

Diverse roles of RNA-protein interactions: From viral antagonism to mammalian development

Thesis by
Abhik Kumar Banerjee

In Partial Fulfillment of the Requirements for
the degree of
Doctor of Philosophy

The logo for the California Institute of Technology (Caltech), featuring the word "Caltech" in a bold, orange, sans-serif font.

CALIFORNIA INSTITUTE OF TECHNOLOGY
Pasadena, California

2021
(Defended on December 14th, 2020)

© 2021

Abhik Kumar Banerjee
ORCID: 0000-0002-9797-0104

ACKNOWLEDGEMENTS

We are a product of our times and the people around us. This collection of work would not have been possible without many different people in my training and life.

I first wanted to thank my Ph.D. advisor and primary research mentor, Mitchell Guttman. Mitch has constantly supported me throughout my graduate training and has always made himself available to troubleshoot experiments, bounce off new ideas, and discuss implications of our work. He has a solid vision for research projects in his laboratory and remains enthusiastically supportive of them, despite myself wavering from time to time from experimental setbacks. Mitch has encouraged and built a culture of discovery in his group, where trainees have permission to fail, identify problem areas, iterate quickly, and try again, which has led to innovations such as RNA Antisense Purification, Covalent Linkage Affinity Purification, and Split-Pool Recognition of Interactions by Tag Extension. There is a sense of mutual, reinforcing excitement, where one labmate's work inspires another and vice versa. Moreover, it is not uncommon for labmates to stay until the late evening hours, waiting to analyze data as soon as they are available because they *have to know*. It is a truly wonderful place to train, and the group is doing and will continue to make great advances in biology.

Next, I would like to thank my second research mentor, Ellen Rothenberg. Ellen was the reason why I initially selected Caltech for graduate training. Ellen is a brilliant scientist, and I admire her genuine, intellectual curiosity. She is a wonderful mentor, and trainees cannot help but feel re-energized and excited to study science following a meeting with her. I aspire to be like her in my future career.

I also wanted to thank my thesis committee for their counsel and advice over my graduate training. Like my research mentors, my committee was invested in my success and taught me many lessons that I needed to learn to ultimately become successful. Although humbling at times, their criticism helped me become the scientist that I am today. In addition to my committee, I also wanted to thank Steve Mittelman, the former program-director of the USC/Caltech MD-PhD Program who “volunteered me” in the American

Physician Scientists Association (APSA), which would later become a major part of my graduate training experience; Julian Martinez, my first research mentor as an undergraduate student at UCLA who exemplifies the ideal model of a physician scientist investigator; Ben Deverman, my first direct research mentor at Caltech; and the late Paul Patterson, who admitted me to the USC-Caltech dual degree program, allowed me to volunteer in his laboratory during my first graduate school rotation, encouraged me to present data and interact with visiting scholars, and told me to join the Guttman laboratory all those years ago.

Over the past six years, I have had the privilege to work with many talented and supportive labmates in the Guttman and Rothenberg groups. Within the Guttman group, I wanted to thank Mario Blanco, Amy Chow, Sofia Quinodoz, Joanna Jachowicz, Anthony Szempruch, Ward Walkup IV, Linlin Chen, Vickie Trinh, Chris Chen, and Jasmine Thai for their support. I also wanted to acknowledge Shawna Hiley, who provided feedback over this collection of work, and Inna-Marie Strazhnik, who designed the figure layouts used in Chapter 2, as well as the cartoons used in Chapters 2-5. Within the Rothenberg group, I wanted to thank Xun Wang, Maile Romero-Wolf, Maria Lerica Quiloan, Hiro Hosokawa, Jonas Ungerbäck, and Wen Zhou.

I also wanted to thank my colleagues and friends outside of the laboratory, including former and fellow MD-PhD trainees in the USC-Caltech program (Jonathan Tucci, Jonathan Liu, David Basta, Emmanuelle Hodara, and Drayton Harvey) and in APSA (Audra Iness-Christovich, Hannah Turbeville and Sarah Groover). I especially wanted to thank Alison Smith, Omar Toubat, and Acacia Hori, three fellow trainees in the USC/Caltech MD-PhD Program that I am incredibly honored to have worked with on many volunteer activities and have the pleasure to call my friends.

Finally and most importantly, I wanted to thank my parents. They have continuously supported me and sacrificed to provide as many opportunities as they could for me. Despite frustrating setback after setback, they always encouraged me to get back up and try again. I love them always. Thank you both for everything.

ABSTRACT

RNA is a widely utilized and integrated component of core cellular function because of its abilities to recognize and hybridize to nucleic acid templates, spatially localize to different compartments within the cell, bind combinatorially to effector molecules, and in some cases directly catalyze chemical reactions. In this thesis, I describe three cases, illustrating the biomolecule's unique importance in several different aspects of cellular homeostasis. Chapter 1 provides historical context for studying RNA-protein interactions within RNA biology and Virology. Chapter 2 details experiments in which we explored RNA as a central target of host cell takeover by SARS-CoV-2. In the process, we highlight the importance of RNA in many integral complexes within the cell, including components of the spliceosome, the eukaryotic ribosome, and signal recognition particle. Chapter 3 presents data from our consideration of RNA within the context of *cis* gene regulation. We specifically focus on a model RNA-binding protein, SMRT/HDAC1 Associated Repressor Protein (SHARP), and the paternally imprinted long non-coding RNA, *Kcnq1ot1*, as case studies. Chapter 4 describes our dissection of a transcriptional circuit involving SHARP and discusses implications of RNA-binding to developmentally sensitive circuits and processes. Finally, Chapter 5 poses new questions raised by these studies. Together these data emphasize the diverse and unique role RNA plays in cellular homeostasis and suggest additional roles in nuclear compartment stabilization and crosstalk.

PUBLISHED CONTENT AND CONTRIBUTIONS

1. Banerjee, A.K., Blanco, M.R., Bruce, E.A., Honson, D.D., Chen, L.M., Chow, A., Bhat, P., Ollikainen, N., Quinodoz, S.A., Loney, C., Thai, J., Miller, Z.D., Lin, A.E., Schmidt, M.M., Stewart, D.G., Goldfarb, D., De Lorenzo, G., Rihn, S. J., Voorhees, R. M., Botten, J.W., Majumdar, D., and Guttman, M. (2020). SARS-CoV-2 Disrupts Splicing, Translation, and Protein Trafficking to Suppress Host Defenses. *Cell*, S0092-8674(20)31310-6. Advance online publication. <https://doi.org/10.1016/j.cell.2020.10.004>

A.K.B. conceived project, performed protein and 5EU purifications, sequencing, NSP1 and 5' leader flow cytometry, and analysis.

2. Quinodoz, S.A., Bhat, P., Ollikainen, N., Jachowicz, J., Banerjee, A.K., Chovanec, P., Blanco, M.R., Chow, A., Markaki, Y., Plath, K. and Guttman, M. (2020). RNA promotes the formation of spatial compartments in the nucleus. bioRxiv 2020.08.25.267435; doi: <https://doi.org/10.1101/2020.08.25.267435> [Submitted]

A.K.B. performed all Kcnq1ot1 biochemical and functional experiments, including CRISPRi knockdowns, TSA treatments, and functional characterizations; worked with A.C. to develop and characterize the inducible Kcnq1ot1 cell line and to generate homozygous deletions of the SHARP Binding Site within Kcn11ot1; worked with M.R.B. to purify SHARP and map it to Kcnq1ot1.

3. Blanco, M.R., Walkup IV, W.G., Bonesteele, G., Banerjee, A.K., Peyda, P., Amaya, E., Guo, J., Chow, A., Trinh, V., and Guttman, M. (2020). Denaturing purifications demonstrate that PRC2 and other chromatin proteins do not bind directly to RNA *in vivo*. [Submitted]

A.K.B performed Chromatin Immunoprecipitation against candidate chromatin regulatory proteins used in study and Covalent Linkage Affinity Purification against SHARP.

4. Jachowicz, J.W., Strehle, M. Banerjee, A.K., Thai, J., Blanco, M.R., and Guttman, M. (2020). Xist-mediated condensates amplify SHARP localization on the X chromosome to promote chromosome-wide transcriptional silencing. [Submitted]

A.K.B performed site directed mutagenesis of SHARP mutants, and designed and performed cloning for Halo-, mEGFP-, and FUS-mCherry-tagged SHARP expression constructs.

5. Markaki, Y., Chong, J.G., Luong, C., Tan, S.X.Y., Wang, Y., Jacobson, E.C., Maestrini, D., Dror, I., Mistry, B.A., Schoeneberg, J., Banerjee, A., Guttman, M., Chou, T., and Plath, K. (2020). Xist-seeded nucleation sites form local

concentration gradients of silencing proteins to inactivate the X-chromosome.
bioRxiv 2020.11.22.393546; doi: <https://doi.org/10.1101/2020.11.22.393546>
[Submitted]

A.K.B cloned the WT- and Delta IDR SPEN constructs.

For the 447,000 American lives lost to the COVID-19 pandemic. We hope our work will help in the battle against newly emerging pathogenic RNA viruses and prevent further losses of life.

TABLE OF CONTENTS

Acknowledgements.....	iii
Abstract.....	v
Published Content and Contributions.....	vi
Table of Contents.....	ix
List of Figures.....	xi
Chapter 1: Establishing Context and Raising Open Questions	1
1.1 Abstract	2
1.2 Introduction	2
1.3 References	8
Chapter 2: SARS-CoV-2 Disrupts Splicing, Translation, and Protein Trafficking to Suppress Host Defenses	20
2.1 Abstract	22
2.2 Introduction	22
2.3 Results	26
2.4.0 Discussion	44
2.4.1 Limitations of Study.....	47
2.5 Acknowledgements.....	48
2.6 Materials and Methods.....	48
2.7 References	78
2.8 Figures.....	91
Chapter 3: Beckwith-Wiedemann Syndrome, Barr Bodies, and Binding Proteins	125
3.1 Abstract	126
3.2 Introduction	127
3.3 Results	130
3.4 Discussion	133
3.5 Materials and Methods.....	134
3.6 References	153
3.7 Figures and Table.....	164
Chapter 4: The SMRT/HDAC1-Associated Repressor Protein Positively Buffers Notch Signaling in Early T-cell Development	173
4.1 Abstract	174
4.2 Introduction	174
4.3 Results	176
4.4 Discussion: Implications for Gene Regulatory Networks.....	181
4.5 Materials and Methods.....	182

4.6 References	190
4.7 Figures and Tables	199
Chapter 5: Conclusion	209
5.1 Discussion and Implications	210
5.2 Future Directions	213
5.3 References	215
5.4 Figures	221

LIST OF FIGURES

<i>Chapter 2</i>	<i>Page</i>
<i>Number</i>	
1. A human and mouse mixing experiment defines RNA-protein interactions that occur in solution after cell lysis.....	91
2. CLAP removes RNA-protein interactions that do not occur in vivo.....	93
3. CLAP accurately identifies RNA-protein interactions that occur in vivo for well-defined RNA binding proteins	95
4. Global RNA binding maps of SARS-CoV-2 proteins.....	97
5. Global RNA binding maps of SARS-CoV-2 proteins, Supplementary Material Related to Figure 4.....	98
6. NSP16 binds to U1 and U2 at their mRNA recognition sites.....	100
7. NSP16 binds to the U1 and U2 components of the spliceosome at their mRNA recognition sites, Supplemental Material Related to Figure 6.....	102
8. NSP16 suppresses host mRNA splicing.....	104
9. NSP16 suppresses host mRNA splicing, Related to Figure 3.....	106
10. NSP1 binds to the 18S ribosomal RNA near the mRNA entry channel to suppress global mRNA translation in cells, Supplemental Material Related to Figure 11...	108
11. NSP1 binds to 18S near the mRNA entry channel to suppress translation.....	111
12. The 5' viral leader sequence protects mRNAs from NSP1-mediated translational inhibition, Supplemental Material Related to Figure 13.....	113
13. The 5' viral leader protects mRNA from NSP1-mediated translational inhibition.....	115
14. NSP8 and NSP9 bind to 7SL RNA of the Signal Recognition Particle.....	117
15. NSP8 and NSP9 bind to the 7SL RNA component of the Signal Recognition Particle, Supplemental Material Related to Figure 14.....	119
16. NSP8 and NSP9 inhibit membrane and secretory protein function, Supplemental Material Related to Figure 17.....	121
17. NSP8 and NSP9 inhibit membrane and secretory protein trafficking.....	123
<i>Chapter 3</i>	<i>Page</i>
<i>Number</i>	
1. SHARP Covalent Linkage Affinity Purification reveals interaction with lncRNA Kcnq1ot.....	164
2. Kcnq1ot1 is necessary and sufficient to seed an RNA-mediated compartment over its imprinted target genes.....	165
3. Kcnq1ot1 transcriptionally silences imprint-target genes in a SHARP-dependent manner.....	167
4. SHARP-RNA interactions are pervasive throughout the nucleus.....	169

	xii
5. SHARP aggregates in a RRM- and IDR-dependent Manner.....	171
6. Table 1: List of High Confidence SHARP-Interacting RNAs.....	172

Chapter 4
Number

Page

1. Classic model of Notch signaling.....	199
2. SHARP deficiency does not functionally phenocopy Notch gain of function...	200
3. SHARP regulates Notch-mediated gene expression in a NRARP-dependent.....	202
4. SHARP exhibits negative auto-regulation.....	204
5. SHARP positively buffers a Notch-mediated Incoherent Feedforward Loop....	206
6. Table 1: List of High Confidence Notch-Dependent Target Genes in SCID.adh.2c2.....	207
7. Table 2: List of High Confidence Notch-Repressed Target Genes in SCID.adh.2c2.....	208

Chapter 5
Number

Page

1. Preliminary evidence of <i>cis</i> gene regulation by SHARP/pre-mRNAs.....	221
2. CLAP reveals a growing number of uncharacterized RNA-protein interactions.....	222
3. Shared effector proteins enable compartment cross-talk.....	223

Chapter 1

ESTABLISHING CONTEXT AND RAISING OPEN QUESTIONS

A.K. Banerjee and M. Guttman

Material included in this chapter were published and adapted from “Blanco, M.R., Walkup IV, W.G., Bonesteele, G., Banerjee, A.K., Peyda, P., Amaya, E., Guo, J., Chow, A., Trinh, V., and Guttman, M. (in submission/review). Denaturing purifications demonstrate that PRC2 and other chromatin proteins do not bind directly to RNA *in vivo*”.

1.0 ABSTRACT

Since the first documented isolation of nucleic acid in 1869 by Friedrich Miescher, RNA has been recognized as a widely utilized and integrated component of core cellular function. In this Chapter, I provide a historical overview of key advances within RNA biology, with emphasis on the diversity of non-coding RNA. I discuss three types of mechanisms used broadly by non-coding RNAs in the cell, including sponge activities, direct RNA-mediated catalytic activity, and protein scaffolding. In addition, I introduce the paradigm of long non-coding RNAs as spatially concentrating protein scaffolds within the cell and discuss limitations in current methods used to evaluate RNA-protein interactions. Finally, I conclude by discussing the relevance of Virology towards advances in RNA biology and overall provide context for the investigation of RNA-protein interactions during SARS-CoV-2 infection in Chapter 2, genomic imprinting in Chapter 3, and transcriptional circuits and auto-regulation in Chapter 4.

2.0 INTRODUCTION

2.1 DISCOVERY, NON-CODING DIVERSITY, AND THE LENS OF CENTRAL DOGMA

Nucleic acid was first isolated in 1869 by Friedrich Miescher, terming the highly acidic material as nuclein¹⁻². Several forms of nucleic acid were reported over the next few decades, varying as a result of purification method, primary source material, and research laboratory, and later resolved to two main groups: phytonucleic acid (yeast nucleic acid, thought to be typical of plants) and thymonucleic acid (also referred to as zoonucleic acid, thought to be typical of animals)³⁻⁴. It was later determined that differences between these two classes of nucleic acid were a result of changes within a pyrimidine base (uracil in yeast nucleic acid versus thymine in

thymonucleic acid) and sugar (pentose in yeast versus what was thought to be a hexose in animal), ultimately resolving to RNA and DNA respectively³⁻⁴.

The connection between RNA and DNA was gradually established over the mid to late 20th century. In the late 1950s, Volkin and Astrachan described RNA as a DNA-like molecule synthesized from DNA^{5-6,2}. Coupled with X-ray crystallography studies from Rosalind Franklin and the publication of the DNA double helix structure by Watson and Crick in 1953, RNA was later proposed to be an intermediate molecule in the flow of biological information from DNA to protein in a model called the Central Dogma of Molecular Biology^{7-9,2}. This model was additionally supported by evidence from Caspersson and Brachet in 1939, who demonstrated that cells producing high amounts of protein also contained abundant RNA, and Jacob and Monod in 1960, who (among others described in Section 2.4) identified the messenger RNA intermediate during protein synthesis^{10,2}.

In parallel to these coding RNA advances, non-coding RNA also emerged to the forefront. Key discoveries included: ribosomal RNA (rRNA), RNA components of the ribosome; transfer RNAs (tRNA), responsible for translation of RNA nucleotides to amino acids during protein assembly first identified in 1958 by Hoagland and Zamecnik; small nuclear RNAs, components of the spliceosome; small nucleolar RNAs, responsible for processing ribosomal RNA within the nucleolus; microRNAs (miRNAs), 20mer RNAs responsible for post-transcriptional gene silencing; and long non-coding RNA (lncRNA), non-coding RNAs greater than 200 nucleotides in length initially associated with gene regulation; among other types^{11-23,2}.

Central Dogma has been an invaluable paradigm for understanding the flow of biological information, and discoveries within non-coding RNA both complement the complexity of

Central Dogma (rRNA and tRNA) and illustrate the diversity of molecular processes within the cell. For better or worse, due to associations between non-coding RNA and regulatory activity (as a result of miRNAs, lncRNAs, and bacterial regulatory RNAs termed small RNAs), Central Dogma has unintentionally established a dichotomy within RNA biology: that is coding mRNA and regulatory non-coding RNA are mutually exclusive^{24,2}. We later question this dichotomy in Chapters 3 and 5.

2.2 REINING IN LONG NON-CODING RNAs AND CONSIDERATIONS OF CAUSALITY

Recent reports have demonstrated that the mammalian genome is pervasively transcribed, producing upwards of 27,000 lncRNAs according to some estimates²⁵⁻²⁹. Coupled with the observations that 93% of the human genome is actively transcribed but only 1% contains protein-coding gene exons, lncRNAs have become an exciting ‘new’ area of RNA biology research^{30-31, 18-20,2}. It is important, however, to take these newly reported lncRNAs with healthy skepticism because several lncRNAs have been reported to exert their function in an RNA-independent manner (that is, through activity as a *cis* regulatory DNA element or through the act of transcription itself)^{27,32}.

For example, *Lockd* (lncRNA downstream of *Cdkn1b*) is a mouse erythroblast lncRNA, associated with positive regulation of its neighbor, *Cdkn1b*³³⁻³⁴. Using a combination of CRISPR/Cas9-mediated excision of the *Lockd* gene body (which reduces neighbor *Cdkn1b* expression), as well as lncRNA transcript truncation via a pre-mature polyA termination signal (which does *not* effect *Cdkn1b* transcription), the authors demonstrated that a *cis* DNA element within the promoter of the *Lockd* locus, rather than the lncRNA itself, was responsible for its associated function. Another example is linc-p21, a p53-associated lncRNA linked to regulation

of *Cdkn1a*³⁶. Through a combination of a *Linc-p21* mouse knockout model, tissue-specific gene expression patterns, and a massively parallel reporter assay for enhancer activity, Groff et al. demonstrated that *in vivo cis* regulatory effects associated with *Linc-p21* are due to DNA enhancer activity within the locus itself rather than the RNA molecule³⁵⁻³⁷.

In contrast to DNA regulatory elements harbored within a lncRNA gene, the act of transcription can be responsible for a given behavior. An example of this is the Ftx RNA, a non-coding RNA associated with *Xist* lncRNA activation³⁸⁻³⁹. In their 2018 study, Furlan et al. demonstrated that deletion of the Ftx promoter resulted in impaired *Xist* activation. They additionally demonstrated that the Ftx transcript was not strictly required for this phenotype through LNA-Gapmer (locked nucleic acid oligonucleotides antisense to target RNAs) knockdown targeting mature Ftx. The authors later demonstrated that transcription of Ftx is functionally responsible for *Xist* activation using CRISPR-interference³⁸. Ftx aside, groups have also documented that intense transcriptional activity can lead to expression of neighboring genes through a process termed transcriptional ripples⁴⁰. Therefore, transcription of the lncRNA itself may lead to permissive changes in chromatin structure, with the produced RNA molecule ultimately dispensable towards an observed behavior in the laboratory³².

Although certain lncRNAs may be ‘merely’ associated with a given behavior or disease, these associations can nevertheless be leveraged for diagnostic or prognostic purposes⁴¹⁻⁴³. That said, lncRNA research demands additional experimental rigor to distinguish between the possibilities of RNA-independent and RNA-dependent mechanisms of action, particularly when claiming the prize of functional causality and possibly identifying a target for therapeutic intervention^{27,32}.

2.3 REINING IN LONG NON-CODING RNAs AND RNA-PROTEIN INTERACTIONS

Assuming a lncRNA molecule is *functionally* responsible for a given behavior, how are lncRNAs thought to work? The first reported class of mechanisms includes decoy and sponge activities². A well-cited example of a lncRNA decoy is the GAS5 (growth arrest-specific 5) lncRNA, responsible for binding to glucocorticoid receptor by mimicking the nucleotide sequence of a glucocorticoid response element (GRE) and preventing the receptor from recognizing GREs within DNA to effect gene expression^{44,2}. Another example includes the pseudogene derived lncRNA PTENP1. PTENP1 was shown to compete for miRNA binding with PTEN, effectively releasing PTEN from miRNA repression in a DICER-dependent manner^{45-47,2}.

A second mechanism, applicable to non-coding RNA more generally, is direct RNA-mediated catalytic activity⁴⁸. Examples include self-cleaving group I introns, intronic RNA structures responsible for binding to guanosine and removal of 5' RNA splice sites; RNase P, a tRNA processing enzyme responsible for cleaving off 5' leader sequences to produce mature tRNAs; and the large subunit of ribosome, responsible for protein synthesis; among other examples⁴⁸⁻⁵⁷.

The third and most prevailing mechanism is protein scaffolding^{51,58,21,27}. One example is the human telomerase RNA; while its 5' terminal domain binds to TERT protein-binding elements, stem loops within the 3' half bind to dyskerin complexes and TCAB1^{59-61,51}. Another well-established example is the Xist lncRNA, responsible for initiation of X-chromosome inactivation (XCI), the process by which one of the two female X-chromosomes is transcriptionally silenced to establish dosage compensation in early mammalian development²¹. During XCI, the Xist lncRNA recruits several regulatory proteins to the inactive X-chromosome, ultimately resulting in stable and heritable chromosome-wide silencing across an organism's lifetime⁶²⁻⁶⁷. Xist

lncRNA remains an important tool for understanding how lncRNAs can directly interact and spatially concentrate effector proteins within the nucleus^{21,27}. As a result of the paradigm established by Xist lncRNA and the repertoire of effector RNA-binding proteins within the cell, this dissertation examines RNA-protein interactions in several different contexts, including SARS-CoV-2 viral infection (Chapter 2), genomic imprinting (Chapter 3), and transcriptional circuits and auto-regulation (Chapter 4).

RNA-protein interaction studies have historically relied on *in vitro* binding techniques or immunoprecipitation-based techniques, which include RNA Immunoprecipitation and Crosslinking Immunoprecipitation (CLIP)⁶⁸⁻⁷⁰. CLIP has been used successfully to identify precise RNA-binding sites for numerous RNA-binding proteins^{71,62}. However, CLIP has also led to claims of direct RNA interactions by non-canonical RNA-binding proteins, including metabolic proteins and chromatin regulators⁷²⁻⁷³. Mili and Steitz previously showed that immunoprecipitation methods can identify RNA-protein interactions that do not occur *in vivo*, but rather form in solution post cell lysis (*in vitro* associations)⁷⁴. Given discrepancies between CLIP-based biochemical evidence (supporting specific RNA-protein interactions) and genetic evidence (demonstrating that these same interactions are often dispensable), CLIP-based methods need to be reevaluated in light of possible *in vitro* association artifacts^{62,75-83}. We discuss this reevaluation in greater detail in Chapter 2.

2.4 LESSONS FROM VIROLOGY

The fields of Virology and RNA biology are fundamentally linked. Insights from viruses have led to key discoveries within basic RNA biology, processing, gene regulation, and disease states⁸⁴. Select examples include: the discovery of mRNA from bacteriophage; the discovery of

how RNA can carry genetic information, from tobacco mosaic virus; discovery of the retroviral reverse transcriptase from Rous sarcoma virus and Rauscher mouse leukaemia virus; the discovery of mRNA capping from simian vacuolating virus 40 (SV40) and vaccinia virus; the discovery of RNA-splicing from adenovirus; the discovery of the polyadenylation signal from SV40; and early discoveries of RNA-interference using various plant-viruses; among others^{10,84-100}. Vice versa as discussed in Chapter 2 with Covalent Linkage Affinity Purification, RNA biology methods development can also lead to insights in Virology. The two go hand-in-hand.

The term ‘virus’ is derived from the Latin word for poison; given their associations with numerous debilitating diseases in man and detrimental effects on agriculture, livestock, and the food-supply, viruses justifiably remain the target of translational research efforts and application¹⁰¹⁻¹¹⁴. Translational application aside, viruses are an ideal basic science model system, owing to their limited coding capacity, rapid growth and replication cycle, and robust associated phenotypes (with notable exceptions in all three categories to be sure)¹¹⁵. Viruses enable researchers to frame every experimental hypothesis within the context of a simpler guiding question: “how does this interaction promote viral fitness and replication?” This interpretation may not necessarily be entirely intuitive or correct however, particularly with systems-level understanding of host-virus dynamics, as later discussed in Chapters 2 and 5. Still, viruses have increased our understanding of basic RNA biology, with clear ramifications for studies of non-coding RNA and RNA-protein interactions.

3.0 REFERENCES

1. Dahm R. (2005). Friedrich Miescher and the discovery of DNA. *Developmental biology*, 278(2), 274–288. <https://doi.org/10.1016/j.ydbio.2004.11.028>

2. Jarroux, J., Morillon, A., and Pinskaya, M. (2017). History, Discovery, and Classification of lncRNAs. *Advances in experimental medicine and biology*, 1008, 1–46. https://doi.org/10.1007/978-981-10-5203-3_1
3. Levene, P.A. (1910). On the biochemistry of nucleic acids. *Journal of the American Chemical Society*, 32(2), 231-240. <https://pubs.acs.org/doi/abs/10.1021/ja01920a010>
4. Frixione, E., and Ruiz-Zamarripa, L. (2019). The "scientific catastrophe" in nucleic acids research that boosted molecular biology. *The Journal of biological chemistry*, 294(7), 2249–2255. <https://doi.org/10.1074/jbc.CL119.007397>
5. Ochoa, S. (1980). The pursuit of a hobby. *Annual review of biochemistry*, 49, 1–30. <https://doi.org/10.1146/annurev.bi.49.070180.000245>
6. Griffiths, A.J., Miller, J.H., Suzuki, D.T., et al. (2000). An Introduction to Genetic Analysis. 7th Edition, *W.H. Freeman*. Transcription and RNA polymerase. Available from: <https://www.ncbi.nlm.nih.gov/books/NBK22085/>
7. Franklin, R.E., and Gosling, R.G. (2003). Molecular configuration in sodium thymonucleate. 1953. *Nature*, 421(6921), 400–396.
8. Crick, F.H. (1958). On protein synthesis. *Symposia of the Society for Experimental Biology*, 12, 138–163.
9. Crick, F. (1970). Central dogma of molecular biology. *Nature*, 227(5258), 561–563. <https://doi.org/10.1038/227561a0>
10. Cobb, M. (2015). Who discovered messenger RNA?. *Current biology : CB*, 25(13), R526–R532. <https://doi.org/10.1016/j.cub.2015.05.032>
11. Palade, G.E. (1955). A small particulate component of the cytoplasm. *The Journal of biophysical and biochemical cytology*, 1(1), 59–68. <https://doi.org/10.1083/jcb.1.1.59>
12. Berk A.J. (2016). Discovery of RNA splicing and genes in pieces. *Proceedings of the National Academy of Sciences of the United States of America*, 113(4), 801–805. <https://doi.org/10.1073/pnas.1525084113>
13. Weinberg, R.A., and Penman, S. (1968). Small molecular weight monodisperse nuclear RNA. *Journal of molecular biology*, 38(3), 289–304. [https://doi.org/10.1016/0022-2836\(68\)90387-2](https://doi.org/10.1016/0022-2836(68)90387-2)
14. Zieve, G., and Penman, S. (1976). Small RNA species of the HeLa cell: metabolism and subcellular localization. *Cell*, 8(1), 19–31. [https://doi.org/10.1016/0092-8674\(76\)90181-1](https://doi.org/10.1016/0092-8674(76)90181-1)

15. Lee, R.C., Feinbaum, R.L., and Ambros, V. (1993). The *C. elegans* heterochronic gene *lin-4* encodes small RNAs with antisense complementarity to *lin-14*. *Cell*, 75(5), 843–854. [https://doi.org/10.1016/0092-8674\(93\)90529-y](https://doi.org/10.1016/0092-8674(93)90529-y)
16. Wightman, B., Ha, I., and Ruvkun, G. (1993). Posttranscriptional regulation of the heterochronic gene *lin-14* by *lin-4* mediates temporal pattern formation in *C. elegans*. *Cell*, 75(5), 855–862. [https://doi.org/10.1016/0092-8674\(93\)90530-4](https://doi.org/10.1016/0092-8674(93)90530-4)
17. Fire, A., Xu, S., Montgomery, M.K., Kostas, S.A., Driver, S.E., and Mello, C.C. (1998). Potent and specific genetic interference by double-stranded RNA in *Caenorhabditis elegans*. *Nature*, 391(6669), 806–811. <https://doi.org/10.1038/35888>
18. Pachnis, V., Belayew, A., and Tilghman, S.M. (1984). Locus unlinked to alpha-fetoprotein under the control of the murine *raf* and *Rif* genes. *Proceedings of the National Academy of Sciences of the United States of America*, 81(17), 5523–5527. <https://doi.org/10.1073/pnas.81.17.5523>
19. Bartolomei, M.S., Zemel, S., and Tilghman, S.M. (1991). Parental imprinting of the mouse H19 gene. *Nature*, 351(6322), 153–155. <https://doi.org/10.1038/351153a0>
20. Borsani, G., Tonlorenzi, R., Simmler, M. C., Dandolo, L., Arnaud, D., Capra, V., Grompe, M., Pizzuti, A., Muzny, D., Lawrence, C., Willard, H.F., Avner, P., and Ballabio, A. (1991). Characterization of a murine gene expressed from the inactive X chromosome. *Nature*, 351(6324), 325–329. <https://doi.org/10.1038/351325a0>
21. Strehle, M., and 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>
22. Hoagland, M.B., Keller, E.B., and Zamecnik, P.C. (2009). The mechanism of amino acid activation: the work of Mahlon Hoagland. 1956. *The Journal of biological chemistry*, 284(25), e7–e8.
23. Hoagland, M.B., Stephenson, M.L., Scott, J.F., Hecht, L.I., and Zamecnik, P.C. (1958). A soluble ribonucleic acid intermediate in protein synthesis. *The Journal of biological chemistry*, 231(1), 241–257.
24. Inouye, M., and Delihias, N. (1988). Small RNAs in the prokaryotes: a growing list of diverse roles. *Cell*, 53(1), 5–7. [https://doi.org/10.1016/0092-8674\(88\)90480-1](https://doi.org/10.1016/0092-8674(88)90480-1)
25. Okazaki, Y., Furuno, M., Kasukawa, T., Adachi, J., Bono, H., Kondo, S., Nikaido, I., Osato, N., Saito, R., Suzuki, H., Yamanaka, I., Kiyosawa, H., Yagi, K., Tomaru, Y., Hasegawa, Y., Nogami, A., Schönbach, C., Gojobori, T., Baldarelli, R., Hill, D.P., ... RIKEN Genome Exploration Research Group Phase I and II Team (2002). Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs. *Nature*, 420(6915), 563–573. <https://doi.org/10.1038/nature01266>

26. Kapranov, P., Cheng, J., Dike, S., Nix, D.A., Dutttagupta, R., Willingham, A.T., Stadler, P. F., Hertel, J., Hackermüller, J., Hofacker, I.L., Bell, I., Cheung, E., Drenkow, J., Dumais, E., Patel, S., Helt, G., Ganesh, M., Ghosh, S., Piccolboni, A., Sementchenko, V., ... Gingeras, T.R. (2007). RNA maps reveal new RNA classes and a possible function for pervasive transcription. *Science (New York, N.Y.)*, 316(5830), 1484–1488. <https://doi.org/10.1126/science.1138341>
27. Engreitz, J.M., Haines, J.E., Perez, E.M., Munson, G., Chen, J., Kane, M., McDonel, P.E., Guttman, M., and Lander, E.S. (2016). Local regulation of gene expression by lncRNA promoters, transcription and splicing. *Nature*, 539(7629), 452–455. <https://doi.org/10.1038/nature20149>
28. Hon, C.C., Ramilowski, J.A., Harshbarger, J., Bertin, N., Rackham, O.J., Gough, J., Denisenko, E., Schmeier, S., Poulsen, T. M., Severin, J., Lizio, M., Kawaji, H., Kasukawa, T., Itoh, M., Burroughs, A.M., Noma, S., Djebali, S., Alam, T., Medvedeva, Y.A., Testa, A.C., ... Forrest, A. R. (2017). An atlas of human long non-coding RNAs with accurate 5' ends. *Nature*, 543(7644), 199–204. <https://doi.org/10.1038/nature21374>
29. Ali, T., and Grote, P. (2020). Beyond the RNA-dependent function of lncRNA genes. *eLife*, 9, e60583. <https://doi.org/10.7554/eLife.60583>
30. ENCODE Project Consortium, Birney, E., Stamatoyannopoulos, J.A., Dutta, A., Guigó, R., Gingeras, T.R., Margulies, E.H., Weng, Z., Snyder, M., Dermitzakis, E.T., Thurman, R.E., Kuehn, M.S., Taylor, C.M., Neph, S., Koch, C.M., Asthana, S., Malhotra, A., Adzhubei, I., Greenbaum, J.A., Andrews, R. M., ... de Jong, P.J. (2007). Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. *Nature*, 447(7146), 799–816. <https://doi.org/10.1038/nature05874>
31. ENCODE Project Consortium (2012). An integrated encyclopedia of DNA elements in the human genome. *Nature*, 489(7414), 57–74. <https://doi.org/10.1038/nature11247>
32. 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., and Ponting, C. P. (2014). Considerations when investigating lncRNA function in vivo. *eLife*, 3, e03058. <https://doi.org/10.7554/eLife.03058>
33. Paralkar, V.R., Mishra, T., Luan, J., Yao, Y., Kossenkov, A.V., Anderson, S.M., Dunagin, M., Pimkin, M., Gore, M., Sun, D., Konuthula, N., Raj, A., An, X., Mohandas, N., Bodine, D. M., Hardison, R.C., and Weiss, M.J. (2014). Lineage and species-specific long noncoding RNAs during erythro-megakaryocytic development. *Blood*, 123(12), 1927–1937. <https://doi.org/10.1182/blood-2013-12-544494>
34. Paralkar, V.R., Tabora, C.C., Huang, P., Yao, Y., Kossenkov, A.V., Prasad, R., Luan, J., Davies, J.O., Hughes, J.R., Hardison, R.C., Blobel, G.A., and Weiss, M.J. (2016). Unlinking an lncRNA from Its Associated cis Element. *Molecular cell*, 62(1), 104–110. <https://doi.org/10.1016/j.molcel.2016.02.029>

35. Groff, A.F., Sanchez-Gomez, D.B., Soruco, M., Gerhardinger, C., Barutcu, A.R., Li, E., Elcavage, L., Plana, O., Sanchez, L.V., Lee, J.C., Sauvageau, M., and Rinn, J.L. (2016). In Vivo Characterization of Linc-p21 Reveals Functional cis-Regulatory DNA Elements. *Cell reports*, 16(8), 2178–2186. <https://doi.org/10.1016/j.celrep.2016.07.050>
36. Sauvageau, M., Goff, L.A., Lodato, S., Bonev, B., Groff, A.F., Gerhardinger, C., Sanchez-Gomez, D.B., Haciosuleyman, E., Li, E., Spence, M., Liapis, S.C., Mallard, W., Morse, M., Swerdel, M.R., D'Ecclessis, M.F., Moore, J.C., Lai, V., Gong, G., Yancopoulos, G.D., Friendewey, D., ... Rinn, J.L. (2013). Multiple knockout mouse models reveal lincRNAs are required for life and brain development. *eLife*, 2, e01749. <https://doi.org/10.7554/eLife.01749>
37. Goff, L.A., Groff, A.F., Sauvageau, M., Trayer-Gibson, Z., Sanchez-Gomez, D.B., Morse, M., Martin, R.D., Elcavage, L.E., Liapis, S.C., Gonzalez-Celeiro, M., Plana, O., Li, E., Gerhardinger, C., Tomassy, G.S., Arlotta, P., and Rinn, J.L. (2015). Spatiotemporal expression and transcriptional perturbations by long noncoding RNAs in the mouse brain. *Proceedings of the National Academy of Sciences of the United States of America*, 112(22), 6855–6862. <https://doi.org/10.1073/pnas.1411263112>
38. Furlan, G., Gutierrez Hernandez, N., Huret, C., Galupa, R., van Bommel, J.G., Romito, A., Heard, E., Morey, C., and Rougeulle, C. (2018). The Ftx Noncoding Locus Controls X Chromosome Inactivation Independently of Its RNA Products. *Molecular cell*, 70(3), 462–472.e8. <https://doi.org/10.1016/j.molcel.2018.03.024>
39. Chureau, C., Chantalat, S., Romito, A., Galvani, A., Duret, L., Avner, P., and Rougeulle, C. (2011). Ftx is a non-coding RNA which affects Xist expression and chromatin structure within the X-inactivation center region. *Human molecular genetics*, 20(4), 705–718. <https://doi.org/10.1093/hmg/ddq516>
40. Ebisuya, M., Yamamoto, T., Nakajima, M., and Nishida, E. (2008). Ripples from neighbouring transcription. *Nature cell biology*, 10(9), 1106–1113. <https://doi.org/10.1038/ncb1771>
41. Fattahi, S., Kosari-Monfared, M., Golpour, M., Emami, Z., Ghasemiyan, M., Nouri, M., and Akhavan-Niaki, H. (2020). LncRNAs as potential diagnostic and prognostic biomarkers in gastric cancer: A novel approach to personalized medicine. *Journal of cellular physiology*, 235(4), 3189–3206. <https://doi.org/10.1002/jcp.29260>
42. Galamb, O., Barták, B.K., Kalmár, A., Nagy, Z.B., Szigeti, K.A., Tulassay, Z., Igaz, P., and Molnár, B. (2019). Diagnostic and prognostic potential of tissue and circulating long non-coding RNAs in colorectal tumors. *World journal of gastroenterology*, 25(34), 5026–5048. <https://doi.org/10.3748/wjg.v25.i34.5026>
43. Dastmalchi, N., Safaralizadeh, R., and Nargesi, M.M. (2020). LncRNAs: Potential Novel Prognostic and Diagnostic Biomarkers in Colorectal Cancer. *Current medicinal chemistry*, 27(30), 5067–5077. <https://doi.org/10.2174/0929867326666190227230024>

44. Kino, T., Hurt, D.E., Ichijo, T., Nader, N., and Chrousos, G.P. (2010). Noncoding RNA gas5 is a growth arrest- and starvation-associated repressor of the glucocorticoid receptor. *Science signaling*, 3(107), ra8. <https://doi.org/10.1126/scisignal.2000568>
45. Poliseno, L., and Pandolfi, P.P. (2015). PTEN ceRNA networks in human cancer. *Methods (San Diego, Calif.)*, 77-78, 41–50. <https://doi.org/10.1016/j.ymeth.2015.01.013>
46. Guo, X., Deng, L., Deng, K., Wang, H., Shan, T., Zhou, H., Liang, Z., Xia, J., and Li, C. (2016). Pseudogene PTENP1 Suppresses Gastric Cancer Progression by Modulating PTEN. *Anti-cancer agents in medicinal chemistry*, 16(4), 456–464. <https://doi.org/10.2174/1871520615666150507121407>
47. An, Y., Furber, K.L., and Ji, S. (2017). Pseudogenes regulate parental gene expression via ceRNA network. *Journal of cellular and molecular medicine*, 21(1), 185–192. <https://doi.org/10.1111/jcmm.12952>
48. Hu, G., Niu, F., Humburg, B.A., Liao, K., Bendi, S., Callen, S., Fox, H. S., and Buch, S. (2018). Molecular mechanisms of long noncoding RNAs and their role in disease pathogenesis. *Oncotarget*, 9(26), 18648–18663. <https://doi.org/10.18632/oncotarget.24307>
49. Bass, B.L., and Cech, T.R. (1984). Specific interaction between the self-splicing RNA of Tetrahymena and its guanosine substrate: implications for biological catalysis by RNA. *Nature*, 308(5962), 820–826. <https://doi.org/10.1038/308820a0>
50. Kruger, K., Grabowski, P. J., Zaug, A.J., Sands, J., Gottschling, D.E., and Cech, T.R. (1982). Self-splicing RNA: autoexcision and autocyclization of the ribosomal RNA intervening sequence of Tetrahymena. *Cell*, 31(1), 147–157. [https://doi.org/10.1016/0092-8674\(82\)90414-7](https://doi.org/10.1016/0092-8674(82)90414-7)
51. Cech, T.R., and 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>
52. Cech T.R. (1990). Nobel lecture. Self-splicing and enzymatic activity of an intervening sequence RNA from Tetrahymena. *Bioscience reports*, 10(3), 239–261. <https://doi.org/10.1007/BF01117241>
53. Guerrier-Takada, C., Gardiner, K., Marsh, T., Pace, N., and Altman, S. (1983). The RNA moiety of ribonuclease P is the catalytic subunit of the enzyme. *Cell*, 35(3 Pt 2), 849–857. [https://doi.org/10.1016/0092-8674\(83\)90117-4](https://doi.org/10.1016/0092-8674(83)90117-4)
54. Wu, S., Mao, G., and Kirsebom, L.A. (2016). Inhibition of Bacterial RNase P RNA by Phenothiazine Derivatives. *Biomolecules*, 6(3), 38. <https://doi.org/10.3390/biom6030038>
55. Brimacombe, R., and Stiege, W. (1985). Structure and function of ribosomal RNA. *The Biochemical journal*, 229(1), 1–17. <https://doi.org/10.1042/bj2290001>

