21*(11), Article 11. https://doi.org/10.1038/s41573-022-00505-4

Miyashita, N., Horie, M., Suzuki, H. I., Saito, M., Mikami, Y., Okuda, K., Boucher, R. C., Suzukawa, M., Hebisawa, A., Saito, A., & Nagase, T. (2020). FOXL1 Regulates Lung Fibroblast Function via Multiple Mechanisms. *American Journal of Respiratory Cell and Molecular Biology*, *63*(6), 831–842. https://doi.org/10.1165/rcmb.2019-0396OC

Németh, A., Conesa, A., Santoyo-Lopez, J., Medina, I., Montaner, D., Péterfia, B., Solovei, I., Cremer, T., Dopazo, J., & Längst, G. (2010). Initial Genomics of the Human Nucleolus. *PLOS Genetics*, *6*(3), e1000889. https://doi.org/10.1371/journal.pgen.1000889

Nishiyama, T., Zhang, Y., Cui, M., Li, H., Sanchez-Ortiz, E., McAnally, J. R., Tan, W., Kim, J., Chen, K., Xu, L., Bassel-Duby, R., & Olson, E. N. (2022). Precise genomic editing of pathogenic mutations in RBM20 rescues dilated cardiomyopathy. *Science Translational Medicine*, *14*(672), eade1633. https://doi.org/10.1126/scitranslmed.ade1633

Norman, M., Rivers, C., Lee, Y.-B., Idris, J., & Uney, J. (2016). The increasing diversity of functions attributed to the SAFB family of RNA-/DNA-binding proteins. *The Biochemical Journal*, *473*(23), 4271–4288. https://doi.org/10.1042/BCJ20160649

Olson, T. M., Illenberger, S., Kishimoto, N. Y., Huttelmaier, S., Keating, M. T., & Jockusch, B. M. (2002). Metavinculin Mutations Alter Actin Interaction in Dilated Cardiomyopathy. *Circulation*, *105*(4), 431–437. https://doi.org/10.1161/hc0402.102930

Olson, T. M., Michels, V. V., Thibodeau, S. N., Tai, Y.-S., & Keating, M. T. (1998). Actin Mutations in Dilated Cardiomyopathy, a Heritable Form of Heart Failure. *Science*, *280*(5364), 750–752. https://doi.org/10.1126/science.280.5364.750

Osborne, R. J., & Thornton, C. A. (2006). RNA-dominant diseases. *Human Molecular Genetics*, *15 Spec No 2*, R162-169. https://doi.org/10.1093/hmg/ddl181

Padovan-Merhar, O., Nair, G. P., Biaesch, A. G., Mayer, A., Scarfone, S., Foley, S. W., Wu, A. R., Churchman, L. S., Singh, A., & Raj, A. (2015). Single mammalian cells compensate for differences in cellular volume and DNA copy number through independent global transcriptional mechanisms. *Molecular Cell*, *58*(2), 339–352. https://doi.org/10.1016/j.molcel.2015.03.005

Pandey, R. R., Mondal, T., Mohammad, F., Enroth, S., Redrup, L., Komorowski, J., Nagano, T., Mancini-Dinardo, D., & Kanduri, C. (2008). Kcnq1ot1 antisense noncoding RNA mediates lineage-specific transcriptional silencing through chromatin-level regulation. *Molecular Cell*, *32*(2), 232–246. https://doi.org/10.1016/j.molcel.2008.08.022

Pandya-Jones, A., Markaki, Y., Serizay, J., Chitiashvili, T., Mancia, W., Damianov, A., Chronis, C., Papp, B., Chen, C.-K., McKee, R., Wang, X.-J., Chau, A., Sabri, S., Leonhardt, H., Zheng, S., Guttman, M., Black, Douglas. L., & Plath, K. (2020). A protein assembly mediates Xist localization and gene silencing. *Nature*, *587*(7832), 145–151. https://doi.org/10.1038/s41586-020-2703-0

Papait, R., Cattaneo, P., Kunderfranco, P., Greco, C., Carullo, P., Guffanti, A., Viganò, V., Stirparo, G. G., Latronico, M. V. G., Hasenfuss, G., Chen, J., & Condorelli, G. (2013). Genome-wide analysis of histone marks identifying an epigenetic signature of promoters and enhancers underlying cardiac hypertrophy. *Proceedings of the National Academy of Sciences of the United States of America*, *110*(50), 20164–20169. https://doi.org/10.1073/pnas.1315155110