56. Noller, H.F., Hoffarth, V., and Zimniak, L. (1992). Unusual resistance of peptidyl transferase to protein extraction procedures. *Science (New York, N.Y.)*, 256(5062), 1416–1419. <https://doi.org/10.1126/science.1604315>
57. Voorhees, R.M., and Ramakrishnan, V. (2013). Structural basis of the translational elongation cycle. *Annual review of biochemistry*, 82, 203–236. <https://doi.org/10.1146/annurev-biochem-113009-092313>
58. Quinodoz, S., and Guttman, M. (2014). Long noncoding RNAs: an emerging link between gene regulation and nuclear organization. *Trends in cell biology*, 24(11), 651–663. <https://doi.org/10.1016/j.tcb.2014.08.009>
59. Tycowski, K. T., Shu, M.D., Kukoyi, A., and Steitz, J.A. (2009). A conserved WD40 protein binds the Cajal body localization signal of scaRNP particles. *Molecular cell*, 34(1), 47–57. <https://doi.org/10.1016/j.molcel.2009.02.020>
60. Venteicher, A.S., Abreu, E.B., Meng, Z., McCann, K.E., Terns, R.M., Veenstra, T.D., Terns, M.P., and Artandi, S.E. (2009). A human telomerase holoenzyme protein required for Cajal body localization and telomere synthesis. *Science (New York, N.Y.)*, 323(5914), 644–648. <https://doi.org/10.1126/science.1165357>
61. Egan, E.D., and Collins, K. (2010). Specificity and stoichiometry of subunit interactions in the human telomerase holoenzyme assembled in vivo. *Molecular and cellular biology*, 30(11), 2775–2786. <https://doi.org/10.1128/MCB.00151-10>
62. 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., and 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>
63. Chu, C., Zhang, Q.C., da Rocha, S.T., Flynn, R.A., Bharadwaj, M., Calabrese, J.M., Magnuson, T., Heard, E., and 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>
64. Moindrot, B., Cerase, A., Coker, H., Masui, O., Grijzenhout, A., Pintacuda, G., Schermelleh, L., Nesterova, T.B., and Brockdorff, N. (2015). A Pooled shRNA Screen Identifies Rbm15, Spen, and Wtap as Factors Required for Xist RNA-Mediated Silencing. *Cell reports*, 12(4), 562–572. <https://doi.org/10.1016/j.celrep.2015.06.053>
65. Monfort, A., Di Minin, G., Postlmayr, A., Freimann, R., Arieti, F., Thore, S., and Wutz, A. (2015). Identification of Spen as a Crucial Factor for Xist Function through Forward Genetic Screening in Haploid Embryonic Stem Cells. *Cell reports*, 12(4), 554–561. <https://doi.org/10.1016/j.celrep.2015.06.067>

66. Chen, C. K., Blanco, M., Jackson, C., Aznauryan, E., Ollikainen, N., Surka, C., Chow, A., Cerase, A., McDonel, P., and Guttman, M. (2016). Xist recruits the X chromosome to the nuclear lamina to enable chromosome-wide silencing. *Science (New York, N.Y.)*, 354(6311), 468–472. <https://doi.org/10.1126/science.aae0047>
67. Wutz, A., and Jaenisch, R. (2000). A shift from reversible to irreversible X inactivation is triggered during ES cell differentiation. *Molecular cell*, 5(4), 695–705. [https://doi.org/10.1016/s1097-2765\(00\)80248-8](https://doi.org/10.1016/s1097-2765(00)80248-8)
68. Peritz, T., Zeng, F., Kannanayakal, T. J., Kilk, K., Eiríksdóttir, E., Langel, U., and Eberwine, J. (2006). Immunoprecipitation of mRNA-protein complexes. *Nature protocols*, 1(2), 577–580. <https://doi.org/10.1038/nprot.2006.82>
69. Keene, J.D., Komisarow, J.M., and Friedersdorf, M.B. (2006). RIP-Chip: the isolation and identification of mRNAs, microRNAs and protein components of ribonucleoprotein complexes from cell extracts. *Nature protocols*, 1(1), 302–307. <https://doi.org/10.1038/nprot.2006.47>
70. Hendrickson, G.D., Kelley, D.R., Tenen, D., Bernstein, B., and Rinn, J.L. (2016). Widespread RNA binding by chromatin-associated proteins. *Genome biology*, 17, 28. <https://doi.org/10.1186/s13059-016-0878-3>
71. Darnell R.B. (2010). HITS-CLIP: panoramic views of protein-RNA regulation in living cells. *Wiley interdisciplinary reviews. RNA*, 1(2), 266–286. <https://doi.org/10.1002/wrna.31>
72. Garcin E.D. (2019). GAPDH as a model non-canonical AU-rich RNA binding protein. *Seminars in cell and developmental biology*, 86, 162–173. <https://doi.org/10.1016/j.semcdb.2018.03.013>
73. Sigova, A.A., Abraham, B.J., Ji, X., Molinie, B., Hannett, N.M., Guo, Y. E., Jangi, M., Giallourakis, C.C., Sharp, P.A., and Young, R.A. (2015). Transcription factor trapping by RNA in gene regulatory elements. *Science (New York, N.Y.)*, 350(6263), 978–981. <https://doi.org/10.1126/science.aad3346>
74. Mili, S., and Steitz, J.A. (2004). Evidence for reassociation of RNA-binding proteins after cell lysis: implications for the interpretation of immunoprecipitation analyses. *RNA (New York, N.Y.)*, 10(11), 1692–1694. <https://doi.org/10.1261/rna.7151404>
75. Zhao, J., Sun, B.K., Erwin, J.A., Song, J.J., and Lee, J.T. (2008). Polycomb proteins targeted by a short repeat RNA to the mouse X chromosome. *Science (New York, N.Y.)*, 322(5902), 750–756. <https://doi.org/10.1126/science.1163045>
76. Cifuentes-Rojas, C., Hernandez, A.J., Sarma, K., and Lee, J.T. (2014). Regulatory interactions between RNA and polycomb repressive complex 2. *Molecular cell*, 55(2), 171–185. <https://doi.org/10.1016/j.molcel.2014.05.009>

77. Schoeftner, S., Sengupta, A.K., Kubicek, S., Mechtler, K., Spahn, L., Koseki, H., Jenuwein, T., and Wutz, A. (2006). Recruitment of PRC1 function at the initiation of X inactivation independent of PRC2 and silencing. *The EMBO journal*, 25(13), 3110–3122. <https://doi.org/10.1038/sj.emboj.7601187>
78. Kalantry, S., and Magnuson, T. (2006). The Polycomb group protein EED is dispensable for the initiation of random X-chromosome inactivation. *PLoS genetics*, 2(5), e66. <https://doi.org/10.1371/journal.pgen.0020066>
79. Plath, K., Fang, J., Mlynarczyk-Evans, S.K., Cao, R., Worringer, K.A., Wang, H., de la Cruz, C. C., Otte, A. P., Panning, B., and Zhang, Y. (2003). Role of histone H3 lysine 27 methylation in X inactivation. *Science (New York, N.Y.)*, 300(5616), 131–135. <https://doi.org/10.1126/science.1084274>
80. da Rocha, S. T., Boeva, V., Escamilla-Del-Arenal, M., Ancelin, K., Granier, C., Matias, N. R., Sanulli, S., Chow, J., Schulz, E., Picard, C., Kaneko, S., Helin, K., Reinberg, D., Stewart, A. F., Wutz, A., Margueron, R., and Heard, E. (2014). Jarid2 Is Implicated in the Initial Xist-Induced Targeting of PRC2 to the Inactive X Chromosome. *Molecular cell*, 53(2), 301–316. <https://doi.org/10.1016/j.molcel.2014.01.002>
81. Kohlmaier, A., Savarese, F., Lachner, M., Martens, J., Jenuwein, T., and Wutz, A. (2004). A chromosomal memory triggered by Xist regulates histone methylation in X inactivation. *PLoS biology*, 2(7), E171. <https://doi.org/10.1371/journal.pbio.0020171>
82. Portoso, M., Ragazzini, R., Brenčić, Ž., Moiani, A., Michaud, A., Vassilev, I., Wassef, M., Servant, N., Sargueil, B., and Margueron, R. (2017). PRC2 is dispensable for *HOTAIR*-mediated transcriptional repression. *The EMBO journal*, 36(8), 981–994. <https://doi.org/10.15252/emboj.201695335>
83. Brockdorff, N. (2013). Noncoding RNA and Polycomb recruitment. *RNA (New York, N.Y.)*, 19(4), 429–442. <https://doi.org/10.1261/rna.037598.112>
84. Enquist, L. W., and Editors of the Journal of Virology (2009). Virology in the 21st century. *Journal of virology*, 83(11), 5296–5308. <https://doi.org/10.1128/JVI.00151-09>
85. Volkin, E., and Astrachan, L. (1956). Phosphorus incorporation in Escherichia coli ribonucleic acid after infection with bacteriophage T2. *Virology*, 2(2), 149–161. [https://doi.org/10.1016/0042-6822\(56\)90016-2](https://doi.org/10.1016/0042-6822(56)90016-2)
86. Brenner, S., Jacob, F., and Meselson, M. (1961). An unstable intermediate carrying information from genes to ribosomes for protein synthesis. *Nature*, 190, 576–581. <https://doi.org/10.1038/190576a0>
87. Fraenkel-Conrat, H. (1956). The role of the nucleic acid in the reconstitution of active tobacco mosaic virus. *Journal of the American Chemical Society*, 78(4), 882–883. <https://doi.org/10.1021/ja01585a055>

88. Temin, H.M., and Mizutani, S. (1970). Viral RNA-dependent DNA Polymerase: RNA-dependent DNA Polymerase in Virions of Rous Sarcoma Virus. *Nature*, 226, 1211-1213. <https://www.nature.com/articles/2261211a0>
89. Baltimore, D. (1970). Viral RNA-dependent DNA Polymerase: RNA-dependent DNA Polymerase in Virions of RNA Tumour Viruses. *Nature*, 226, 1209-1211.
90. Lavi, S., and Shatkin, A.J. (1975). Methylated simian virus 40 specific RNA from nuclei and cytoplasm of infected BSC 1 cells. *Proceedings of the National Academy of Sciences of the United States of America*, 72(6), 2012-2016.
91. Wei, C.M., and Moss, B. (1975). Methylated nucleotides block 5' terminus of vaccinia virus messenger RNA. *Proceedings of the National Academy of Sciences of the United States of America*, 72(1), 318-322.
92. Chow, L.T., Roberts, J.M., Lewis, J.B., and Broker, T.R. (1977). A map of cytoplasmic RNA transcripts from lytic adenovirus type 2, determined by electron microscopy of RNA:DNA hybrids. *Cell*, 11(4), 819–836. [https://doi.org/10.1016/0092-8674\(77\)90294-x](https://doi.org/10.1016/0092-8674(77)90294-x)
93. Berk, A.J., and Sharp, P.A. (1977). Sizing and mapping of early adenovirus mRNAs by gel electrophoresis of S1 endonuclease-digested hybrids. *Cell*, 12(3), 721–732. [https://doi.org/10.1016/0092-8674\(77\)90272-0](https://doi.org/10.1016/0092-8674(77)90272-0)
94. Berget, S.M., Moore, C., and Sharp, P.A. (1977). Spliced segments at the 5' terminus of adenovirus 2 late mRNA. *Proceedings of the National Academy of Sciences of the United States of America*, 74(8), 3171-3175. <https://doi.org/10.1073/pnas.74.8.3171>
95. Berk A.J. (2016). Discovery of RNA splicing and genes in pieces. *Proceedings of the National Academy of Sciences of the United States of America*, 113(4), 801–805. <https://doi.org/10.1073/pnas.1525084113>
96. Fitzgerald, M., and Shenk, T. (1981). The sequence 5'-AAUAAA-3' forms parts of the recognition site for polyadenylation of late SV40 mRNAs. *Cell*, 24(1), 251–260. [https://doi.org/10.1016/0092-8674\(81\)90521-3](https://doi.org/10.1016/0092-8674(81)90521-3)
97. Napoli, C., Lemieux, C., and Jorgensen, R. (1990). Introduction of a Chimeric Chalcone Synthase Gene into Petunia Results in Reversible Co-Suppression of Homologous Genes in trans. *The Plant cell*, 2(4), 279–289. <https://doi.org/10.1105/tpc.2.4.279>
98. Angell, S.M., and Baulcombe, D.C. (1997). Consistent gene silencing in transgenic plants expressing a replicating potato virus X RNA. *The EMBO journal*, 16(12), 3675–3684. <https://doi.org/10.1093/emboj/16.12.3675>
99. Ratcliff, F., Harrison, B.D., and Baulcombe, D.C. (1997). A similarity between viral defense and gene silencing in plants. *Science (New York, N.Y.)*, 276(5318), 1558–1560. <https://doi.org/10.1126/science.276.5318.1558>

100. Eamens, A., Wang, M.B., Smith, N.A., and Waterhouse, P.M. (2008). RNA silencing in plants: yesterday, today, and tomorrow. *Plant physiology*, 147(2), 456–468. <https://doi.org/10.1104/pp.108.117275>
101. Lustig, A., and Levine, A.J. (1992). One hundred years of virology. *Journal of virology*, 66(8), 4629–4631. <https://doi.org/10.1128/JVI.66.8.4629-4631.1992>
102. Taylor, M.W. (2014). Viruses and Man: A History of Interactions. Introduction: A Short History of Virology, 1-22. doi: [10.1007/978-3-319-07758-1_1](https://doi.org/10.1007/978-3-319-07758-1_1)
103. Dürst, M., Gissmann, L., Ikenberg, H., and zur Hausen, H. (1983). A papillomavirus DNA from a cervical carcinoma and its prevalence in cancer biopsy samples from different geographic regions. *Proceedings of the National Academy of Sciences of the United States of America*, 80(12), 3812–3815. <https://doi.org/10.1073/pnas.80.12.3812>
104. Poiesz, B.J., Ruscetti, F.W., Gazdar, A.F., Bunn, P.A., Minna, J.D., and Gallo, R.C. (1980). Detection and isolation of type C retrovirus particles from fresh and cultured lymphocytes of a patient with cutaneous T-cell lymphoma. *Proceedings of the National Academy of Sciences of the United States of America*, 77(12), 7415–7419. <https://doi.org/10.1073/pnas.77.12.7415>
105. Barré-Sinoussi, F., Chermann, J.C., Rey, F., Nugeyre, M.T., Chamaret, S., Gruest, J., Dauguet, C., Axler-Blin, C., Vézinet-Brun, F., Rouzioux, C., Rozenbaum, W., and Montagnier, L. (1983). Isolation of a T-lymphotropic retrovirus from a patient at risk for acquired immune deficiency syndrome (AIDS). *Science (New York, N.Y.)*, 220(4599), 868–871. <https://doi.org/10.1126/science.6189183>
106. Gallo, R.C., Salahuddin, S.Z., Popovic, M., Shearer, G.M., Kaplan, M., Haynes, B.F., Palker, T.J., Redfield, R., Oleske, J., and Safai, B. (1984). Frequent detection and isolation of cytopathic retroviruses (HTLV-III) from patients with AIDS and at risk for AIDS. *Science (New York, N.Y.)*, 224(4648), 500–503. <https://doi.org/10.1126/science.6200936>
107. Levy, J.A., Hoffman, A.D., Kramer, S.M., Landis, J.A., Shimabukuro, J.M., and Oshiro, L.S. (1984). Isolation of lymphocytopathic retroviruses from San Francisco patients with AIDS. *Science (New York, N.Y.)*, 225(4664), 840–842. <https://doi.org/10.1126/science.6206563>
108. Choo, Q.L., Kuo, G., Weiner, A.J., Overby, L.R., Bradley, D.W., and Houghton, M. (1989). Isolation of a cDNA clone derived from a blood-borne non-A, non-B viral hepatitis genome. *Science (New York, N.Y.)*, 244(4902), 359–362. <https://doi.org/10.1126/science.2523562>
109. Andersen, K.G., Rambaut, A., Lipkin, W.I., Holmes, E.C., and Garry, R.F. (2020). The proximal origin of SARS-CoV-2. *Nature medicine*, 26(4), 450–452. <https://doi.org/10.1038/s41591-020-0820-9>

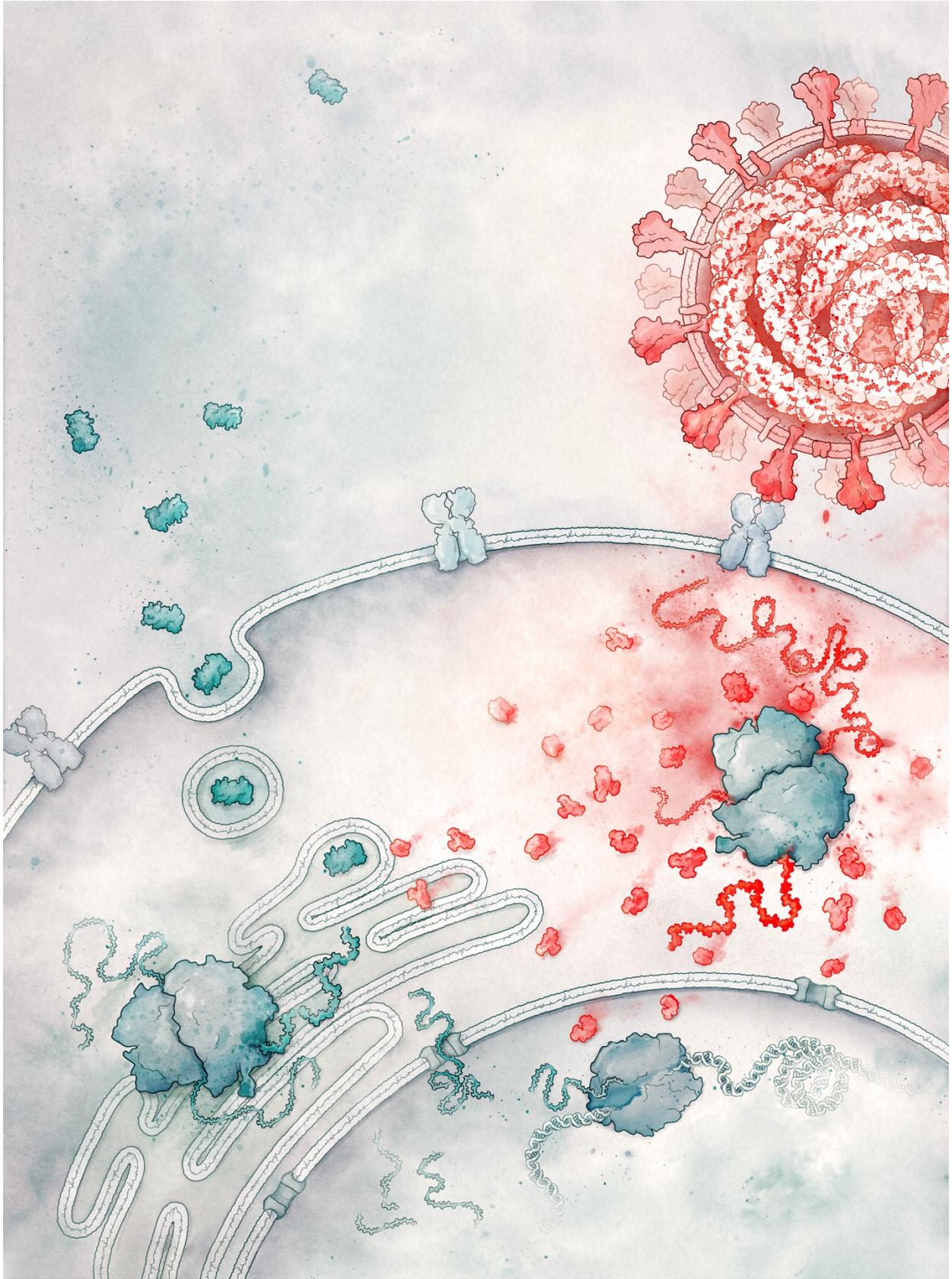
110. Zou, L., Ruan, F., Huang, M., Liang, L., Huang, H., Hong, Z., Yu, J., Kang, M., Song, Y., Xia, J., Guo, Q., Song, T., He, J., Yen, H.L., Peiris, M., and Wu, J. (2020). SARS-CoV-2 Viral Load in Upper Respiratory Specimens of Infected Patients. *The New England journal of medicine*, 382(12), 1177–1179. <https://doi.org/10.1056/NEJMc2001737>
111. Nathues, H., Alarcon, P., Rushton, J., Jolie, R., Fiebig, K., Jimenez, M., Geurts, V., and Nathues, C. (2017). Cost of porcine reproductive and respiratory syndrome virus at individual farm level - An economic disease model. *Preventive veterinary medicine*, 142, 16–29. <https://doi.org/10.1016/j.prevetmed.2017.04.006>
112. Chi, J., VanLeeuwen, J. A., Weersink, A., & Keefe, G. P. (2002). Direct production losses and treatment costs from bovine viral diarrhoea virus, bovine leukosis virus, Mycobacterium avium subspecies paratuberculosis, and Neospora caninum. *Preventive veterinary medicine*, 55(2), 137–153. [https://doi.org/10.1016/s0167-5877\(02\)00094-6](https://doi.org/10.1016/s0167-5877(02)00094-6)
113. Rubio, L., Galipienso, L., & Ferriol, I. (2020). Detection of Plant Viruses and Disease Management: Relevance of Genetic Diversity and Evolution. *Frontiers in plant science*, 11, 1092. <https://doi.org/10.3389/fpls.2020.01092>
114. Mumford, R.A., Macarthur, R., and Boonham, N. (2016). The role and challenges of new diagnostic technology in plant biosecurity. *Food Sec.* 8, 103–109. <https://doi.org/10.1007/s12571-015-0533-y>
115. Compans, R.W., & Roberts, P.C. (1994). Viruses as model systems in cell biology. *Methods in cell biology*, 43 Pt A, 3–42. [https://doi.org/10.1016/s0091-679x\(08\)60596-8](https://doi.org/10.1016/s0091-679x(08)60596-8)

*Chapter 2***SARS-CoV-2 DISRUPTS SPLICING, TRANSLATION, AND PROTEIN
TRAFFICKING TO SUPPRESS HOST DEFENSES**

Material included in this chapter were published as “Banerjee, A.K., Blanco, M.R., Bruce, E.A., Honson, D.D., Chen, L.M., Chow, A., Bhat, P., Ollikainen, N., Quinodoz, S.A., Loney, C., Thai, J., Miller, Z.D., Lin, A.E., Schmidt, M.M., Stewart, D.G., Goldfarb, D., De Lorenzo, G., Rihn, S. J., Voorhees, R.M., Botten, J.W., Majumdar, D., and Guttman, M. (2020). SARS-CoV-2 Disrupts Splicing, Translation, and Protein Trafficking to Suppress Host Defenses. *Cell*, S0092-8674(20)31310-6. Advance online publication. <https://doi.org/10.1016/j.cell.2020.10.004>” ;

AND

“Blanco, M.R., Walkup IV, W.G., Bonesteele, G., Banerjee, A.K., Peyda, P., Amaya, E., Guo, J., Chow, A., Trinh, V., and Guttman, M. (in submission/review). Denaturing purifications demonstrate that PRC2 and other chromatin proteins do not bind directly to RNA *in vivo*”.



1.0 ABSTRACT

SARS-CoV-2 is a novel coronavirus that causes the respiratory disease known as COVID-19. Despite the urgent need, we still do not fully understand the molecular basis of SARS-CoV-2 pathogenesis. Here, we comprehensively define the interactions between SARS-CoV-2 proteins and human RNAs using Covalent Linkage Affinity Purification, a technique capable of identifying *bona fide* RNA-protein interactions while reducing deleterious *in vitro* association artifacts. We show that NSP16 binds to the mRNA recognition domains of the U1 and U2 splicing RNAs and acts to suppress global mRNA splicing upon SARS-CoV-2 infection. We find that NSP1 binds to 18S ribosomal RNA in the mRNA entry channel of the ribosome and leads to global inhibition of mRNA translation upon infection. Finally, we find that NSP8 and NSP9 bind to the 7SL RNA in the Signal Recognition Particle and interfere with protein trafficking upon infection. Disruption of these cellular functions suppresses the interferon response to viral infection. Our results uncover a multipronged strategy utilized by SARS-CoV-2 to antagonize essential cellular processes to suppress host defenses.

2.0 INTRODUCTION

Coronaviruses are a family of viruses with notably large single-stranded RNA genomes and broad species tropism among mammals¹. Recently, a new coronavirus, SARS-CoV-2, was discovered to cause the severe respiratory disease known as COVID-19. It is highly transmissible within human populations and its spread has resulted in a global pandemic with nearly a million deaths to date²⁻³. We do not fully understand the molecular basis of infection and pathogenesis of

this virus in human cells. Accordingly, there is an urgent need to understand these mechanisms to guide the development of therapeutics.

SARS-CoV-2 encodes 27 proteins with diverse functional roles in viral replication and packaging⁴⁻⁵. These include 4 structural proteins: the nucleocapsid (N, which binds the viral RNA), and the envelope (E), membrane (M), and spike (S) proteins, which are integral membrane proteins. In addition, there are 16 non-structural proteins (NSP1-16) which encode the RNA-directed RNA polymerase, helicase, and other components required for viral replication⁶. Finally, there are 7 accessory proteins (ORF3a-8) whose function in viral replication or packaging remain largely uncharacterized⁷⁻⁸.

As obligate intracellular parasites, viruses require host cell components to translate and transport their proteins and to assemble and secrete viral particles⁹. Upon viral infection, the mammalian innate immune system acts to rapidly detect and block viral infection at all stages of the viral life cycle¹⁰⁻¹². The primary form of intracellular viral surveillance engages the interferon pathway, which amplifies signals resulting from detection of intracellular viral components to induce a systemic type I interferon response upon infection¹³. Specifically, cells contain various RNA sensors (such as RIG-I and MDA5) that detect the presence of viral RNAs, promote nuclear translocation of the transcription factor IRF3 leading to transcription, translation, and secretion of interferon (e.g. IFN-a and IFN-b). Binding of interferon to cognate cell-surface receptors leads to transcription and translation of hundreds of antiviral genes. In order to successfully replicate, viruses employ a range of strategies to counter host antiviral responses¹⁴.

In addition to their essential roles in the viral life cycle, many viral proteins also antagonize core cellular functions in human cells to evade host immune responses. For example, human cytomegalovirus (HCMV) encodes proteins that inhibit class 1 Major Histocompatibility (MHC) display on the cell surface by retaining MHC proteins in the endoplasmic reticulum¹⁵, polioviruses encode proteins that degrade translation initiation factors (eIF4G) to prevent translation of 5'-capped host mRNAs¹⁶⁻¹⁷, and influenza A encodes a protein that modulates mRNA splicing to degrade the mRNA that encodes RIG-I¹⁸⁻¹⁹.

Suppression of the interferon response has recently emerged as a major clinical determinant of COVID-19 severity²⁰, with almost complete loss of secreted IFN characterizing the most severe cases²¹. The extent to which SARS-CoV-2 suppresses the interferon response is a key characteristic that distinguishes COVID-19 from SARS and MERS²². Several strategies have been proposed for how the related SARS- and MERS-causing viruses may hijack host cell machinery and evade immune detection, including repression of host mRNA transcription in the nucleus²³, degradation of host mRNA in the nucleus and cytoplasm²⁴⁻²⁵, and inhibition of host translation²⁶. Nonetheless, the extent to which SARS-CoV-2 uses these or other strategies, and how they may be executed at a molecular level remains unclear.

Understanding the interactions between viral proteins and components of human cells is essential for elucidating their pathogenic mechanisms and for development of effective therapeutics.

Because SARS-CoV-2 is an RNA virus and many of its encoded proteins are known to bind RNA²⁷, we reasoned that these viral proteins may interact with specific human mRNAs (critical

intermediates in protein production) or non-coding RNAs (critical structural components of diverse cellular machines) to promote viral propagation.

To date, most studies of RNA-protein interactions have relied on *in vitro* binding assays or immunoprecipitation experiments (e.g. RNA Immunoprecipitation, RIP, or Crosslinking Immunoprecipitation, CLIP)²⁸⁻³⁰. However in a classic experiment, Mili and Steitz showed that immunoprecipitation methods can identify RNA-protein interactions that do not occur *in vivo*, but rather form in solution after cell lysis, which may be problematic for screens with potential therapeutic implications³¹. To address this issue, we developed a new method called Covalent Linkage and Affinity Purification (CLAP) that enables purification of RNA-protein interactions using fully-denaturing conditions.

Using CLAP, we comprehensively define the interactions between each SARS-CoV-2 protein and human RNAs. We show that 10 viral proteins form highly specific interactions with mRNAs or ncRNAs, including those involved in progressive steps of host cell protein production. We show that NSP16 binds to the mRNA recognition domains of the U1 and U2 RNA components of the spliceosome and acts to suppress global mRNA splicing in SARS-CoV-2-infected human cells. We find that NSP1 binds to a precise region on the 18S ribosomal RNA that resides in the mRNA entry channel of the initiating 40S ribosome. This interaction leads to global inhibition of mRNA translation upon SARS-CoV-2 infection of human cells. Finally, we find that NSP8 and NSP9 bind to discrete regions on the 7SL RNA component of the Signal Recognition Particle (SRP) and interfere with protein trafficking to the cell membrane upon infection. We show that disruption of each of these essential cellular functions acts to suppress the type I interferon

response to viral infection. Together, our results uncover a multipronged strategy utilized by SARS-CoV-2 to antagonize essential cellular processes and robustly suppress host immune defenses.

3.0 RESULTS

3.1 CLIP IDENTIFIES MANY PRC2-RNA INTERACTIONS THAT FORM IN SOLUTION AFTER CELL LYSIS

CLIP methods are the gold-standard for defining *in vivo* RNA-protein interactions and have been successfully used to define the precise RNA binding sites of numerous RNA binding proteins³². Briefly, CLIP utilizes UV crosslinking to form covalent interactions in cells between directly interacting RNA and protein, followed by immunoprecipitation, subsequent separation through a denaturing SDS-PAGE gel, transfer to a nitrocellulose membrane, size extraction of the RNA-protein complex, and sequencing of the associated RNAs³²⁻³⁴. Given critical discrepancies between biochemical evidence (supporting specific chromatin-RNA interactions) and genetic evidence (demonstrating that these same interactions are often dispensable), we wanted to examine the specificity of CLIP and if it were subject to *in vitro* association artifacts prior to examining RNA-protein interactions in SARS-CoV-2³⁵⁻⁴⁴.

Based on the abundance of literature characterizing interactions between RNA and the chromatin regulatory complex Polycomb Repressive Complex 2 (PRC2), we specifically focused on PRC2 complex components EZH2, EED, and SUZ12 as test cases^{45-48,36}. To determine whether any observed PRC2-RNA interactions might represent associations that occur in solution, we designed an experiment modeled after the Mili and Steitz experiment, where authors expressed

an RNA binding protein (HuR) in one cell, its known RNA target (fos mRNA) in a separate cell, and measured RNA-protein associations that occur after mixing these distinct cells³¹. We extended this framework to enable quantitative measurements of in solution association by generating V5-tagged versions of all three PRC2 components (EED, EZH2, and SUZ12) and transfecting them into human HEK293T cells, followed by UV-crosslinking these cells to form covalent interactions between RNA and proteins that directly interact *in vivo* (+tag sample). We then mixed these human cells with UV-crosslinked mouse ES cells that do not express the V5-tagged protein (-tag sample). We performed CLIP in these mixed samples using an antibody against the V5-tagged PRC2 proteins and only analyzed sequencing reads that mapped uniquely and unambiguously to either the human or mouse genomes (**Figure 1A**). In this system, any detected mouse RNA must represent an RNA-protein interaction that occurred after cell lysis because the immunoprecipitated V5-tagged protein is not expressed in mouse cells.

We find that the majority of expressed RNAs are significantly enriched for all 3 PRC2 components in the +tag samples (~65%, $p < 10^{-6}$, **Figure 1B**). For example, we observe strong enrichment for all 3 PRC2 components across several lncRNAs that have previously been reported to bind to PRC2 including XIST, HOTAIR, KCNQ1ot1, and TUG1 (**Figure 1B**)^{49-52,48,39,43-44}. We also observed PRC2 binding in the -tag samples, suggesting CLIP methods are subject to *in vitro* association artifacts (**Figure 1B**).

In order to directly compare protein binding to the same RNAs in the +tag and -tag experiments, we transfected each of the 3 PRC2 components into a human cell line (+tag^[Human]) and mixed them with untransfected mouse cells. In parallel, we transfected these same proteins into a mouse

cell line and mixed them with untransfected human cells (-tag^[Human]). We performed CLIP in these two sets of mixed samples and directly compared PRC2 binding to the same human RNAs in the +tag^[Human] and -tag^[Human] experiments. We observed a strong global correlation between RNA regions that are highly enriched in the -tag^[Human] samples and those that are enriched in the +tag^[Human] samples (Pearson correlation = 0.43, **Figures 1C and 1E**). For example, when focusing on XIST, we observed that the 3 PRC2 components showed highly comparable profiles in both the +tag^[Human] and -tag^[Human] samples and display broad enrichment across the RNA with the strongest enrichment being over the A-repeat region as previously reported (**Figures 1D and 1E**)^{35,50}. Notably, these binding profiles are highly similar to those observed across XIST when performing CLIP using antibodies that recognize the endogenous PRC2 components (**Figure 1D**).

These results demonstrate that thousands of PRC2-RNA interactions can be detected by CLIP even when they do not occur *in vivo*.

3.2 COVALENT LINKAGE AFFINITY PURIFICATION REMOVES RNA-PROTEIN INTERACTIONS THAT DO NOT OCCUR *IN VIVO*

While the presence of strong PRC2-RNA binding in solution does not preclude the possibility that these PRC2 components also bind to RNA *in vivo*, it highlights the challenge in accurately determining which of the detected PRC2-RNA represent *bona fide* interactions that occur *in vivo*. Because CLIP is the current gold standard approach for studying RNA-protein interactions, there are currently no methods available that allow us to confidently assess *in vivo* interactions between RNA and PRC2 or other putative non-canonical RBPs.

We considered several possibilities that could lead to the detection of RNA-protein interactions that do not occur *in vivo*: (i) the captured protein may interact directly with RNA in solution and this RNA could even be crosslinked to a distinct protein, (ii) the captured protein may associate in solution with other proteins that are crosslinked to RNA, or (iii) other proteins that are crosslinked to RNA may still be retained after immunoprecipitation (**Figure 2A**). Any of these non-specific interactions that remain after immunoprecipitation would be detected because the protein purification (immunoprecipitation) and denaturation steps (gel electrophoresis) are decoupled in the CLIP procedure (**Figure 2B**). These sources of protein-specific background binding would be especially problematic when the captured protein does not actually bind to RNA, or binds to rare RNA targets *in vivo*, because non-specific RNA targets will be present in vast excess relative to *bona fide* targets.

To address these issues, we developed a new method called Covalent Linkage and Affinity Purification (CLAP) that enables purification of RNA-protein interactions using fully-denaturing conditions (**Figure 2B**). CLAP integrates an epitope tag into a protein of interest that enables covalent coupling of the tagged protein to a resin (e.g. HaloTag, SpyTag)⁵³⁻⁵⁴. Because the tagged protein is covalently coupled to the resin, rather than captured through an antibody, we can use a purification procedure that employs fully denaturing conditions – including high temperatures, high concentrations of denaturants and detergents, and chaotropic salts – that disrupt protein folding and RNA folding. This procedure directly couples protein purification and denaturation and accordingly the only RNA-protein interactions that should remain are those that represent the protein of interest where the RNA is covalently crosslinked *in vivo*.

To test whether CLAP reduces RNA-protein associations that occur in solution, we performed a mixing experiment where we expressed proteins fused to both Halo and V5 tags in human cells (+tag) and mixed them with untransfected mouse cells (-tag). We then split the lysate and performed CLIP and CLAP captures from the same mixture allowing us to directly compare the contribution of in solution associations in each experiment (**Figure 2B**). In the -tag samples, CLAP led to greatly reduced levels of background associated RNA for all 3 of the PRC2 proteins relative to V5 CLIP (**Figure 2C and 2D**). These results demonstrate that CLAP accurately removes RNA-protein interactions that do not occur *in vivo*.

3.3 CLAPS ACCURATELY MAPS RNA-PROTEIN INTERACTIONS FOR A WIDE-RANGE OF RBPS *IN VIVO*

To ensure that CLAP can identify *bona fide* RNA-protein interactions that occur *in vivo*, we performed CLAP on seven well-characterized RNA binding proteins (RBPs) that are known to interact with distinct classes of RNAs to mediate well-defined functions (e.g. mRNA splicing, translational regulation). These include (i) proteins that bind predominately within intronic regions and have high selectivity towards precise RNA sequence motifs (PTBP1, hnRNPC, hnRNPH1), (ii) proteins that bind to nascent pre-mRNA with broad affinity and a promiscuous binding profile (SAF-A and SRSF9), and (iii) proteins that bind to mature mRNAs in the cytoplasm (FMR1 and IGF2BP1)⁵⁵⁻⁶¹. In all cases, we observed RNA binding profiles by CLAP that are highly comparable to those observed in previously reported CLIP experiments (**Figures 3A-C**). For example, we find that hnRNPC binds specifically to intronic regions, SRSF9 binds broadly across nascent pre-mRNAs, and FMR1 binds predominately to spliced mRNAs with a

strong bias towards the 3' end (**Figures 3A-B**). Importantly, for all proteins, the observed RNA binding profiles are markedly different from the distribution of RNA present in the total input sample (**Figure 3A**).

Because these 7 well-characterized RBPs bind to many different RNAs, we wanted to ensure that the CLAP method would have the sensitivity to define RNA-protein interactions that might occur if a protein only binds to very few, highly specific RNA targets *in vivo*. To test this, we expressed a Halo-tagged GFP fused to a λ N bacteriophage RNA binding protein, which does not have any endogenous RNA targets in animal cells, but is known to interact with high affinity *in vivo* to RNAs containing a BoxB RNA aptamer⁶². We co-transfected these cells with an MBP-BoxB RNA. Using CLAP, we found that GFP- λ N was enriched exclusively over the co-expressed MBP-BoxB RNA, but not over any endogenously expressed RNAs (**Figure 3D**). CLAP reads were enriched specifically over the BoxB containing portion of the RNA and comparatively depleted over the MBP RNA (**Figure 3E**). Moreover, by exploiting the well-described tendency for reverse transcriptase to preferentially terminate at the site of a UV-crosslinked RNA-protein binding site, we find that the cDNA induced truncation sites correspond precisely to the location of the BoxB RNA sequences (**Figure 3E**)³³.

These results demonstrate that CLAP accurately identifies RNA-protein interactions that are crosslinked *in vivo* with high sensitivity and specificity across many different types of RNA-protein interactions.

3.4 COMPREHENSIVE MAPPING OF SARS-COV-2 PROTEIN BINDING TO HUMAN RNAS

We cloned all 27 of the known SARS-CoV-2 viral proteins into mammalian expression vectors containing an N-terminal HaloTag⁵³, expressed each in HEK293T cells, and exposed them to UV light to covalently crosslink proteins to their bound RNAs. We then lysed the cells and purified each viral protein using stringent, denaturing conditions to disrupt any non-covalent associations and capture those with a UV-mediated interaction (**Figure 4A**, Methods). As positive and negative controls, we purified a known human RNA binding protein (PTBP1) and a metabolic protein (GAPDH) (**Figures 5A-E**). We successfully purified 26 of the 27 viral proteins (**Figure 5A**; full-length Spike was not soluble when expressed). We found that 10 viral proteins (NSP1, NSP4, NSP8, NSP9, NSP12, NSP15, NSP16, ORF3b, N, and E protein) bind to specific host RNAs (p-value < 0.001, **Figure 4B**, **Table S1**), including 6 structural ncRNAs and 142 mRNAs (**Table S1**). These include mRNAs involved in protein translation (e.g. COPS5, EIF1, and RPS12), protein transport (ATP6V1G1, SLC25A6, and TOMM20), protein folding (HSPA5, HSPA6, and HSPA1B), transcriptional regulation (YY1, ID4, and IER5), and immune response (JUN, AEN, and RACK1) (FDR < 0.05, **Figures 4B and 5F**). Importantly, the observed interactions are highly specific for each viral protein, and each protein binds to a precise region within each RNA (**Figures 4C and 5F**).

Using these data, we identified several viral proteins that interact with structural ncRNA components of the spliceosome (U1 and U2 snRNA), the ribosome (18S and 28S rRNA), and the Signal Recognition Particle (7SL) (**Figure 4B**). Because these molecular machines are essential for three essential steps of protein production – mRNA splicing, translation, and protein

trafficking – we focused on their interactions with viral proteins to understand their functions and mechanisms in SARS-CoV-2 pathogenesis.

3.5 NSP16 BINDS TO THE PRE-MRNA RECOGNITION DOMAINS OF THE U1 AND U2 SNRNAS

After transcription in the nucleus, nascent pre-mRNAs are spliced to generate mature mRNAs which are translated into protein. Splicing is mediated by a complex of ncRNAs and proteins known as the spliceosome. Specifically, the U1 small nuclear RNA (snRNA) hybridizes to the 5' splice site at the exon-intron junction and the U2 snRNA hybridizes to the branchpoint site within the intron to initiate splicing of virtually all human mRNAs⁶⁵. We identified a highly specific interaction between the NSP16 viral protein and the U1 and U2 snRNAs (**Figure 4B**).

Because U1 and U2 are small RNAs (164 and 188 nucleotides, respectively), we noticed strong enrichment of NSP16-associated reads across the entire length of each. To more precisely define the binding sites, we exploited the well-described tendency of reverse transcriptase to preferentially terminate when it encounters a UV-crosslinked protein on RNA (**Figures 4A and 5D**)³³. We determined that NSP16 binds to the 5' splice site recognition sequence of U1 (**Figures 6A-B and 7A-B**) and the branch point recognition site of U2 (**Figures 6C-D and 7C-D**). These binding sites are highly specific to NSP16 relative to all of the other viral and human proteins (**Figures 4B, 7A, and 7C**). Consistent with its interaction with U1/U2, we observed that NSP16 localizes within the nucleus upon SARS-CoV-2 infection (**Figures 6E and 7E-F**) and when expressed in human cells (**Figure 7G**).

3.6 NSP16 DISRUPTS GLOBAL MRNA SPLICING UPON SARS-COV-2 INFECTION

Based on the locations of the NSP16 binding sites relative to the mRNA recognition domains of the U1/U2 spliceosomal components, we hypothesized that NSP16 might disrupt splicing of newly transcribed genes (**Figure 6F**). To test this, we co-expressed NSP16 in human cells along with a splicing reporter derived from IRF7 (an exon-intron-exon minigene) fused to GFP⁶⁶. In this system, if the reporter is spliced, then GFP is made; if not, translation is terminated (via a stop codon present within the first intron) and GFP is not produced (**Figure 8A**). We observed a >3-fold reduction in GFP levels in the presence of NSP16 compared to a control human protein (**Figures 8B and 9A**).

To explore whether NSP16 has a global impact on splicing of endogenous mRNAs, we measured the splicing ratio of each gene using nascent RNA sequencing. Specifically, we metabolically labeled nascent RNA by feeding cells for 20 minutes with 5-ethynyl uridine (5EU), purified and sequenced 5EU-labeled RNA, and quantified the proportion of unspliced fragments spanning the 3' splice site of each gene (**Figures 8C, 9B**). We observed a global increase in the fraction of unspliced genes in the presence of NSP16 compared to controls (**Figures 8D, 9C, 9D**).