Pederson, T., & Politz, J. C. (2000). The nucleolus and the four ribonucleoproteins of translation. *The Journal of Cell Biology*, *148*(6), 1091–1095. https://doi.org/10.1083/jcb.148.6.1091

Pertea, M., Pertea, G. M., Antonescu, C. M., Chang, T.-C., Mendell, J. T., & Salzberg, S. L. (2015). StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nature Biotechnology*, *33*(3), 290–295. https://doi.org/10.1038/nbt.3122

Phelps, W. A., Carlson, A. E., & Lee, M. T. (2021). Optimized design of antisense oligomers for targeted rRNA depletion. *Nucleic Acids Research*, *49*(1), e5. https://doi.org/10.1093/nar/gkaa1072

*Picard Tools—By Broad Institute*. (n.d.). Retrieved May 1, 2023, from http://broadinstitute.github.io/picard/

Pilotte, J., Larocque, D., & Richard, S. (2001). Nuclear translocation controlled by alternatively spliced isoforms inactivates the QUAKING apoptotic inducer. *Genes & Development*, *15*(7), 845–858. https://doi.org/10.1101/gad.860301

Pintacuda, G., Wei, G., Roustan, C., Kirmizitas, B. A., Solcan, N., Cerase, A., Castello, A., Mohammed, S., Moindrot, B., Nesterova, T. B., & Brockdorff, N. (2017). HnRNPK Recruits PCGF3/5-PRC1 to the Xist RNA B-Repeat to Establish Polycomb-Mediated Chromosomal Silencing. *Molecular Cell*, *68*(5), 955-969.e10. https://doi.org/10.1016/j.molcel.2017.11.013

Ponting, C. P., Oliver, P. L., & Reik, W. (2009). Evolution and Functions of Long Noncoding RNAs. *Cell*, *136*(4), 629–641. https://doi.org/10.1016/j.cell.2009.02.006

Puente, B. N., Kimura, W., Muralidhar, S. A., Moon, J., Amatruda, J. F., Phelps, K. L., Grinsfelder, D., Rothermel, B. A., Chen, R., Garcia, J. A., Santos, C. X., Thet, S., Mori, E., Kinter, M. T., Rindler, P. M., Zacchigna, S., Mukherjee, S., Chen, D. J., Mahmoud,

A. I., … Sadek, H. A. (2014). The Oxygen Rich Postnatal Environment Induces Cardiomyocyte Cell Cycle Arrest Through DNA Damage Response. *Cell*, *157*(3), 565–579. https://doi.org/10.1016/j.cell.2014.03.032

Quinlan, A. R., & Hall, I. M. (2010). BEDTools: A flexible suite of utilities for comparing genomic features. *Bioinformatics (Oxford, England)*, *26*(6), 841–842. https://doi.org/10.1093/bioinformatics/btq033

Raap, A. K., van de Corput, M. P., Vervenne, R. A., van Gijlswijk, R. P., Tanke, H. J., & Wiegant, J. (1995). Ultra-sensitive FISH using peroxidase-mediated deposition of biotin- or fluorochrome tyramides. *Human Molecular Genetics*, *4*(4), 529–534. https://doi.org/10.1093/hmg/4.4.529

Raj, A., & Rinn, J. L. (2019). Illuminating Genomic Dark Matter with RNA Imaging. *Cold Spring Harbor Perspectives in Biology*, *11*(5), a032094. https://doi.org/10.1101/cshperspect.a032094

Ramaccini, D., Montoya-Uribe, V., Aan, F. J., Modesti, L., Potes, Y., Wieckowski, M. R., Krga, I., Glibetić, M., Pinton, P., Giorgi, C., & Matter, M. L. (2021). Mitochondrial Function and Dysfunction in Dilated Cardiomyopathy. *Frontiers in Cell and Developmental Biology*, *8*, 624216. https://doi.org/10.3389/fcell.2020.624216

Ramanathan, M., Majzoub, K., Rao, D. S., Neela, P. H., Zarnegar, B. J., Mondal, S., Roth, J. G., Gai, H., Kovalski, J. R., Siprashvili, Z., Palmer, T. D., Carette, J. E., & Khavari, P. A. (2018). RNA-protein interaction detection in living cells. *Nature Methods*, *15*(3), 207–212. https://doi.org/10.1038/nmeth.4601

Ramanathan, M., Porter, D. F., & Khavari, P. A. (2019). Methods to study RNA-protein interactions. *Nature Methods*, *16*(3), 225–234. https://doi.org/10.1038/s41592-019-0330-1

Ramírez, F., Dündar, F., Diehl, S., Grüning, B. A., & Manke, T. (2014). deepTools: A flexible platform for exploring deep-sequencing data. *Nucleic Acids Research*, *42*(Web Server issue), W187–W191. https://doi.org/10.1093/nar/gku365

Redrup, L., Branco, M. R., Perdeaux, E. R., Krueger, C., Lewis, A., Santos, F., Nagano, T., Cobb, B. S., Fraser, P., & Reik, W. (2009). The long noncoding RNA Kcnq1ot1 organises a lineage-specific nuclear domain for epigenetic gene silencing. *Development (Cambridge, England)*, *136*(4), 525–530. https://doi.org/10.1242/dev.031328

Rees, J. S., Li, X.-W., Perrett, S., Lilley, K. S., & Jackson, A. P. (2015). Selective Proteomic Proximity Labeling Assay Using Tyramide (SPPLAT): A Quantitative Method for the Proteomic Analysis of Localized Membrane-Bound Protein Clusters. *Current Protocols in Protein Science*, *80*, 19.27.1-19.27.18. https://doi.org/10.1002/0471140864.ps1927s80

Refaat, M. M., Lubitz, S. A., Makino, S., Islam, Z., Frangiskakis, J. M., Mehdi, H., Gutmann, R., Zhang, M. L., Bloom, H. L., MacRae, C. A., Dudley, S. C., Shalaby, A. A., Weiss, R., McNamara, D. M., London, B., & Ellinor, P. T. (2012). Genetic variation in the alternative splicing regulator RBM20 is associated with dilated cardiomyopathy. *Heart Rhythm*, *9*(3), 390–396. https://doi.org/10.1016/j.hrthm.2011.10.016

Rhee, H.-W., Zou, P., Udeshi, N. D., Martell, J. D., Mootha, V. K., Carr, S. A., & Ting, A. Y. (2013). Proteomic Mapping of Mitochondria in Living Cells via Spatially-Restricted Enzymatic Tagging. *Science (New York, N.Y.)*, *339*(6125), 1328–1331. https://doi.org/10.1126/science.1230593

Rinn, J., & Guttman, M. (2014). RNA Function. RNA and dynamic nuclear organization. *Science (New York, N.Y.)*, *345*(6202), 1240–1241. https://doi.org/10.1126/science.1252966

*Rmats2sashimiplot*. (2023). [Python]. Xinglab. https://github.com/Xinglab/rmats2sashimiplot (Original work published 2015)

Robert-Finestra, T., Tan, B. F., Mira-Bontenbal, H., Timmers, E., Gontan, C., Merzouk, S., Giaimo, B. D., Dossin, F., van IJcken, W. F. J., Martens, J. W. M., Borggrefe, T., Heard, E., & Gribnau, J. (2021). SPEN is required for Xist upregulation during initiation of X chromosome inactivation. *Nature Communications*, *12*(1), Article 1. https://doi.org/10.1038/s41467-021-27294-5

Roca, X. (2023, May 23). *SRRM2 confers phase-separation properties to regulate GC-rich splicing events in the nucleus center* [Oral]. RNA 2023, Singapore, Suntec Convention Center. https://www2.rnasociety.org/conferences/rna-2023/tentative-schedule/

Roeder, R. G., & Rutter, W. J. (1970). Specific Nucleolar and Nucleoplasmic RNA Polymerases*. *Proceedings of the National Academy of Sciences*, *65*(3), 675–682. https://doi.org/10.1073/pnas.65.3.675