Given that NSP16 is sufficient to suppress global mRNA splicing, we expect that its expression in SARS-CoV-2-infected cells would result in a global mRNA splicing deficit. To test this, we infected human lung epithelial cells (Calu3) with SARS-CoV-2 and measured splicing levels of newly transcribed mRNAs compared to a mock infected control. As expected, we observed a global increase in the fraction of unspliced transcripts upon SARS-CoV-2 infection, with ~90%

of measured genes showing increased intron retention (**Figures 8E, 9E**). Together these results indicate that NSP16 binds to the splice site and branch point sites of U1/U2 to suppress global mRNA splicing in SARS-CoV-2 infected cells (**Figure 8F**). Although NSP16 is known to act as an enzyme that deposits 2'-O-methyl modifications on viral RNAs⁶⁷, our results demonstrate that it also acts as a host virulence factor. Global disruption of mRNA splicing may act to decrease host protein and mRNA levels by triggering nonsense-mediated decay of improperly spliced mRNAs⁶⁸. Consistent with this, we observed a strong global decrease in steady-state mRNA levels (relative to ncRNA levels) upon SARS-CoV-2 infection (**Figure 9F**).

3.7 INHIBITION OF MRNA SPLICING SUPPRESSES HOST INTERFERON RESPONSE TO VIRAL INFECTION

Because many of the key genes stimulated by interferon (IFN) are spliced, we reasoned that mRNA splicing would be critical for a robust IFN response. To test this, we utilized a reporter line engineered to express alkaline phosphatase upon IFN signaling (mimicking an antiviral response gene). This IFN Stimulated Gene (ISG) reporter line can be stimulated using IFN- β and assayed for reporter induction. We observed strong repression of this IFN responsive gene upon expression of NSP16 (**Figure 8G**) and upon addition of a small molecule that interferes with spliceosomal assembly (**Figure 9G**). These results demonstrate that one outcome of NSP16-mediated inhibition of mRNA splicing is to reduce the host cells' innate immune response to viral recognition. Consistent with such a role, we observed an increase in intron retention within multiple IFN-responsive genes (such as ISG15 and RIG-I) upon SARS-CoV-2 infection (**Figures 8H and 9H-I**).

3.8 NSP1 BINDS TO 18S RIBOSOMAL RNA IN THE MRNA ENTRY CHANNEL OF THE 40S SUBUNIT

Once exported to the cytoplasm, spliced mRNA is translated into protein on the ribosome. Initiation of translation begins with recognition of the 5' cap by the small 40S subunit (which scans the mRNA to find the first start codon). We observed that NSP1 binds exclusively to the 18S ribosomal RNA (**Figures 4B and 10A**) – the structural RNA component of the 40S ribosomal subunit.

Several roles for NSP1 have been reported in SARS-CoV and MERS-CoV including roles in viral replication, translational inhibition, transcriptional inhibition, mRNA degradation, and cell cycle arrest^{69-70, 24-25}. One of the reported roles for NSP1 in SARS-CoV is that it can associate with the 40S ribosome to inhibit host mRNA translation^{24,71}, yet it remains unknown whether this association is due to interaction with the ribosomal RNA, protein components of the ribosome, or other auxiliary ribosomal factors. Accordingly, the mechanisms by which NSP1 acts to suppress protein production remain elusive.

We mapped the location of NSP1 binding to a 37 nucleotide region corresponding to Helix 18 (**Figure 11A**), adjacent to the mRNA entry channel (**Figure 11B**)⁷². The interaction would position NSP1 to disrupt 40S mRNA scanning and prevent translation initiation (**Figure 11B**), and disrupt tRNA recruitment to the 80S ribosome and block protein production (**Figure 10B**). Interestingly, the NSP1 binding site includes the highly conserved G626 nucleotide which monitors the minor groove of the codon-anticodon helix for tRNA binding fidelity⁷³. We noticed that the C-terminal region of NSP1 has similar structural regions to SERBP1⁷⁴ and Stm1⁷⁵, two

known ribosome inhibitors that bind within the mRNA entry channel to preclude mRNA access (**Figure 10C**). Consistent with this, a recent cryo-EM structure confirms that NSP1 binds to these same nucleotides of 18S within the mRNA entry channel⁷⁶.

3.9 NSP1 SUPPRESSES GLOBAL TRANSLATION OF HOST MRNAS UPON SARS-COV-2 INFECTION

Given the location of NSP1 binding on the 40S ribosome, we hypothesized that it could suppress global initiation of mRNA translation. To test this, we performed in vitro translation assays of a GFP reporter in HeLa cell lysates and found that addition of NSP1 led to potent inhibition of translation (**Figure 10D**). We observed a similar NSP1-mediated translational repression when we co-expressed NSP1 and a GFP reporter gene in HEK293T cells (**Figures 11C-D**). In contrast, we did not observe this inhibition when we expressed other SARS-CoV-2 proteins (NSP8, NSP9, M) or human proteins (GAPDH) (**Figure 11D**).

To determine if NSP1 leads to translational inhibition of endogenous proteins in human cells, we used a technique called Surface Sensing of Translation (SUnSET) to measure global protein production levels⁷⁷. In this assay, translational activity is measured by the level of puromycin incorporation into elongating polypeptides (**Figure 10E**). We observed a strong reduction in the level of global puromycin integration in cells expressing NSP1 compared to cells expressing GFP (**Figures 10F-G**).

Because NSP1 expression is sufficient to suppress global mRNA translation in human cells, we hypothesized that SARS-CoV-2 infection would also suppress global translation. To test this, we

infected a human lung epithelial (Calu3) or monkey kidney (Vero) cell line with SARS-CoV-2 and measured nascent protein synthesis levels using SUnSET. We observed a strong reduction of global puromycin integration upon SARS-CoV-2 infection in both cell types (**Figures 11E-F, 10H-I**).

To explore whether NSP1 binding to 18S rRNA is critical for translational repression, we generated a mutant NSP1 in which two positively charged amino acids (K164 and H165) in the C-terminal domain were replaced with alanine residues (**Figure 10C**)⁷⁸. We observed a complete loss of in vivo contacts with 18S (**Figure 11G**); because this mutant disrupts ribosome contact, we refer to it as NSP1 Delta RC. We co-expressed GFP and NSP1 Delta RC in HEK293T cells and found that the mutant fails to inhibit translation (**Figures 11H and 10J**). In contrast, mutations to the positively charged amino acids at positions 124/125 do not impact 18S binding (**Figure 11G**) or the ability to inhibit translation (**Figure 11H**).

Together, these results demonstrate that NSP1 binds within the mRNA entry channel of the ribosome and that this interaction is required for translational inhibition of host mRNAs upon SARS-CoV-2 infection.

3.10 NSP1-MEDIATED TRANSLATIONAL INHIBITION SUPPRESSES HOST INTERFERON RESPONSE

We explored whether NSP1 binding to 18S rRNA suppresses the ability of cells to respond to IFN- β stimulation upon viral infection. We transfected ISG reporter cells with NSP1, stimulated with IFN- β , and observed robust repression of the IFN responsive gene (>6-fold, **Figure 11I**). To confirm that this NSP1-mediated repression occurs in human cells upon activation of double

stranded RNA (dsRNA)-sensing pathways typically triggered by viral infection, we treated a human lung epithelial cell line (A549) with poly(I:C), a molecule that is structurally similar to dsRNA and known to induce an antiviral innate immune response(**Figure 10K**)⁷⁹⁻⁸⁰. We observed a marked downregulation of IFN- β protein and endogenous IFN- β responsive mRNAs in the presence of NSP1, but not in the presence of NSP1 Delta RC (**Figures 10L-M**). These results demonstrate that NSP1, through its interaction with 18S rRNA, suppresses the innate immune response to viral recognition (**Figure 11J**).

3.11 THE VIRAL 5' LEADER PROTECTS MRNA FROM NSP1-MEDIATED TRANSLATIONAL INHIBITION

Because NSP1 blocking the mRNA entry channel would impact both host and viral mRNA translation, we explored how translation of viral mRNAs is protected from NSP1-mediated translational inhibition. Many viruses contain 5' untranslated regions that regulate viral gene expression and translation⁸¹; all SARS-CoV-2 encoded subgenomic RNAs contain a common 5' leader sequence that is added during negative strand synthesis⁸². We explored whether the leader sequence protects viral mRNAs from translational inhibition by fusing the viral leader sequence to the 5' end of GFP or mCherry reporter genes (**Figure 12A**). We found that NSP1 fails to suppress translation of these leader-containing mRNAs (**Figures 13A-B, 12B**). We dissected the leader sequence and found that the first stem loop (SL1) is sufficient to prevent translational suppression upon NSP1 expression (**Figure 13C**) or SARS-CoV-2 infection (**Figure 13D**).

We considered three models for how the leader could protect viral mRNAs: (i) it could compete with the ribosome for NSP1 binding, (ii) it could directly recruit free ribosomes or (iii) NSP1

could bind to the leader independently of its ribosome interaction to allosterically modulate the NSP1-ribosome interaction. We reasoned that if the leader competes for NSP1 binding or directly recruits free ribosomes, then the presence of SL1 should be sufficient for protection, regardless of its precise position in the 5' UTR. In contrast, if the leader allosterically modulates ribosome binding then the spacing between the 5' cap (which is bound to NSP1-40S) and SL1 would be critical for protection. To distinguish between these models, we swapped the location of SL1 and SL2 in the 5' leader or inserted 5 nucleotides between the 5' cap and SL1 (Figure S5C) and found that both mutants ablate protection (**Figures 13E and 12D**).

These results indicate that an mRNA requires the 5' leader to be precisely positioned relative to the NSP1-bound 40S ribosome to enable translational initiation (**Figure 13F**). While many aspects of this allosteric model remain to be explored, it would explain how leader-mediated protection can occur on an mRNA only when present in cis. Moreover, this model suggests that NSP1 might also act to further increase viral mRNA translation by actively recruiting the ribosome to its own mRNAs. Consistent with this, we observe a consistent ~20% increase in translation of leader-containing reporter levels upon viral infection (**Figure 13D**) or expression of NSP1 (**Figure 12E**).

3.12 NSP8 AND NSP9 BIND TO THE 7SL RNA COMPONENT OF THE SIGNAL RECOGNITION PARTICLE

Upon engaging the start codon in an mRNA, the 60S subunit of the ribosome is recruited to form the 80S ribosome which translates mRNA. The Signal Recognition Particle (SRP) is a universally conserved complex that binds to the 80S ribosome and acts to co-translationally scan

the nascent peptide to identify hydrophobic signal peptides present in integral membrane proteins and proteins secreted from the plasma membrane⁸³. When these are identified, SRP triggers ribosome translocation to the endoplasmic reticulum (ER) to ensure proper folding and trafficking of these proteins to the cell membrane⁸³.

We identified two viral proteins – NSP8 and NSP9 – that bind at distinct and highly specific regions within the S-domain of the 7SL RNA scaffold of SRP (**Figures 14A and 15A**). NSP8 interacts with 7SL in the region bound by SRP54 (the protein responsible for signal peptide recognition, SRP-receptor binding, and ribosome translocation) (**Figure 14B**)⁸³⁻⁸⁴. NSP9 binds to 7SL in the region that is bound by the SRP19 protein (**Figure 14B**), which is required for proper folding and assembly of SRP (including proper loading of SRP54)⁸³.

Because SRP scans nascent peptides co-translationally, we were intrigued to find that NSP8 also forms a highly specific interaction with 28S rRNA (the structural component of the 60S subunit) (**Figures 14C and 15B**). The binding site on 28S rRNA corresponds to the largest human-specific expansion segment within the ribosome, referred to as ES27⁸⁵. ES27 is highly dynamic, and thus has not been resolved in most ribosome structures⁸⁶. However, when engaged by specific factors, ES27 can become ordered, and was recently shown to be capable of interacting with the ribosome exit tunnel, adjacent to the 60S binding site of SRP (**Figures 14D and 15C**)⁸⁷.

Together, these observations suggest that NSP8 and NSP9 bind to the co-translational SRP complex. Consistent with this, we find that NSP8 and NSP9 localize broadly throughout the cytoplasm when expressed in human cells (**Figure 15D**) or upon SARS-CoV-2 infection (Figure

15E-F).

3.13 NSP8 AND NSP9 SUPPRESS PROTEIN INTEGRATION INTO THE CELL MEMBRANE

Because NSP8 and NSP9 binding on 7SL are positioned to disrupt SRP function, we hypothesized that they may alter translocation of secreted and integral membrane proteins (**Figure 16A**). To test this, we expressed an SRP-dependent membrane protein (Nerve Growth Factor Receptor, NGFR⁸⁸) fused via an Internal Ribosome Entry Site (IRES) to a non-membrane GFP (**Figure 16F**). In this system, if a perturbation specifically affects membrane protein levels we expect to see a decrease in the ratio of membrane to non-membrane protein levels. To ensure that the NGFR reporter accurately reports on SRP function, we treated HEK293T cells with siRNAs against SRP54 or SRP19 and found that both lead to a dramatic reduction of the NGFR membrane protein relative to the non-membrane GFP protein (**Figure 16B**). Similarly, we found that expression of NSP8 and NSP9 (alone or together) lead to a striking reduction in expression of NGFR relative to GFP (**Figure 17A**). Expression of control proteins did not specifically impact NGFR levels (**Figures 17A and 16B**).

To determine if there is a global effect on membrane protein levels, we utilized the SUnSET method to measure puromycin levels in membrane proteins using flow cytometry (see Methods). We confirmed that disruption of SRP leads to a global reduction in puromycin levels in the cell membrane (**Figure 16C**). We observed a comparable global reduction of puromycin-labeled membrane proteins upon expression of NSP8 or NSP9 individually or together, but not with control proteins (**Figure 17B-C**).

3.14 SARS-COV-2 INFECTION SUPPRESSES PROTEIN INTEGRATION INTO THE CELL MEMBRANE

Because NSP8 and NSP9 are each sufficient to suppress protein integration into the cell membrane, we anticipate that SARS-CoV-2 infection would lead to similar suppression. However, determining whether SARS-CoV-2 infection specifically impacts membrane protein expression is confounded by the fact that NSP1 inhibits translation of membrane and nonmembrane proteins upon infection.

To address this, we co-expressed a membrane protein reporter (NGFR) containing the 5' viral leader along with a non-membrane GFP reporter containing the viral leader. Upon viral infection, we observed a strong reduction of membrane protein levels (**Figure 17C**), but no reduction in non-membrane GFP levels (**Figure 13D**). To ensure that these effects are specific to SARS-CoV-2 infected cells, we separated individual cells within the infected population into those expressing the viral Spike protein (S+) and those not expressing the protein (S-). We found that the shift in membrane protein levels only occurs in S+ cells (**Figure 17D**), while the S- population resembled the mock infected samples (**Figure 17C**). We observed a strong relationship between the level of Spike protein – likely reflecting the amount of viral replication within each cell – and the level of membrane protein suppression (**Figure 17C**). We observed this membrane protein-specific decrease upon infection of human lung epithelial (Calu3, **Figure 16D**) and monkey kidney (Vero, **Figures 17C-D**) cell lines.

Together, these results demonstrate that NSP8 and NSP9 bind to 7SL to disrupt SRP function and suppress membrane protein trafficking in SARS-CoV-2 infected cells. Although NSP8 and

NSP9 are thought to be components of the viral replication machinery⁸⁹, our results indicate that they play an additional role as host virulence factors. Because viral membrane proteins also require trafficking to the ER, viral disruption of SRP might negatively impact viral propagation, unless viral proteins are trafficked in an SRP-independent manner (**Figure 16E**) or if NSP8/9 selectively impacts host (but not viral) proteins.

3.15 VIRAL DISRUPTION OF PROTEIN TRAFFICKING SUPPRESSES INTERFERON RESPONSE

Next we explored how disruption of SRP might be advantageous for viral propagation. Because secretion of IFN and other cytokines is dependent on the SRP complex for secretion (**Figure 16F**), a central component of the IFN response is dependent on SRP. Accordingly, we hypothesized that NSP8/9-mediated viral suppression of SRP would act to suppress the IFN response upon infection. To test this, we co-expressed NSP8 and NSP9 and observed a significant reduction in the IFN response relative to a control protein (**Figure 16G**).

Together, these results suggest that SARS-CoV-2 mediated suppression of SRP-dependent protein secretion enables suppression of host immune defenses (**Figure 17E**). Interestingly, many proteins involved in anti-viral immunity – including most cytokines and class I major histocompatibility complex – are membrane-anchored or secreted, and are known to use the SRP pathway for transport (**Figure 16F**), suggesting that there may be other effects of SRP pathway inhibition on SARS-CoV-2 pathogenesis⁹⁰.

4.0 DISCUSSION

We identified several novel pathogenic functions of SARS-CoV-2 in human cells – including

global inhibition of host mRNA splicing, protein translation, and membrane protein trafficking – and described the molecular mechanisms by which the virus acts to disrupt these essential cell processes. Interestingly, all of the viral proteins involved (NSP1, NSP8, NSP9, and NSP16) are produced in the first stage of the viral life cycle, prior to generation of double stranded RNA (dsRNA) products during viral genome replication. Because dsRNA is detected by host immune sensors and triggers the type I interferon response, disruption of these cellular processes would allow the virus to replicate its genome while minimizing the host innate immune response. Disruption of these three non-overlapping steps of protein production may represent a multipronged mechanism that synergistically acts to suppress the host antiviral response (**Figure 17F**). Specifically, the IFN response is usually boosted >1,000-fold upon viral detection (through amplification and feedback, **Figure 11K**), yet each individual mechanism impacts IFN levels on the order of ~5-10-fold. Accordingly, if each independent mechanism impacts IFN levels moderately, the three together may be able to achieve dramatic suppression of IFN ($10^3=1,000$ -fold). This multi-pronged mechanism may explain the molecular basis for the potent suppression of IFN observed in severe COVID-19 patients.

Interferon is emerging not only as a determinant of disease severity, but also a potential treatment option⁹¹. As such, our work identifies several therapeutic opportunities for boosting IFN levels upon SARS-CoV-2 infection. For example, disrupting the interaction between NSP1 and 18S rRNA could allow cells to detect and respond to viral infection. Because many small-molecule drugs target ribosomal RNAs⁹², it may be possible to develop drugs to block the NSP1-18S and other interactions. Additionally, disrupting the 5' viral leader may be a potent antiviral strategy since it is critical for translation of all viral proteins. Because SL1 is a structured RNA,

it may be possible to design small molecules that specifically bind this structure to suppress viral protein production⁹³.

Viral suppression of these cellular functions is not exclusive to the IFN response and will also impact other spliced, translated, secreted, and membrane proteins. Many proteins involved in anti-viral immunity are spliced and/or membrane-anchored or secreted. For example, class I major histocompatibility complex (MHC), which is critical for antigen presentation to CD8 T cells at the cell surface of infected cells⁹⁴. By antagonizing membrane trafficking, SARS-CoV-2 may prevent viral antigens from being presented on MHC and allow infected cells to escape T-cell recognition and clearance. In this way, interference with these essential cellular processes might further aid SARS-CoV-2 in evading the host immune response.

More generally, we expect that insights gained from the SARS-CoV-2 protein-RNA binding maps will be critical for exploring additional viral mechanisms. Specifically, we identified many other interactions, including highly specific interactions with mRNAs. For example, NSP12 binds to the JUN mRNA (**Figure 5E**) which encodes the critical immune transcription factor c-Jun which is activated in response to multiple cytokines and immune signaling pathways⁹⁵. We also identified an interaction between NSP9 and the start codon of the mRNA that encodes COPS5 (**Figure 4C**), the enzymatic subunit of the COP9 Signalosome complex which regulates protein homeostasis⁹⁶, suggesting that it might disrupt its translation. Interestingly, COPS5 (also known as JAB1) is known to bind and stabilize c-Jun protein levels⁹⁷ and several viruses are known to disrupt this protein⁹⁸⁻¹⁰⁰. While it remains unknown what, if any, role these interactions

play in virally infected cells, the specificity suggests that they may provide a selective advantage for viral propagation.

Together, our results demonstrate that global mapping of RNA binding by viral proteins could enable rapid characterization of mechanisms for newly emerging pathogenic RNA viruses.

4.1 LIMITATIONS OF STUDY

We note several limitations of our current study that will need to be explored in future work. (i) Our mapping experiments were performed in uninfected human cells expressing tagged viral proteins. Accordingly, it remains possible that our maps may not fully capture all of the interactions that occur when human cells are infected, such as interactions that occur with viral induced RNAs, in specific viral compartments, or that require multiple viral proteins. (ii) While we characterized the functional and mechanistic roles of several viral proteins and structural ncRNAs, we did not explore what roles viral protein interactions with mRNAs might play. (iii) How the virus disrupts fundamental cellular processes while still maintaining its own production is still largely undefined. While we showed that the 5' leader is sufficient to relieve translational inhibition by NSP1, we still do not fully understand how this protection occurs and specifically how NSP1 might interact with the viral leader or allosterically modulate ribosome binding. Similarly, viral membrane proteins are dependent on trafficking to the ER and how NSP8/9 might selectively impact ER translocation of host – but not viral – proteins remains to be explored. (iv) While we showed that viral disruption of these essential cellular functions can suppress IFN, what other roles host cell shutdown might play in viral pathogenesis and in suppressing other aspects of anti-viral immunity, including possible roles in adaptive immune

responses, have not been explored.

5.0 ACKNOWLEDGEMENTS

We thank Fritz Roth for clones; Marko Jovanovic, Bil Clemons, Shu-ou Shan, and Jamie Wangen for guidance and discussion; Joanna Jachowicz for analysis; Shawna Hiley for editing; Inna-Marie Strazhnik for figures; Scott Tighe, Pheobe Laaguiby, Roxana del Rio-Guerra, Joyce Oetjen, Philip Eisenhauer, Arvind H. Patel and Arthur Wickenhagen for technical help. This work was supported by NIH F30HL136080 (A.K.B); P20GM125498, P30GM118228-04 (E.A.B); NIH GM7616-40, F30CA247447 (P.B.); UM1HL120877, U54GM115516 (D.M.); R41AI132047, U01AI1141997 (J.W.B); U01 DA040612, U01 HL130007 (M.G.) and USC MD/PhD Program (A.K.B); UCLA-Caltech MSTP (P.B.); American Cancer Society PF-17-240-01 (N.O.); Wellcome Trust 201366/Z/16/Z (S.J.R.); Office of the Vice President for Research at UVM (J.W.B); Heritage Medical Research Institute (M.G. and R.V.); NYSCF, CZI, and Caltech (M.G.). M.G. is a NYSCF-Robertson Investigator.

6.0 MATERIALS AND METHODS

Cell lines and culture conditions

Cell lines used in this study. We used the following cell lines in this study: (i) HEK293T, a female human embryonic kidney cell line obtained from ATCC. (ii) HEK-Blue™ ISG, Interferon regulatory factor (IRF)-inducible Secreted Alkaline Phosphatase (SEAP) reporter HEK293 cells of female origin (Invivogen). (iii) A549, a male human lung epithelial cell line obtained from ATCC. (iii) Calu3, a male human lung epithelial cell line obtained from ATCC, (iv) Vero E6, a female African green monkey kidney cell line, kindly provided by J.L. Whitton and Michele Bouloy.

Cell culture conditions. A549s, HEK293T cells and derivatives were cultured in complete media consisting of DMEM (GIBCO, Thermo Fisher Scientific) supplemented with 10% FBS (Seradigm Premium Grade HI FBS, VWR), 1X penicillin-streptomycin (GIBCO, Thermo Fisher Scientific), 1X MEM non-essential amino acids (GIBCO, Thermo Fisher Scientific), 1 mM sodium pyruvate (GIBCO, Thermo Fisher Scientific) and maintained at 37C under 5% CO₂. For maintenance, 800,000 cells were seeded into 10 mL of complete media every 3-4 days in 10 cm dishes. Vero E6 cells were maintained in complete DMEM (Thermo Fisher Scientific, 11965–092) containing 10% fetal bovine serum (FBS) (Thermo Fisher Scientific, 16140–071), 1% HEPES Buffer Solution (Thermo Fisher Scientific, 15630–130), and 1% penicillin-streptomycin (Thermo Fisher Scientific, 15140–122). Calu3 cells were maintained in Eagles’s Minimal Essential Medium (ATCC) containing 10% FBS and 1% penicillin-streptomycin purchased from Thermo Fisher Scientific. All cell lines were maintained at 37C under 5% CO₂. Cells were grown in a humidified incubator at 37C with 5% CO₂.

SARS-CoV-2 Viral Infection conditions

All experiments using infectious SARS-CoV-2 conducted at the UVM BSL-3 facility were performed under an approved Institutional Biosafety protocol. SARS-CoV-2 strain 2019-nCoV/USA_USA WA1/2020 (WA1) was generously provided by Kenneth Plante and the World Reference Center for Emerging Viruses and Arboviruses (WRCEVA) at the University of Texas Medical Branch and propagated in Vero E6 cells. Viral infections were performed at the indicated multiplicity of infection in a low volume of normal cellular maintenance media containing 2% FBS for one hour at 37.C, inoculum was removed and then overlaid in the respective cellular maintenance media containing 10% FBS for the indicated time periods.

Experiments performed to visualize the location of viral NSP proteins (and associated antibody validation) were performed in a Containment Level 3 facility at the MRC-University of Glasgow Centre for Virus Research using SARS-CoV-2 strain England-02 (from Public Health England [now called National Institute for Health Protection], GISAID: EPI_ISL_407073) using a MOI of 0.1 or 1 (as indicated).

Crosslinking and Immunoprecipitation (CLIP)

Purifications. Cells were lysed in 1 ml lysis buffer (50 mM Tris pH 7.5, 100mM NaCl, 1% NP-40, 0.5% Sodium Deoxycholate, 1x Promega protease inhibitor cocktail). RNA was digested with Ambion RNase I (1:3000 dilution) to achieve a size range of 100-500 nucleotides in length. Lysate preparations were precleared by mixing with Protein G beads for 30 min at 4C. Target proteins were immunoprecipitated from 5 million cells with 10µg of antibody and 75µl of Protein G beads in 100µL lysis buffer. The antibodies were pre-coupled to the beads for 1 hour at room temperature with mixing and unbound antibodies removed with 3 washes of lysis buffer. The precleared lysate was added to the Protein G coupled antibody beads overnight at 4C. After the immunoprecipitation, the beads were washed four times with High salt wash buffer (50 mM TrisHCl pH 7.4, 1 M NaCl, 1 mM EDTA, 1% NP-40, 0.1% SDS, 0.5% sodium deoxycholate) and four times with Wash buffer (20 mM Tris-HCl pH 7.4, 10 mM MgCl₂, 0.2% Tween-20). RNA and protein were eluted by incubating at 50C in NLS elution buffer (20 mM Tris-HCl pH 7.5, 10 mM EDTA, 2% N-lauroylsacrosine, 2.5 mM TCEP) supplemented with 100 mM DTT for 20 minutes. Samples were then run through an SDS-PAGE gel and transferred to a nitrocellulose membrane using the iBLOT transfer system, and a region 70 kDa above the molecular size of the protein of interest was isolated and treated with Proteinase K (NEB)

followed by buffer exchange and concentration with RNA Clean & Concentrator™-5 (Zymo). RNA sequencing libraries from these samples were constructed as previously described^{102,119-121}. We used the following antibodies: V5 antibody (Bethyl, A190-120A), EZH2 (Active Motif, 39933), SUZ12 (Active Motif, 39357), PTBP1 (Abcam, ab5642), and SAFA/hnRNPU (Santa Cruz, SC-32315).

CLIP Library Construction. CLIP samples were treated as previously described^{102,120}. Briefly, after immunoprecipitation and wash steps, the RNA was dephosphorylated (Fast AP) and cyclic phosphates removed (T4 PNK) and then ligated on Protein G beads with an RNA adapter containing a RT primer binding site. The ligated protein-bound RNA was then run through a denaturing PAGE gel and transferred to nitrocellulose membrane (as described above). The RNA was then extracted by proteinase K and purified using a spin column (Zymo). The RNA was reverse transcribed into single stranded cDNA. After RT, the RNA was degraded and a second adapter was ligated to the single stranded DNA. PCR amplification is achieved using primers that target the 3' and 5' adapters. Input total RNA libraries were constructed using the same steps as outlined above except that the dephosphorylation, cyclic phosphate removal, and ligation were performed in solution rather than on Protein-G beads.

Read processing and Alignment. Sequencing reads were trimmed to remove adaptor sequences and any bases containing a quality scores <10 using Trimmomatic¹²². We filtered out all read-pairs where either read was trimmed to <25 nucleotides. We excluded PCR duplicates using the FastUniq tool¹²³. The remaining reads were then aligned to Ribosomal RNAs (rRNAs) using the Tagdust program¹²⁴ with a database of 18S, 28S, 45S, 5S, 5.8S sequences. TagDust was chosen

because it allowed more permissive alignments to rRNA reads that contained mismatches and indels due to RT errors induced by rRNA post-transcriptional modifications. The remaining reads were then aligned to a combined genome reference containing the mouse (mm9) and human (hg19) genomes using STAR aligner¹²⁵. Only reads that mapped uniquely in the genome and unambiguously to the human or mouse genomes were kept for further analysis.

Gene Window Enrichment calculations. All human (hg19) and mouse (mm9) annotated genes (RefSeq, downloaded from UCSC Hg19 and MM9, respectively) were used as a reference set except for the genes encoding the 6 transfected proteins. In addition, we added all human lncRNAs as annotated by Genecode (release 26). For each gene, we enumerated 100 nucleotide windows that span across the exons and introns of each gene. For each window, we calculated the enrichment by computing the number of reads overlapping the window in the protein elution sample divided by the number of reads in the input sample. Because all windows overlapping a gene should have the same expression level in the input sample, we estimated the number of reads in the input as the maximum of either (i) the number of reads over the window or (ii) the median read count over all windows within the gene. This approach provides a conservative estimation of enrichment because it prevents windows from being scored as enriched if the input values over a given window are artificially low, while at the same time accounting for any non-random issues that lead to increases in read counts over a given window (i.e. alignment artifacts leading to non-random assignment or pileups).

We normalized this observed ratio by the expected number of reads in a window defined as the total number of reads in the protein elution sample divided by the number of windows covered in

the sample. This normalization estimates the expected read coverage for each window and accounts for the redistribution of reads that occur because of the fixed sequencing depth used and possible “drop out” of specific RNA regions during the enrichment process. The total number of reads in the protein elution or input samples was calculated by adding the total number of human-specific reads, mouse-specific reads, and ribosomal RNA reads. Nominal p -values were calculated for each window using a binomial test where k (number of successes) is defined as the number of reads in the protein elution samples within the window, N (number of trials) is the sum of the number of reads in the protein elution and input samples, and p (probability of success) is the expected number of reads per window in the elution divided by the sum of the expected number of reads per window in elution and input samples. (The expected number of reads is defined as the total number of reads divided by the number of windows). For plotting and reporting purposes, we considered all regions with a nominal binomial p -value $< 10^{-6}$ as significant. However, the overall results reported are robust to the precise p -value cutoff used.

Plotting and visualization. IGV plots for specific RNAs were generated by computing enrichments (as described above) across 100 nucleotide windows and the enrichment value was plotted at the midpoint of each window.

Generation of SARS-CoV-2 RNA binding maps

Cloning of expression constructs. SARS-CoV-2 protein constructs (with the exception of

Nsp11) were a gift from Fritz Roth (see Table S3 for Addgene information)¹⁰¹ and were LR-cloned (Invitrogen Gateway Cloning, Thermo Fisher Scientific) into mammalian expression destination vector pCAG-Halo-TEV-DEST-V5-IRES-puroR. Note that following LR cloning, proteins were not V5-tagged because all entry clones contained stop codons. For NSP11, an entry clone was generated by BP cloning (Invitrogen Gateway Cloning, Thermo Fisher Scientific) a PCR amplicon (primers: ggGGACAAGTTTGTACAAAAAAGCAGGCTTTtcagctgatgcacaatcgttttaaacgg and gGGGACCACTTTGTACAAGAAAGCTGGGTTtacaccgcaaaccgtttaaaaacgattg; template: pGBW-m4133457 a gift from Ginkgo Bioworks) into pDONR221.

Expression and lysis. For each viral protein capture, we transfected 10 µg of these expression vectors into HEK293T cells grown on a 15cm dish using BioT transfection reagent (Bioland) according to manufacturer's recommendations. 24-48 hours post-transfection, cells were washed once with PBS and then crosslinked on ice using 0.25 J cm⁻² (UV2.5k) of UV at 254 nm in a Spectrolinker UV Crosslinker. Cells were then scraped from culture dishes, washed once with PBS, pelleted by centrifugation at 1,000 X g for 5 min, and flash-frozen in liquid nitrogen for storage at -80C. We lysed batches of 5 million cells by completely resuspending frozen cell pellets in 1 mL of ice cold lysis buffer (50 mM Hepes, pH 7.4, 100 mM NaCl, 1% NP-40, 0.1% SDS, 0.5% Sodium Deoxycholate) supplemented with 1X Protease Inhibitor Cocktail (Promega), 200 U of Ribolock (Thermo Fisher Scientific), 20 U Turbo DNase (Ambion), and 1X Manganese/Calcium Mix (0.5mM CaCl₂, 2.5 mM MnCl₂). Samples were incubated on ice for 10 minutes to allow lysis to proceed. The lysates were then incubated at 37C for 10 minutes at 700 rpm shaking on a Thermomixer (Eppendorf). Lysates were cleared by centrifugation at 15,000 X

g for 2 minutes. The supernatant was collected and kept on ice until bound to the HaloLink Resin (Promega). Of the 1mL lysis volume, 50uL was set aside for input, 20uL used for protein expression confirmation, and the rest for capture on HaloLink Resin as described below.

Protein capture. We used 200 μ L of 25% HaloLink Resin slurry (50 μ L of HaloLink Resin total) per 5 million cells. Resin was washed three times with 2 mL of 1X PBS-T (1x PBS + 0.1% Triton X-100) and incubated in 1X Blocking Buffer (50 mM HEPES, pH 7.4, 100 μ g/mL BSA) for 20 minutes at room temperature with continuous rotation. After the incubation, resin was washed three times with 1X PBS-T. The cleared lysate was mixed with 50 μ l of HaloLink Resin and incubated at 4C for 3-16 hrs with continuous rotation. The captured protein bound to resin was washed three times with lysis buffer at room temperature and then washed three times at 90C for 3 minutes while shaking on a Thermomixer at 1200 rpm with each of the following buffers: 1X NLS buffer (1xPBS, 2% NLS, 10 mM EDTA), High Salt Buffer (50 mM HEPES, pH 7.4, 0.1% NP-40, 1M NaCl), 8M Urea Buffer (50 mM HEPES, pH 7.5, 0.1% NP-40, 8 M Urea), Tween buffer (50 mM HEPES, pH 7.4, 0.1% Tween 20) and TEV buffer (50 mM HEPES, pH 7.4, 1 mM EDTA, 0.1% NP-40). The extended incubation of the bound RNA with the wash buffers leads to chemical fragmentation of the RNA yielding sizes that are suitable for RNA library preparation and binding site resolution. Between each wash, samples were centrifuged at 1,000 X g for 30 seconds and supernatant was removed. After the last wash, samples were centrifuged at 7,500 X g for 30 seconds and supernatant was discarded. For elution, the resin was resuspended in 100 μ L of NLS Buffer and 10 μ L of Proteinase K (NEB) and the sample was incubated at 50C for 30 minutes while shaking at 1200 rpm. Input samples were similarly digested. Capture reactions were transferred to microspin cups (Pierce, Thermo Fisher

Scientific), centrifuged at 2,000 X g for 30 seconds, and elutions used for RNA purification by RNA Clean and Concentrate-5 kits (Zymo, >17nt protocol).

For qPCR analysis, cDNA was generated from purified RNA using Maxima H- reverse transcriptase (Thermo Fisher Scientific) following manufacturer's recommendations.

Amplification reactions were assembled with primer sets indicated in Table S2 and LightCycler. 480 SYBR Green I Master (Roche) following manufacturer's protocols and read out in a Roche Lightcycler 480.

Library construction. RNA-Seq libraries were constructed from purified RNA as previously described¹⁰². Briefly, after proteinase K elution, the RNA was dephosphorylated (Fast AP) and cyclic phosphates removed (T4 PNK) and then cleaned using Silane beads as previously described¹⁰². An RNA adapter containing a RT primer binding site was ligated to the 3' end of the cleaned and end-repaired RNA. The ligated RNA was reverse transcribed (RT) into cDNA, the RNA was degraded using NaOH, and a second adapter was ligated to the single stranded cDNA. Library preparation was the same for input samples except that an initial chemical fragmentation step (90C for 2 min 30 s in 1X FastAP buffer) was included prior to FastAP treatment. This chemical fragmentation step was designed to be similar to the fragmentation conditions used for purified Halo bound samples. The DNA was amplified and Illumina sequencing adaptors were added by PCR using primers that are complementary to the 3' and 5' adaptors. The molarity of PCR amplified libraries were measured by Agilent TapeStation High Sensitivity DNA screentapes and all samples were pooled at equal molarity. The pool was then purified and size selected on a 2% agarose gel and cut between 150- 700 nts. The final libraries

were measured by Agilent Bioanalyzer and Qubit high sensitivity DNA to determine the loading density of the final pooled sample. Pooled samples were paired-end sequenced on an Illumina HiSeq 2500 with read length 35 x 35nts.

Sequence alignment and analysis. For Halo purifications and RNA binding mapping sequencing reads were aligned to a combined genome reference containing the sequences of structural RNAs (ribosomal RNAs, snRNAs, snoRNAs, 45S pre-rRNA) and annotated mRNAs (RefSeq hg38) using Bowtie2. To distinguish between the nascent pre-ribosomal RNA and mature 18S, 28S, and 5.8S rRNA, we separated each of the components of the 45S into separate sequence units for alignment (e.g. ITS, ETS). We excluded all low quality alignments ($\text{MAPQ} < 2$) from the analysis. For mRNA analysis, we removed PCR duplicates using the Picard MarkDuplicates function (<https://broadinstitute.github.io/picard/>).

For each RNA, we enumerated 100 nucleotide windows across the entire RNA. For each window, we calculated the enrichment by computing the number of reads overlapping the window in the protein elution sample divided by the total number of reads within the protein elution sample. We normalized this ratio by the number of reads in the input sample divided by the total number of reads in the input sample. Because all windows overlapping a gene should have the same expression level in the input sample (which represents RNA expression), we estimated the number of reads in the input as the maximum of either (i) the number of reads over the window or (ii) the median read count over all windows within the gene. This approach provides a conservative estimation of enrichment because it prevents windows from being scored as enriched if the input values over a given window are artificially low, while at the same time

accounting for any non-random issues that lead to increases in read counts over a given window (e.g. fragmentation biases or alignment artifacts leading to non-random assignment or pileups).

We calculated a multiple testing corrected p-value using a scan statistic, as previously described¹⁰³⁻¹⁰⁴. Briefly, n was defined as the number of reads in the protein elution plus the number of reads in the control sample. p was defined as the total number of reads in the protein elution sample divided by the sum of the protein elution sample total reads and total reads in the control sample. w was the size of the window used for the analysis (100 nucleotides). The scan statistic p-value was defined using the Poisson estimations based on standard distributions previously described¹⁰⁵.

Because RNA within input samples are fragmented differently than the protein elution samples, we noticed that the overall positional distribution of protein elution samples was distinct from Input distributions. Accordingly, we used the remaining protein elution samples (rather than Input) as controls for each protein. Specifically, this enabled us to test whether a given protein is enriched within a given window relative to all other viral and control proteins. Enrichments were computed as described above. These values are plotted in Figure 1 and Table 1.

Plotting and visualization. Enrichment plots for specific RNAs were visualized in IGV¹⁰⁷ and were generated by either: (i) computing the enrichment for each nucleotide as described above. In this case, the read count for each nucleotide was computed as the total number of reads that overlapped the nucleotide. (ii) Counting the number of RT stop sites at a given nucleotide. In this case, we compute the alignment start position of the second in pair read and computed a count of

each nucleotide. We normalized this count by the total number of reads in the sample to account for sequencing depth generated. We then normalized this ratio by the same ratio computed for the control sample (merge of all other protein samples) for each nucleotide. Heatmaps were generated using Morpheus (<https://software.broadinstitute.org/morpheus/>). All values were included if they contained a significant 100nt window with a p-value<0.001 (see above) and minimum enrichment of 3-fold above the control sample.

Gene ontology analysis. The 66 non-N enriched mRNAs were analyzed against the Gene Ontology Biological Processes and Reactome gene sets using the Molecular Signatures Database (MSigDB)¹⁰⁷. Significantly enriched gene sets with an FDR<0.05 were used. To ensure that significant gene sets were not being driven by the multiple ribosomal proteins or histone proteins, these analyses were also carried out excluding these proteins.

Antibody Generation

To generate the sheep polyclonal anti-NSP1, anti-NSP8, anti-NSP9, and anti-NSP16 antibodies utilized in this study, NSP1, NSP8, NSP9 and NSP16 (using QHD43415.1 as reference) were cloned into pGex (GST-tagged) and pMex (MBP-tagged), in order to produce GST- and MBP-tagged respective NSP proteins. The N-terminal GST fusions were then used as antigens to immunize sheep. A bleed from the sheep was taken 7 days later, after which the MBP-tagged NSP proteins were used for serum affinity purification of the antibodies. To validate expression of the antibodies, Vero E6 cells were uninfected (mock) or infected with SARS-CoV-2 England-02 using a MOI of 0.1 or 1 (as indicated). At 72 hours post infection, the samples were harvested and the resulting whole cell lysates were probed by western blot with either sheep anti-

NSP or mouse anti-actin (Developmental Studies Hybridoma Bank JLA20, antibody registry ID: [AB_528068](#)) primary antibodies.

Microscopy imaging

Cells were seeded on gelatin/laminin and poly-D-lysine (Sigma) coated coverslips or chamber slides (Nunc, Thermo Fisher Scientific) and transfected with mammalian expression vectors for Halo-tagged viral proteins. After 16-24 hours, cells were incubated with TMR-HaloTag. Ligand (Promega) according to manufacturer's instructions, washed with PBS and fixed in 4% Formaldehyde (Pierce, Thermo Fisher Scientific). Cells were subsequently incubated in DAPI for 10 min and washed with PBS. For chamber slides, samples were imaged directly. For coverslips, samples were washed with water and mounted with ProLong Gold + DAPI (Molecular Probes, Thermo Fisher Scientific). We acquired images on a Nikon TS100-F widefield microscope or a Zeiss LSM800 inverted confocal microscope, collecting in line-scanning mode with 4x line averaging using a 63x oil objective.

For staining of infected cells, cells were fixed and permeabilized in 8% formaldehyde 1% Triton, and subsequently labelled with primary antibodies raised in sheep to SARS-CoV-2 at 1/500 dilution, followed by incubation with a rabbit anti-sheep Alexa 555 secondary antibody (Abcam, ab150182) at 1/1000 dilution and mounted with DAPI in the medium (Thermo Fisher Scientific, cat# P36395). Cells were imaged with a Zeiss LSM 880 confocal microscope, with 1 Airy unit pinhole for all primary antibody channel acquisitions and pixel size 0.07 μm x 0.07 μm . The objective lens used was a Zeiss Plan-Apochromatic 63x/1.4NA M27.

Structure modeling

NSP1 homology model. The predicted model of SARS-CoV-2 NSP1 was generated using the transform-restrained Rosetta (trRosetta) algorithm, a deep learning-based modeling method based on the Rosetta energy minimization pipeline with additional distance and interaction restraints generated from co-evolution¹⁰⁸. All figures were generated using Pymol (www.pymol.org).

NSP1-ribosome model. The model of NSP1 bound to the ribosome was generated using Modeller version 9.24¹⁰⁹. The C-terminal sequence of NSP1 (KHSSGVTRRLMRELNGG) was modeled using the structure of SERBP1 bound to the ribosome (PDB ID: 6MTE, chain w) as a template. The default Modeller parameters were used to create an alignment of NSP1 and SERBP1 and to generate the model, and all atoms within 6 Å of SERBP1 were included in the model to define the neighboring environment. Twenty models were generated and the model with the lowest DOPE score was selected to visualize with Pymol¹¹⁰.

Structural analysis of protein-RNA interactions. X-ray crystal structures and cryo-electron microscopy structures were obtained from the Protein Data Bank (www.rcsb.org)¹¹¹ and visualized with PyMOL¹¹⁰. For U1 and U2 structural analysis, we used a cryo-EM structure of the pre-catalytic human spliceosome (PDB ID: 6QX9). For 7SL structural analysis, we used an X-ray crystal structure of the human signal recognition particle (PDB ID: 1MFQ). To examine human SRP in the context of the ribosome, we used a cryo-EM structure of the mammalian SRP-ribosome complex (PDB ID: 3JAJ). To analyze the ribosomal ES27 expansion segment, we

superimposed a cryo-EM structure of the expansion segment (PDB ID: 6SXO) onto the complete ribosome structure (PDB ID: 3JAJ) using the PyMOL command “super.” Finally, for NSP1–18S rRNA structural analysis, we used multiple structures of the ribosome, including structures of the pre-40S subunit (PDB ID: 6G5H), 48S late-stage initiation complex (PDB ID: 6YAL), 80S in complex with SERBP1 (PDB ID: 6MTE), and 80S in complex with Stm1 (PDB ID: 4V88).

Recombinant NSP1 production

NSP1 was cloned into a bacterial expression vector resulting in N-terminally tagged Halo-6xHis-tagged Nsp1. The NSP1 sequence was PCR amplified from Addgene Nsp1 entry vector to add a N-terminal 6X HIS tag and restriction enzyme sites for digestion and ligation into N-terminal Halo bacterial expression vector. This construct was transformed into BL21 DE3 E. coli (Agilent), expanded to a 500mL liquid culture, and grown until OD₆₀₀ reached 1.0. IPTG was added to a final concentration of 1mM. After 3 hours of IPTG induction, bacteria was centrifuged for 15 min at 5000 X g. Pellet was lysed with binding buffer (50mM HEPES, pH 7.5, 20mM MgCl₂, 600mM NaCl, 2mM TCEP, 10mM Imidazole, 2mM ATP, 1% Triton X-100) supplemented with ATP (2mM), protease inhibitor cocktail (Promega), Benzonase (Sigma) and Triton-X 100 (Sigma) using 5mL of lysis mix per gram of wet cell paste. Cell suspension was rocked for 20 min at room temperature and then centrifuged at 16,000 X g for 20 min at 4C. Supernatant was incubated with washed iMAC resin (Bio-Rad) and rocked for 20 min at room temperature. We loaded the resin-lysate mixture into an appropriately-sized column and washed with 5 column volumes of binding buffer (50mM HEPES, pH 7.5, 20mM MgCl₂, 600mM NaCl, 2mM TCEP, 10mM Imidazole, 2mM ATP, 1% Triton X-100) followed by 10 column volumes of wash buffer (50mM HEPES, pH 7.5, 600mM NaCl, 2mM TCEP, 20mM Imidazole, pH 8).

Recombinant NSP1 (rNSP1) was eluted with 5 column volumes of elution buffer by adding 1 column volume at a time with column flow stopped, collecting eluate after each addition, and waiting 15 min between each elution buffer addition. We dialyzed these eluates with a 10mL Spectra-Por. Float-A-Lyzer. G2 (Spectrum Laboratories) into storage buffer (50mM HEPES, pH 7.5, 150mM NaCl, 10% glycerol) at 4C using 2 exchanges, one after 2 hours and then overnight.

***In vitro* translation assays**

Pierce 1-Step Human Coupled IVT-DNA (Thermo Fisher Scientific) *in vitro* translation kit was used to measure rNsp1-dependent translation inhibition. Bovine Serum Albumin (BSA), and buffer only controls were used to control for the addition of excess protein or changes in buffer composition. To measure translation inhibition, 5 μ L *in vitro* translation reactions were assembled, scaled according to manufacturer's recommendations. The included control plasmid pCFE-GFP was used to measure translational output of the reactions. GFP fluorescence was measured on a BioTek Cytation3 plate reader using emission filters for GFP fluorescence. 1.5 μ M stock dilutions of rNsp1 and BSA were made in storage buffer (50mM HEPES, pH 7.5.,150mM NaCl,10% glycerol). Subsequent 10 fold dilutions were made in storage buffer to span a concentration range of 1000 nM to 1 nM for each protein in the final reaction. 10 μ L of the diluted protein solution was added to the 5 μ L translation reactions, and incubated for 5 minutes at room temperature prior to the addition of the GFP reporter plasmid. Duplicate reactions were made to measure variability for each condition. In addition, a buffer only control was included to measure the effect of dilution of the translation reaction by the storage buffer. After the 5 minute incubation, 50 ng of GFP reporter plasmid was added to each reaction and incubated at 30C for

4 hours prior to fluorescence detection. Two microliters from each reaction was measured in duplicate on a Biotek Cytation3 microplate reader using excitation and emission filters for GFP. Sample readings were blanked by subtracting values obtained from the buffer only control. Promega's Rabbit Reticulocyte Lysate System was also used to assay translation inhibition. To measure translation inhibition, 10 μ L *in vitro* translation reactions were assembled, scaled according to manufacturer's recommendations. For each translation reaction, either 10 μ L of recombinant protein storage buffer or rNSP1 was added, followed by 500ng of mRNA. After 4 hours of incubation at 30C, luciferase was read out using the Bright-Glo luciferase assay (Promega) or GFP fluorescence was measured, both on a Biotek Cytation3 plate reader.

***In vivo* translation assays**

We assayed translation in HEK293T cells transfected with mammalian expression vectors, mRNAs, or combinations of these. For mRNA transfections of fluorescence protein translation reporters (including unmodified, +SARS-CoV2 leader sequence, +SL1, +SL2-SL1, and +5nts), DNA templates for *in vitro* transcription were generated with sequences appended to the 5' end of GFP and mCherry (see Tables S4 and S5 for primers and templates, respectively) and transcribed using HiScribe™ T7 ARCA mRNA Kit with tailing (New England Biolabs). For Nsp1 mRNA transfection, indicated primers from Table S4 were used to add restriction enzyme sites for cloning into pT7CFE1-CHis backbone provided in the Pierce Human 1-step Coupled IVT Kit and HiScribe™ T7 High Yield RNA Synthesis Kit (New England Biolabs) was used for *in vitro* transcription.

Using BioT transfection reagent, mammalian expression vectors for a GFP reporter and for SARS-CoV-2 viral proteins were transfected into HEK293T cells seeded for imaging, as described above, or seeded in 24 well plate format. To transfect only mRNA, Lipofectamine messengerMax (Invitrogen, Thermo Fisher Scientific) or TransIT-mRNA Transfection Kit (Mirus Bio) was used. For transfections that included both mRNA and plasmid, Lipofectamine 2000 (Invitrogen, Thermo Fisher Scientific) was used.

To measure fluorescence at 24 hours (leader-mCherry, no leader-GFP) or 48 hours (leader GFP, no-leader mCherry) post-transfection, cells were trypsinized and processed for flow cytometry or transferred into black 96 well plates (Nunc) for fluorescence detection on a Biotek Cytation 3 plate reader. For flow cytometry, lifted cells were washed with CBH buffer (10mM HEPES, pH 7.4, 0.5% BSA, Hank's Balanced Salt Solution (GIBCO, Thermo Fisher Scientific)), resuspended with a viability dye (7-AAD or DAPI) and analyzed on a MACSQuant Vyb. Acquisition files were analyzed with FlowJo analysis software.

SUnSET assay

To assay global protein translation, a SUnSET assay was performed as previously described⁷⁷. Mammalian expression vectors were exchanged for versions that did not confer puromycin resistance and thus, for these experiments, LR reactions were carried out with destination vector pB-Halo-DEST-IRES-NGFR. Resulting expression vectors drive protein expression by a dox-inducible promoter, contain the rtTA needed for dox induction, and produce an N-terminally-tagged Halo fusion protein. Generation of this destination vector made use of the pB-TAG-ERN backbone (a gift from Knut Woltjen; Addgene plasmid # 80476; <http://n2t.net/addgene:80476> ;

RRID:Addgene_80476)¹¹² and the NGFR (Truncated Human Nerve Growth Factor Receptor) coding sequence from Addgene plasmid #27489 (a gift from Warren Pear; <http://n2t.net/addgene:27489> ; RRID:Addgene_27489)⁸⁸.

We transfected these mammalian expression vectors for NSP1 and GFP into HEK293T using BioT transfection reagent. After 3 hours, doxycycline (Sigma) was added to a final concentration of 2µg/mL. After 24 hours, cells were incubated with puromycin (10µg/mL) for 10 min, then washed with fresh media, and harvested with cold PBS. Pelleted cells were lysed for 10 min on ice (mixing after 5 min) with 100uL RIPA buffer supplemented with protease inhibitor cocktail (Promega). Insoluble debris was pelleted by centrifuging at 12,500 X g for 2.5 minutes and supernatant was run on a Bolt™ 4-12% Bis-Tris Plus Gel (Thermo Fisher Scientific). Proteins were then transferred to nitrocellulose using the iBlot transfer system (Thermo Fisher Scientific) and Western blotting carried out using an anti-puro antibody (clone 12D10, EMD Millipore).

SUnSET in SARS-CoV-2 infected cells

SUnSET in SARS-CoV-2 infection was performed as above with the following modifications. Cells were infected or not (mock) with SARS-CoV-2, and 48 hpi cells were incubated with puromycin (10µg/mL) for 20 min. Media was aspirated and cells lysed directly in 2X Laemmli's buffer (Biorad), heated at 95C for ten minutes and run on a 4-12% NuPAGE Gel (Thermo Fisher Scientific). Proteins were transferred to nitrocellulose using the iBlot transfer system and probed as above.

Membrane protein reporter experiments

To assay SRP-dependent membrane protein transport to the cell surface, we monitored surface arrival of exogenously expressed Neuronal Growth Factor Receptor (NGFR) by flow cytometry in the presence of NSPs. Mammalian expression vectors were exchanged for versions that contained an IRES-NGFR to co-express a membrane reporter and thus, for these experiments, LR reactions were carried out with destination vector pB-6xHis-GFP-DEST-IRES-NGFR. Resulting expression vectors drive protein expression by a dox-inducible promoter, contain the rtTA needed for dox induction, and produce an N-terminally-tagged His-GFP fusion protein and a co-expressed NGFR. The GFP here is an enhanced GFP containing an amino acid substitution (A205K) to generate a monomeric variant based on previous literature¹¹³.

We transfected these mammalian expression vectors for NSP8, NSP9, NSP1DRC mutant and EED into HEK293T using BioT transfection reagent, induced expression with 2 μ g/ml doxycycline 24 hours after transfection, and assessed surface arrival of NGFR 24 hours after induction. To carry out flow cytometric analysis, cells were lifted with 1mM EDTA, washed once with PBS and stained with PE-labeled anti-NGFR antibody (Biolegend; 1/600 dilution in PBS, 0.5%BSA) and analyzed on a MACSQuant Vyb. Fluorescence intensity measurements were taken for GFP and PE and analyzed using FloJo analysis software.

siRNA experiments for SRP19 and SRP54

To knockdown SRP19 and SRP54, siRNAs targeting each (Dharmacon cat# L-019729-01-0005 and L-005122-01-0005, respectively) were transfected into HEK293T cells using Lipofectamine RNAiMAX (Invitrogen) according to manufacturer's protocols. To validate knockdown,

transfected cells were assayed by qPCR using primer sets (Table S2) to amplify each target as well as normalizer CALM3. Transfections were carried out 48 hours prior to assaying cells, either by qPCR, membrane reporter, or membrane SUnSET (see below) experiments.

Leader-NGFR measurements

Calu3 and Vero cells were transfected with mRNAs encoding leader-NGFR and leader-GFP using TransIT-mRNA Transfection Kit (Mirus) and subsequently infected with SARS-CoV-2 at an MOI of 0.1. After 24 hours, cells were washed with PBS, trypsinized and fixed in 4% PFA for 20 minutes before staining with biotinylated anti-NGFR (BioLegend) and anti-SARS-Cov-2 Spike Antibody (Sino) and subsequently stained with PE-labeled anti-Rabbit (Thermo, P-2771MP) and PacBlue-labeled streptavidin (Thermo, S1222). FACS was performed on a MACSquant Flow cytometer and analyzed using FloJo analysis software; FACS distributions were compared using a 2-tailed Kolmogorov-Smirnov test. For these experiments, RNA was transcribed from a PCR template (see Table S4) using the HiScribe T7 ARCA mRNA kit (with tailing).

Membrane SUnSET assay

To assay transport to the cell surface of all plasma membrane proteins, the SUnSET assay was adapted to puoro-label surface proteins as previously described⁷⁷, and read out by flow cytometry. Briefly, cells were incubated with puromycin as described above, followed by two quick washes and a chase with fresh complete media for 50 min. Cells were lifted with 1mM EDTA as described above and stained with an anti-puro antibody (clone 12D10, EMD Millipore) conjugated to Alexa-647. For these experiments, NSP was expressed from the same vector

described above for membrane reporter assays. Fluorescence intensity measurements were taken for GFP and Alexa-647 on a MACSquant Flow cytometer and analyzed using FloJo analysis software; distributions were compared using a 2-tailed Kolmogorov-Smirnov.

Splicing assessment experiments

IRF7-GFP splicing reporter. To assess splicing efficiency, exons 5-6 of mouse IRF7 (ENMUST00000026571.10) containing its endogenous intron were fused upstream of 2A self-cleaving peptide and eGFP and cloned into an MSCV vector (PIG, Addgene)¹¹⁸. This construct was co-transfected into HEK293Ts with NSP16 or GFP and measured 24 hours after transfection by flow cytometry (Macsqant) and analyzed using FloJo analysis software.

5EU labeling of RNA. SARS-Cov2 or mock infected Calu3 cells and Nsp16- or GAPDH-expressing HEK293Ts were labeled with 5-Ethynyl-uridine (5EU; Jena Bioscience) by adding 5EU containing media to cells for 20 min at a final concentration of 1mM, as previously described¹¹⁴. After the pulse label, cells were washed with warm PBS and lysed in RLT buffer (Qiagen). Total RNA was isolated from cells using manufacturer's protocols for Qiashtredder and RNeasy RNA isolation (both Qiagen), followed by Turbo DNase treatment (Ambion, Thermo Scientific), and Zymo RNA Clean and Concentrate. For each sample, 2µg of RNA was used for ligation of a unique barcoded RNA adaptor, following the relevant steps in the protocol described above in Library Construction of RNA-seq libraries. Samples were then pooled before proceeding to biotinylation steps.

Biotinylation of 5EU labeled RNA. To biotinylate 5EU-labeled RNA, samples were first mixed, in order, with water, HEPES (100 mM), biotin picolyl azide (1 mM; Click Chemistry Tools) and Ribolock RNase inhibitor, then added to premixed CuSO₄ (2 mM) and THPTA (10mM), and finally added to freshly prepared sodium ascorbate (12mM), as previously described¹¹⁵. The click reaction was incubated for 1 hour at 25C with 1000rpm shaking on an Eppendorf thermomixer followed by RNA purification using >17nt protocol for Zymo Clean and Concentrate.

Sequential capture of biotinylated RNA. We completed three rounds of sequential capture on streptavidin beads to isolate nascent transcripts (**see Figure 9B**). To capture biotinylated RNA, MyOne Streptavidin C1 Dynabeads (ThermoFisher Scientific) were first washed three times in Urea buffer (10mM HEPES, pH 7.5, 10mM EDTA, 0.5M LiCl, 0.5% Triton X-100, 0.2% SDS, 0.1% sodium deoxycholate, 2.5mM TCEP, 4M Urea) followed by three additional washes in M2 buffer (20mM Tris, pH 7.5, 50mM NaCl, 0.2% Triton X-100, 0.2% sodium deoxycholate, 0.2% NP-40). Washed beads were mixed with 3 parts 4M Urea buffer and 1 part biotinylated RNA and incubated for 60 min with 900rpm thermomixer shaking at room temperature.

After magnetic separation, beads were washed 3 times with M2 buffer followed by 3 washes with Urea buffer at 37C at 750rpm for 5 min. RNA was eluted from beads in 2 sequential elutions by incubating with elution buffer (5.7M guanidine thiocyanate , 1% N-lauroylsarcosine; both Sigma) at 65C for 2 minutes, repeating with more elution buffer for a second elution. The elutions were pooled, diluted with Urea buffer, incubated with pre-washed streptavidin beads, washed, and eluted for 2 additional rounds exactly as described above for a total of 3

sequential captures. Final elutions were pooled, cleaned with Zymo RNA Clean and Concentrate following manufacturer's protocols, and carried through RNA-seq library preparation as described above starting with the reverse transcription step.

Splicing analysis of 5EU data. Sequenced reads were demultiplexed according to barcoded RNA adaptor sequences ligated to each respective sample. Trimmomatic (<https://github.com/timflutre/trimmomatic>) was used to remove any contaminating Illumina primer sequences in the reads and low quality reads. Demultiplexed and trimmed files were then aligned to a hg19 reference genome using the spliceaware STAR aligner (<https://github.com/alexdobin/STAR>). Alignments were then deduplicated for PCR duplicates using PICARD MarkDuplicates (<https://broadinstitute.github.io/picard/>).

Aligned read-fragments were defined as read1 and read2 contained within a paired-end read fragment along with the insert between these two reads. We defined a set of high-quality represented isoforms per gene using the APPRIS database¹¹⁶. All read fragments that spanned any 3' splice site within an isoform of one of these genes was retained. For each 3' splice site spanning fragment, we classified the read-fragment as a spliced fragment if it spanned an exon-exon junction (e.g. aligned entirely within 2 distinct exons) or an unspliced fragment if it spanned an intron-exon junction (e.g. one of the reads was contained, or partially contained, within the intron). For each isoform, we computed an unspliced ratio by counting the total number of reads that were classified as unspliced divided by the total number of read fragments spanning 3' splice sites within that gene. To ensure that the splicing ratio that we measured is a reliable metric and not inflated/deflated due to low read counts, we only included genes that

contained at least 10 read-fragments in each sample and where the total number of reads in the control and sample conditions (when merged together) contained a significant number of reads to reliably measure a difference between the two groups as measured by a hypergeometric test ($p < 0.01$).

Because different genes contain different baseline splicing ratios due to gene length and coverage, we computed a change in the splicing ratio for each gene independently. To do this, we subtracted the unspliced ratio for each sample from the average unspliced ratio for that gene in all of the control samples. We plotted the overall distribution of these differences in splicing ratios as violin plots for each sample. If there is no change in splicing ratio, we would expect that some genes would have higher splicing ratios and others lower splicing ratios but that the overall distribution would be centered around 0.

Total RNA analysis of infected samples. Total RNA-Seq libraries were generated from the same mock infected and SARS-CoV-2 virally infected Calu3 samples treated with 5EU. Prior to 5EU purification, total RNA was taken and an RNA-Seq library constructed as described above using barcoded RNA adapters. Cytoplasmic ribosomal RNAs (18S and 28S) were depleted using NEBNext ribosomal RNA depletion kit (NEB E6310L) per manufacturers recommendations. Demultiplexed reads were aligned using Bowtie2 (<http://bowtie-bio.sourceforge.net/bowtie2/index.shtml>) to custom genomes encoding classical noncoding RNAs (ncRNAs) or human messenger RNAs (mRNAs). Expression levels were computed for each mRNA by counting the total number of sequencing reads aligned to the mature mRNA. To normalize across the different libraries, we computed the read counts for each sample that align

to non-spliced structural non-coding RNAs – excluding rRNA but including snRNAs, 7SL, 7SK, etc. We then divided each mRNA count by the sum of all ncRNA counts. This normalized value for each gene per sample was then converted into a fold-change by dividing this normalized value to the mean value for both mock infected samples. The fold change of each gene relative to mock was plotted across all mRNAs as a violin plot.

Interferon stimulation experiments

HEK-Blue™ ISG cells were seeded in 96 well plates, transfected with Nsp1 mammalian expression vectors using BioT and stimulated with 50 ng/ml human IFN-B (R&D Systems). Supernatants were assayed for alkaline phosphatase as per manufacturer instructions using QUANTI-Blue reagent (Invivogen).

HEK-293T cells were seeded in 6 well plates, transfected with either Halo-tagged GapdH, Nsp1, NSP8 and NSP9 in combination, or NSP16 mammalian expression vectors using BioT. 24 hours later, the media was replaced with media containing 50 ng/ml human IFN-b (R&D Systems). Expression was assayed using live cell Halo-imaging. Halo-TMR ligand was diluted 1:200 in media and added to the culture for a 1:1000 final dilution. Samples were incubated 30 minutes at 37C, 5% CO₂ and then the media was aspirated. Wells were rinsed twice with PBS, then media was added back to the wells. Samples were incubated 30 minutes at 37C, 5% CO₂ to allow uncoupled ligand to diffuse out of the cells. Media was then aspirated and replaced, and cells were imaged by widefield fluorescence microscopy. Cultures were ultimately harvested for RNA 24 hours later, or 48 hours post transfection.

A549s were seeded in 6 well plates, transfected with NSP1 mammalian expression vectors using Lipofectamine 2000 and stimulated with 1 µg/ml HMW poly(I:C) (Invivogen) 24h after transfection. Supernatant was assayed for secreted IFN- β by ELISA (Human IFN Beta ELISA, High Sensitivity, PBL) 24 hours after stimulation, and RNA from cells was purified and assessed for ISG gene expression as normalized to GAPDH expression (SYBR Green Master Mix, Bio-Rad). Primers used for qPCR are listed in Table S2.

5' viral leader experiments

Sars-CoV-2 Leader sequence was appended to the 5' end of GFP and mCherry reporter templates via PCR. PCR templates were then transcribed using HiScribe T7 ARCA mRNA kit (with tailing). Leader mutants, including SL1 only, SL1/SL2 swap, and +5nts mutants were likewise appended to the 5' end of fluorescent reporter templates via PCR and transcribed using Hiscribe T7 ARCA kit. mRNA reporters were transfected in HEK-293T cells with Lipofectamine MessengerMax. To measure fluorescence of mCherry and GFP reporters, 24 hours post transfection cells were either lifted with PBS and transferred into black 96 well plates for fluorescence readout on a Biotek Cytation 3 or trypsinized and processed for flow cytometry.

Alignments and phylogeny reconstructions

Alignments were performed with MAFFTT (v7.407) using a local alignment (`linsi --ep 0.123 --reorder [in.fasta] > [out.aln.fasta]`). Resulting alignments were visualized with Geneious.

Pairwise distance matrices were visualized with Morpheus. Phylogeny reconstructions were performed with IQTREE multicore (v1.6.12), model selection with 1000 bootstrap pseudoreplicates (`iqtree -s [out.aln.fasta] -m TEST -bb 1000 -nt 4 -o [outgroup]`). Phylogenies

were visualized with FigTree.

Table S1: RNA binding map of SARS-CoV-2 proteins, Related to Figure 4B can be found online¹¹⁷

Table S2: qPCR primers

Target	Forward Primer	Reverse Primer
GAPDH	ATTCCACCCATGGCAAATTC	TGGGATTTCCATTGATGACAAG
IFNbeta	TTCAGTGTGAGAAGCTCCTGTGG	CTGCTTAATCTCCTCAGGGATGTCA
ISG54	CTTCCCAGTCTATCATCAACCTT	CCGTCGCTTCTAGCTATGTATCT
18S_1	CATTCGAACGTCTGCCCTAT	ACCCTGATTCCCCGTCAC
18S_2	ATTGGAGGGCAAGTCTGGT	CCAAGATCCAACACTACGAGCTTT
LBR	AGTATAGCCTTCGTCCAAGAAGA	CAAAGGTTCTCACTGCCAGTT
CALM3	GACCATTGACTTCCCGGAGTT	GATGTAGCCATTCCCATCCTTG
SRP19	GCCGACCAGGACAGGTTTATT	TCTCTGTAGCTGTAGGATTTTCAAC
SRP54	TGGATTGCAAGGGAGTGGTAA	GGTCAAAAGCCCCTGCTCT

Table S3: Addgene ORFs, can be found online¹¹⁷.

Table S4: Primers used in mRNA generation.

mRNA	Forward Primer	Reverse Primer	Template
- Leader mCherry	TAATACGACTCACTATAGGCTAGCAC C ATGGTGA	CTACTACTACTTGT A CAGCTCGT	Leader mCherry Gblock
- Leader GFP	TAATACGACTCACTATAGGCTAGCAC C ATGGCCACATCGTGATGGT	CTTGTACAGCTCGT C CATGCCGA	Spytag- eGFP cDNA
+ Leader mCherry	TAATACGACTCACTATAGATTAAAG GT TTATACCTTCCCAGG	CTACTACTACTTGT A CAGCTCGT	Leader mCherry Gblock
+ Leader GFP	TAATACGACTCACTATAGATTAAAG GT TTATACCTTCCCAGGTAACAAACCAA C	CTTGTACAGCTCGT C	Spytag- eGFP cDNA

	CAACTTTCGATCTCTTGTAGATCGCA CC ATGGCCCACATCGTGATG	CATGCCGA	
+5nt- SL1 mCherry	TAATACGACTCACTATAGAATTATTA A AGGTTTATACCTTCCCAGGTAACAAA C CAACCAACTTTCGATCTCTTGTAGAT CG GCTAGCACCATGgtga	CTACTACTACTTGT A CAGCTCGT	Leader mCherry Gblock
+SL1 mCherry	TAATACGACTCACTATAGATTAAG GT TTATACCTTCCCAGGTAACAAACCGG C TAGCACCATGgtga	CTACTACTACTTGT A CAGCTCGT	Leader mCherry Gblock
+SL2- SL1 mCherry	TAATACGACTCACTATAGATTAAG AT CTCTTGTAGATCAACCAACTTTCGGT TT ATACCTTCCCAGGTAACAAACCGGCT A GCACCATGgtga	CTACTACTACTTGT A CAGCTCGT	Leader mCherry Gblock
Nsp1	AAAAGGATCCATGGAGAGCCTGGTG CC	AAAACTCGAGTTAG C CACCGTTCAGTTCA C	Nsp1 mammalian expression vector
+ Leader NGFR	ATTAAAGGTTTATACCTTCCCAGGTA A CAAACCGGCTAGCACCATGGCCACA AC CATGGA	TTATCGATAAGCTT G GCTGCAG	pB-Halo- eGFPIRES - NGFR mammalian expression vector

Table S5: Templates used in mRNA generation, Related to STAR Methods.

Leader mCherry Gblock	TCAGTCGCGATCGAACCCTGGAATTCATTAAAGGTTTATACCTTCC CAGGTAACAAACCAACCAACTTTTCGATCTCTTGTAGATCTGTTCTC TAAACGAACGGCTAGCACCATGGTGAGCAAGGGCGAGGAGGATA ACATGGCCATCATCAAGGAGTTCATGCGCTTCAAGGTGCACATGG AGGGCTCCGTGAACGGCCACGAGTTCGAGATCGAGGGCGAGGGC GAGGGCCGCCCTACGAGGGCACCCAGACCGCCAAGCTGAAGGT GACCAAGGGTGGCCCCCTGCCCTTCGCCTGGGACATCCTGTCCCCT CAGTTCATGTACGGCTCCAAGGCCTACGTGAAGCACCCCGCCGAC ATCCCCGACTACTTGAAGCTGTCCTTCCCCGAGGGCTTCAAGTGG GAGCGCGTGATGAACTTCGAGGACGGCGGCGTGTTGACCGTGACC CAGGACTCCTCCCTGCAGGACGGCGAGTTCATCTACAAGGTGAAG CTGCGCGGCACCAACTTCCCCTCCGACGGCCCCGTAATGCAGAAG AAGACCATGGGCTGGGAGGCCTCCTCCGAGCGGATGTACCCCGAG GACGGCGCCCTGAAGGGCGAGATCAAGCAGAGGCTGAAGCTGAA GGACGGCGGCCACTACGACGCTGAGGTCAAGACCACCTACAAGG CCAAGAAGCCCGTGCAGCTGCCCGGCGCCTACAACGTCAACATCA AGTTGGACATCACCTCCCACAACGAGGACTACCCATCGTGGAAC AGTACGAACGCGCCGAGGGCCGCCACTCCACCGGCGGCATGGAC GAGCTGTACAAGTAGTAGTAGGGCGGCCGCATAAAAATTAAGA
Spytag-eGFP cDNA	AGCGGATAACAATTTACACAGGAAACAGCTATGACCATGATTAC GCCAAGCTTATACGAAATTAATACGACTCACTATAGGGAGACCAC AACGGTTTCCCGAATTGTGAGCGGATAACAATAGAAATAATTTTG TTAACTTTAAGAAGGAGATATATCCATGGCCACATCGTGATGG TGGACGCCTACAAGCCTACTAAGGGCAGCGGCTCTAGAATGGCCA CAACCATGGTGAGCAAGGGCGAGGAGCTGTTACCGGGGTGGTG CCCATCCTGGTCGAGCTGGACGGCGACGTAAACGGCCACAAGTTC AGCGTGTCTGGCGAGGGCGAGGGCGATGCCACCTACGGCAAGCT GACCCTGAAGTTCATCTGCACCACCGCAAGCTGCCCGTGCCCTG GCCACCCTCGTGACCACCCTGACCTACGGCGTGACGTGCTTCAG CCGCTACCCCGACCACATGAAGCAGCACGACTTCTTCAAGTCCGC CATGCCC GAAGGCTACGTCCAGGAGCGCACCATCTTCTTCAAGGA CGACGGCAACTACAAGACCCGCGCCGAGGTGAAGTTCGAGGGCG ACACCCTGGTGAACCGCATCGAGCTGAAGGGCATCGACTTCAAGG AGGACGGCAACATCCTGGGGCACAAGCTGGAGTACAACACTACAAC AGCCACAACGTCTATATCATGGCCGACAAGCAGAAGAACGGCATC AAGGCGAACTTCAAGATCCGCCACAACATCGAGGACGGCAGCGT GCAGCTCGCCGACCACTACCAGCAGAACACCCCATCGGCGACGG CCCCGTGCTGCTGCCCGACAACCACTACCTGAGCACCCAGTCCGC CCTGAGCAAAGACCCCAACGAGAAGCGCGATCACATGGTCTCTGCT

	GGAGTTCGTGACCGCCGCGGGATCACTCTCGGCATGGACGAGCT GTACAAGGGATTCGGATCCGAACAAAACTCATCTCAGAAGAGG ATCTGGCGGCCGCTGGATCTGGTGGCCAGAAGCAAGCTGAAGAG GCGGCAGCGAAAGCGGCGGCAGATGCTAAAGCGAAGGCCGAAGC AGATGCTAAAGCTGCGGAAGAAGCAGCGAAGAAAGCGGCTGCAG ACGCAAAGAAAAAAGCAGAAGCAGAAGCCGCCAAAGCCGCAGCC GAAGCGCAGAAAAAAGCCGAGGCAGCCGCTGCGGCACTGAAGAA GAAAGCGGAAGCGGCAGAAGCAGCTGCAGCTGAAGCAAGAAAGA AAGCGGCAACTGAAACCGCACACCTTACTGGTGTGCGGGGCATCC CGACATGGACGTTGAATTATGAATTCCTGGCCGTCGTTTTACA
--	---

7.0 REFERENCES

1. Graham, R.L., and Baric, R.S. (2010). Recombination, reservoirs, and the modular spike: mechanisms of coronavirus cross-species transmission. *Journal of virology*, 84(7), 3134–3146. <https://doi.org/10.1128/JVI.01394-09>
2. Andersen, K.G., Rambaut, A., Lipkin, W.I., Holmes, E.C., and Garry, R. F. (2020). The proximal origin of SARS-CoV-2. *Nature medicine*, 26(4), 450–452. <https://doi.org/10.1038/s41591-020-0820-9>
3. Zou, L., Ruan, F., Huang, M., Liang, L., Huang, H., Hong, Z., Yu, J., Kang, M., Song, Y., Xia, J., Guo, Q., Song, T., He, J., Yen, H. L., Peiris, M., and Wu, J. (2020). SARS-CoV-2 Viral Load in Upper Respiratory Specimens of Infected Patients. *The New England journal of medicine*, 382(12), 1177–1179. <https://doi.org/10.1056/NEJMc2001737>
4. Bar-On, Y. M., Flamholz, A., Phillips, R., and Milo, R. (2020). SARS-CoV-2 (COVID-19) by the numbers. *eLife*, 9, e57309. <https://doi.org/10.7554/eLife.57309>
5. Wang, H., Li, X., Li, T., Zhang, S., Wang, L., Wu, X., and Liu, J. (2020). The genetic sequence, origin, and diagnosis of SARS-CoV-2. *European journal of clinical microbiology & infectious diseases : official publication of the European Society of Clinical Microbiology*, 39(9), 1629–1635. <https://doi.org/10.1007/s10096-020-03899-4>
6. da Silva, S., Alves da Silva, C.T., Mendes, R., and Pena, L. (2020). Role of nonstructural proteins in the pathogenesis of SARS-CoV-2. *Journal of medical virology*, 10.1002/jmv.25858. Advance online publication. <https://doi.org/10.1002/jmv.25858>
7. Chen, L., and Zhong, L. (2020). Genomics functional analysis and drug screening of SARS-CoV-2. *Genes & diseases*, 10.1016/j.gendis.2020.04.002. Advance online publication. <https://doi.org/10.1016/j.gendis.2020.04.002>

8. Finkel, Y., Mizrahi, O., Nachshon, A., Weingarten-Gabbay, S., Morgenstern, D., Yahalom-Ronen, Y., Tamir, H., Achdout, H., Stein, D., Israeli, O., Beth-Din, A., Melamed, S., Weiss, S., Israely, T., Paran, N., Schwartz, M., and Stern-Ginossar, N. (2020). The coding capacity of SARS-CoV-2. *Nature*, 10.1038/s41586-020-2739-1. Advance online publication. <https://doi.org/10.1038/s41586-020-2739-1>
9. Maier, H.J., Neuman, B.W., Bickerton, E., Keep, S.M., Alrashedi, H., Hall, R., and Britton, P. (2016). Extensive coronavirus-induced membrane rearrangements are not a determinant of pathogenicity. *Scientific reports*, 6, 27126. <https://doi.org/10.1038/srep27126>
10. Chow, K.T., Gale, M., Jr, and Loo, Y.M. (2018). RIG-I and Other RNA Sensors in Antiviral Immunity. *Annual review of immunology*, 36, 667–694. <https://doi.org/10.1146/annurev-immunol-042617-053309>
11. Jensen, S., and Thomsen, A.R. (2012). Sensing of RNA viruses: a review of innate immune receptors involved in recognizing RNA virus invasion. *Journal of virology*, 86(6), 2900–2910. <https://doi.org/10.1128/JVI.05738-11>
12. Wilkins, C., and Gale, M., Jr (2010). Recognition of viruses by cytoplasmic sensors. *Current opinion in immunology*, 22(1), 41–47. <https://doi.org/10.1016/j.coi.2009.12.003>
13. Stetson, D.B., and Medzhitov, R. (2006). Type I interferons in host defense. *Immunity*, 25(3), 373–381. <https://doi.org/10.1016/j.immuni.2006.08.007>
14. Beachboard, D.C., and Horner, S.M. (2016). Innate immune evasion strategies of DNA and RNA viruses. *Current opinion in microbiology*, 32, 113–119. <https://doi.org/10.1016/j.mib.2016.05.015>
15. Miller, D.M., Rahill, B.M., Boss, J.M., Lairmore, M.D., Durbin, J.E., Waldman, J.W., and Sedmak, D.D. (1998). Human cytomegalovirus inhibits major histocompatibility complex class II expression by disruption of the Jak/Stat pathway. *The Journal of experimental medicine*, 187(5), 675–683. <https://doi.org/10.1084/jem.187.5.675>
16. Kempf, B.J., and Barton, D.J. (2008). Poliovirus 2A(Pro) increases viral mRNA and polysome stability coordinately in time with cleavage of eIF4G. *Journal of virology*, 82(12), 5847–5859. <https://doi.org/10.1128/JVI.01514-07>
17. Lloyd, R.E. (2006). Translational control by viral proteinases. *Virus research*, 119(1), 76–88. <https://doi.org/10.1016/j.virusres.2005.10.016>
18. Kochs, G., García-Sastre, A., and Martínez-Sobrido, L. (2007). Multiple anti-interferon actions of the influenza A virus NS1 protein. *Journal of virology*, 81(13), 7011–7021. <https://doi.org/10.1128/JVI.02581-06>

19. Zhang, L., Wang, J., Muñoz-Moreno, R., Kim, M., Sakthivel, R., Mo, W., Shao, D., Anantharaman, A., García-Sastre, A., Conrad, N.K., and Fontoura, B. (2018). Influenza Virus NS1 Protein-RNA Interactome Reveals Intron Targeting. *Journal of virology*, 92(24), e01634-18. <https://doi.org/10.1128/JVI.01634-18>
20. Zhang, X., Tan, Y., Ling, Y., Lu, G., Liu, F., Yi, Z., Jia, X., Wu, M., Shi, B., Xu, S., Chen, J., Wang, W., Chen, B., Jiang, L., Yu, S., Lu, J., Wang, J., Xu, M., Yuan, Z., Zhang, Q., ... Lu, H. (2020). Viral and host factors related to the clinical outcome of COVID-19. *Nature*, 583(7816), 437–440. <https://doi.org/10.1038/s41586-020-2355-0>
21. Hadjadj, J., Yatim, N., Barnabei, L., Corneau, A., Boussier, J., Smith, N., Péré, H., Charbit, B., Bondet, V., Chenevier-Gobeaux, C., Breillat, P., Carlier, N., Gauzit, R., Morbieu, C., Pène, F., Marin, N., Roche, N., Szwebel, T.A., Merklings, S.H., Treluyer, J.M., ... Terrier, B. (2020). Impaired type I interferon activity and inflammatory responses in severe COVID-19 patients. *Science (New York, N.Y.)*, 369(6504), 718–724. <https://doi.org/10.1126/science.abc6027>
22. Lokugamage, K.G., Hage, A., de Vries, M., Valero-Jimenez, A.M., Schindewolf, C., Dittmann, M., Rajsbaum, R., and Menachery, V.D. (2020). Type I Interferon Susceptibility Distinguishes SARS-CoV-2 from SARS-CoV. *Journal of virology*, 94(23), e01410-20. <https://doi.org/10.1128/JVI.01410-20>
23. Canton, J., Fehr, A.R., Fernandez-Delgado, R., Gutierrez-Alvarez, F.J., Sanchez-Aparicio, M.T., García-Sastre, A., Perlman, S., Enjuanes, L., and Sola, I. (2018). MERS-CoV 4b protein interferes with the NF- κ B-dependent innate immune response during infection. *PLoS pathogens*, 14(1), e1006838. <https://doi.org/10.1371/journal.ppat.1006838>
24. Kamitani, W., Huang, C., Narayanan, K., Lokugamage, K.G., and Makino, S. (2009). A two-pronged strategy to suppress host protein synthesis by SARS coronavirus Nsp1 protein. *Nature structural & molecular biology*, 16(11), 1134–1140. <https://doi.org/10.1038/nsmb.1680>
25. Lokugamage, K.G., Narayanan, K., Nakagawa, K., Terasaki, K., Ramirez, S.I., Tseng, C.T., and Makino, S. (2015). Middle East Respiratory Syndrome Coronavirus nsp1 Inhibits Host Gene Expression by Selectively Targeting mRNAs Transcribed in the Nucleus while Sparing mRNAs of Cytoplasmic Origin. *Journal of virology*, 89(21), 10970–10981. <https://doi.org/10.1128/JVI.01352-15>
26. Nakagawa, K., Narayanan, K., Wada, M., and Makino, S. (2018). Inhibition of Stress Granule Formation by Middle East Respiratory Syndrome Coronavirus 4a Accessory Protein Facilitates Viral Translation, Leading to Efficient Virus Replication. *Journal of virology*, 92(20), e00902-18. <https://doi.org/10.1128/JVI.00902-18>
27. Sola, I., Mateos-Gomez, P.A., Almazan, F., Zuñiga, S., and Enjuanes, L. (2011). RNA-RNA and RNA-protein interactions in coronavirus replication and transcription. *RNA biology*, 8(2), 237–248. <https://doi.org/10.4161/rna.8.2.14991>