Rom, A., Melamed, L., Gil, N., Goldrich, M. J., Kadir, R., Golan, M., Biton, I., Perry, R. B.-T., & Ulitsky, I. (2019). Regulation of CHD2 expression by the Chaserr long noncoding RNA gene is essential for viability. *Nature Communications*, *10*(1), 5092. https://doi.org/10.1038/s41467-019-13075-8

Saitoh, N., Spahr, C. S., Patterson, S. D., Bubulya, P., Neuwald, A. F., & Spector, D. L. (2004). Proteomic analysis of interchromatin granule clusters. *Molecular Biology of the Cell*, *15*(8), 3876–3890. https://doi.org/10.1091/mbc.e04-03-0253

Schindelin, J., Arganda-Carreras, I., Frise, E., Kaynig, V., Longair, M., Pietzsch, T., Preibisch, S., Rueden, C., Saalfeld, S., Schmid, B., Tinevez, J.-Y., White, D. J., Hartenstein, V., Eliceiri, K., Tomancak, P., & Cardona, A. (2012). Fiji: An open-source platform for biological-image analysis. *Nature Methods*, *9*(7), Article 7. https://doi.org/10.1038/nmeth.2019

Schneider, C. A., Rasband, W. S., & Eliceiri, K. W. (2012). NIH Image to ImageJ: 25 years of image analysis. *Nature Methods*, *9*(7), Article 7. https://doi.org/10.1038/nmeth.2089

Schnell, U., Dijk, F., Sjollema, K. A., & Giepmans, B. N. G. (2012). Immunolabeling artifacts and the need for live-cell imaging. *Nature Methods*, *9*(2), 152–158. https://doi.org/10.1038/nmeth.1855

Sexton, T., & Cavalli, G. (2015). The role of chromosome domains in shaping the functional genome. *Cell*, *160*(6), 1049–1059. https://doi.org/10.1016/j.cell.2015.02.040

Sha, K., & Boyer, L. A. (2008). The chromatin signature of pluripotent cells. In *StemBook*. Harvard Stem Cell Institute. http://www.ncbi.nlm.nih.gov/books/NBK27041/

Shechner, D. M., & Bartel, D. P. (2011). The structural basis of RNA-catalyzed RNA polymerization. *Nature Structural & Molecular Biology*, *18*(9), Article 9. https://doi.org/10.1038/nsmb.2107

Shen, S., Park, J. W., Lu, Z., Lin, L., Henry, M. D., Wu, Y. N., Zhou, Q., & Xing, Y. (2014). rMATS: Robust and flexible detection of differential alternative splicing from replicate RNA-Seq data. *Proceedings of the National Academy of Sciences of the United States of America*, *111*(51), E5593-5601. https://doi.org/10.1073/pnas.1419161111

Shopland, L. S., Johnson, C. V., Byron, M., McNeil, J., & Lawrence, J. B. (2003). Clustering of multiple specific genes and gene-rich R-bands around SC-35 domains: Evidence for local

euchromatic neighborhoods. *The Journal of Cell Biology*, *162*(6), 981–990. https://doi.org/10.1083/jcb.200303131

Sibbing, D., Pfeufer, A., Perisic, T., Mannes, A. M., Fritz-Wolf, K., Unwin, S., Sinner, M. F., Gieger, C., Gloeckner, C. J., Wichmann, H.-E., Kremmer, E., Schäfer, Z., Walch, A., Hinterseer, M., Näbauer, M., Kääb, S., Kastrati, A., Schömig, A., Meitinger, T., … von Beckerath, N. (2011). Mutations in the mitochondrial thioredoxin reductase gene TXNRD2 cause dilated cardiomyopathy. *European Heart Journal*, *32*(9), 1121–1133. https://doi.org/10.1093/eurheartj/ehq507

Simon, M. D. (2013). Capture hybridization analysis of RNA targets (CHART). *Current Protocols in Molecular Biology*, *Chapter 21*, Unit 21.25. https://doi.org/10.1002/0471142727.mb2125s101

Simon, M. D., & Machyna, M. (2019). Principles and Practices of Hybridization Capture Experiments to Study Long Noncoding RNAs That Act on Chromatin. *Cold Spring Harbor Perspectives in Biology*, *11*(11), a032276. https://doi.org/10.1101/cshperspect.a032276

Simon, M. D., Pinter, S. F., Fang, R., Sarma, K., Rutenberg-Schoenberg, M., Bowman, S. K., Kesner, B. A., Maier, V. K., Kingston, R. E., & Lee, J. T. (2013). High-resolution Xist binding maps reveal two-step spreading during X-chromosome inactivation. *Nature*, *504*(7480), 465–469. https://doi.org/10.1038/nature12719

Simon, M. D., Wang, C. I., Kharchenko, P. V., West, J. A., Chapman, B. A., Alekseyenko, A. A., Borowsky, M. L., Kuroda, M. I., & Kingston, R. E. (2011). The genomic binding sites of a noncoding RNA. *Proceedings of the National Academy of Sciences of the*

*United States of America*, *108*(51), 20497–20502. https://doi.org/10.1073/pnas.1113536108

Sleutels, F., Zwart, R., & Barlow, D. P. (2002). The non-coding Air RNA is required for silencing autosomal imprinted genes. *Nature*, *415*(6873), 810–813. https://doi.org/10.1038/415810a

Smith, K. P., Hall, L. L., & Lawrence, J. B. (2020). Nuclear Hubs Built on RNAs and Clustered Organization of the Genome. *Current Opinion in Cell Biology*, *64*, 67–76. https://doi.org/10.1016/j.ceb.2020.02.015

Spector, D. L., & Lamond, A. I. (2011). Nuclear speckles. *Cold Spring Harbor Perspectives in Biology*, *3*(2), a000646. https://doi.org/10.1101/cshperspect.a000646

Statello, L., Guo, C.-J., Chen, L.-L., & Huarte, M. (2021). Gene regulation by long non-coding RNAs and its biological functions. *Nature Reviews. Molecular Cell Biology*, *22*(2), 96–118. https://doi.org/10.1038/s41580-020-00315-9

Stovner, E. B., & Sætrom, P. (2019). Epic2 efficiently finds diffuse domains in ChIP-seq data. *Bioinformatics (Oxford, England)*, *35*(21), 4392–4393. https://doi.org/10.1093/bioinformatics/btz232

Strehle, M., & Guttman, M. (2020). Xist drives spatial compartmentalization of DNA and protein to orchestrate initiation and maintenance of X inactivation. *Current Opinion in Cell Biology*, *64*, 139–147. https://doi.org/10.1016/j.ceb.2020.04.009

Subramanian, A., Tamayo, P., Mootha, V. K., Mukherjee, S., Ebert, B. L., Gillette, M. A., Paulovich, A., Pomeroy, S. L., Golub, T. R., Lander, E. S., & Mesirov, J. P. (2005). Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide

expression profiles. *Proceedings of the National Academy of Sciences of the United States of America*, *102*(43), 15545–15550. https://doi.org/10.1073/pnas.0506580102

Tang, L., Nogales, E., & Ciferri, C. (2010). Structure and Function of SWI/SNF Chromatin Remodeling Complexes and Mechanistic Implications for Transcription. *Progress in Biophysics and Molecular Biology*, *102*(2–3), 122–128. https://doi.org/10.1016/j.pbiomolbio.2010.05.001

Tsue, A. F., Kania, E. E., Lei, D. Q., Fields, R., McGann, C. D., Hershberg, E., Deng, X., Kihiu, M., Ong, S.-E., Disteche, C. M., Kugel, S., Beliveau, B. J., Schweppe, D. K., & Shechner, D. M. (2023). Oligonucleotide-directed proximity-interactome mapping (O-MAP): A unified method for discovering RNA-interacting proteins, transcripts and genomic loci in situ. *BioRxiv: The Preprint Server for Biology*, 2023.01.19.524825. https://doi.org/10.1101/2023.01.19.524825