28. Peritz, T., Zeng, F., Kannanayakal, T.J., Kilk, K., Eiríksdóttir, E., Langel, U., Eberwine, J. (2006). Immunoprecipitation of mRNA-protein complexes. *Nature protocols*, 1(2), 577–580. <https://doi.org/10.1038/nprot.2006.82>
29. Keene, J.D., Komisarow, J.M., and Friedersdorf, M.B. (2006). RIP-Chip: the isolation and identification of mRNAs, microRNAs and protein components of ribonucleoprotein complexes from cell extracts. *Nature protocols*, 1(1), 302–307. <https://doi.org/10.1038/nprot.2006.47>
30. Hendrickson, G.D., Kelley, D.R., Tenen, D., Bernstein, B., and Rinn, J.L. (2016). Widespread RNA binding by chromatin-associated proteins. *Genome biology*, 17, 28. <https://doi.org/10.1186/s13059-016-0878-3>
31. Mili, S., and Steitz, J.A. (2004). Evidence for reassociation of RNA-binding proteins after cell lysis: implications for the interpretation of immunoprecipitation analyses. *RNA (New York, N.Y.)*, 10(11), 1692–1694. <https://doi.org/10.1261/rna.7151404>
32. Darnell, R.B. (2010). HITS-CLIP: panoramic views of protein-RNA regulation in living cells. *Wiley interdisciplinary reviews. RNA*, 1(2), 266–286. <https://doi.org/10.1002/wrna.31>
33. König, J., Zarnack, K., Rot, G., Curk, T., Kayikci, M., Zupan, B., Turner, D.J., Luscombe, N.M., and Ule, J. (2010). iCLIP reveals the function of hnRNP particles in splicing at individual nucleotide resolution. *Nature structural & molecular biology*, 17(7), 909–915. <https://doi.org/10.1038/nsmb.1838>
34. Hafner, M., Landthaler, M., Burger, L., Khorshid, M., Hausser, J., Berninger, P., Rothballer, A., Ascano, M., Jr, Jungkamp, A.C., Munschauer, M., Ulrich, A., Wardle, G.S., Dewell, S., Zavolan, M., and Tuschl, T. (2010). Transcriptome-wide identification of RNA-binding protein and microRNA target sites by PAR-CLIP. *Cell*, 141(1), 129–141. <https://doi.org/10.1016/j.cell.2010.03.009>
35. Zhao, J., Sun, B.K., Erwin, J.A., Song, J.J., and Lee, J.T. (2008). Polycomb proteins targeted by a short repeat RNA to the mouse X chromosome. *Science (New York, N.Y.)*, 322(5902), 750–756. <https://doi.org/10.1126/science.1163045>
36. Cifuentes-Rojas, C., Hernandez, A.J., Sarma, K., and Lee, J.T. (2014). Regulatory interactions between RNA and polycomb repressive complex 2. *Molecular cell*, 55(2), 171–185. <https://doi.org/10.1016/j.molcel.2014.05.009>
37. Schoeftner, S., Sengupta, A.K., Kubicek, S., Mechtler, K., Spahn, L., Koseki, H., Jenuwein, T., and Wutz, A. (2006). Recruitment of PRC1 function at the initiation of X inactivation independent of PRC2 and silencing. *The EMBO journal*, 25(13), 3110–3122. <https://doi.org/10.1038/sj.emboj.7601187>
38. Kalantry, S., and Magnuson, T. (2006). The Polycomb group protein EED is dispensable for the initiation of random X-chromosome inactivation. *PLoS genetics*, 2(5), e66. <https://doi.org/10.1371/journal.pgen.0020066>

39. Plath, K., Fang, J., Mlynarczyk-Evans, S.K., Cao, R., Worringer, K.A., Wang, H., de la Cruz, C.C., Otte, A.P., Panning, B., and Zhang, Y. (2003). Role of histone H3 lysine 27 methylation in X inactivation. *Science (New York, N.Y.)*, 300(5616), 131–135.
<https://doi.org/10.1126/science.1084274>
40. da Rocha, S.T., Boeva, V., Escamilla-Del-Arenal, M., Ancelin, K., Granier, C., Matias, N. R., Sanulli, S., Chow, J., Schulz, E., Picard, C., Kaneko, S., Helin, K., Reinberg, D., Stewart, A. F., Wutz, A., Margueron, R., and Heard, E. (2014). Jarid2 Is Implicated in the Initial Xist-Induced Targeting of PRC2 to the Inactive X Chromosome. *Molecular cell*, 53(2), 301–316.
<https://doi.org/10.1016/j.molcel.2014.01.002>
41. 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., and 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>
42. Kohlmaier, A., Savarese, F., Lachner, M., Martens, J., Jenuwein, T., and Wutz, A. (2004). A chromosomal memory triggered by Xist regulates histone methylation in X inactivation. *PLoS biology*, 2(7), E171. <https://doi.org/10.1371/journal.pbio.0020171>
43. Portoso, M., Ragazzini, R., Brenčić, Ž., Moiani, A., Michaud, A., Vassilev, I., Wassef, M., Servant, N., Sargueil, B., and Margueron, R. (2017). PRC2 is dispensable for *HOTAIR*-mediated transcriptional repression. *The EMBO journal*, 36(8), 981–994.
<https://doi.org/10.15252/embj.201695335>
44. Brockdorff, N. (2013). Noncoding RNA and Polycomb recruitment. *RNA (New York, N.Y.)*, 19(4), 429–442. <https://doi.org/10.1261/rna.037598.112>
45. Davidovich, C., Zheng, L., Goodrich, K.J., and Cech, T.R. (2013). Promiscuous RNA binding by Polycomb repressive complex 2. *Nature structural & molecular biology*, 20(11), 1250–1257. <https://doi.org/10.1038/nsmb.2679>
46. Wang, X., Goodrich, K.J., Gooding, A.R., Naeem, H., Archer, S., Paucek, R.D., Youmans, D.T., Cech, T.R., and Davidovich, C. (2017). Targeting of Polycomb Repressive Complex 2 to RNA by Short Repeats of Consecutive Guanines. *Molecular cell*, 65(6), 1056–1067.e5.
<https://doi.org/10.1016/j.molcel.2017.02.003>
47. Beltran, M., Yates, C.M., Skalska, L., Dawson, M., Reis, F.P., Viiri, K., Fisher, C.L., Sibley, C.R., Foster, B.M., Bartke, T., Ule, J., and Jenner, R.G. (2016). The interaction of PRC2 with RNA or chromatin is mutually antagonistic. *Genome research*, 26(7), 896–907.
<https://doi.org/10.1101/gr.197632.115>

48. Kaneko, S., Bonasio, R., Saldaña-Meyer, R., Yoshida, T., Son, J., Nishino, K., Umezawa, A., and Reinberg, D. (2014). Interactions between JARID2 and noncoding RNAs regulate PRC2 recruitment to chromatin. *Molecular cell*, 53(2), 290–300. <https://doi.org/10.1016/j.molcel.2013.11.012>
49. Khalil, A.M., Guttman, M., Huarte, M., Garber, M., Raj, A., Rivea Morales, D., Thomas, K., Presser, A., Bernstein, B.E., van Oudenaarden, A., Regev, A., Lander, E.S., and Rinn, J.L. (2009). Many human large intergenic noncoding RNAs associate with chromatin-modifying complexes and affect gene expression. *Proceedings of the National Academy of Sciences of the United States of America*, 106(28), 11667–11672. <https://doi.org/10.1073/pnas.0904715106>
50. Zhao, J., Ohsumi, T.K., Kung, J.T., Ogawa, Y., Grau, D.J., Sarma, K., Song, J.J., Kingston, R.E., Borowsky, M., and Lee, J.T. (2010). Genome-wide identification of polycomb-associated RNAs by RIP-seq. *Molecular cell*, 40(6), 939–953. <https://doi.org/10.1016/j.molcel.2010.12.011>
51. Kaneko, S., Son, J., Shen, S.S., Reinberg, D., and Bonasio, R. (2013). PRC2 binds active promoters and contacts nascent RNAs in embryonic stem cells. *Nature structural & molecular biology*, 20(11), 1258–1264. <https://doi.org/10.1038/nsmb.2700>
52. Pandey, R.R., Mondal, T., Mohammad, F., Enroth, S., Redrup, L., Komorowski, J., Nagano, T., Mancini-Dinardo, D., and 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>
53. Los, G.V., Encell, L.P., McDougall, M.G., Hartzell, D.D., Karassina, N., Zimprich, C., Wood, M.G., Learish, R., Ohana, R.F., Urh, M., Simpson, D., Mendez, J., Zimmerman, K., Otto, P., Vidugiris, G., Zhu, J., Darzins, A., Klaubert, D.H., Bulleit, R.F., and Wood, K.V. (2008). HaloTag: a novel protein labeling technology for cell imaging and protein analysis. *ACS chemical biology*, 3(6), 373–382. <https://doi.org/10.1021/cb800025k>
54. Zakeri, B., Fierer, J. O., Celik, E., Chittock, E.C., Schwarz-Linek, U., Moy, V.T., and Howarth, M. (2012). Peptide tag forming a rapid covalent bond to a protein, through engineering a bacterial adhesin. *Proceedings of the National Academy of Sciences of the United States of America*, 109(12), E690–E697. <https://doi.org/10.1073/pnas.1115485109>
55. Ray, D., Kazan, H., Chan, E.T., Peña Castillo, L., Chaudhry, S., Talukder, S., Blencowe, B.J., Morris, Q., and Hughes, T.R. (2009). Rapid and systematic analysis of the RNA recognition specificities of RNA-binding proteins. *Nature biotechnology*, 27(7), 667–670. <https://doi.org/10.1038/nbt.1550>
56. Sokolowski, M., and Schwartz, S. (2001). Heterogeneous nuclear ribonucleoprotein C binds exclusively to the functionally important UUUUU-motifs in the human papillomavirus type-1 AU-rich inhibitory element. *Virus research*, 73(2), 163–175. [https://doi.org/10.1016/s0168-1702\(00\)00238-0](https://doi.org/10.1016/s0168-1702(00)00238-0)

57. Uren, P.J., Bahrami-Samani, E., de Araujo, P.R., Vogel, C., Qiao, M., Burns, S.C., Smith, A. D., and Penalva, L.O. (2016). High-throughput analyses of hnRNP H1 dissects its multi-functional aspect. *RNA biology*, 13(4), 400–411. <https://doi.org/10.1080/15476286.2015.1138030>
58. Xiao, R., Tang, P., Yang, B., Huang, J., Zhou, Y., Shao, C., Li, H., Sun, H., Zhang, Y., and Fu, X.D. (2012). Nuclear matrix factor hnRNP U/SAF-A exerts a global control of alternative splicing by regulating U2 snRNP maturation. *Molecular cell*, 45(5), 656–668. <https://doi.org/10.1016/j.molcel.2012.01.009>
59. Simard, M.J., and Chabot, B. (2002). SRp30c is a repressor of 3' splice site utilization. *Molecular and cellular biology*, 22(12), 4001–4010. <https://doi.org/10.1128/mcb.22.12.4001-4010.2002>
60. Siomi, H., Choi, M., Siomi, M.C., Nussbaum, R.L., and Dreyfuss, G. (1994). Essential role for KH domains in RNA binding: impaired RNA binding by a mutation in the KH domain of FMR1 that causes fragile X syndrome. *Cell*, 77(1), 33–39. [https://doi.org/10.1016/0092-8674\(94\)90232-1](https://doi.org/10.1016/0092-8674(94)90232-1)
61. Yisraeli, J.K. (2005). VICKZ proteins: a multi-talented family of regulatory RNA-binding proteins. *Biology of the cell*, 97(1), 87–96. <https://doi.org/10.1042/BC20040151>
62. Baron-Benhamou, J., Gehring, N.H., Kulozik, A.E., and Hentze, M.W. (2004). Using the lambdaN peptide to tether proteins to RNAs. *Methods in molecular biology (Clifton, N.J.)*, 257, 135–154. <https://doi.org/10.1385/1-59259-750-5:135>
65. Séraphin, B., Kretzner, L., and Rosbash, M. (1988). A U1 snRNA:pre-mRNA base pairing interaction is required early in yeast spliceosome assembly but does not uniquely define the 5' cleavage site. *The EMBO journal*, 7(8), 2533–2538.
66. Majumdar, D., Frankiw, L., Burns, C., Garcia-Flores, Y., and Baltimore, D. (2018). Programmed Delayed Splicing: A Mechanism for Timed Inflammatory Gene Expression. *BioRxiv* 443796.
67. Decroly, E., Debarnot, C., Ferron, F., Bouvet, M., Coutard, B., Imbert, I., Gluais, L., Papageorgiou, N., Sharff, A., Bricogne, G., Ortiz-Lombardia, M., Lescar, J., and Canard, B. (2011). Crystal structure and functional analysis of the SARS-coronavirus RNA cap 2'-O-methyltransferase nsp10/nsp16 complex. *PLoS pathogens*, 7(5), e1002059. <https://doi.org/10.1371/journal.ppat.1002059>
68. Kurosaki, T., Popp, M.W., and Maquat, L.E. (2019). Quality and quantity control of gene expression by nonsense-mediated mRNA decay. *Nature reviews. Molecular cell biology*, 20(7), 406–420. <https://doi.org/10.1038/s41580-019-0126-2>

69. Brockway, S.M., and Denison, M.R. (2005). Mutagenesis of the murine hepatitis virus nsp1-coding region identifies residues important for protein processing, viral RNA synthesis, and viral replication. *Virology*, 340(2), 209–223. <https://doi.org/10.1016/j.virol.2005.06.035>
70. Narayanan, K., Ramirez, S.I., Lokugamage, K.G., and Makino, S. (2015). Coronavirus nonstructural protein 1: Common and distinct functions in the regulation of host and viral gene expression. *Virus research*, 202, 89–100. <https://doi.org/10.1016/j.virusres.2014.11.019>
71. Tanaka, T., Kamitani, W., DeDiego, M.L., Enjuanes, L., and Matsuura, Y. (2012). Severe acute respiratory syndrome coronavirus nsp1 facilitates efficient propagation in cells through a specific translational shutoff of host mRNA. *Journal of virology*, 86(20), 11128–11137. <https://doi.org/10.1128/JVI.01700-12>
72. Simonetti, A., Guca, E., Bochler, A., Kuhn, L., and Hashem, Y. (2020). Structural Insights into the Mammalian Late-Stage Initiation Complexes. *Cell reports*, 31(1), 107497. <https://doi.org/10.1016/j.celrep.2020.03.061>
73. Ogle, J.M., Brodersen, D.E., Clemons, W.M., Jr, Tarry, M.J., Carter, A.P., and Ramakrishnan, V. (2001). Recognition of cognate transfer RNA by the 30S ribosomal subunit. *Science (New York, N.Y.)*, 292(5518), 897–902. <https://doi.org/10.1126/science.1060612>
74. Brown, A., Baird, M.R., Yip, M.C., Murray, J., and Shao, S. (2018). Structures of translationally inactive mammalian ribosomes. *eLife*, 7, e40486. <https://doi.org/10.7554/eLife.40486>
75. Ben-Shem, A., Garreau de Loubresse, N., Melnikov, S., Jenner, L., Yusupova, G., and Yusupov, M. (2011). The structure of the eukaryotic ribosome at 3.0 Å resolution. *Science (New York, N.Y.)*, 334(6062), 1524–1529. <https://doi.org/10.1126/science.1212642>
76. Thoms, M., Buschauer, R., Ameismeier, M., Koepke, L., Denk, T., Hirschenberger, M., Kratzat, H., Hayn, M., Mackens-Kiani, T., Cheng, J., Straub, J.H., Stürzel, C.M., Fröhlich, T., Berninghausen, O., Becker, T., Kirchhoff, F., Sparrer, K., and Beckmann, R. (2020). Structural basis for translational shutdown and immune evasion by the Nsp1 protein of SARS-CoV-2. *Science (New York, N.Y.)*, 369(6508), 1249–1255. <https://doi.org/10.1126/science.abc8665>
77. Schmidt, E.K., Clavarino, G., Ceppi, M., and Pierre, P. (2009). SUnSET, a nonradioactive method to monitor protein synthesis. *Nature methods*, 6(4), 275–277. <https://doi.org/10.1038/nmeth.1314>
78. Narayanan, K., Huang, C., Lokugamage, K., Kamitani, W., Ikegami, T., Tseng, C.T., and Makino, S. (2008). Severe acute respiratory syndrome coronavirus nsp1 suppresses host gene expression, including that of type I interferon, in infected cells. *Journal of virology*, 82(9), 4471–4479. <https://doi.org/10.1128/JVI.02472-07>

79. Alexopoulou, L., Holt, A.C., Medzhitov, R., and Flavell, R.A. (2001). Recognition of double-stranded RNA and activation of NF-kappaB by Toll-like receptor 3. *Nature*, *413*(6857), 732–738. <https://doi.org/10.1038/35099560>
80. Kato, H., Takeuchi, O., Sato, S., Yoneyama, M., Yamamoto, M., Matsui, K., Uematsu, S., Jung, A., Kawai, T., Ishii, K.J., Yamaguchi, O., Otsu, K., Tsujimura, T., Koh, C.S., Reis e Sousa, C., Matsuura, Y., Fujita, T., and Akira, S. (2006). Differential roles of MDA5 and RIG-I helicases in the recognition of RNA viruses. *Nature*, *441*(7089), 101–105. <https://doi.org/10.1038/nature04734>
81. Gaglia, M.M., Covarrubias, S., Wong, W., and Glaunsinger, B.A. (2012). A common strategy for host RNA degradation by divergent viruses. *Journal of virology*, *86*(17), 9527–9530. <https://doi.org/10.1128/JVI.01230-12>
82. Kim, D., Lee, J.Y., Yang, J.S., Kim, J.W., Kim, V.N., and Chang, H. (2020). The Architecture of SARS-CoV-2 Transcriptome. *Cell*, *181*(4), 914–921.e10. <https://doi.org/10.1016/j.cell.2020.04.011>
83. Akopian, D., Shen, K., Zhang, X., and Shan, S.O. (2013). Signal recognition particle: an essential protein-targeting machine. *Annual review of biochemistry*, *82*, 693–721. <https://doi.org/10.1146/annurev-biochem-072711-164732>
84. Holtkamp, W., Lee, S., Bornemann, T., Senyushkina, T., Rodnina, M.V., and Wintermeyer, W. (2012). Dynamic switch of the signal recognition particle from scanning to targeting. *Nature structural & molecular biology*, *19*(12), 1332–1337. <https://doi.org/10.1038/nsmb.2421>
85. Parker, M.S., Balasubramaniam, A., Sallee, F.R., and Parker, S.L. (2018). The Expansion Segments of 28S Ribosomal RNA Extensively Match Human Messenger RNAs. *Frontiers in genetics*, *9*, 66. <https://doi.org/10.3389/fgene.2018.00066>
86. Zhang, Y., Ma, C., Yuan, Y., Zhu, J., Li, N., Chen, C., Wu, S., Yu, L., Lei, J., and Gao, N. (2014). Structural basis for interaction of a cotranslational chaperone with the eukaryotic ribosome. *Nature structural & molecular biology*, *21*(12), 1042–1046. <https://doi.org/10.1038/nsmb.2908>
87. Wild, K., Aleksić, M., Lapouge, K., Juaire, K.D., Flemming, D., Pfeffer, S., and Sinning, I. (2020). MetAP-like Ebp1 occupies the human ribosomal tunnel exit and recruits flexible rRNA expansion segments. *Nature communications*, *11*(1), 776. <https://doi.org/10.1038/s41467-020-14603-7>
88. Izon, D.J., Punt, J.A., Xu, L., Karnell, F.G., Allman, D., Myung, P.S., Boerth, N.J., Pui, J.C., Koretzky, G.A., and Pear, W.S. (2001). Notch1 regulates maturation of CD4+ and CD8+ thymocytes by modulating TCR signal strength. *Immunity*, *14*(3), 253–264. [https://doi.org/10.1016/s1074-7613\(01\)00107-8](https://doi.org/10.1016/s1074-7613(01)00107-8)

89. Sutton, G., Fry, E., Carter, L., Sainsbury, S., Walter, T., Nettleship, J., Berrow, N., Owens, R., Gilbert, R., Davidson, A., Siddell, S., Poon, L.L., Diprose, J., Alderton, D., Walsh, M., Grimes, J.M., and Stuart, D.I. (2004). The nsp9 replicase protein of SARS-coronavirus, structure and functional insights. *Structure (London, England : 1993)*, 12(2), 341–353. <https://doi.org/10.1016/j.str.2004.01.016>
90. Vermeire, K., Bell, T.W., Van Puyenbroeck, V., Giraut, A., Noppen, S., Liekens, S., Schols, D., Hartmann, E., Kalies, K.U., and Marsh, M. (2014). Signal peptide-binding drug as a selective inhibitor of co-translational protein translocation. *PLoS biology*, 12(12), e1002011. <https://doi.org/10.1371/journal.pbio.1002011>
91. Zhou, Q., Chen, V., Shannon, C.P., Wei, X.S., Xiang, X., Wang, X., Wang, Z.H., Tebbutt, S. J., Kollmann, T.R., and Fish, E.N. (2020). Interferon- α 2b Treatment for COVID-19. *Frontiers in immunology*, 11, 1061. <https://doi.org/10.3389/fimmu.2020.01061>
92. Liaud, N., Horlbeck, M.A., Gilbert, L.A., Gjoni, K., Weissman, J.S., and Cate, J. (2019). Cellular response to small molecules that selectively stall protein synthesis by the ribosome *PLoS genetics*, 15(3), e1008057. <https://doi.org/10.1371/journal.pgen.1008057>
93. Hermann, T. (2016). Small molecules targeting viral RNA. *Wiley interdisciplinary reviews. RNA*, 7(6), 726–743. <https://doi.org/10.1002/wrna.1373>
94. Hansen, T.H., and Bouvier, M. (2009). MHC class I antigen presentation: learning from viral evasion strategies. *Nature reviews. Immunology*, 9(7), 503–513. <https://doi.org/10.1038/nri2575>
95. Weston, C.R., and Davis, R.J. (2007). The JNK signal transduction pathway. *Current opinion in cell biology*, 19(2), 142–149. <https://doi.org/10.1016/j.ceb.2007.02.001>
96. Cope, G.A., and Deshaies, R.J. (2003). COP9 signalosome: a multifunctional regulator of SCF and other cullin-based ubiquitin ligases. *Cell*, 114(6), 663–671. [https://doi.org/10.1016/s0092-8674\(03\)00722-0](https://doi.org/10.1016/s0092-8674(03)00722-0)
97. Claret, F.X., Hibi, M., Dhut, S., Toda, T., and Karin, M. (1996). A new group of conserved coactivators that increase the specificity of AP-1 transcription factors. *Nature*, 383(6599), 453–457. <https://doi.org/10.1038/383453a0>
98. Lungu, G.F., Stoica, G., and Wong, P.K. (2008). Down-regulation of Jab1, HIF-1 α , and VEGF by Moloney murine leukemia virus-ts1 infection: a possible cause of neurodegeneration. *Journal of neurovirology*, 14(3), 239–251. <https://doi.org/10.1080/13550280802093919>
99. Oh, W., Yang, M.R., Lee, E.W., Park, K.M., Pyo, S., Yang, J.S., Lee, H.W., and Song, J. (2006). Jab1 mediates cytoplasmic localization and degradation of West Nile virus capsid protein. *The Journal of biological chemistry*, 281(40), 30166–30174. <https://doi.org/10.1074/jbc.M602651200>

100. Tanaka, Y., Kanai, F., Ichimura, T., Tateishi, K., Asaoka, Y., Guleng, B., Jazag, A., Ohta, M., Imamura, J., Ikenoue, T., Ijichi, H., Kawabe, T., Isobe, T., and Omata, M. (2006). The hepatitis B virus X protein enhances AP-1 activation through interaction with Jab1. *Oncogene*, 25(4), 633–642. <https://doi.org/10.1038/sj.onc.1209093>
101. Kim, D.K., Knapp, J.J., Kuang, D., Chawla, A., Cassonnet, P., Lee, H., Sheykhkarimli, D., Samavarchi-Tehrani, P., Abdouni, H., Rayhan, A., Li, R., Pogoutse, O., Coyaud, É., van der Werf, S., Demeret, C., Gingras, A.C., Taipale, M., Raught, B., Jacob, Y., and Roth, F.P. (2020). A Comprehensive, Flexible Collection of SARS-CoV-2 Coding Regions. *G3 (Bethesda, Md.)*, 10(9), 3399–3402. <https://doi.org/10.1534/g3.120.401554>
102. Van Nostrand, E.L., Pratt, G.A., Shishkin, A.A., Gelboin-Burkhart, C., Fang, M.Y., Sundararaman, B., Blue, S.M., Nguyen, T.B., Surka, C., Elkins, K., Stanton, R., Rigo, F., Guttman, M., and Yeo, G.W. (2016). Robust transcriptome-wide discovery of RNA-binding protein binding sites with enhanced CLIP (eCLIP). *Nature methods*, 13(6), 508–514. <https://doi.org/10.1038/nmeth.3810>
103. Guttman, M., Amit, I., Garber, M., French, C., Lin, M.F., Feldser, D., Huarte, M., Zuk, O., Carey, B.W., Cassady, J.P., Cabili, M.N., Jaenisch, R., Mikkelsen, T.S., Jacks, T., Hacohen, N., Bernstein, B.E., Kellis, M., Regev, A., Rinn, J.L., and Lander, E.S. (2009). Chromatin signature reveals over a thousand highly conserved large non-coding RNAs in mammals. *Nature*, 458 (7235), 223–227. <https://doi.org/10.1038/nature07672>
104. Guttman, M., Garber, M., Levin, J.Z., Donaghey, J., Robinson, J., Adiconis, X., Fan, L., Koziol, M.J., Gnirke, A., Nusbaum, C., Rinn, J.L., Lander, E.S., and Regev, A. (2010). Ab initio reconstruction of cell type-specific transcriptomes in mouse reveals the conserved multi-exonic structure of lincRNAs. *Nature biotechnology*, 28(5), 503–510. <https://doi.org/10.1038/nbt.1633>
105. Naus, J.I. (1982). Approximations for Distributions of Scan Statistics. *J. Am. Stat. Assoc.* 77 (177).
106. Robinson, J.T., Thorvaldsdóttir, H., Winckler, W., Guttman, M., Lander, E.S., Getz, G., and Mesirov, J.P. (2011). Integrative genomics viewer. *Nature biotechnology*, 29(1), 24–26. <https://doi.org/10.1038/nbt.1754>
107. Liberzon, A., Birger, C., Thorvaldsdóttir, H., Ghandi, M., Mesirov, J.P., and Tamayo, P. (2015). The Molecular Signatures Database (MSigDB) hallmark gene set collection. *Cell systems*, 1(6), 417–425. <https://doi.org/10.1016/j.cels.2015.12.004>
108. Yang, J., Anishchenko, I., Park, H., Peng, Z., Ovchinnikov, S., and Baker, D. (2020). Improved protein structure prediction using predicted interresidue orientations. *Proceedings of the National Academy of Sciences of the United States of America*, 117(3), 1496–1503. <https://doi.org/10.1073/pnas.1914677117>

109. Webb, B., and Sali, A. (2016). Comparative Protein Structure Modeling Using MODELLER. *Current protocols in bioinformatics*, 54, 5.6.1–5.6.37. <https://doi.org/10.1002/cpbi.3>
110. Delano, W. (2002). PyMOL molecular graphics system.
111. Berman, H.M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T.N., Weissig, H., Shindyalov, I.N., and Bourne, P.E. (2000). The Protein Data Bank. *Nucleic acids research*, 28(1), 235–242. <https://doi.org/10.1093/nar/28.1.235>
112. Kim, S.I., Ocegüera-Yanez, F., Sakurai, C., Nakagawa, M., Yamanaka, S., and Woltjen, K. (2016). Inducible Transgene Expression in Human iPS Cells Using Versatile All-in-One piggyBac Transposons. *Methods in molecular biology (Clifton, N.J.)*, 1357, 111–131. https://doi.org/10.1007/7651_2015_251
113. Alberti, S., Saha, S., Woodruff, J.B., Franzmann, T.M., Wang, J., and Hyman, A.A. (2018). A User's Guide for Phase Separation Assays with Purified Proteins. *Journal of molecular biology*, 430(23), 4806–4820. <https://doi.org/10.1016/j.jmb.2018.06.038>
114. Jao, C.Y., and Salic, A. (2008). Exploring RNA transcription and turnover in vivo by using click chemistry. *Proceedings of the National Academy of Sciences of the United States of America*, 105(41), 15779–15784. <https://doi.org/10.1073/pnas.0808480105>
115. Hong, V., Presolski, S.I., Ma, C., and Finn, M.G. (2009). Analysis and optimization of copper-catalyzed azide-alkyne cycloaddition for bioconjugation. *Angewandte Chemie (International ed. in English)*, 48(52), 9879–9883. <https://doi.org/10.1002/anie.200905087>
116. Rodriguez, J.M., Maietta, P., Ezkurdia, I., Pietrelli, A., Wesselink, J.J., Lopez, G., Valencia, A., and Tress, M.L. (2013). APPRIS: annotation of principal and alternative splice isoforms. *Nucleic acids research*, 41(Database issue), D110–D117. <https://doi.org/10.1093/nar/gks1058>
117. Banerjee, A.K., Blanco, M.R., Bruce, E.A., Honson, D.D., Chen, L.M., Chow, A., Bhat, P., Ollikainen, N., Quinodoz, S.A., Loney, C., Thai, J., Miller, Z.D., Lin, A.E., Schmidt, M.M., Stewart, D.G., Goldfarb, D., De Lorenzo, G., Rihn, S.J., Voorhees, R.M., Botten, J.W., ... Guttman, M. (2020). SARS-CoV-2 Disrupts Splicing, Translation, and Protein Trafficking to Suppress Host Defenses. *Cell*, S0092-8674(20)31310-6. Advance online publication. <https://doi.org/10.1016/j.cell.2020.10.004>
118. Mayr, C., and Bartel, D.P. (2009). Widespread shortening of 3'UTRs by alternative cleavage and polyadenylation activates oncogenes in cancer cells. *Cell*, 138(4), 673–684. <https://doi.org/10.1016/j.cell.2009.06.016>

119. Chen, C.K., Blanco, M., Jackson, C., Aznauryan, E., Ollikainen, N., Surka, C., Chow, A., Cerase, A., McDonel, P., and Guttman, M. (2016). Xist recruits the X chromosome to the nuclear lamina to enable chromosome-wide silencing. *Science (New York, N.Y.)*, 354(6311), 468–472. <https://doi.org/10.1126/science.aae0047>
120. Cirillo, D., Blanco, M., Armaos, A., Bunes, A., Avner, P., Guttman, M., Cerase, A., and Tartaglia, G.G. (2016). Quantitative predictions of protein interactions with long noncoding RNAs. *Nature methods*, 14(1), 5–6. <https://doi.org/10.1038/nmeth.4100>
121. Shishkin, A.A., Giannoukos, G., Kucukural, A., Ciulla, D., Busby, M., Surka, C., Chen, J., Bhattacharyya, R.P., Rudy, R.F., Patel, M.M., Novod, N., Hung, D.T., Gnirke, A., Garber, M., Guttman, M., and Livny, J. (2015). Simultaneous generation of many RNA-seq libraries in a single reaction. *Nature methods*, 12(4), 323–325. <https://doi.org/10.1038/nmeth.3313>
122. Bolger, A.M., Lohse, M., and Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics (Oxford, England)*, 30(15), 2114–2120. <https://doi.org/10.1093/bioinformatics/btu170>
123. Xu, H., Luo, X., Qian, J., Pang, X., Song, J., Qian, G., Chen, J., and Chen, S. (2012). FastUniq: a fast de novo duplicates removal tool for paired short reads. *PloS one*, 7(12), e52249. <https://doi.org/10.1371/journal.pone.0052249>
124. Lassmann, T., Hayashizaki, Y., and Daub, C.O. (2009). TagDust--a program to eliminate artifacts from next generation sequencing data. *Bioinformatics (Oxford, England)*, 25(21), 2839–2840. <https://doi.org/10.1093/bioinformatics/btp527>
125. Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., and Gingeras, T.R. (2013). STAR: ultrafast universal RNA-seq aligner. *Bioinformatics (Oxford, England)*, 29(1), 15–21. <https://doi.org/10.1093/bioinformatics/bts635>
126. Charenton, C., Wilkinson, M.E., and Nagai, K. (2019). Mechanism of 5' splice site transfer for human spliceosome activation. *Science (New York, N.Y.)*, 364(6438), 362–367. <https://doi.org/10.1126/science.aax3289>
127. Ameisemeier, M., Cheng, J., Berninghausen, O., and Beckmann, R. (2018). Visualizing late states of human 40S ribosomal subunit maturation. *Nature*, 558(7709), 249–253. <https://doi.org/10.1038/s41586-018-0193-0>
128. Kuglstatter, A., Oubridge, C., and Nagai, K. (2002). Induced structural changes of 7SL RNA during the assembly of human signal recognition particle. *Nature structural biology*, 9(10), 740–744. <https://doi.org/10.1038/nsb843>
129. Voorhees, R.M., and Hegde, R.S. (2015). Structures of the scanning and engaged states of the mammalian SRP-ribosome complex. *eLife*, 4, e07975. <https://doi.org/10.7554/eLife.07975>

8.0 FIGURES

Figure 1

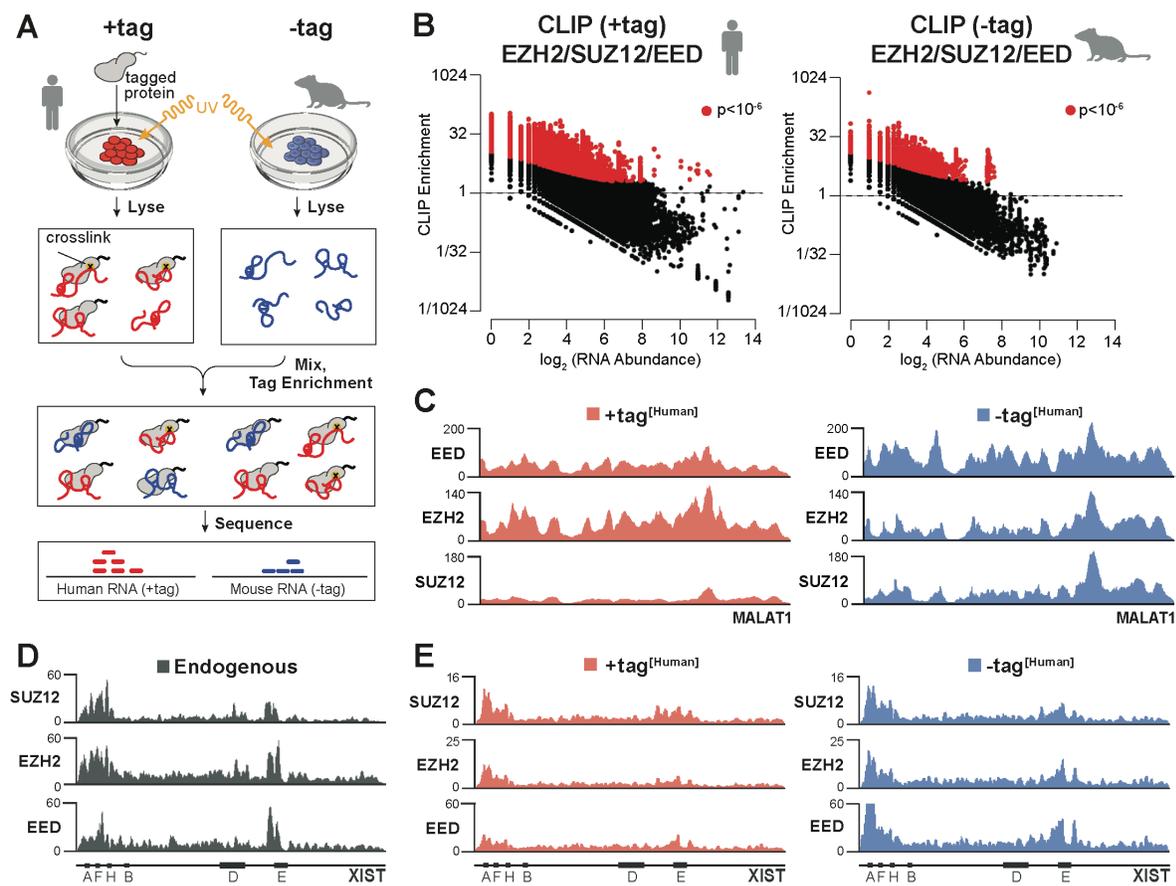


Figure 1. A human and mouse mixing experiment defines RNA-protein interactions that occur in solution after cell lysis. (A) A schematic overview of the human and mouse mixing experiment. An epitope-tagged protein is expressed in human cells (+tag) and UV-crosslinked, lysed, and mixed with UV-crosslinked cell lysate from mouse cells not expressing the tagged protein (-tag). The tagged protein is specifically enriched using an antibody against the epitope tag and its associated RNAs are sequenced and aligned to the human and mouse transcriptomes to quantify the amount of signal associated with human RNAs and mouse RNAs, respectively.

(B) Scatter plot of RNA abundance (log scale, x-axis) compared to CLIP enrichment (log scale, y-axis) for all the PRC2 components (EZH2, SUZ12, EED) components across all 100-nucleotide windows of all annotated human RNAs (+tag, left) and mouse RNAs (-tag, right). Windows with significant enrichment (binomial $p < 10^{-6}$) are shown in red. **(C)** CLIP enrichments for each PRC2 protein in the +tag (red) or -tag (blue) samples are plotted across the human lncRNA MALAT1. **(D)** CLIP enrichment profiles for PRC2 components captured using endogenous antibodies in untagged cells plotted across human XIST. **(E)** CLIP enrichment profiles for V5-tagged PRC2 components across human XIST from +tag (red) or -tag (blue) samples.

Figure 2

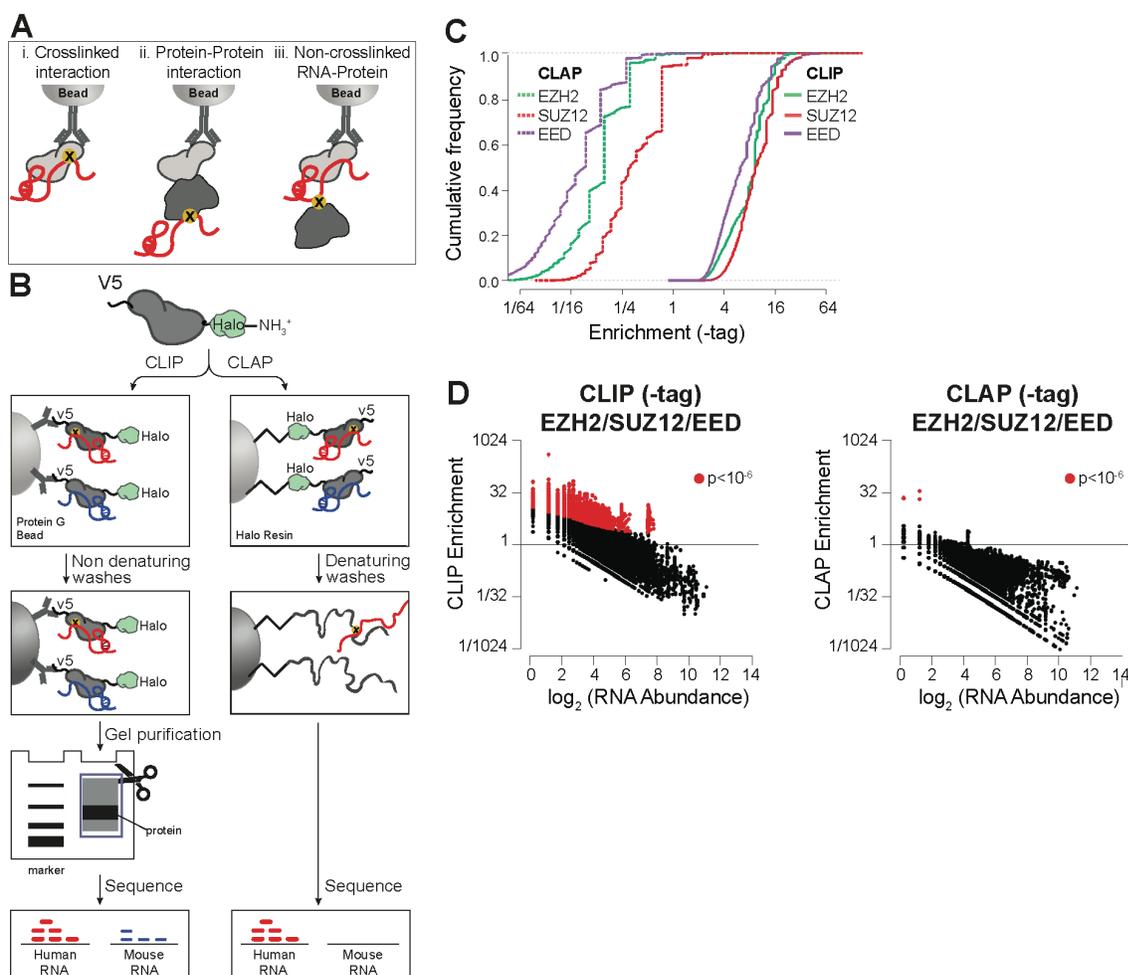


Figure 2. CLAP removes RNA-protein interactions that do not occur *in vivo*. (A) A schematic illustrating several potential ways in which RNA-protein interactions that form in solution may still be detected by CLIP. (B) A schematic overview comparing the CLIP (left) and CLAP (right) methods. A protein is tagged with both a HaloTag and V5 epitope and expressed in human cells (+tag) which are then mixed with mouse cells not expressing the tagged protein (-tag). After the human and mouse lysates are mixed, the sample is split and CLIP and CLAP are performed on each. CLIP is performed with an anti-V5 antibody followed by standard CLIP washes, gel electrophoresis, transfer to a nitrocellulose membrane, and size selected prior to

RNA sequencing. CLAP covalently binds the protein to a Halo capture resin and washes in fully denaturing conditions prior to RNA sequencing. **(C)** Cumulative Distribution Function (CDF) plots for all 100 nucleotide windows across -tag RNAs comparing CLIP and CLAP data for the same proteins. **(D)** Scatter plot of RNA abundance compared to the CLIP enrichment (left) or CLAP enrichment (right) for all the PRC2 (EZH2,SUZ12,EED) proteins across all 100-nucleotide windows of annotated mouse RNAs (-tag). Windows with significant enrichment (binomial $p < 10^{-6}$) are shown in red.

Figure 3

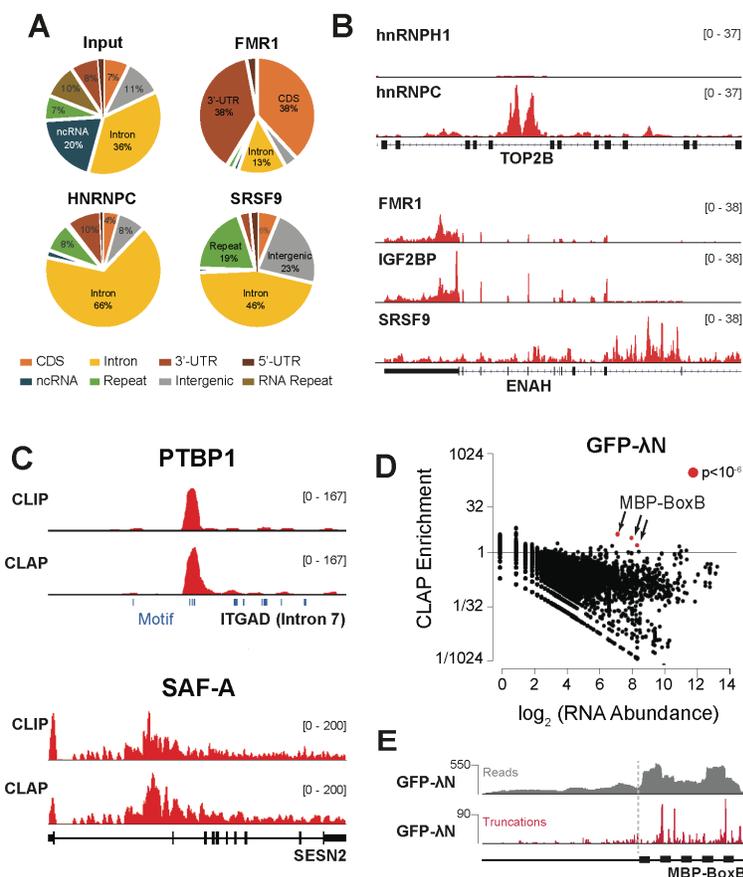


Figure 3. CLAP accurately identifies RNA-protein interactions that occur *in vivo* for well-defined RNA binding proteins. (A) Read distribution profiles are shown for three distinct RNA binding proteins (FMR1, hnRNP C, and SRSF9) as well as the distribution of features in input. (B) Examples of CLAP enrichments are plotted for hnRNPH1 and hnRNPC over an intronic region of TOP2B gene (top). FMR1, IGF2BP1, and SRSF9 CLAP enrichment profiles over the entire length of ENAH gene are shown. (bottom) Exons are denoted by boxes and introns by connecting lines in the schematic. (C) Examples of CLIP (top) and CLAP (bottom) enrichments are plotted for PTBP1 over intronic region of ITGAD (+tag) human pre-mRNA. The locations of the corresponding PTBP1 recognition motif (blue boxes) are shown. Examples of CLIP (top) and

CLAP (bottom) enrichments are plotted for SAF-A over pre-mRNA of SESN2 (+tag). Exons are denoted by boxes and introns by connecting lines in the schematic. **(D)** A scatter plot for cells co-expressing GFP- λ N and MBP-BoxB RNA. All 100 nucleotide windows from the GFP- λ N CLAP experiment show that most windows are depleted, except for a few significant windows (red) that correspond to a co-expressed MBP-BoxB RNA control for which λ N has known binding affinity. **(E)** Distribution of reads (gray) and crosslinked induced truncations (red) are plotted for CLAP captures of GFP- λ N in cells expressing MBP-BoxB.