Uszczynska-Ratajczak, B., Lagarde, J., Frankish, A., Guigó, R., & Johnson, R. (2018). Towards a complete map of the human long non-coding RNA transcriptome. *Nature Reviews. Genetics*, *19*(9), 535–548. https://doi.org/10.1038/s41576-018-0017-y

van der Velde, A., Fan, K., Tsuji, J., Moore, J. E., Purcaro, M. J., Pratt, H. E., & Weng, Z. (2021). Annotation of chromatin states in 66 complete mouse epigenomes during development. *Communications Biology*, *4*, 239. https://doi.org/10.1038/s42003-021-01756-4

van Koningsbruggen, S., Gierliński, M., Schofield, P., Martin, D., Barton, G. J., Ariyurek, Y., den Dunnen, J. T., & Lamond, A. I. (2010). High-Resolution Whole-Genome Sequencing Reveals That Specific Chromatin Domains from Most Human Chromosomes Associate

with Nucleoli. *Molecular Biology of the Cell*, *21*(21), 3735–3748. https://doi.org/10.1091/mbc.E10-06-0508

van Steensel, B., & Belmont, A. S. (2017). Lamina-Associated Domains: Links with Chromosome Architecture, Heterochromatin, and Gene Repression. *Cell*, *169*(5), 780–791. https://doi.org/10.1016/j.cell.2017.04.022

Wang, K. C., & Chang, H. Y. (2011). Molecular mechanisms of long noncoding RNAs. *Molecular Cell*, *43*(6), 904–914. https://doi.org/10.1016/j.molcel.2011.08.018

Waskom, M. L. (2021). seaborn: Statistical data visualization. *Journal of Open Source Software*, *6*(60), 3021. https://doi.org/10.21105/joss.03021

Watanabe, T., Kimura, A., & Kuroyanagi, H. (2018). Alternative Splicing Regulator RBM20 and Cardiomyopathy. *Frontiers in Molecular Biosciences*, *5*, 105. https://doi.org/10.3389/fmolb.2018.00105

Wickham, H. (2009). *ggplot2: Elegant Graphics for Data Analysis*. Springer. https://doi.org/10.1007/978-0-387-98141-3

Wu, G., Liu, J., Liu, M., Huang, Q., Ruan, J., Zhang, C., Wang, D., Sun, X., Jiang, W., Kang, L., Wang, J., & Song, L. (2021). Truncating Variants in OBSCN Gene Associated With Disease-Onset and Outcomes of Hypertrophic Cardiomyopathy. *Circulation: Genomic and Precision Medicine*, *14*(5), e003401. https://doi.org/10.1161/CIRCGEN.121.003401

Wu, M., Xu, G., Han, C., Luan, P.-F., Xing, Y.-H., Nan, F., Yang, L.-Z., Huang, Y., Yang, Z.-H., Shan, L., Yang, L., Liu, J., & Chen, L.-L. (2021). LncRNA SLERT controls phase separation of FC/DFCs to facilitate Pol I transcription. *Science (New York, N.Y.)*, *373*(6554), 547–555. https://doi.org/10.1126/science.abf6582

Wu, Y., Jin, Y., Yamamoto, N., Takeuchi, A., Miwa, S., Tsuchiya, H., & Yang, Z. (2021). MSX2 inhibits the growth and migration of osteosarcoma cells by repressing SOX2. *American Journal of Translational Research*, *13*(6), 5851–5865.

Wutz, A. (2011). Gene silencing in X-chromosome inactivation: Advances in understanding facultative heterochromatin formation. *Nature Reviews. Genetics*, *12*(8), 542–553. https://doi.org/10.1038/nrg3035

Wyles, S. P., Li, X., Hrstka, S. C., Reyes, S., Oommen, S., Beraldi, R., Edwards, J., Terzic, A., Olson, T. M., & Nelson, T. J. (2016). Modeling structural and functional deficiencies of RBM20 familial dilated cardiomyopathy using human induced pluripotent stem cells. *Human Molecular Genetics*, *25*(2), 254–265. https://doi.org/10.1093/hmg/ddv468

Yap, K., Chung, T. H., & Makeyev, E. V. (2022). Hybridization-proximity labeling reveals spatially ordered interactions of nuclear RNA compartments. *Molecular Cell*, *82*(2), 463-478.e11. https://doi.org/10.1016/j.molcel.2021.10.009

Yi, W., Li, J., Zhu, X., Wang, X., Fan, L., Sun, W., Liao, L., Zhang, J., Li, X., Ye, J., Chen, F., Taipale, J., Chan, K. M., Zhang, L., & Yan, J. (2020). CRISPR-assisted detection of RNA–protein interactions in living cells. *Nature Methods*, *17*(7), Article 7. https://doi.org/10.1038/s41592-020-0866-0

Yoon, J.-H., Srikantan, S., & Gorospe, M. (2012). MS2-TRAP (MS2-tagged RNA affinity purification): Tagging RNA to identify associated miRNAs. *Methods (San Diego, Calif.)*, *58*(2), 81–87. https://doi.org/10.1016/j.ymeth.2012.07.004

Zhang, B., Arun, G., Mao, Y. S., Lazar, Z., Hung, G., Bhattacharjee, G., Xiao, X., Booth, C. J., Wu, J., Zhang, C., & Spector, D. L. (2012). The lncRNA Malat1 is dispensable for mouse

development but its transcription plays a cis-regulatory role in the adult. *Cell Reports*, *2*(1), 111–123. https://doi.org/10.1016/j.celrep.2012.06.003

Zhang, Z., Sun, W., Shi, T., Lu, P., Zhuang, M., & Liu, J.-L. (2020). Capturing RNA–protein interaction via CRUIS. *Nucleic Acids Research*, *48*(9), e52–e52. https://doi.org/10.1093/nar/gkaa143

Zheng, H., & Xie, W. (2019). The role of 3D genome organization in development and cell differentiation. *Nature Reviews Molecular Cell Biology*, *20*(9), Article 9. https://doi.org/10.1038/s41580-019-0132-4

Zhou, Y., Zhou, B., Pache, L., Chang, M., Khodabakhshi, A. H., Tanaseichuk, O., Benner, C., & Chanda, S. K. (2019). Metascape provides a biologist-oriented resource for the analysis of systems-level datasets. *Nature Communications*, *10*(1), 1523. https://doi.org/10.1038/s41467-019-09234-6

# VITA

**Evan Erik Kania**

2019 B.S. in Pharmaceutical Sciences, The Ohio State University

2023 PhD in Pharmacology, University of Washington

PUBLICATIONS

1. **Kania EE,** Fenix AM, Ong SE, Bertero A, Murry CE, Shechner DM. Nascent transcript-targeted proximity omics enables high-resolution microdissection of splicing speckles and reveals new molecular players in Dilated Cardiomyopathy. In preparation.
2. **Kania EE,** Deng X, Disteche CM, Shechner DM. lncRNA targeted proximity omics enables new insights on the interactome of gene-silencing lncRNAs. In preparation.
3. Tsue AF*, **Kania EE***, Lei DQ, Fields R, McGann CD, Hershberg E, Deng X, Kihiu M, Ong SE, Disteche CM, Kugel S, Beliveau BJ, Schweppe DK, Shechner DM. Oligonucleotide-directed proximity-interactome mapping (O-MAP): A unified method for discovering RNA-interacting proteins, transcripts and genomic loci in situ. bioRxiv. 2023. In review at Nature Methods. doi:10.1101/2023.01.19.524825. (*–Co-first authors)
4. **Kania EE**, Moreno JC, Hernandez VA, English A, Papa JL, Shkolnikov N, Ozer HG, Yilmaz AS, Yalowich JC, Elton TS. hsa-miR-9-3p and hsa-miR-9-5p as Post-Transcriptional Modulators of DNA Topoisomerase IIα in Human Leukemia K562 Cells with Acquired Resistance to Etoposide. Molecular Pharmacology. 2020;97(3):159-170. doi:10.1124/mol.119.118315.
5. Kanagasabai R, Karmahapatra S, Kietz CA, Yu Y, Hernandez VA, **Kania EE**, Yalowich JC, Elton TS. The Novel C-terminal Truncated 90-kDa Isoform of Topoisomerase IIα (TOP2α/90) Is a Determinant of Etoposide Resistance in K562 Leukemia Cells via Heterodimerization with the TOP2α/170 Isoform. Molecular Pharmacology. 2018; 93: 515-525. doi:10.1124/mol.117.111567.

PATENTS

Shechner DM, Beliveau BJ, Akilesh S**,** Attar S, Camplisson C, **Kania EE**, and Tsue AF. "Oligo-mediated deposition of functional molecules for spatially resolved genomics, transcriptomics, and proteomics." Patent Application Serial Number PCT/US23/60620, filed 01/13/2023.