Figure 4

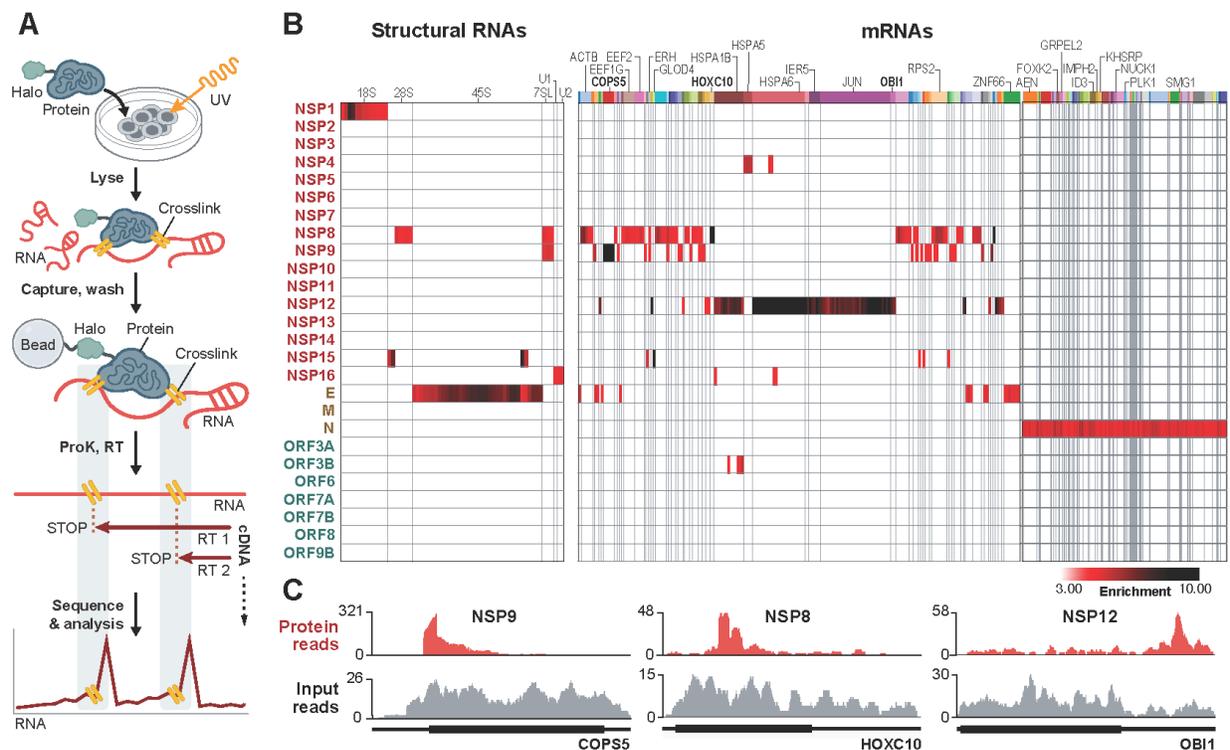


Figure 4. Global RNA binding maps of SARS-CoV-2 proteins. (A) Schematic of our approach. **(B)** Enrichment heatmap of each SARS-CoV-2 protein (rows) by significantly enriched 100 nucleotide RNA bins (columns, p -value <0.001 and enrichment > 3 -fold, Methods). Shared colored bars indicate multiple bins within the same mRNA. For spacing reasons, the 82 mRNAs bound by N-protein are displayed separately. **(C)** Examples of sequencing reads over specific mRNAs for viral proteins (red) relative to input RNA coverage (gray) are shown. Coding regions (thick lines) and untranslated regions (thin lines) are shown for each mRNA. See also Table S1.

Figure 5

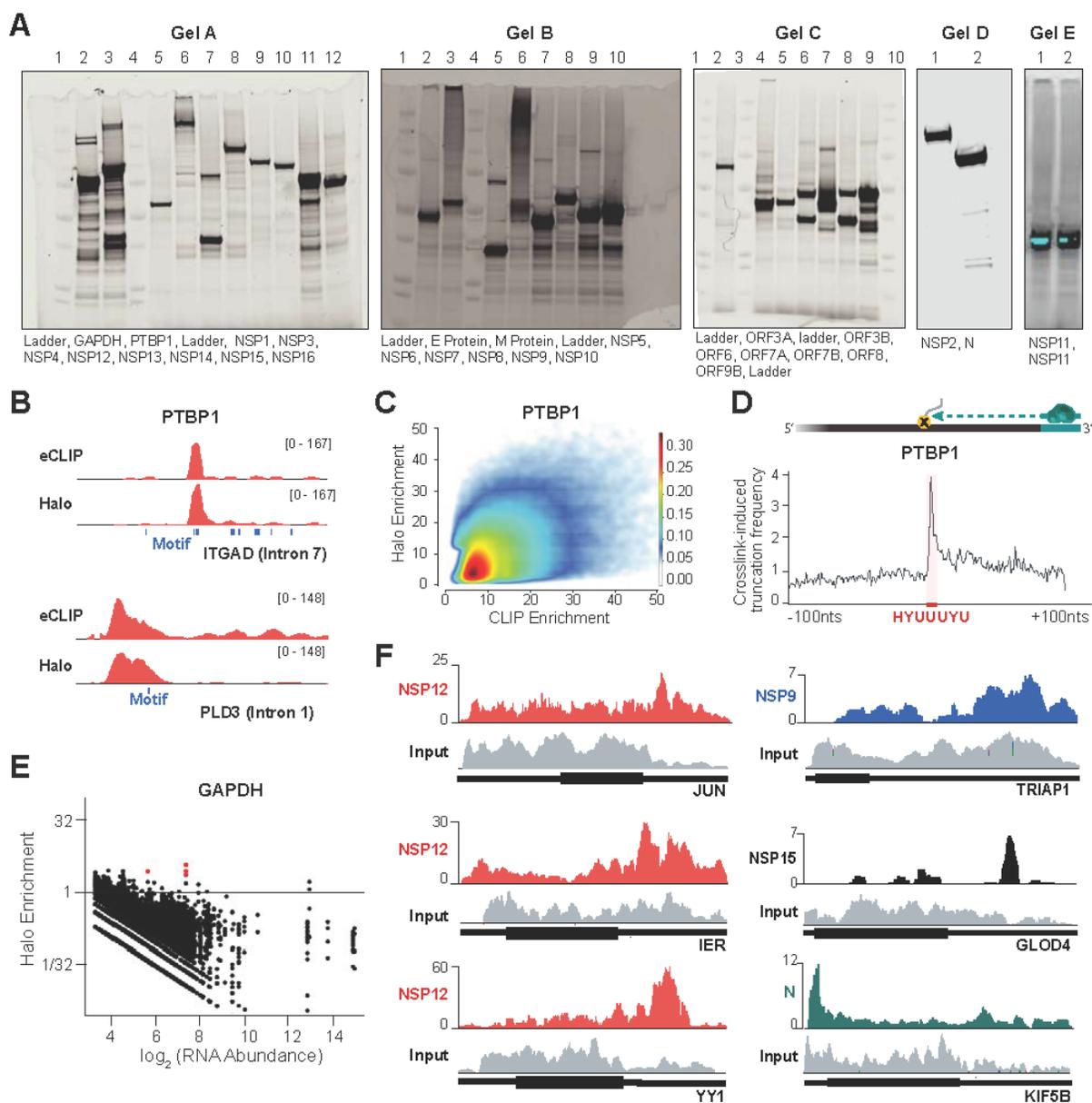


Figure 5. Global RNA binding maps of SARS-CoV-2 proteins, Supplementary Material Related to Figure 4. (A) Protein expression gels of Halo-tagged SARS-CoV-2 proteins. Expression is visualized via AlexaFluor-660 conjugated Halo-ligand. (B) Example of eCLIP

(top) and Halo (bottom) enrichments are plotted for PTBP1 over intronic regions of ITGAD mRNA. The location of the corresponding PTBP1 recognition motif (blue boxes) are shown. **(C)** Density scatter plot of the enrichment levels of PTBP1 over all human RNA regions as measured by eCLIP (x-axis) compared to the enrichment levels as measured by Halo (y-axis) for all RNAs identified as significantly enriched by eCLIP. **(D)** Cartoon illustrating protein-adduct mediated reverse transcriptase read stops at binding motifs (top). PTBP1 crosslink-induced truncation frequency relative to known PTBP1 motif (HYUUUYU, shown in red). **(E)** Scatter plot of RNA abundance (log scale, x-axis) compared to Halo enrichment (log scale, y-axis) for the GAPDH protein across all 100-nucleotide windows of all annotated human RNAs (exon and introns) are plotted. Windows with significant enrichment are shown in red. **(F)** Representative tracks illustrating different mRNA binding patterns in Halo captures of NSP12 (red), NPS9 (blue), NSP15 (black), and N-protein (blue). Input tracks are presented for each mRNA (gray).

Figure 6

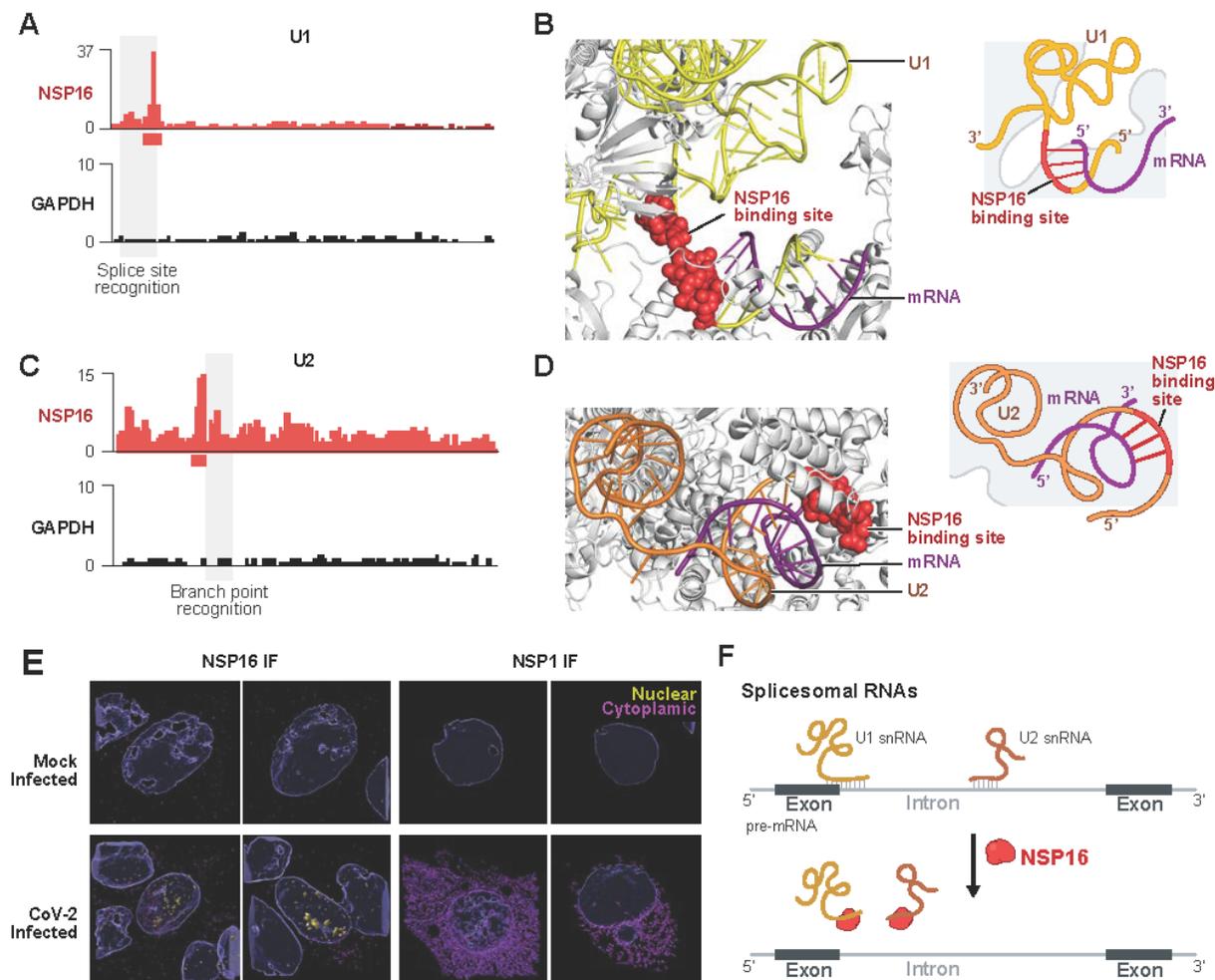


Figure 6. NSP16 binds to U1 and U2 at their mRNA recognition sites. (A) NSP16 enrichment of reverse transcription stop positions across each nucleotide of U1 (red) compared to a control protein (GAPDH - black). Red box (below the x-axis) represents most enriched nucleotide positions (U1:9-13 nts). Gray shaded box (overlay) outlines the position of the splice site recognition sequence. (B) **Left:** Structure of the pre-catalytic human spliceosome (PDB:6QX9)¹²⁶ highlighting the location of NSP16 binding site (red spheres) relative to U1 (yellow ribbon) and mRNA (purple ribbon). **Right:** Schematic of the structure. (C) Enrichment across each nucleotide of U2 for NSP16 (red) and GAPDH (black). Red box demarcates most

enriched nucleotide positions (U2:27-34 nts). Gray shaded box outlines the location of the branchpoint recognition sequence. **(D)** Structure of the pre-catalytic human spliceosome (PDB:6QX9)¹²⁶ displaying NSP16 binding site (red spheres), U2 (orange), and mRNA (purple). **(E)** Mock (top) or SARS-CoV-2 infected (bottom) Vero E6 cells immunostained with a polyclonal antibody to NSP16 (left) or NSP1 (right). Imaris 3D reconstruction of DAPI (nucleus) and NSP16 or NSP1 signal are shown for each protein. Signal contained within the 3D nuclear volume (blue) is shown in yellow and cytoplasmic signal in purple. Size bars are 3 μ m. **(F)** **Model:** NSP16 binding to U1/U2 can impact mRNA recognition during splicing.

Figure 7

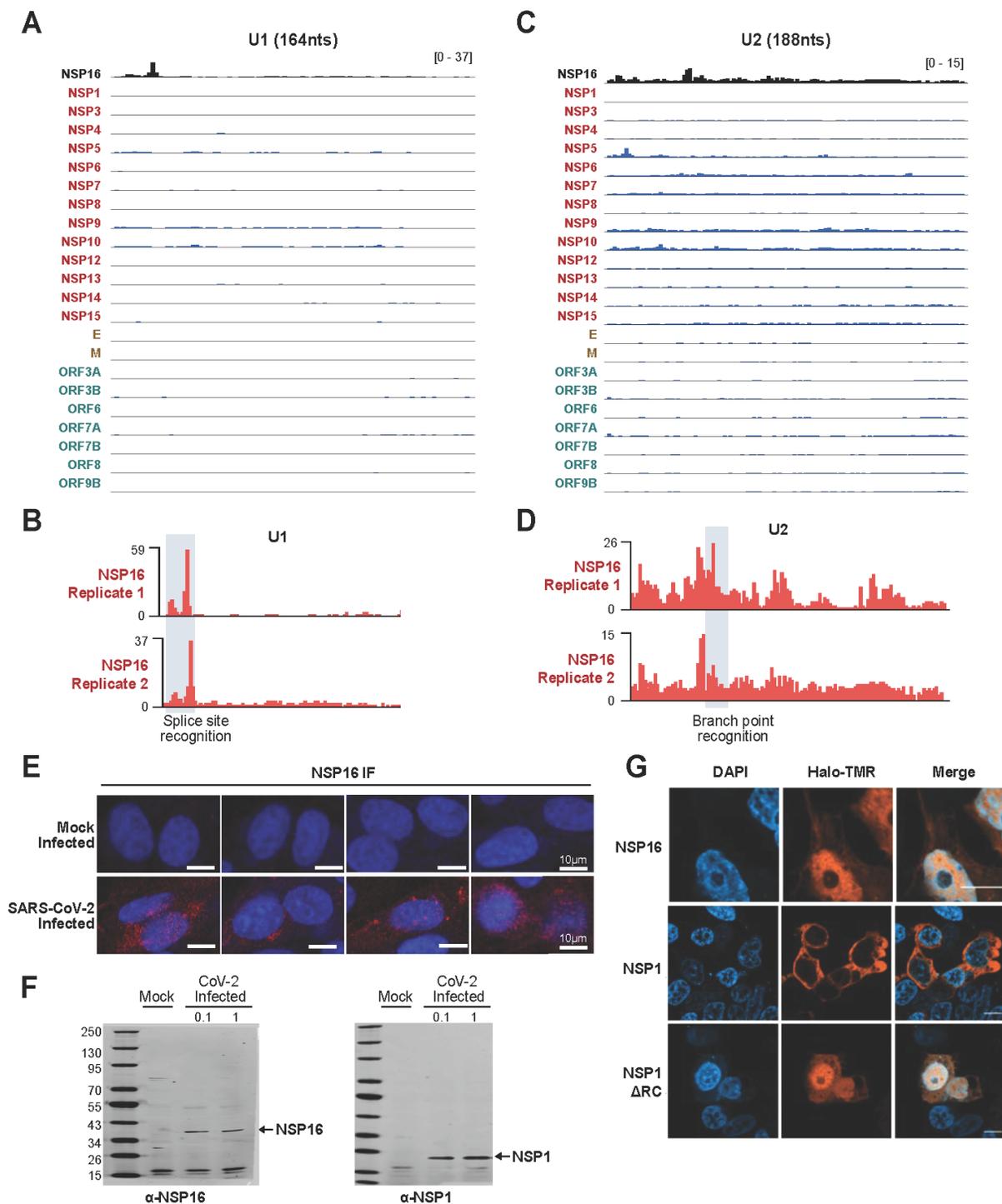


Figure 7. NSP16 binds to the U1 and U2 components of the spliceosome at their mRNA recognition sites, Supplemental Material Related to Figure 6. (A) Comparison of U1 RNA

enrichment across SARS-CoV-2 Halo capture datasets. **(B)** NSP16 binding traces along U1 RNA between two separate captures. Splice site recognition domain is highlighted in gray. **(C)** Comparison of U2 RNA enrichment across SARS-CoV-2 Halo capture datasets. **(D)** NSP16 binding traces along U2 RNA between two separate captures. Branch point recognition domain is highlighted in gray. **(E)** NSP16 immunofluorescence in Vero E6 cells infected (or mock infected) with SARS-CoV-2 at an MOI of 0.1 for 48h. Four representative fields are displayed, with size bar indicating 10 microns. **(F)** Western blot confirmation of NSP16 and NSP1 antibodies used to generate images in **(E)**. Vero cells were infected (or mock infected) with SARS-CoV-2 at an MOI of 0.1 or 1; 72 hpi cells were lysed and probed by western blot with antibodies raised against NSP1 or NSP16. **(G)** Imaging of HEK 293T cells transfected with Halo-tagged NSP16, NSP1, and NSP1 Delta RC plasmids. Proteins are visualized using TMR-conjugated Halo-ligand (orange) and counter-stained with DAPI (blue). Scale bars indicate 10 microns.

Figure 8

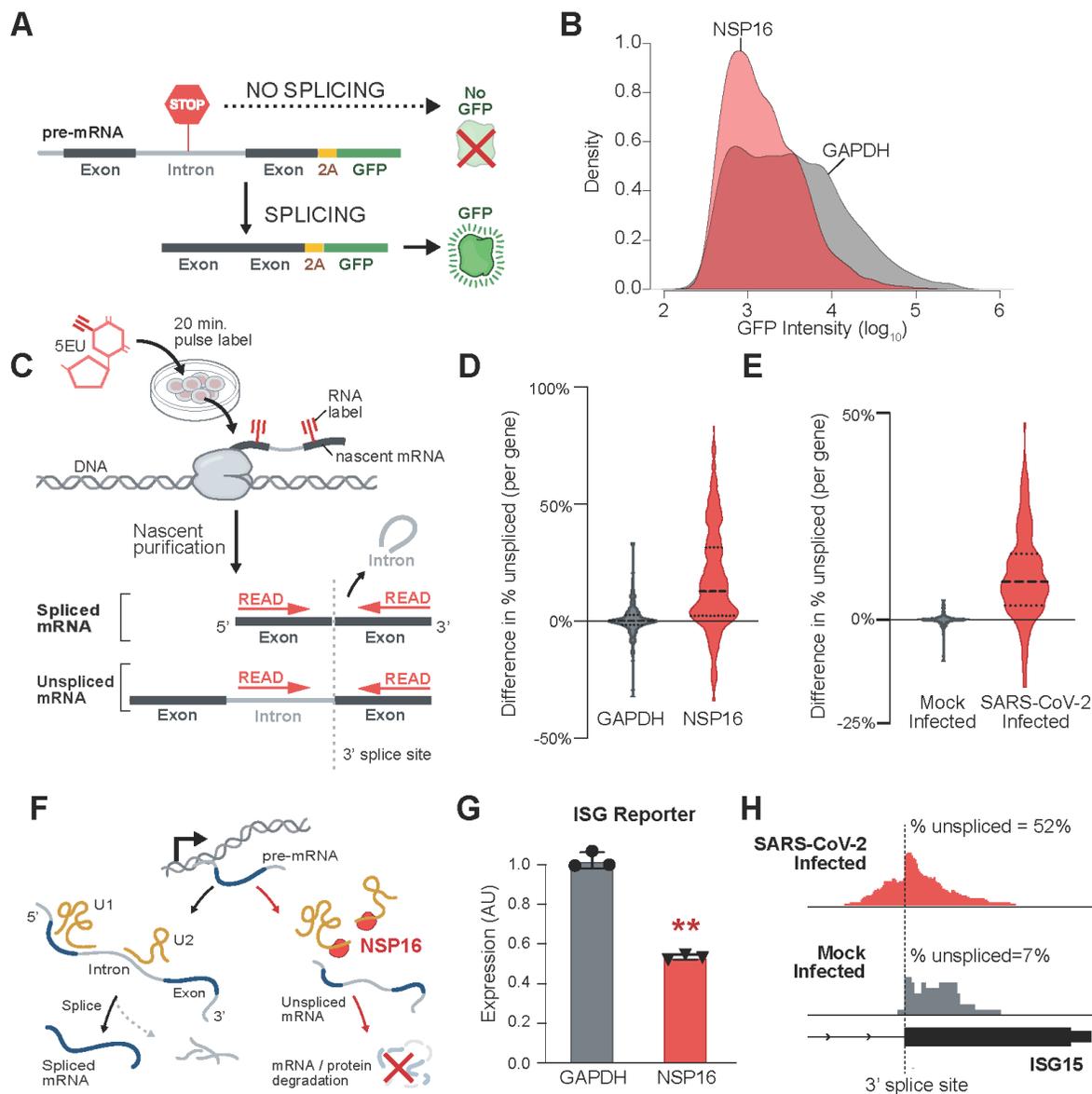


Figure 8. NSP16 suppresses host mRNA splicing. (A) Schematic of fluorescence reporter used to assay mRNA splicing. (B) GFP density plot of HEK293T cells expressing the GFP splicing reporter and either GAPDH (gray) or NSP16 (red). (C) Schematic of nascent RNA purification method. (D) The % unspliced difference for each gene between HEK293T cells transfected with GAPDH (gray) or NSP16 (red). Plot represents the merge of four independent biological

replicates; replicates plotted in Figure S4C. **(E)** Violin plot for SARS-CoV-2 infected human lung epithelial cells (MOI=0.01, 48 hours) compared to mock. Plot are merge of two biological replicates; replicates plotted in Figure S4E. **(F) Model:** NSP16 binding to U1 and U2 can reduce overall mRNA and protein levels. **(G)** Expression of an interferon stimulated gene (ISG) reporter upon transfection with GAPDH (gray) or NSP16 (red) after stimulation with IFN- β . Three independent biological replicates, ** indicates p-value<0.01. **(H)** Example of nascent RNA sequencing at the intron of ISG15 (intron = line, exon = box) upon SARS-CoV-2 (red) or mock (gray) infection.

Figure 9

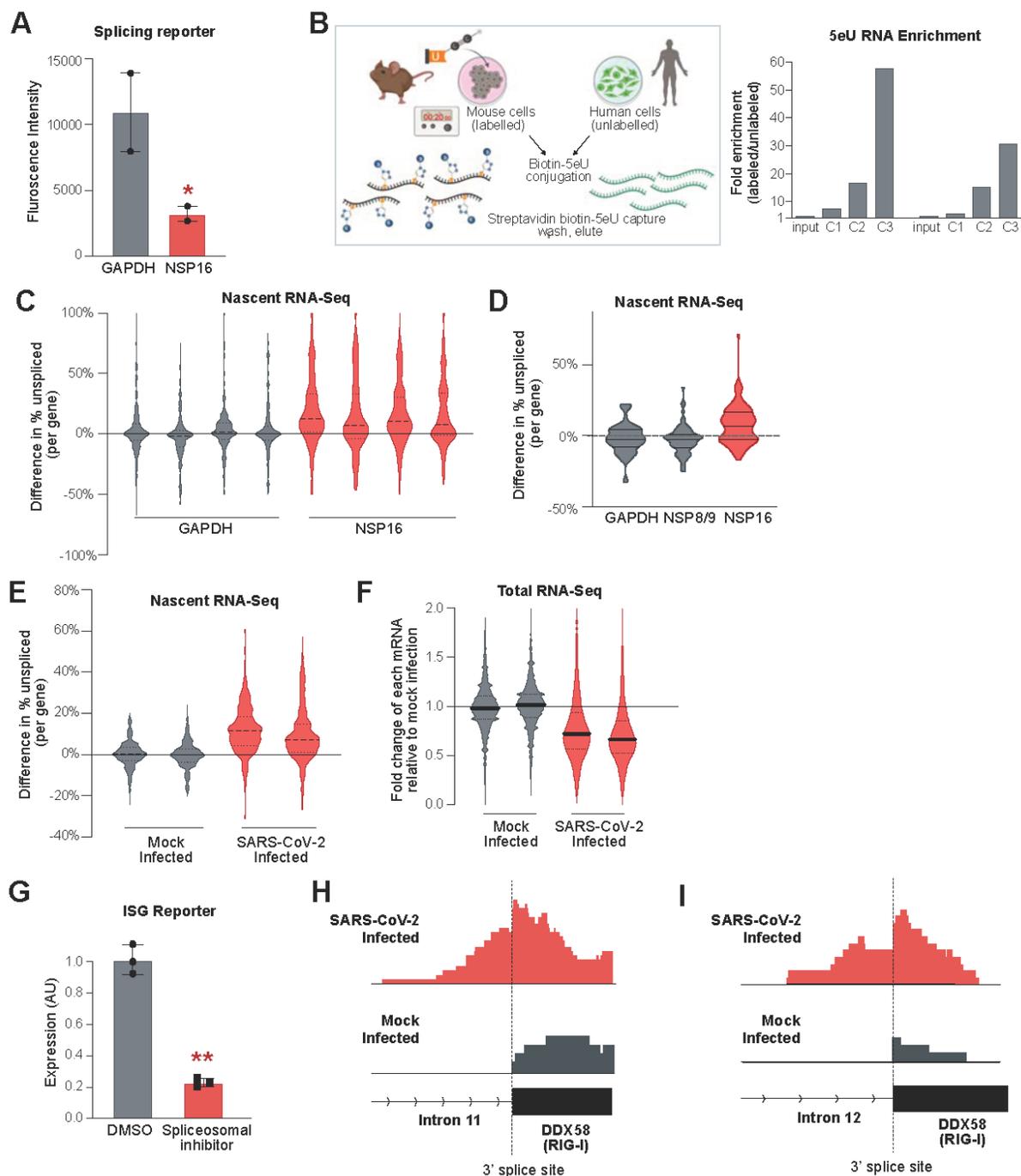


Figure 9. NSP16 suppresses host mRNA splicing, Related to Figure 3. (A) Median of raw GFP fluorescence measured in splicing reporter assay performed in HEK293T cells expressing

either Halo-GAPDH (gray) or Halo-NSP16 (red). Two independent biological replicates per condition. **(B)** Overview of nascent RNA-sequencing method, including 5eU nucleotide feeding, biotin click chemistry conjugation, and biotin-streptavidin-based iterative capture methods. Human/mouse mixing experimental data illustrates selective enrichment of labeled material over unlabeled material after three sequential captures. (C1= capture 1 enrichment, C2= capture 2 enrichment, etc.) **(C)** Violin plot depicting difference in unspliced reads per gene (defined as the difference between number of unspliced fragment divided by total fragments spanning the 3' splice site between condition and median of all control samples) for HEK293T cells transfected with either GAPDH (gray) or NSP16 (red) for 48hrs. All four individual replicates are presented. **(D)** Violin plot depicting difference in unspliced reads per gene (relative to median of GAPDH) for HEK 293T cells transfected with either GAPDH, NSP9, or NSP16 (red) for 48hrs. **(E)** Violin plot depicting difference in unspliced reads per gene (relative to median of the mock condition) for Calu3 cells infected with SARS-CoV-2 virus at an MOI of 0.01 for 48 hrs (red) or uninfected (gray). Biological replicates are presented. **(F)** Violin plot depicting fold change in total steady-state mRNA levels (mRNA initially normalized to ncRNA and ratio is fold normalized mock treatment) for SARS-CoV-2 infected (red) compared to mock infected (gray) samples. Data is presented for two biological replicates for each condition. **(G)** Normalized expression of an interferon signaling reporter upon stimulation with IFN- β and treatment with madrasin spliceosomal inhibitor (red) or DMSO vehicle (gray). Three independent biological replicates were measured for each condition. **(H-I)** Representative nascent RNA tracks from SARS-Cov-2 infected (red) and mock-treated cells (gray) along Intron 11 and Intron 12 of interferon stimulated gene, RIG-I.

Figure 10

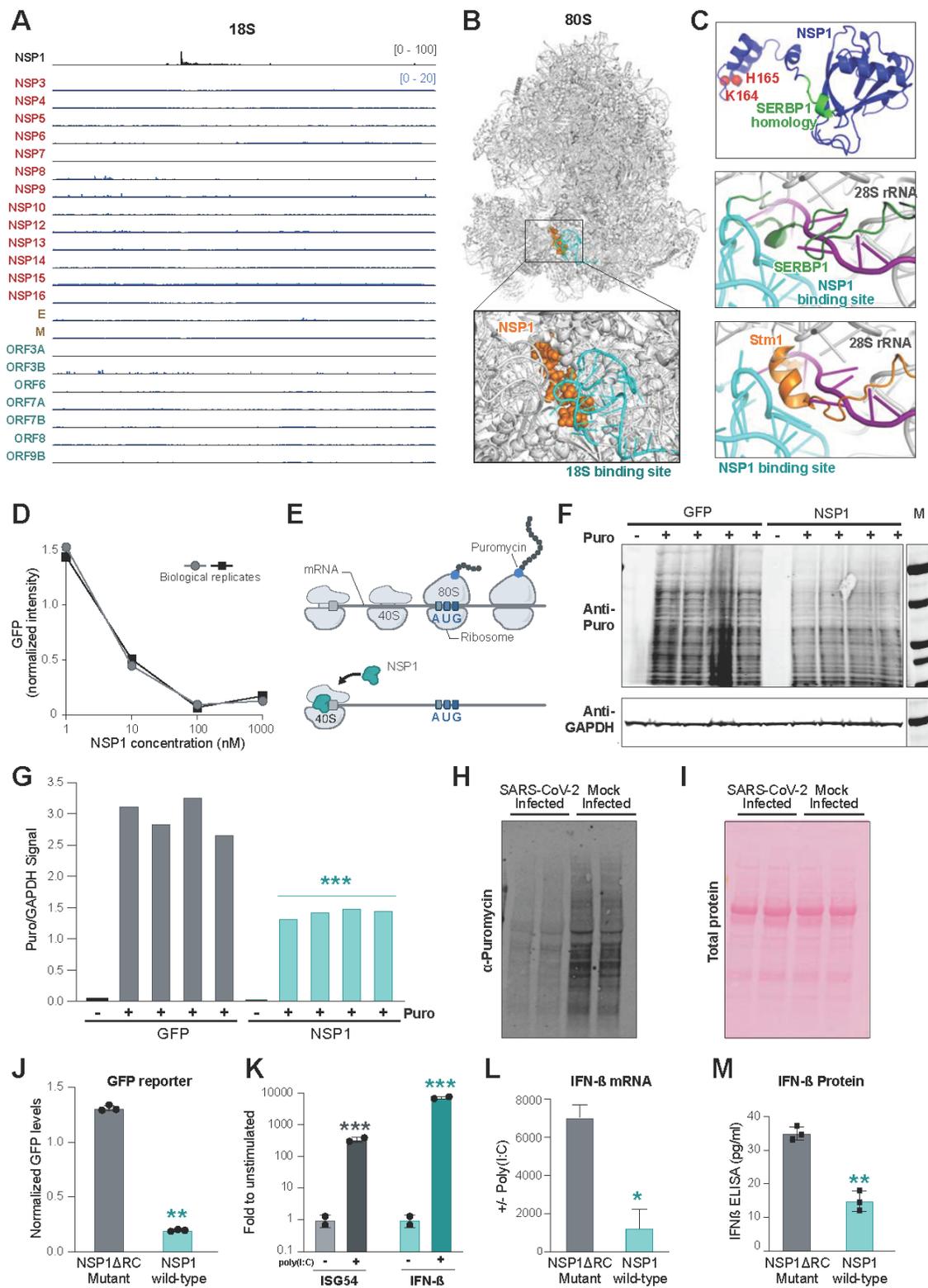


Figure 10. NSP1 binds to the 18S ribosomal RNA near the mRNA entry channel to suppress global mRNA translation in cells, Supplemental Material Related to Figure 11.

(A) Comparison of 18S RNA enrichment across SARS-CoV-2 Halo capture datasets. **(B)** The location of NSP1 binding (orange spheres) relative to 18S binding site (cyan) upon known structure of the 80S ribosome (gray). **(C)** Predicted structure of NSP1 based on Robetta modeling. The critical C-terminal amino acids required for binding 18S (K164 and H165) are indicated as red spheres. The region of homology with SERBP1 is shown in green. The observed NSP1 binding sites on the 18S rRNA are demarcated in cyan on the structure of the human 40S ribosome (PDB: 6G5H; gray)¹²⁷, relative to the mRNA path (purple; 6YAL)⁷² and known clogging factors **(E)** SERBP1 (green; 6MTE)⁷⁴ and **(F)** Stm1 (orange; 4V88)⁷⁵. **(D)** An mRNA encoding GFP was added to HeLa cell extracts along with different concentrations of purified NSP1 protein (x-axis). The amount of GFP protein measured relative to the median of replicates for a buffer only control is shown (y-axis). Two independent dose titrations were performed and are shown on top of each other. **(E)** Schematic illustrating puromycin tagging of newly translated proteins via the SuNSET method. If the level of ongoing translation is high, we expect to detect a large amount of newly generated proteins containing puromycin; if global translation is suppressed, we expect to observe a decrease in the amount of puromycin integrated into proteins. **(F)** Western blot of global puromycin incorporation into proteins of HEK293T cells transfected with either Halo-GFP (**left**) or Halo-NSP1 (**right**). GAPDH levels were measured in the same lysates to normalize for total protein in the sample (bottom). (-) puro lanes indicate transfected samples that were not treated with puromycin. **(G)** Quantification of puromycin intensity across each lane of the gel in Panel F. The ratio of puromycin signal over total GAPDH signal is plotted for individual replicates. **(H)** Vero E6 cells were infected (or mock infected) with SARS-CoV-2

at an MOI of 0.01. 48hpi cells were labelled with media containing puromycin, and lysates were probed by western blot. **(I)** As a control for total protein levels, after samples in **(G)** were run on a SDS-PAGE gel, transferred to nitrocellulose, and total proteins were stained with PONCEAU before blocking/antibody detection of puromycin signal. **(J)** Normalized GFP fluorescence intensity of GFP reporter co-transfected in HEK293T cells in the presence of the NSP1DRC mutant that does not bind to 18S (gray) or NSP1 (cyan) proteins. Three independent biological replicates were measured for each sample.

Note: This experiment was performed alongside the various controls displayed in Figure 4D and are plotted on the same scale. **(K)** mRNA levels of ISG54 and IFN- β following stimulation with poly(I:C) normalized to levels in unstimulated A549 cells. **(L and M)** mRNA and protein levels of IFN- β following stimulation with poly(I:C) normalized to levels in unstimulated A549 cells transfected with NSP1 Delta RC mutant (gray) or NSP1 (cyan). Two independent biological replicates were measured for each condition. In all panels, error bars represent standard deviation across replicates, and dots represent individual values for each replicate. * indicates $p < 0.05$, ** $p < 0.01$, and *** $p < 0.001$.

Figure 11

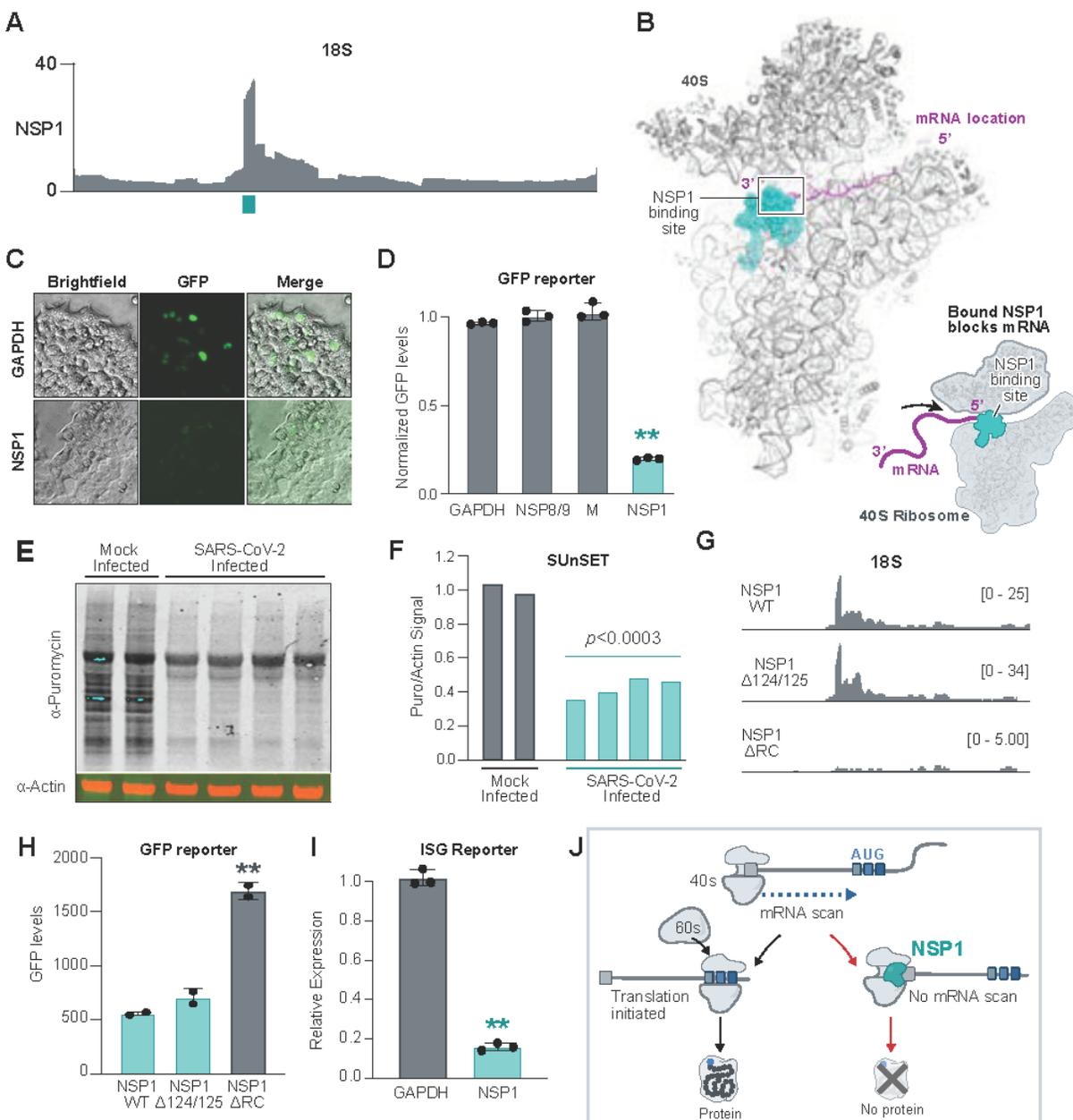


Figure 11. NSP1 binds to 18S near the mRNA entry channel to suppress translation. (A)

NSP1 enrichment across each nucleotide of 18S. Cyan box indicates the most enriched nucleotides of NSP1 binding (18S:607-644 nts). **(B)** The location of NSP1 binding (cyan spheres) relative to the known structure of 40S (gray) and mRNA (purple ribbon). **Right:**

Schematic illustrating structure¹²⁷ and how NSP1 binding would block mRNA entry. **(C)** Images of HEK293T cells co-expressing GFP reporter and GAPDH (top) or NSP1 (bottom). **(D)** Flow cytometry quantification (mean intensity) of GFP in the presence of GAPDH, NSP8/9, M, or NSP1 proteins. Three independent biological replicates per condition. **(E)** Puromycin incorporation (top) or total actin levels (bottom) measured in Calu3 cells infected with SARS-CoV-2 (MOI=0.01, 48h) or mock control (left 2 lanes). **(F)** The ratio of puromycin signal over total actin signal is plotted for each individual replicate. **(G)** Read enrichment on 18S for an independent replicate of NSP1 Wild Type, NSP1 R124A/K125A Mutant, and NSP1 K164A/H165A (Delta RC) Mutant. **(H)** Flow cytometry analysis of HEK293T cells transfected with GFP and NSP1 Delta RC mutant (gray), wild-type NSP1, or NSP1 R124A/K125A (cyan). **(I)** Quantification of IFN- β response in the presence of GAPDH (gray) or NSP1 (cyan). **(J)** Schematic of how NSP1 acts to suppress mRNA translation. In all panels, error bars represent standard deviation across biological replicates, dots represent individual values for each replicate, * $p < 0.05$ and ** $p < 0.01$.

Figure 12

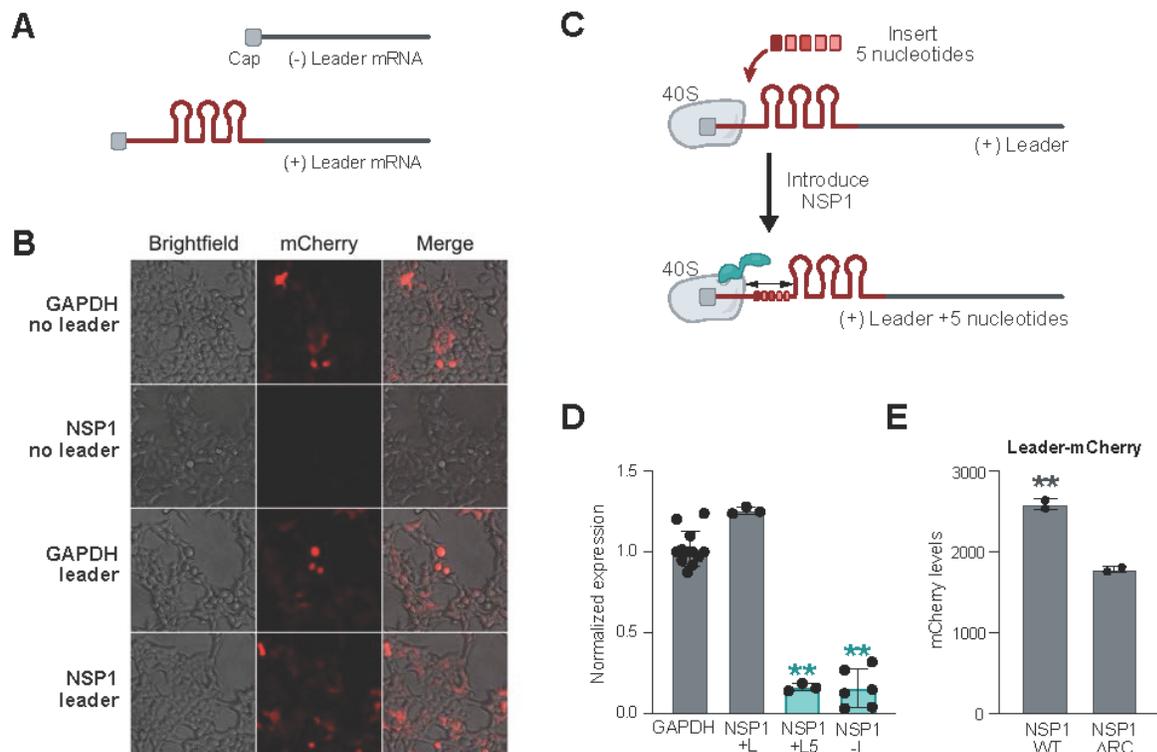


Figure 12. The 5' viral leader sequence protects mRNAs from NSP1-mediated translational inhibition, Supplemental Material Related to Figure 13. (A) A schematic of the experimental design containing two reporter RNAs encoding fluorescent proteins, without the viral leader (**top**) and with the viral leader sequence appended to the 5' end of the mRNA (**bottom**). Viral leader represented by three stem-loops in red. (B) Representative images of HEK 293T cells co-transfected with GAPDH or NSP1 along with mCherry RNA with or without SARS-CoV-2 leader sequence. (C) Schematic illustrating the insertion of 5 nucleotides between the 5' cap and the viral leader sequence. NSP1 protein represented in red. (D) Quantification of mCherry expression in HEK 293T cells transfected with mCherry RNAs, fused to different 5' leader variants, and either GAPDH or NSP1. Values are normalized to the median values of mCherry

levels from control condition (GAPDH with + mCherry). At least 3 independent biological replicates per condition. Dots represent value for each independent replicate (e.g. NSP1 -L contains 6 independent replicates). **(E)** Quantification of mCherry expression from HEK 293T cells transfected with Halo-tagged NSP1 WT or NSP1 Delta RC mutant, along with leader-mCherry expressing plasmids. Values are normalized to the median values of mCherry levels in control sample (NSP1 with + leader-mCherry). Two independent biological replicates were measured per condition. In all panels, error bars represent standard deviation across replicates, and dots represent individual values for each replicate. * indicates $p < 0.05$ and ** $p < 0.01$.

Figure 13

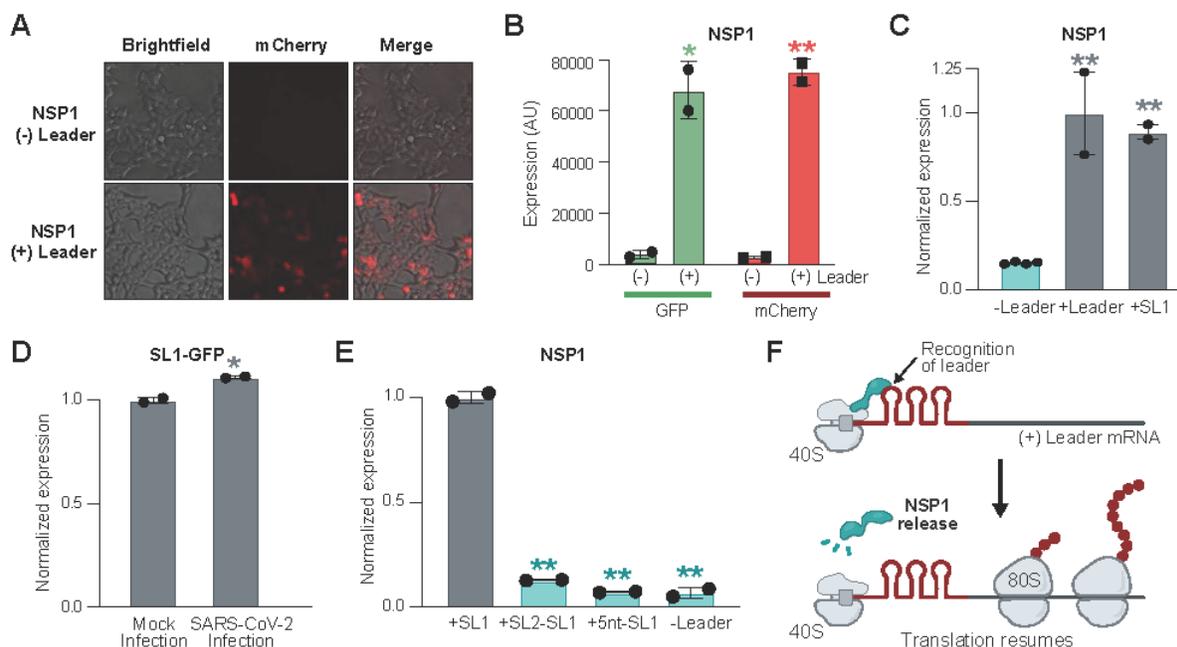


Figure 13. The 5' viral leader protects mRNA from NSP1-mediated translational inhibition. (A) Images of cells co-transfected with NSP1 and mCherry alone (- leader, top) or mCherry fused to the SARS-CoV-2 leader (+ leader, bottom). (B) GFP (green) or mCherry (red) levels when fused to the viral leader (+leader, right) or lacking viral leader (-leader, left). (C) GFP reporter with no leader (left), full leader (middle), or stem loop 1 (SL1) upon NSP1 expression. (D) Calu3 cells expressing SL1 fused to GFP. Cells were mock or SARS-CoV-2 infected (MOI=0.1) and GFP expression was measured 24 hours post infection by flow cytometry. (E) GFP reporter containing SL1 (left), a swap of SL2 and SL1 (SL2-SL1), insertion of 5 nucleotides between the 5' end and SL1 (+5nt-SL1), or no leader. GFP protein level was measured for each condition upon expression of NSP1. (F) Proposed model for how NSP1 binding to the viral leader can allosterically modulate NSP1 structure to protect mRNAs in cis. In

all panels, error bars represent standard deviation across biological replicates, dots represent individual replicate values, * indicates $p < 0.05$ and ** $p < 0.01$.

Figure 14

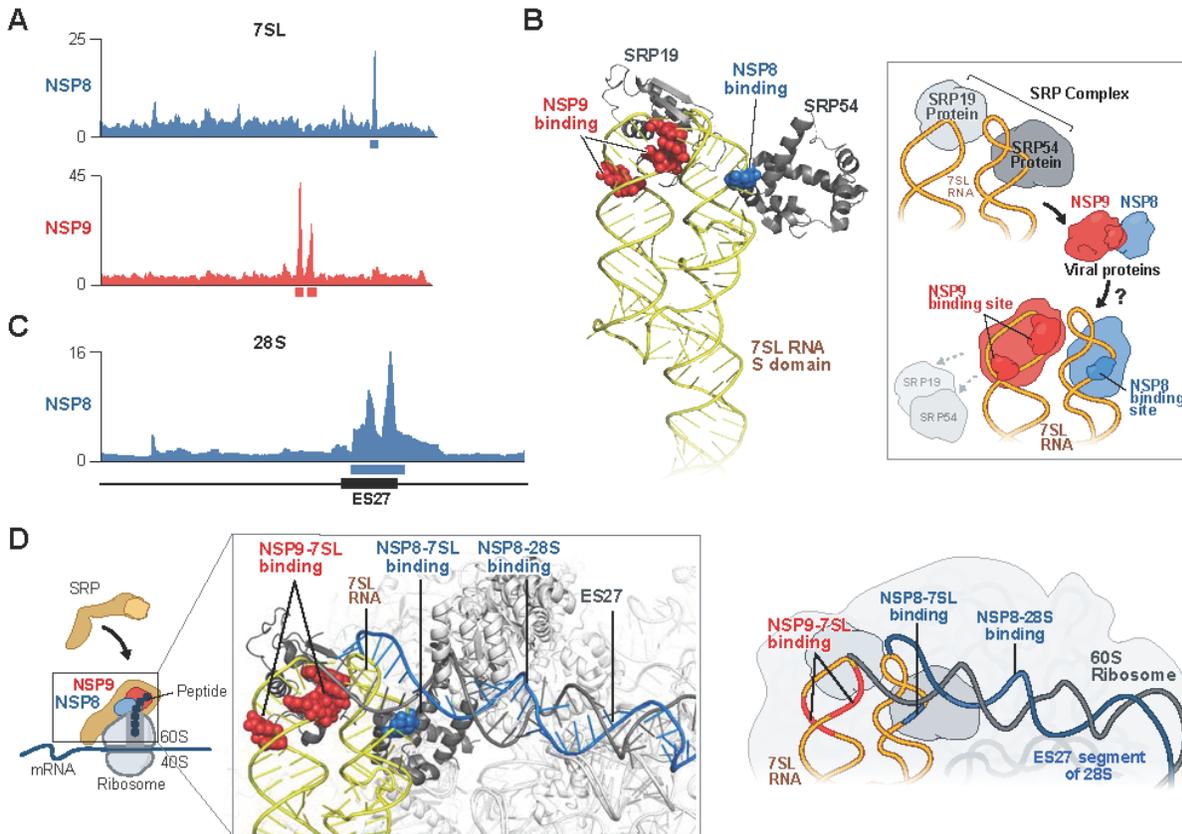


Figure 14. NSP8 and NSP9 bind to 7SL RNA of the Signal Recognition Particle. (A) Enrichment of reverse transcription stop positions across each nucleotide of 7SL is shown for NSP8 (blue) and NSP9 (red). Red (7SL:142-143 and 7SL:149-151 nts) and blue (7SL:193-194 nts) boxes demarcate the most enriched nucleotide positions. **(B)** The locations of the NSP8 (blue spheres) and NSP9 (red spheres) binding sites on the S domain of 7SL (yellow ribbon) structure relative to SRP54 and SRP19 (gray) (PDB:1MFQ)¹²⁸. Right: Schematic of the structure and model of how NSP8/9 binding to 7SL could impact SRP protein binding. **(C)** Read enrichment across each nucleotide of 28S for NSP8 (blue) is shown. Black box indicates the

location of the ES27 expansion sequence (28S:2889-3551 nts). Blue box indicates the most enriched nucleotide position on 28S rRNA (28S:3017-3529 nts). **(D)** The locations of the NSP8 (blue) and NSP9 (red) binding sites relative to the structure of SRP ribosome complex (PDB:3JAJ)¹²⁹ superimposed with the structure of the ES27 region of 28S (Ebp1-ribosome complex; PDB:6SXO)⁸⁷. The observed NSP8 binding site within the ES27 region of 28S (gray) is demarcated in blue, and the NSP8 (blue) and NSP9 (red) binding sites on 7SL (yellow) are each highlighted. **Right:** Schematic illustrating the interaction between the ribosome and SRP.

Figure 15

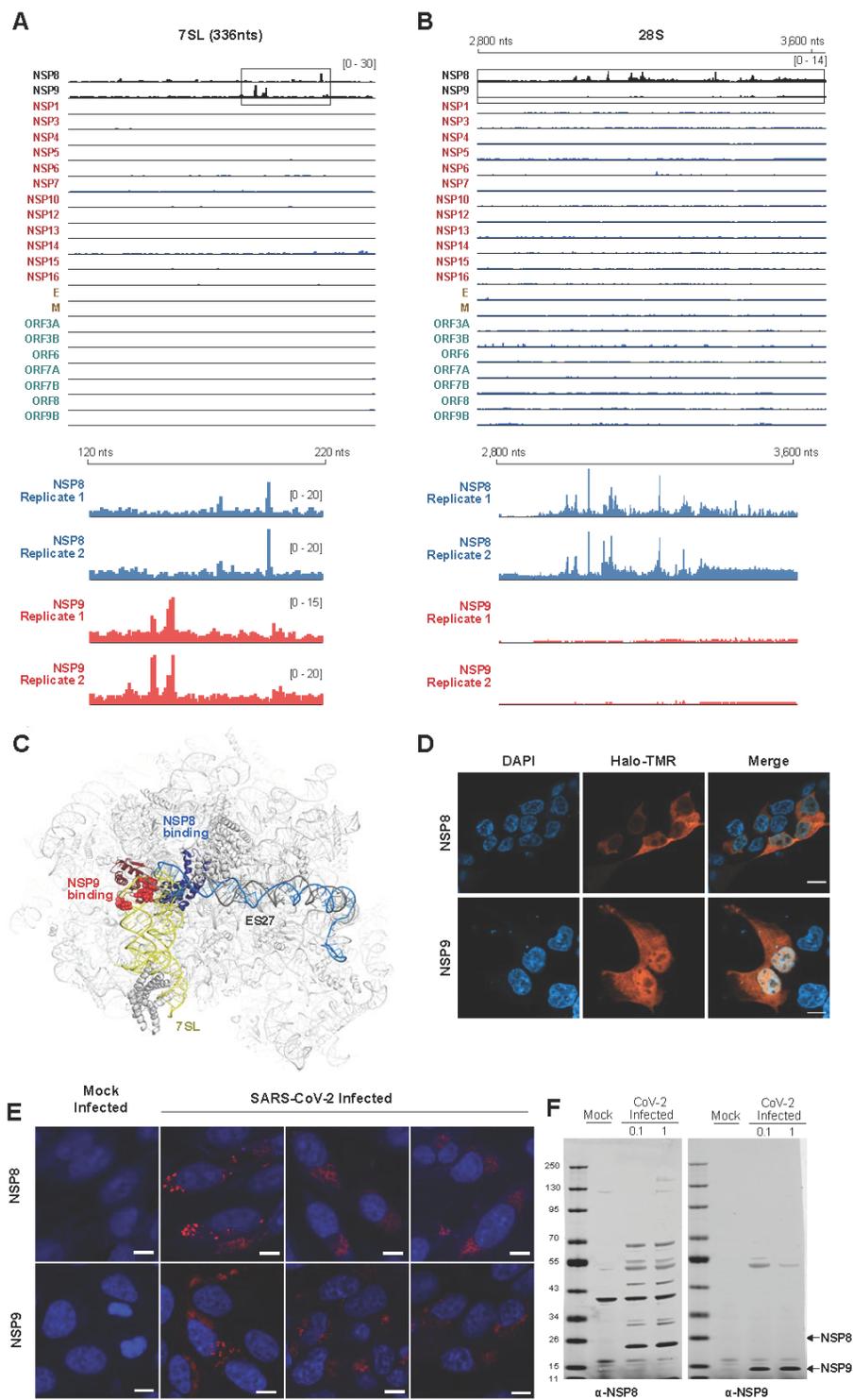


Figure 15. NSP8 and NSP9 bind to the 7SL RNA component of the Signal Recognition

Particle, Supplemental Material Related to Figure 14. (A) Comparison of 7SL RNA second read enrichment across viral protein capture datasets (top) with region of highest enrichment for NSP8/9 boxed. Independent expression, purification, and sequencing experiments for NSP8 and NSP9 were performed and are shown. **(B)** Comparison of 28S RNA enrichment across SARS-CoV-2 Halo capture datasets (top). Replicate representative tracks of NSP8 (blue) and NSP9 (red) on 28S rRNA are presented below. **(C)** Full view of 80S ribosome structure, interfaced with SRP (7SL RNA, yellow line), NSP9 binding sites on 7SL (red circles), and NSP8 binding sites on 7SL (dark blue circles) and on ES27 expansion segment on the 28S ribosomal RNA (light blue line). **(D)** Imaging of HEK 293T cells transfected with Halo-NSP8 or Halo-NSP9 plasmids. Proteins are visualized using TMR-conjugated Halo-ligand (orange) and counterstained with DAPI (blue) nuclear stain. Size bars indicate 10 microns. **(E)** Vero E6 cells were infected (or mock infected) with SARS-CoV-2 at an MOI of 0.1 for 48h, before fixing and staining with an antibody raised against NSP8 or NSP9. Cells are counter-stained with DAPI. Size bars indicate 10 microns. **(F)** Western blot confirmation of NSP8 and NSP9 antibodies used to generate images in (E). Vero cells were infected (or mock infected) with SARS-CoV-2 at an MOI of 0.1 or 1; 72 hpi cells were lysed and probed by western blot with antibodies raised against NSP8 or NSP9.

Figure 16

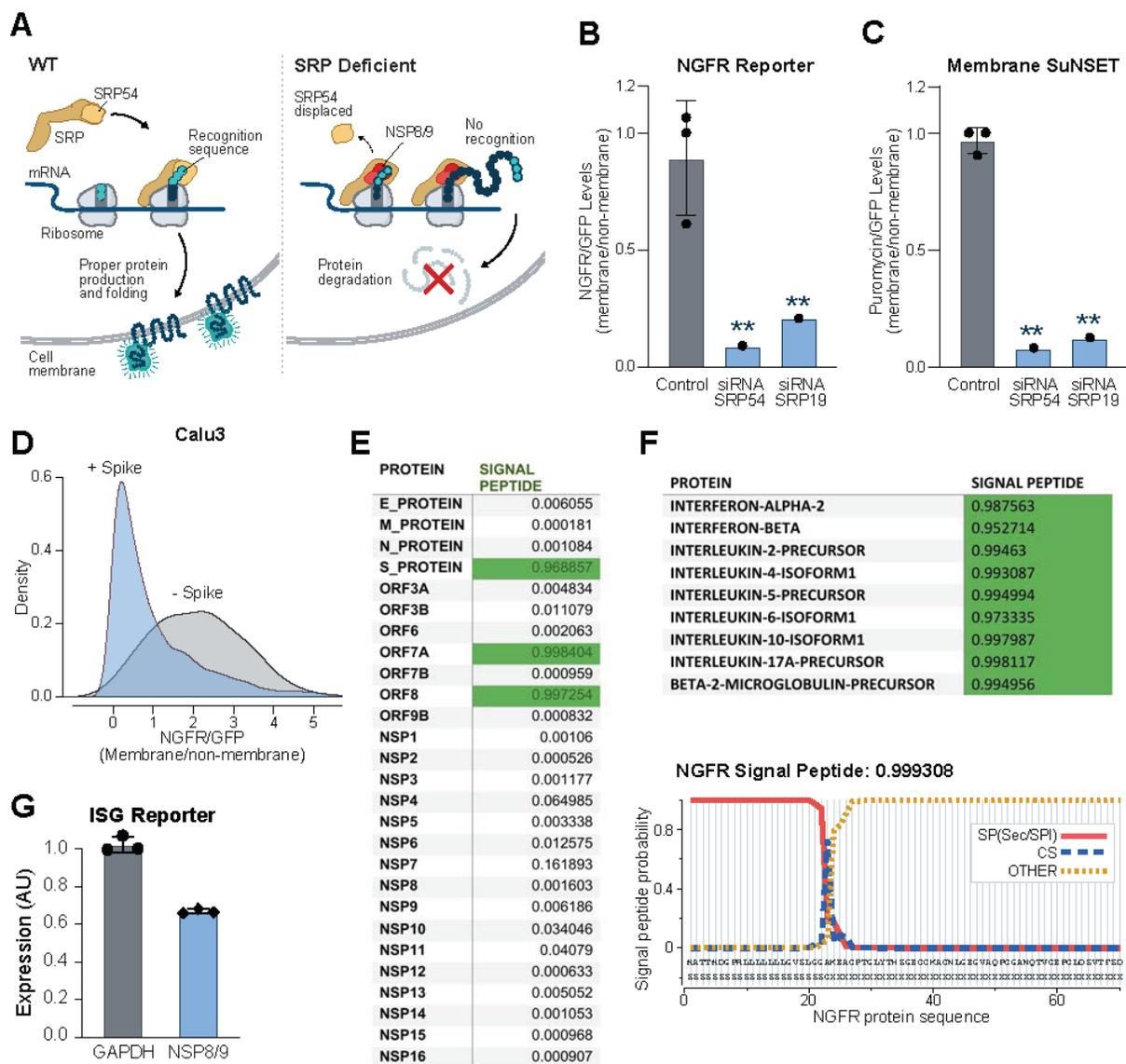


Figure 16. NSP8 and NSP9 inhibit membrane and secretory protein function,

Supplemental Material Related to Figure 17. (A) Schematic illustrating Signal Recognition Particle-mediated recognition and translocation of nascent membrane and secreted proteins (left). Upon SRP dysfunction, membrane and secreted proteins are predicted to be mis-localized and degraded (right). **(B)** Quantification of truncated Nerve Growth Factor Receptor (NGFR)

fluorescence normalized to eGFP fluorescence (NGFR:GFP) from HEK 293T cells transfected with control EED plasmid together with siRNAs targeting protein components of Signal Recognition Particle, SRP54 and SRP19. **(C)** Quantification of Membrane SuNSET puromycin staining fluorescence normalized to eGFP fluorescence (Puromycin:GFP) from HEK 293T cells transfected with control EED plasmid together with with siRNAs targeting protein components of Signal Recognition Particle, SRP54 and SRP19. Three independent replicates for control and one replicate for siRNA treatments within this experiment. **(D)** NGFR:GFP ratio from Calu3 human lung epithelial cells infected with SARS-CoV-2 for 24 hrs at an MOI of 0.1. Density comparison between Spike positive cells in virally infected condition to Spike negative cells in virally infected condition. **(E)** Signal P analysis of open reading frames of SARS-CoV-2 expressed proteins utilized in study. Proteins with greater than 0.95 predicted probability indicated Signal P algorithm are highlighted in green. **(F) Top:** Signal P analysis of open reading frames of various immunoregulatory cytokines and proteins, including Interferon Beta and Beta-2-Microglobulin- Precursor. **Bottom:** Signal P analysis of NGFR (membrane reporter) amino acid sequence and plot of signal peptide probability along the first 70 amino acids of NGFR sequence. In all panels, error bars represent standard deviation across replicates, and dots represent individual values for each replicate. * indicates $p < 0.05$ and ** $p < 0.01$. **(G)** Expression of an interferon stimulated gene reporter upon transfection with GAPDH or NSP8 and NSP9 (in combination), followed by stimulation with IFN- β . We note that because this assay measures intensity across a population of cells, any cells that are not transfected by NSP8/9 would not show this effect and would lead to a smaller overall difference than might occur within individual cells. In contrast, NGFR and SUNSET flow cytometry measurements (B-C) represent analysis of cells expressing NSP8/9.

Figure 17

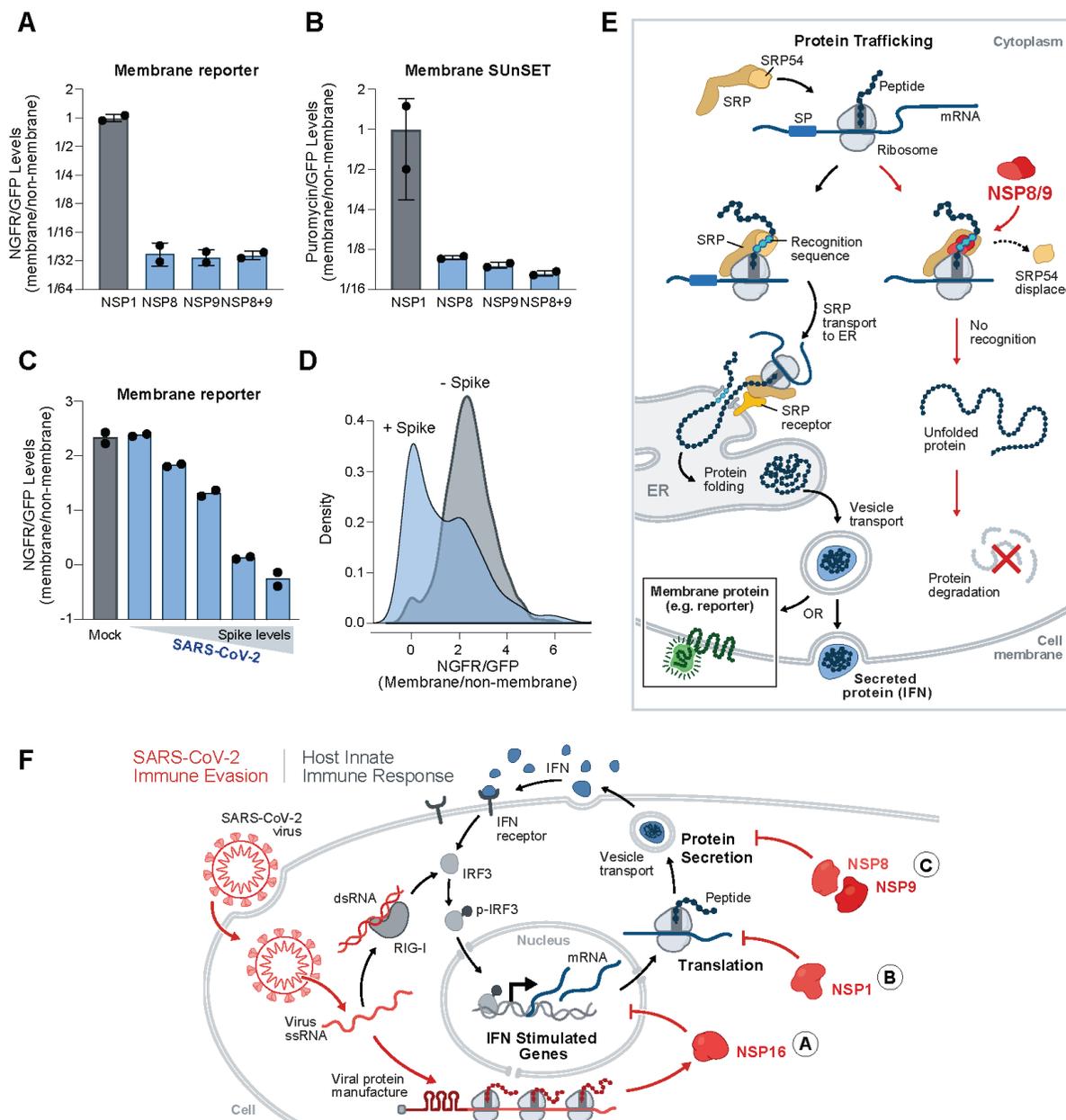


Figure 17. NSP8 and NSP9 inhibit membrane and secretory protein trafficking. (A)

Quantification of HEK293T cells transfected with plasmids co-expressing GFP-tagged NSPs and the NGFR membrane protein. Plotted is the ratio of NGFR to GFP levels for each condition. **(B)**

The ratio of puromycin-containing proteins at the cell membrane normalized to GFP expression for each condition. **(C)** Quantification of two mRNA reporters containing SL1 fused to either GFP (leader-GFP) or NGFR (leader-NGFR) in Vero cells infected with SARS-CoV-2 or mock for 24 hrs (MOI 0.1). Plotted is ratio of leader-NGFR to leader-GFP, binned by increasing amounts of Spike protein. **(D)** Density plot for leader-NGFR to leader-GFP ratios in virally infected Vero cells or mock treated controls. Replicate conditions were merged for display. **(E)** Model of how NSP8/9 act to suppress SRP-dependent protein trafficking upon viral infection. In all panels, error bars represent standard deviation across independent biological replicates, dots represent individual values for each replicate, * indicates $p < 0.05$ and ** $p < 0.01$. **(F)** A model of how SARS-CoV-2 suppresses host immune responses through multi-pronged inhibition of core cellular functions. Cellular mechanisms are shown in gray and viral mechanisms in red.

*Chapter 3*BECKWITH-WIEDEMANN SYNDROME, BARR BODIES, AND BINDING
PROTEINS

A.K. Banerjee, M.R. Blanco, S. Quinodoz, J. Jachowicz, J. Wangen, J. Thai, V. Trinh, A. Chow,
and M. Guttman

Material included in this chapter was adapted from “Quinodoz, S.A., Bhat, P., Ollikainen, N., Jachowicz, J., Banerjee, A.K., Chovanec, P., Blanco, M.R., Chow, A., Markaki, Y., Plath, K. and Guttman, M. (2020). RNA promotes the formation of spatial compartments in the nucleus. *bioRxiv* 2020.08.25.267435; doi: <https://doi.org/10.1101/2020.08.25.267435>” ;

“Blanco, M.R., Walkup IV, W.G., Bonesteele, G., Banerjee, A.K., Peyda, P., Amaya, E., Guo, J., Chow, A., Trinh, V., and Guttman, M. (in submission/review). Denaturing purifications demonstrate that PRC2 and other chromatin proteins do not bind directly to RNA *in vivo*.” ; AND

“Jachowicz, J.W., Strehle, M. Banerjee, A.K., Thai, J., Blanco, M.R., and Guttman, M. (in submission/review). Xist-mediated condensates amplify SHARP localization on the X chromosome to promote chromosome-wide transcriptional silencing.”

1. ABSTRACT

Through its abilities to recognize and hybridize to nucleic acid templates, spatially localize to different compartments within the cell, bind combinatorically to effector molecules, and in some cases directly catalyze chemical reactions, RNA is a widely utilized and integrated component of core cellular function. RNA has been reported to play key roles in many developmentally regulated processes, with one well-characterized example being the role of the Xist long non-coding RNA (lncRNA) in the initiation of X-chromosome silencing in mammalian development. Xist-mediated silencing, in particular the recruitment of effector binding proteins including SMRT/HDAC1 Associated Repressor Protein (SHARP), is an established paradigm for RNA-mediated gene regulation *in cis*. To explore how prevalent this paradigm is, we used Covalent Linkage Affinity Purification to identify additional SHARP-RNA interactions and revealed an interaction within a critical region of the Kcnq1ot1 lncRNA. Kcnq1ot1 lncRNA is responsible for paternal imprinting of a conserved one megabase domain in mouse and human.

Dysregulation of this imprinting cluster has been linked to an overgrowth disorder called Beckwith-Wiedemann Syndrome, which presents with a variety of findings, including visceromegaly of intra-abdominal organs, abdominal wall defects, hemihyperplasia, and a predisposition for embryonal malignancies. Here we demonstrate that Kcnq1ot1 silencing of imprinting target genes is SHARP-dependent. Additionally, we identify several more SHARP-lncRNA interactions and surprisingly, many SHARP/pre-mRNA interactions. Coupled with super resolution microscopy of full length and mutant forms of SHARP, our data reveals RNA-mediated compartmentalization of SHARP as a widely utilized mechanism for gene regulation *in cis*.

2.0 INTRODUCTION

2.1 RNA-MEDIATED REGULATION OF GENE EXPRESSION

RNA is a widely utilized and integrated component of core cellular function because of its abilities to recognize and hybridize to nucleic acid templates, spatially localize to different compartments within the cell, bind combinatorically to effector molecules, and in some cases directly catalyze chemical reactions¹⁻⁸. As discussed in Chapter 2 within the context of host cell takeover by SARS-CoV-2, nonstructural RNAs can be utilized as scaffolds for many integral complexes within the cell, including subunits of the eukaryotic ribosome as well as the Signal Recognition Particle⁸. RNA can also play a key role in regulating gene expression, with discovered mechanisms inspiring novel treatment modalities such as Antisense Oligonucleotides or Small Interfering RNA⁹⁻¹⁴.

Eukaryotic genomes are pervasively transcribed, with clusters of genes coordinately regulated in *cis*¹⁵⁻²¹; this coordination is integral to cellular homeostasis as disruptions can lead to compromised viability and disease pathogenesis²²⁻²⁵. Despite the ever-expanding catalog of tissue- and cell-type specific gene expression profiles, we still do not completely understand how complex transcriptional programs are coordinately established and maintained, nor the complete molecular repertoire responsible for coordinated gene regulation in *cis*²⁶.

2.2 XIST-MEDIATED SILENCING AS A PARADIGM OF RNA-MEDIATED GENE REGULATION

How can cells coordinate gene expression *in cis*? While DNA *cis* regulatory elements (promoters or enhancers) offer exquisite cell type- and stage-specific gene regulation, these elements must

be encoded within each individual gene to allow *trans* acting factors (transcription factors) to recognize and coordinate a regulatory program²⁷⁻²⁸. Chromatin relaxation, in contrast, can result in deleterious upregulation of bystander genes¹⁷⁻¹⁸. Long non-coding RNAs (lncRNA) are an ideal answer; lncRNAs can leverage their preferential localization in the nucleus, spread *in cis* across local chromatin environments, and spatially concentrate effector chromatin effector molecules to seed defined regulatory compartments within the nucleus^{2,28-29}.

A well-established model for studying lncRNA-mediated gene regulation is X-chromosome inactivation (XCI), the process by which one of the two female X-chromosomes is transcriptionally silenced to establish dosage compensation in early mammalian development²⁹. XCI is initiated through expression of a lncRNA called Xist, which was originally identified as the sole gene expressed from the inactive X-chromosome and later demonstrated to be both necessary and sufficient for this process^{30-35, 29}. Xist-mediated silencing coordinates several different processes, including recruitment of chromatin-modifying proteins to the inactive X-chromosome, chromosome-wide DNA compaction, and translocation of the entire chromosome to the nuclear lamina, to form a nuclear compartment called the Barr Body and ultimately establish stable and heritable chromosome-wide silencing across an organism's lifetime^{3-6,36-46,34}.

2.3 SHARP AS A MODEL RNA-BINDING PROTEIN

One RNA-binding protein required for Xist-mediated silencing is SMRT/HDAC1 Associated Repressor Protein (SHARP), also known as Split Ends Transcriptional Repressor (Spen) or Msx2-Interacting Nuclear Target Protein (Mint)³⁻⁶. SHARP is a member of the Split Ends Protein Family (which consists of SHARP, RBM15, and RBM15B) and is simultaneously the functional ortholog of the *Drosophila* Hairless protein with respect to regulation of Notch signal

transduction (discussed further in Chapter 4)⁴⁷⁻⁴⁸. SHARP is an approximately 3600 amino acid protein which contains four N-terminal RNA-Recognition Motifs (RRM), a C-terminal Spen Paralog and Ortholog C-terminal (SPOC) domain (demonstrated to be responsible for recruitment of SMRT (Silencing Mediator for Retinoid or Thyroid-Hormone Receptor, also known as Nuclear Receptor Co-Repressor 2 or NCOR2) and transcriptional silencing), as well as a large intrinsically disordered region (**Figure 1A**)^{50,54,29}. Despite its documented interaction with direct chromatin regulators, transcription factors and non-coding RNA, SHARP lacks a defined DNA-binding domain^{51-54,3}. Interestingly, the RRM domains of SHARP have been shown to be both necessary for Xist-mediated silencing and recruitment to the inactive X-chromosome compartment, as well as sufficient for interactions with the Xist A-repeat^{55,50}. Likewise, the SPOC domain has been shown to be necessary for Xist-mediated silencing and sufficient once localized to the compartment⁵⁰.

SHARP is a model RNA-binding protein for studying RNA-mediated chromatin remodeling and transcriptional regulation because of its ability to directly recognize and bind to RNA and its clear effector function when recruited to chromatin^{50,53-54}. Despite the clear mechanistic dissection of SHARP's role in XCI, it remains unclear if Xist-mediated recruitment of SHARP to regulate gene expression *in cis* is an anomaly or a more widely applied mechanism. That is, is Xist special or emblematic of a more broadly utilized regulatory strategy in the nucleus?

3.0 RESULTS

3.1 RESULTS: COVALENT LINKAGE AFFINITY PURIFICATION REVEALS INTERACTION BETWEEN SHARP AND LNCRNA KCNQ1OT1

To address this question, we performed Covalent Linkage Affinity Purification against SHARP in pSM33-derived mouse embryonic stem cells (mESC) expressing N-terminal Halo-epitope tagged SHARP protein⁵⁶. We induced cells to express Xist lncRNA using doxycycline induction and were able to recover the SHARP/Xist-Repeat interaction using this method (**Figure 1B**).

Surprisingly, we observed an interaction between SHARP and the *Kcnq1ot1* lncRNA, which we subsequently verified using Immunostaining with RNA Fluorescence In Situ Hybridization (IF-FISH) (**Figure 1C and 1D**).

3.2 LNCRNA KCNQ1OT1 REGULATES IMPRINTING TARGET GENES IN TOPOLOGICAL ASSOCIATED DOMAIN

Kcnq1ot1 lncRNA (also known as KCNQ1 Opposite Strand Transcript 1, KCNQ1 overlapping transcript 1, or Long QT Intronic Transcript 1) is expressed in antisense orientation from within an intronic region of the *Kcnq1* protein coding gene on chromosome 7 in mouse and chromosome 11 in human respectively^{22,57}. *Kcnq1ot1* is responsible for paternal imprinting of the *Kcnq1* gene cluster, which includes Cyclin-Dependent Kinase Inhibitor 1C (CDKN1C), Solute Carrier Family 22 Member 18 (Slc22a18), and Pleckstrin Homology Like Domain Family A Member 2 (Phlda2)⁵⁸⁻⁶¹. Dysregulation of *Kcnq1ot1*-mediated imprinting is linked to a pediatric overgrowth syndrome called Beckwith Wiedemann Syndrome, which presents with findings ranging from visceromegaly of intra-abdominal organs, abdominal wall defects,

hemihyperplasia, and a predisposition for embryonal malignancies, including Wilm's Tumor and hepatoblastoma, among other sequelae^{22,62-63}.

Interestingly, Kcnq1ot1 lncRNA localizes to the same topologically associated domain on chromatin as its imprinting targets within the Kcnq1 gene cluster (**Figure 2A**). Given that lncRNAs may exert their function in an RNA-independent manner (that is, through the action of transcription or activity as a DNA *cis* regulatory element), we tested if Kcnq1ot1 is necessary and sufficient for transcriptional repression of its targets *in vivo*^{17,64}. Using CRISPR-interference, we observed upregulation of imprint targets within the Kcnq1 gene cluster compared to control (**Figure 2B**). Moreover, using an mESC line engineered to express Kcnq1ot1 upon doxycycline administration, we observed specific repression of targets within the Kcnq1 gene cluster upon Kcnq1ot1 induction (**Figure 2C**). These data demonstrate that Kcnq1ot1 lncRNA is both necessary and sufficient for transcriptional repression *in cis*, consistent with findings from previous studies using deletion of the Kcnq1ot1 promoter, pre-mature transcriptional truncation, and RNA-destabilization^{58-59,65-68}.

3.3 KCNQ1OT1 REGULATES IMPRINTING GENES IN SHARP-DEPENDENT MANNER

SHARP binds to Kcnq1ot1 lncRNA within a region previously demonstrated to be required for its silencing function, leading us to hypothesize that like Xist, Kcnq1ot1 transcriptionally silences genes *in cis* in a SHARP-dependent manner^{59,69}. To test this hypothesis, we generated mESC lines harboring homozygous deletions of the SHARP binding site within Kcnq1ot1 and observed upregulation of imprint targets in this line compared to the control (**Figure 3A**). Given that SHARP interacts with SMRT to ultimately recruit histone de-acetylases, we independently tested our hypothesis by using measuring Kcnq1ot1-mediated transcriptional silencing in the

presence of Trichostatin A⁷⁰. We observed blunting of Kcnq1ot1-mediated silencing in the presence of Trichostatin A compared to control (**Figure 3B**). Finally, we generated SHARP-knockout mESC lines using the CRISPR-Cas9 system and observed upregulation of CDKN1C in knockout lines compared to parent; this upregulation reverted upon re-introduction of Full Length SHARP (**Figure 3C and D**)⁷¹. These data demonstrate that Kcnq1ot1 lncRNA transcriptionally silences imprint target genes *in cis* in a SHARP-dependent manner, similar to Xist. (**Figure 3E**). It should be noted that unlike Xist, which is capable of establishing RNA-independent transcriptional memory, Kcnq1ot1 is required for maintenance of transcriptional silencing, suggesting additional complexity during Barr Body formation^{60,65,46,34}.

3.4 EXAMINING SHARP-RNA INTERACTIONS ON A GLOBAL SCALE

Beyond our identification of SHARP as a critical protein for Kcnq1ot1-mediated silencing, we asked if SHARP interacted with additional lncRNA targets. Using our SHARP CLAP-Seq dataset, we identified two additional lncRNA targets: Dleu2 and lncRNA 2610035D17Rik (named *Embryonic stem cell SHARP-Associated Long Coding Transcript 1*, eSalt1). We subsequently verified these interactions using IF-FISH (**Figure 4A**).

We also identified several SHARP/pre-mRNA interactions, which actually outnumbered the number of lncRNA interactions (**Figure 4B and 4C, Table 1**). To test if these pre-mRNA interactions were conserved between mouse and human, we performed CLAP-Seq against SHARP in HEK293T cells and identified similar binding profiles in several pre-mRNA targets, including Zc3h11a, Jarid2, and SHARP's mRNA transcript (Spen) (**Figure 4D**). Similar to SHARP-interacting lncRNAs, SHARP-interacting pre-mRNAs display broad localization *in cis* over chromatin (**Figure 4E and 4F**).

Given the number of SHARP-interacting RNAs and SHARP's role in establishing repressive transcriptional compartments on both the inactive X-chromosome and within the *Kcnq1* gene cluster, we wanted to examine RNA-mediated localization of the protein within the nucleus. Using 3-Dimensional Structural Illumination Microscopy, we observed punctate localization of Full Length SHARP protein, which was lost upon deletion of the RRM domain (**Figure 5A**). SHARP also contains an Intrinsically Disordered Region, which based on previous phase separation studies has been speculated to play a role in cooperative self-aggregation and RNA-mediated super-stoichiometric silencing of the X-chromosome^{72-73,29}. To test if the IDR domains are required for SHARP aggregation in the nucleus, we over-expressed mutant forms of the protein tagged monomeric eGFP-tagged SHARP in HEK 293T and performed live cell imaging⁷⁴. In comparison to Full Length Protein or deletion of the SPOC domain, we observed loss of punctate localization upon deletion of the RRM and IDR domains (**Figure 5B**).

4.0 DISCUSSION

Here we demonstrate that *Kcnq1ot1* recruits SHARP to a defined gene cluster to silence imprint target genes, which supports a broader model of RNA-mediated gene regulation in the nucleus, where RNA can recruit and concentrate proteins to defined chromatin territories to spatially amplify local regulatory signals *in cis*²⁸⁻²⁹. *Kcnq1ot1* recruits SHARP to a defined gene cluster to silence target genes. Through its avidity to RNA and its intrinsically disordered domains, SHARP can then aggregate and compartmentalize over this chromatin territory. Control over the entire 1 megabase gene cluster and RNA-mediated compartment can therefore be coordinated from a single 3.6 kilobase *cis* regulatory DNA element: the *Kcnq1ot1* promoter (also known as the *Kcnq1* Imprint Control Region)^{57,75}.

It is important to note that although SHARP acts *in trans* (being spatially recruited from a nucleoplasmic pool to given location on DNA) and synthetic recruitment of a 3kb Kcnq1ot1 fragment (harboring the SHARP binding site) is sufficient to silence ectopic genes, endogenous Kcnq1ot1 is spatially limited *in cis* to a defined location on chromatin (**Figure 3F and 2A**). Therefore, Kcnq1ot1 lncRNA interacts with *trans*-acting factors to spatially amplify and coordinate gene regulation *in cis*²⁸⁻²⁹.

What was most surprising to us from this dataset was the identification of SHARP/pre-mRNA interactions, which actually outnumber the SHARP/lncRNA interactions observed. Many of these pre-mRNAs localize broadly to chromatin in a similar manner to lncRNAs. Together, these data combined with our mechanistic dissection of Kcnq1ot1 suggest that SHARP-RNA interactions may be more pervasive than previously understood, and responsible for establishing repressive compartments across the nucleus to coordinate gene regulation *in cis*.

5.0 MATERIALS AND METHODS

Cell line generation, cell culture, and drug treatments

Cell lines used in this study. We used the following cell lines in this study: (i) male ES cells (*pSM33* ES cell line) previously engineered to express Xist lncRNA under the control of a tetracycline-inducible promoter and stably transfected with an episomally-maintained plasmid expressing Halo-tagged SHARP from a constitutive CAG promoter (Halo-SHARP *pSM33*)⁵⁶. (ii) Female ES cells containing dCas9 fused to 4 copies of the SID transcriptional repression domain integrated into a single locus in the genome (dCas9-4XSID). (iii) Female ES cells where we replaced the endogenous Kcnq1ot1 promoter with a tetracycline-inducible promoter on both alleles (*Kcnq1ot1-inducible* ES cell line). In the absence of doxycycline, these cells do not

express *Kcnq1ot1* from either allele, and in the presence of doxycycline they express *Kcnq1ot1* biallelically (**Figure 2C**). (iv) male e14-derived ES cells containing clonal insertions of PB-3xUAS-PGK-tGFP reporter, Gal4-Lambda-N insertion into the *Rosa* locus, and BoxB-tagged RNA expressed by stable insertion via piggybac transposon⁸⁹⁻⁹². (v) HEK293T, a female human embryonic kidney cell line obtained from ATCC.

Cell culture conditions. All mouse ES cell lines were cultured in serum-free 2i/LIF medium as previously described⁷⁶. Halo-SHARP pSM33 cells were additionally maintained in 0.5µg/ml puromycin to promote stable selection of the episomal plasmid. HEK293T cells were cultured in complete media consisting of DMEM (GIBCO, Life Technologies) supplemented with 10% FBS (Seradigm Premium Grade HI FBS, VWR), 1X penicillin-streptomycin (GIBCO, Life Technologies), 1X MEM non-essential amino acids (GIBCO, Life Technologies), 1 mM sodium pyruvate (GIBCO, Life Technologies) and maintained at 37C under 5% CO₂. For maintenance, 800,000 cells were seeded into 10 mL of complete media every 3-4 days in 10 cm dishes.

Doxycycline Inducible *Kcnq1ot1* lines. Female ES cells (F1 2-1 line, provided by K. Plath) were CRISPR-targeted (nicking gRNA pairs TGGGCGGGAGTCTTCTGGGCAGG and GGATTCTCCCAGGCCAGGGCGG) to express the Tet transactivator (M2rtTA) from the *Rosa26* locus using R26P-M2rtTA, a gift from Rudolf Jaenisch (Addgene plasmid #47381). The endogenous promoter of *Kcnq1ot1* was CRISPR-targeted (nicking gRNA pairs TCGTGGCTGCCACGTCACCA and CAGATGCTGAATAATGACTA) to insert a TRE and

minimal CMV promoter. Clones were screened for ablation of endogenous Kcnq1ot1 expression and biallelic upregulation of expression upon administration of doxycycline (**Figure 2C**).

CRISPRi: dCas9-4XSID cell line generation. A catalytically dead Cas9 (dCas9) fused to 4 copies of the SID repressive domain (4XSID) expressed from an Efla promoter was integrated into a single copy locus in the genome (mm10 - chr6:86,565,487-86,565,506; gRNA sequence AATCTTAGTACTACTGCTGC) using CRISPR (cells hereby referred to as dCas9-4XSID).

Doxycycline induction. Expression of Xist and Kcnq1ot1 was induced in their respective cell lines by treating cells with 2 µg/ml doxycycline (Sigma). Xist was induced for 24 hours prior to crosslinking and analysis. Kcnq1ot1 was induced for 12-16hrs prior to harvesting RNA.

Trichostatin (TSA) treatment. For HDAC inhibitor experiments, cells were treated cells with either DMSO (control) or 5µM TSA in fresh 2i media or 2µg/ml doxycycline in standard 2i.

3D-SIM SHARP-Halo cell culture conditions. pSM33 cells were seeded in 4-well imaging chambers (ibidi) equipped with a high precision glass bottom and plasmids were transfected with lipofectamine 3000 24 hours prior to imaging according to the manufacturer's instructions. SHARP expression was induced by the addition of doxycycline 8hrs prior to imaging. 1µM JF646 Halo ligand was introduced to the media for 30 min, washed off twice with PBS and exchanged with fresh media, following which cells were incubated for another 15 min. Live-cell 3D-SIM imaging was performed at 37C and 5% CO₂ in media without phenol red.

Transfection of SHARP constructs in HEK293T. For those experiments that required high expression of proteins (CLAP, FRAP, live-cell imaging), we used human HEK293T cells instead of mESCs because they show much higher overexpression levels (**Supplemental Note 2,3**). We

transfected HEK293T cells using BioT transfection reagent (Bioland) according to manufacturer's recommendations. We used ~10 μg of DNA when cells were grown on a 15 cm dish and ~1 μg of DNA when cells were grown on 3 cm glass-bottom dishes (Matek), adjusting DNA concentrations to match mole numbers across constructs. 48 hours post-transfection, cells were used for live-cell imaging (**Figure 5B**).

Note: We observed that high and prolonged overexpression of SHARP and its mutants in HEK293T cells leads to rapid cell death. We found that processing cells 48 hours post-transfection is an optimal time, providing both high protein expression and cell number. Another feature of the overexpression experiment that we had to take into consideration was heterogeneity in the level of protein expression between different cells and constructs. To correct for this and compare cells with similar levels of SHARP protein, we set laser powers and gains based on the lowest expressing samples (FL-SHARP) and used these settings for all other samples. As a consequence, some cells with extremely high overexpression levels (from Delta IDR-SHARP, Delta RRM-IDR, eGFP groups) were not imaged or analyzed. When lower laser powers and gains were used (too low to detect FL-SHARP), we observed the formation of condensates in Delta IDR-SHARP sample, but not in the eGFP control. This suggests that SHARP mutant proteins can form condensates; however, they require much higher concentration than the full-length protein.

Microscopy imaging

RNA-FISH and IF. For immunostaining combined with *in situ* RNA visualization, we used the ViewRNA Cell Plus (Thermo Fisher Scientific, 88-19000-99) kit per the manufacturer's protocol with minor modifications. Cells were grown on coverslips and rinsed with 1xPBS, pre-extracted with 0.5% Triton X-100 in PBS for 5 minutes at room temperature, fixed in 4% paraformaldehyde in PBS for 15 minutes at room temperature, and rinsed in 1xPBS, and permeabilized with 0.5% Triton X-100 in 4% paraformaldehyde for 10 minutes at room temperature. Cells were either stored at -20C in 70% ethanol or used directly for immunostaining and incubated in blocking solution (0.2% BSA in PBS) for at least 1 h. If stored in 70% ethanol, cells were re-hydrated prior to staining by washing 3 times in 1xPBS and incubated in blocking solution (0.2% BSA in PBS) for at least 1 hour. Primary antibody against SHARP (Bethyl # A301-119A, Lot # 2) was diluted in blocking solution and added to coverslips for 3-5 hours at room temperature. Cells were washed three times with 0.01% Triton X-100 in PBS for 5 minutes each and then incubated in blocking solution containing corresponding secondary antibodies labeled with Alexa fluorophores (Invitrogen) for 1 hour at room temperature. Next, cells were washed 3 times in 1xPBS for 5 minutes at room temperature. After the last wash in 1xPBS, cells were subject to post-fixation for 10 min in 2% paraformaldehyde on 1xPBS at room temperature, washed 3 times in 1XPBS, and then RNA-FISH, based on the ViewRNA ISH (Thermo Fisher Scientific, QVC0001) protocol, was performed. Coverslips were mounted with ProLong Gold with DAPI (Invitrogen, P36935) and stored at 4C until acquisition. All probes used in the study were custom made by Thermofisher. To test their specificity, we either utilized RNase treatment prior to RNA-FISH or two different probes targeting the same RNA. Images were acquired on a Zeiss LSM800 confocal microscope with a 100x glycerol immersion objective lens. Z-sections

were taken every 0.3 μm . Image visualization and analysis were performed with Icy and ImageJ software respectively.

3D-Structured Illumination Microscopy (3D-SIM): 3D-SIM 1000 super-resolution imaging was performed on a DeltaVision OMX-SR system (Cytiva, Marlborough, MA, USA) equipped with a 60x/1.42 NA Plan Apo oil immersion objective (Olympus, Tokyo, Japan), sCMOS cameras (PCO, Kelheim, Germany) and 642 nm diode laser. Image stacks were acquired with z-steps of 125 nm and with 15 raw images per plane. The raw data were computationally reconstructed with the soft-WoRx 7.0.0 software package (Cytiva, Marlborough, MA, USA) using a wiener filter set to 0.002 and channel-specifically measured optical transfer functions (OTFs) using an immersion oil with a 1.518 refractive index (RI). 32-bit raw datasets were imported to ImageJ and converted to 16-bit stacks.

HEK 293T Live Cell Imaging: Live samples were imaged using the Leica Stellaris microscope with 63x water objective ($\sim 80\text{nm}$ xy, $\sim 300\text{nm}$ z), and 16 Z-stacks were collected every 60 seconds for 5 minutes. The microscope was equipped with a stage incubator to keep cells at 37°C and 5% CO₂. Image analysis was performed using ICY or FIJI (ImageJ v2.1.0/1.53c) software.

Molecular Biology Protocols

SHARP Covalent Linkage Affinity Purification. An expression vector containing full-length SHARP with an N-terminal Halo-FLAG (HF) fusion protein was transfected into mouse ES cells containing a doxycycline inducible Xist gene. Cells were washed once with PBS and then crosslinked on ice using 0.25 J cm⁻² (UV2.5k) of UV at 254 nm in a Spectrolinker UV Crosslinker. Cells were then scraped from culture dishes, washed once with PBS, pelleted by centrifugation at 1,500g for 4 min, and flash-frozen in liquid nitrogen for storage at -80C. We

lysed batches of 5 million cells by completely resuspending frozen cell pellets in 1 mL of ice cold iCLIP lysis buffer (50 mM Hepes, pH 7.4, 100 mM NaCl, 1% NP-40, 0.1% SDS, 0.5% Sodium Deoxycholate) supplemented with 1X Protease Inhibitor Cocktail (Promega), 200 U of Murine RNase Inhibitor (New England Biolabs), 20 U Turbo DNase (Ambion), and 1X Manganese/Calcium Mix (0.5mM CaCl₂, 2.5 mM MnCl₂). Samples were incubated on ice for 10 minutes to allow lysis to proceed. The lysates were then incubated at 37C for 10 minutes at 1150 rpm shaking on a Thermomixer (Eppendorf). Lysates were cleared by centrifugation at 15,000g for 2 minutes. The supernatant was collected and kept on ice until bound to the HaloLink Resin.

We used 200 μ L of 25% HaloLink Resin (50 μ L of HaloLink Resin total) per 5 million cells. Resin was washed three times with 2 mL of 1X TBS (50 mM Tris pH 7.5, 150 mM NaCl) and incubated in 1X Blocking Buffer (50 mM HEPES, pH 7.5, 10 μ g/mL Random 9-mer, 100 μ g/mL BSA) for 20 minutes at room temperature with continuous rotation. After the incubation, resin was washed three times with 1X TBS. The cleared lysate was mixed with 50 μ L of HaloLink Resin and incubated at 4C for 3-16 hrs with continuous rotation. The captured protein bound to resin was washed three times with iCLIP lysis buffer at room temperature and then washed three times at 90C for 2 minutes while shaking at 1200 rpm with each of the following buffers: 1X ProK/NLS buffer (50 mM HEPES, pH 7.5, 2% NLS, 10 mM EDTA, 0.1% NP-40, 10 mM DTT), High Salt Buffer (50 mM HEPES, pH 7.5, 10 mM EDTA, 0.1% NP-40, 1M NaCl), 8M Urea Buffer (50 mM HEPES, pH 7.5, 10 mM EDTA, 0.1% NP-40, 8 M Urea), and Tween buffer (50 mM HEPES, pH 7.5, 0.1% Tween 20, 10 mM EDTA). Finally, we adjusted the buffer by washing with Elution Buffer (50 mM HEPES, pH 7.5, 0.5 mM EDTA, 0.1% NP-40) three times at 30C. The resin was resuspended in 83 μ L of Elution Buffer and split into a 75 μ L (ProK

elution) and 8 μ L (TEV elution) reaction. 25 μ L of 4X ProK/NLS Buffer and 10 μ L of ProK were added to the ProK elution tube and the sample was incubated at 50C for 30 minutes while shaking at 1200 rpm. 2.3 μ L of ProTEV Plus Protease (Promega) was added to the TEV Elution and the sample was incubated at 30C for 30 minutes while shaking at 1200 rpm.

For each experiment, we ensured that we successfully purified the Halo-tagged protein. To do this, the TEV elution sample was mixed with 1X LDS Sample Buffer (Invitrogen) and 1X Reducing Agent (Invitrogen) and heated for 6 minutes at 70C. The sample was run on a 3-8% Tris Acetate Gel (Invitrogen) for 1 hour at 150 V. The gel was transferred to a nitrocellulose membrane using an iBlot Transfer Device (Invitrogen). The nitrocellulose membrane was blocked with Odyssey Blocking Buffer (LI-COR) for 30 minutes. We incubated the membrane in Anti-FLAG mouse monoclonal Antibody (Sigma, F3166) and V5 rabbit polyclonal antibody (Santa Cruz, sc-83849-R) at a 1:2500 dilution for 2 hours at room temperature to detect the protein. We visualized the protein by incubating the membrane in 1:17,500 dilution of both IRDye 800CW Goat anti-Rabbit IgG (LI-COR, 925-32210) and IRDYE 680DR Goat anti-Mouse IgG (LI-COR, 925-68070) for 1 hour at room temperature followed by imaging on a LICOR Odyssey.

RNA was purified from the ProK elution sample and an RNA-Seq library was constructed as previously described⁷⁷. Briefly, after proK elution, the RNA was dephosphorylated (Fast AP) and cyclic phosphates removed (T4 PNK) and then cleaned up on Silane beads as previously described⁷⁷. The RNA was then ligated to an RNA adapter containing a RT primer binding site. The ligated RNA was reverse transcribed (RT) into cDNA, the RNA was degraded using NaOH, and a second adapter was ligated to the single stranded cDNA. The DNA was amplified and Illumina sequencing adaptors were added by PCR using primers complementary to the 3' and 5'

adapters. The molarity of PCR amplified libraries was measured by Agilent TapeStation High Sensitivity DNA screentapes and all samples were pooled at equal molarity. The pool was then purified and size selected on a 2% agarose gel and cut between 150-700 nts. The final libraries were measured by Agilent Bioanalyzer and Qubit high sensitivity DNA to determine the loading density of the final pooled sample. Pooled samples were paired-end sequenced on an Illumina HiSeq 2500 with read length 35 x 35nts.

Sequencing reads were trimmed to remove adaptor sequences and any bases containing quality scores <10 using Trimmomatic⁷⁸. Read-pairs where either read was trimmed to <25 nucleotides were filtered. PCR duplicates were excluded using the FastUniq tool⁸³. The remaining reads were then aligned to Ribosomal RNAs (rRNAs) using the Tagdust program⁸⁴ with a database of 18S, 28S, 45S, 5S, 5.8S sequences. TagDust was chosen because it allowed more permissive alignments to rRNA reads that contained mismatches and indels due to RT errors induced by rRNA post-transcriptional modifications. The remaining reads were then aligned to the mouse (mm9) genomes using STAR aligner⁷⁹. Only reads that mapped uniquely in the genome were kept for further analysis. SHARP CLAP performed from transiently transfected HEK293T cells was processed in a similar manner. Peaks and input-normalized enrichments were called using CLIPCLAPEnrichment javascript pipeline. Targets ultimately presented in Figure 4C were selected based on initial screen of peaks based on window-normalized p-value cut off of $< e^{-5}$ that were pooled between two biological replicates and additionally curated based on visual examination of pile-ups to remove repetitive element artifacts and comparisons between cell type datasets. Enrichment figures were made from mm10 aligned (mouse) or hg19 aligned files (human).

SPRITE 2.0: RNA and DNA SPRITE

SPRITE 2.0 is an adaptation of our initial SPRITE protocol with significant improvements to the RNA molecular biology steps that enable generation of higher complexity RNA libraries^{76,80}. It was performed as follows:

Crosslinking, chromatin isolation, and chromatin digestion. Cells were crosslinked at room temperature with 2mM disuccinimidyl glutarate (DSG) for 45 minutes followed by 3% formaldehyde for 10 minutes to preserve RNA and DNA interactions *in situ*. After crosslinking, the formaldehyde crosslinker was quenched with addition of 2.5M glycine for final concentration of 0.5M for 5 minutes, cells were spun down, and isolated in 1x PBS + 0.5% RNase Free BSA (AmericanBio #AB01243-00050) and flash frozen at -80C for storage. When using BSA to isolate cells, we found that using RNase-Free BSA was critical to avoid RNA degradation. RNase Inhibitor (1:40, NEB Murine RNase Inhibitor or Thermofisher Ribolock) was added to all lysis buffers to avoid degradation. After lysis, cells were sonicated at 4-5W of power for 1 minute (pulses 0.7 second on, 3.3 seconds off) using the Branson Sonicator and chromatin was fragmented using DNase digestion to obtain DNA of approximately ~150-1kb in length.

Estimating molarity. After DNase digestion, crosslinks were reversed on approximately 10 μ l of lysate in 82 μ L of 1X Proteinase K Buffer (20 mM Tris pH 7.5, 100 mM NaCl, 10 mM EDTA, 10 mM EGTA, 0.5% Triton-X, 0.2% SDS) with 8 μ L Proteinase K (NEB) at 65C for 1 hour. RNA and DNA were purified using Zymo RNA Clean and Concentrate columns per the manufacturer's specifications (>17nt protocol) with minor adaptations, such as binding to the column with 2X volume RNA Binding Buffer combined with by 1X volume 100% EtOH to improve yield. Molarities of the RNA and DNA were calculated by measuring the RNA and

DNA concentration with the Qubit Fluorometer (HS RNA kit, HS dsDNA kit) and the average RNA and DNA sizes were estimated using the RNA High Sensitivity TapeStation and Agilent Bioanalyzer (High Sensitivity DNA kit).

NHS bead coupling. RNA and DNA molarity estimated in the lysate was used to calculate the total number of RNA and DNA molecules per microliter of lysate. We coupled the lysate to NHS-activated magnetic beads (Pierce) in 1x PBS + 0.1% SDS combined with 1:40 dilution of NEB Murine RNase Inhibitor overnight at 4C as previously described⁷⁶. We coupled at a ratio of 0.5 molecules per bead to reduce the probability of simultaneously coupling multiple independent complexes to the same bead (which would lead to their association during the split-pool barcoding process). Because multiple molecules of DNA and RNA can be crosslinked in a single complex, this estimate is a more conservative estimate of the number of molecules to avoid collisions on individual beads. After NHS coupling overnight, the reaction was quenched in 0.5M Tris pH 7.5 and beads were washed as previously described⁷⁶.

Because the crosslinked complexes are immobilized on NHS magnetic beads, several enzymatic steps can be performed by adding buffers and enzymes directly to the beads and performing rapid buffer exchange between each step on a magnet. All enzymatic steps were performed with shaking at 1200 rpm (Eppendorf Thermomixer) to avoid bead settling and aggregation, and all enzymatic steps were inactivated by adding 1 mL of SPRITE Wash buffer supplemented with 50 mM EDTA and 50 mM EGTA to the NHS beads instead of Modified RLT buffer.

DNA End Repair and dA-tailing. DNA ends are repaired to enable ligation of tags to each molecule. Specifically, we blunt end and phosphorylate the 5' ends of double-stranded DNA using two enzymes. First, T4 Polynucleotide Kinase (NEB) treatment is performed at 37C for 1

hr, the enzyme is quenched using 1 mL Modified RLT buffer, and then buffer is exchanged with two washes of 1 mL SPRITE Detergent Buffer to beads at room temperature. Next, the NEBNext End Repair Enzyme cocktail (containing T4 DNA Polymerase and T4 PNK) and 1x NEBNext End Repair Reaction Buffer is added to beads and incubated at 20C for 1 hr, and inactivated and buffer exchanged as specified above. DNA was then dA-tailed using the Klenow fragment (5'-3' exo-, NEBNext dA-tailing Module) at 37C for 1 hr, inactivated, and buffer exchanged as specified above. Note, we do not use the combined End Repair/dA tailing modules as the temperatures in the protocol are not compatible with SPRITE (the higher temperature will reverse crosslinks). To prevent degradation of RNA, each enzymatic step is performed with the addition of 1:40 NEB Murine RNase Inhibitor or ThermoFisher Ribolock.

Ligation of the DNA Phosphate Modified (“DPM”) Tag. After end repair and dA-tailing of DNA, we performed a pooled ligation with “DNA Phosphate Modified” (DPM) tag that contains certain modifications that we found to be critical for the success of SPRITE 2.0. Specifically, (i) the phosphothiorate modification prevents enzymatic digestion by Exo1 in subsequent RNA steps and (ii) the internal biotin modification facilitates an on-bead library prep post reverse-crosslinking. The DPM adaptor also contains a 5' phosphorylated sticky end overhang to ligate tags during split-pool barcoding. Ligation was performed as previously described using Instant Sticky End Mastermix (NEB) except that all ligations were supplemented with 1:40 RNase inhibitor (ThermoFisher Ribolock or NEB Murine RNase Inhibitor) to prevent RNA degradation. Because the T4 DNA Ligase included in Instant Sticky End Mastermix ligates double-stranded DNA, the unique DPM sequence enables identification between RNA and DNA molecules during sequencing.

Ligation of the RNA Phosphate Modified (“RPM”) Tag. To map RNA and DNA interactions simultaneously, we ligated an RNA adaptor containing the same 7nt 5’phosphorylated sticky end overhang as the DPM adaptor used to ligate tags to both RNA and DNA during split-pool barcoding. To do this, the 3’end of RNA is first modified to have a 3’OH for ligation of RPM. Specifically, RNA overhangs are repaired with T4 Polynucleotide Kinase (NEB) with no ATP at 37C for 20 min. RNA is subsequently ligated with a “RNA Phosphate Modified” (RPM) as previously described using High Concentration T4 RNA Ligase I⁸¹. Because T4 RNA Ligase 1 ligates single-stranded RNA, ligation of RNA with the unique RPM sequence enables identification between RNA and DNA molecules during sequencing. After RPM ligation, RNA was converted to cDNA using Superscript III at 42C for 1 hour using the “RPM bottom” RT primer that contains an internal biotin to facilitate on-bead library prep as above and a 5’end sticky end to ligate tags during SPRITE. Excess primer was digested with Exonuclease 1. All ligations were supplemented with 1:40 RNase inhibitor (ThermoFisher Ribolock or NEB Murine RNase Inhibitor) to prevent RNA degradation.

Split-and-pool barcoding to identify RNA and DNA interactions. The beads were then repeatedly split-and-pool ligated over four rounds with a set of “Odd,” “Even,” and “Terminal” tags (see [SPRITE Tag Design](#) in reference 76). Both DPM and RPM contain the same 7 nucleotide sticky end that will ligate to all subsequent split-pool barcoding rounds. All split-pool ligation steps and reverse crosslinking were performed for 45min to 1 hour at 20C as previously described⁷². All ligations were supplemented with 1:40 RNase inhibitor (ThermoFisher Ribolock or NEB Murine RNase Inhibitor) to prevent RNA degradation.

Reverse crosslinking. After multiple rounds of SPRITE split-and-pool barcoding, the tagged RNA and DNA molecules were eluted from NHS beads by reverse crosslinking overnight (~12-

13 hours) at 50C in NLS Elution Buffer (20mM Tris-HCl pH 7.5, 10mM EDTA, 2% N-Lauroylsarcosine, 50mM NaCl) with added 5M NaCl to 288mM NaCl final combined with 5uL Proteinase K (NEB).

Post reverse-crosslinking library preparation. AEBSF (Gold Biotechnology CAS#30827-99-7) is added to the Proteinase K (NEB Proteinase K #P8107S; ProK) reactions to inactivate the ProK prior to coupling to streptavidin beads. Biotinylated barcoded RNA and DNA are bound to streptavidin beads. To improve recovery, the supernatant is bound again to 20 ul of streptavidin beads and combined with the first capture. Beads are washed in 1X PBS + RNase inhibitor and then resuspended in 1x First Strand buffer to prevent any melting of the RNA:cDNA hybrid. Beads were pre-incubated at 40C for 2 min to prevent any sticky barcodes from annealing. A second reverse transcription is performed by adding Superscript III (without primer) to extend the cDNA through the areas which were previously crosslinked. The second RT ensures that cDNA recovery is maximal, particularly if RT terminated at a crosslinked site prior to reverse crosslinking. After generating cDNA, the RNA is degraded by addition of RNaseH and RNase cocktail, and the 3' end of the resulting cDNA is ligated to attach an dsDNA oligo containing library amplification sequences for subsequent library amplification.

Previously, we performed cDNA (ssDNA) to ssDNA primer ligation which relies on the two reactants coming together for conversion to a product that can then be amplified for library preparation. To improve the efficiency of cDNA molecules ligated with the Read1 Illumina priming sequence, here we performed a “splint” ligation, which involves a partially ssDNA partially dsDNA primer that contains a random 6mer annealing to the 3' end of the cDNA and brings the 5' phosphorylated end of the cDNA adapter directly together with the cDNA via

annealing. This ligation is performed with 1x NEB Instant Sticky End Master Mix at 20C for 1 hour. This greatly improves the cDNA tagging and overall RNA yield.

Libraries were amplified using Q5 Hot-Start Mastermix (NEB) with primers that add the full Illumina adaptor sequences. After amplification, the libraries are cleaned up using 0.8X SPRI (AMPure XP) and then gel cut using the Zymo Gel Extraction Kit selecting for sizes between 280 bp - 1.3 kb.

Sequencing. Sequencing was performed on an Illumina NovaSeq S4 paired-end 150x150 cycle run. For the mES RNA-DNA SPRITE 2.0 data in this experiment, 144 different SPRITE libraries generated from two replicate SPRITE experiments were sequenced. Each SPRITE library corresponds to a distinct aliquot during the Proteinase K reverse crosslinking step which is separately amplified with a different barcoded primer to provide an additional round of SPRITE barcoding.

Primers Used for RPM, DPM, and Splint Ligation (IDT):

1. RPM top: /5Phos/rArUrCrArGrCrACTTAGCG TCAG/3SpC3/
2. RPM bottom (internal biotin): /5Phos/TGACTTGC/iBiodT/GACGCTAAGTGCTGAT
3. DPM Phosphorothioate top: /5Phos/AAGACCACCAGATCGGAAGAGCGTCGTG*T*
A*G*G* /32MOErG/ *Denotes Phosphorothioate bonds
4. DPM bottom (internal biotin):
/5Phos/TGACTTGTCATGTCT/iBioT/CCGATCTGGTGGTCTT
5. 2Puni splint top: TACACGACGCTCTTCCGATCT NNNNNN/3SpC3/
6. 2Puni splint bottom: /5Phos/AGA TCG GAA GAG CGT CGT GTA/3SpC3/

Annealing of adaptors. A double-stranded DPM oligo and 2P universal splint oligo were generated by annealing the complementary top and bottom strands at equimolar concentrations. Specifically, the oligos were annealed in 1x Annealing Buffer (0.2 M LiCl₂, 10 mM Tris-HCl pH 7.5) by heating to 95C and then slowly cooling to room temperature (-1C every 10 sec) using a thermocycler.

Assessing molecule to bead ratio. First, we ensured that SPRITE clusters represent *bona fide* interactions that occur within a cell by mixing human and mouse cells and ensuring that virtually all SPRITE clusters (~99%) represent molecules exclusively from a single species. Specifically, we separately crosslinked HEK293T cells performed a human-mouse mixing SPRITE 2.0 experiment and identified conditions with low interspecies mixing (molecules = RNA+DNA instead of DNA). Specifically, for SPRITE clusters containing 2-1000 reads, the percent of interspecies contacts was: 2 beads:molecule = 0.9% interspecies contacts, 4 beads:molecule = 1.1% interspecies contacts, 8 beads:molecule = 1.1% interspecies contacts. We used the 2 beads:molecule ratio for the SPRITE 2.0 data set generated in this paper.

SPRITE 2.0 processing pipeline

Adapter trimming. Adapters were trimmed from raw paired-end fastq files using Trim Galore! v0.6.2 and assessed with Fastqc v0.11.9. Subsequently, the DPM (GATCGGAAGAG) and RPM (ATCAGCACTTA) sequences are trimmed using Cutadapt v2.5 from 5' end of R1 along with the 3' end DPM sequences that result from short reads being read through into the barcode (GGTGGTCTTT, GCCTCTTGTT, CCAGGTATTT, TAAGAGAGTT, TTCTCCTCTT, ACCCTCGATT)⁸². The additional trimming improves read mapping in the end-to-end alignment mode. The SPRITE barcodes of trimmed reads are later identified using Barcode ID v1.2.0 and

the ligation efficiency is assessed. Reads with an RPM or a DPM barcode are split into two separate files, to process RNA and DNA reads individually downstream, respectively.

Processing RNA reads. RNA reads were aligned to GRCm38.p6 with the Ensembl GRCm38 v95 gene model annotation using Hisat2 v2.1.0 with a high penalty for soft-clipping --sp 1000,1000. Unmapped and reads with a low MapQ score (samtools view -bq 20) were filtered out for downstream realignment. Mapped reads were annotated for gene exons and introns with the featureCounts tool from the subread package v1.6.4 using [Ensembl GRCm38 v95](#) gene model annotation and the Repeat and Transposable element annotation from the Hammel lab⁸³. Filtered reads were subsequently realigned to our custom collection of repeat sequences using Bowtie v2.3.5, only keeping mapped and primary alignment reads⁸⁴.

Processing DNA reads. DNA reads were aligned to GRCm38.p6 using Bowtie2 v2.3.5, filtering out unmapped and reads with a low MapQ score (samtools view -bq 20). Data generated in F1 hybrid cells (psm33 - C57BL and 129S1 or psm44 – CAST and 129S1) were assigned the allele of origin using SNPsplit v0.3.4⁸⁵. RepeatMasker regions with milliDev ≤ 140 and blacklisted v2 regions were filtered out using Bedtools v2.29.0⁸⁶⁻⁸⁷. RNA and DNA reads were merged, and a cluster file was generated for all downstream analysis. MultiQC v1.6 was used to aggregate all reports⁸⁸.

Masked bins. In addition to known repeat containing bins, we manually masked the following bins (mm10 genomic regions: chr2:79490000-79500000, chr11:3119270-3192250, chr15:99734977-99736026, chr3:5173978-5175025, chr13:58176952-58178051) because we observed a major overrepresentation of reads in the input samples.

Kcnq1ot1 protein binding, perturbations, and gene expression measurements

Kcnq1ot1 CRISPR interference. dCas9-4XSID cells were transfected using multiplexed gRNA vector constructs containing an episomal polyoma origin of replication, puromycin resistance driven by a PGK promoter and four tandem U6-gRNA cassettes (allowing for simultaneous expression of four sgRNAs). Negative control gRNA sequences recognizing the *Saccharomyces cerevisiae* Upstream Activation Sequence (UAS) and the Tetracycline Response Element (TRE) were multiplexed together (referred to as sgTUUT; gRNAs are follows:

TCTCTATCACTGATAGGGAG, GAGGACAGTACTCCGCTCGG, GCGGAGTACTGTCCTCCGAG, and TCTCTATCACTGATAGGGAG). Four gRNA sequences targeting the Kcnq1ot1 promoter were multiplexed together (referred to as sgKcnq1ot1; gRNAs are as follows: GCCTAGCCGTTGTCGCTAGG, GCCCTGTACTGCATTGAGGT, GCCTGCACAGTAGGATTCCA, and GGAGGATGGGTCGAGTGGCT).

dCas9-4XSID cells were transfected with either sgTUUT or sgKcnq1ot1 and selected for three days with 1µg/ml of puromycin in standard 2i culture conditions. Cells were subsequently passaged and maintained in 0.5µg/ml puromycin for an additional 7 days prior to RNA harvesting. Data presented are from two separate transfections and biological replicates.

Genetic deletion of SHARP Binding Site. F1 2-1 line cells were CRISPR-targeted with gRNAs targeting the SHARP-Binding Site identified previously via Covalent Linkage Affinity Purification Sequencing (SHARP Binding Site Coordinates: mm10 – chr7:143,295,789-143,296,455; gRNA sequences were ATGCACCATCATAGACCACG and TCATAGCCTCCCCCTCCTCG). Following selection using 1µg/ml of puromycin in standard 2i

culture conditions, transfected cells were allowed to recover in standard 2i media prior to sub-cloning. Clones were subsequently screened using genomic DNA PCR with primers flanking the deletion region (CAGCATCTGTCCAATCAACAG and GCAAATACGAGAACTGAGCC respectively). In contrast to the wild type 1048bp band, successfully targeted alleles produced a 305bp band. Sub-clones homozygous for the targeted allele were subject to RT-qPCR and GAPDH-normalized gene expression was further normalized to the F1 parent line).

SHARP Knockout Line Generation. Mouse ES cell lines with doxycycline-inducible Xist expression were CRISPR-targeted using pool of gRNAs selected against the *Spn* gene body. Subclones were screened using RT-qPCR using primer pairs selected against regions upstream, within, and downstream of gRNA targeting (gRNAs were GGAGACCGAGACCTCCGCA, TATGAGCGGAGACTCGATG, and GACTGGGAGAACTAACACA, TTCTGCCGATACGGACCTG). Subclones were identified based on the combination of absence of targeted region and upregulation of flanking upstream and downstream regions, matching the previously observed targeting signature observed by Montfort et al.⁶

Primer pairs used for RT-qPCR screening were as follows: (SHARP Exon 1-2:

CATTTCAAACGATATGGCCG and CTTTCTGTGCACTTTTGATGTC; SHARP Exon 3-5:

CACTGACAGTAGCGACTCC and TAAGGCTTGTGTCTGTAGAGC; SHARP Exon 8-10:

CTGTGGTAAAGGTGGTGTGTTG and TGCAAAGTCCACCTTAATCTTATTC; and SHARP

Exon 14-16: GAGAATGACAGTGGAAACCGA and GTAGGCAGGCTGATTGGAG.)

Knockout subclones B1 and D5 were used for subsequent experiments.

SHARP Rescue Experiments. Knockout subclones B1 and D5 were transfected with membrane-mCherry plasmid alone (gift from J. Jachowicz) or co-transfected with membrane-mCherry and

CAG-Halo-SHARP, sorted for mCherry⁺ cells 48 hours post transfection, and harvested for RNA, PolyA-selection RNA-library preparation, and ultimately high-throughput sequencing. Transfection and sorting experiments were performed in biological duplicate, beginning at transfection.

HDAC inhibitor treatment. The inducible Kcnq1ot1 cell line was treated with either DMSO (control) or 5 μ M TSA in fresh 2i media or 2 μ g/ml doxycycline in standard 2i. RNA was extracted, reverse transcribed, and qPCR was performed. CT values were normalized to GAPDH to compare gene expression differences between induced and non-induced samples within the same pharmacologic condition (i.e. GAPDH-normalized “Induced DMSO” to GAPDH-normalized “Non-Induced DMSO Vehicle”) to generate fold gene expression ratios. RT-qPCR data presented is summarized from two separate replicate experiments.

Kcnq1ot1 Synthetic Recruitment. Piggy-bac transposon modified e14 ES cells were analyzed using mean fluorescence intensity of tGFP reporter using flow cytometry. Mean fluorescence activity was normalized to mCherry control, with standard deviation representing inter-experimental variability between two replicate experiments.

6.0 REFERENCES

1. Séraphin, B., Kretzner, L., and Rosbash, M. (1988). A U1 snRNA:pre-mRNA base pairing interaction is required early in yeast spliceosome assembly but does not uniquely define the 5' cleavage site. *The EMBO journal*, 7(8), 2533–2538.
2. Quinodoz, S., and Guttman, M. (2014). Long noncoding RNAs: an emerging link between gene regulation and nuclear organization. *Trends in cell biology*, 24(11), 651–663.
<https://doi.org/10.1016/j.tcb.2014.08.009>

3. 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., and 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>
4. Chu, C., Zhang, Q.C., da Rocha, S.T., Flynn, R.A., Bharadwaj, M., Calabrese, J.M., Magnuson, T., Heard, E., and 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>
5. Moindrot, B., Cerase, A., Coker, H., Masui, O., Grijzenhout, A., Pintacuda, G., Schermelleh, L., Nesterova, T.B., and Brockdorff, N. (2015). A Pooled shRNA Screen Identifies Rbm15, Spen, and Wtap as Factors Required for Xist RNA-Mediated Silencing. *Cell reports*, *12*(4), 562–572. <https://doi.org/10.1016/j.celrep.2015.06.053>
6. Monfort, A., Di Minin, G., Postlmayr, A., Freimann, R., Arieti, F., Thore, S., and Wutz, A. (2015). Identification of Spen as a Crucial Factor for Xist Function through Forward Genetic Screening in Haploid Embryonic Stem Cells. *Cell reports*, *12*(4), 554–561. <https://doi.org/10.1016/j.celrep.2015.06.067>
7. Kruger, K., Grabowski, P.J., Zaug, A.J., Sands, J., Gottschling, D.E., and Cech, T.R. (1982). Self-splicing RNA: autoexcision and autocyclization of the ribosomal RNA intervening sequence of Tetrahymena. *Cell*, *31*(1), 147–157. [https://doi.org/10.1016/0092-8674\(82\)90414-7](https://doi.org/10.1016/0092-8674(82)90414-7)
8. Steitz, T.A., and Steitz, J.A. (1993). A general two-metal-ion mechanism for catalytic RNA. *Proceedings of the National Academy of Sciences of the United States of America*, *90*(14), 6498–6502. <https://doi.org/10.1073/pnas.90.14.6498>
8. Cech, T.R., and 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>
9. Pasquinelli, A.E., Reinhart, B.J., Slack, F., Martindale, M.Q., Kuroda, M.I., Maller, B., Hayward, D.C., Ball, E.E., Degan, B., Müller, P., Spring, J., Srinivasan, A., Fishman, M., Finnerty, J., Corbo, J., Levine, M., Leahy, P., Davidson, E., and Ruvkun, G. (2000). Conservation of the sequence and temporal expression of let-7 heterochronic regulatory RNA. *Nature*, *408*(6808), 86–89. <https://doi.org/10.1038/35040556>
10. Lagos-Quintana, M., Rauhut, R., Lendeckel, W., & Tuschl, T. (2001). Identification of novel genes coding for small expressed RNAs. *Science (New York, N.Y.)*, *294*(5543), 853–858. <https://doi.org/10.1126/science.1064921>
11. Lau, N.C., Lim, L.P., Weinstein, E.G., and Bartel, D.P. (2001). An abundant class of tiny RNAs with probable regulatory roles in *Caenorhabditis elegans*. *Science (New York, N.Y.)*, *294*(5543), 858–862. <https://doi.org/10.1126/science.1065062>
12. Lee, R.C., and Ambros, V. (2001). An extensive class of small RNAs in *Caenorhabditis elegans*. *Science (New York, N.Y.)*, *294*(5543), 862–864. <https://doi.org/10.1126/science.1065329>

13. Gurtan, A.M., and Sharp, P.A. (2013). The role of miRNAs in regulating gene expression networks. *Journal of molecular biology*, 425(19), 3582–3600.
<https://doi.org/10.1016/j.jmb.2013.03.007>
14. Coutinho, M.F., Matos, L., Santos, J.I., and Alves, S. (2019). RNA Therapeutics: How Far Have We Gone?. *Advances in experimental medicine and biology*, 1157, 133–177.
https://doi.org/10.1007/978-3-030-19966-1_7
15. Okazaki, Y., Furuno, M., Kasukawa, T., Adachi, J., Bono, H., Kondo, S., Nikaido, I., Osato, N., Saito, R., Suzuki, H., Yamanaka, I., Kiyosawa, H., Yagi, K., Tomaru, Y., Hasegawa, Y., Nogami, A., Schönbach, C., Gojobori, T., Baldarelli, R., Hill, D.P., ... RIKEN Genome Exploration Research Group Phase I & II Team (2002). Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs. *Nature*, 420(6915), 563–573.
<https://doi.org/10.1038/nature01266>
16. Kapranov, P., Cheng, J., Dike, S., Nix, D. A., Duttagupta, R., Willingham, A. T., Stadler, P. F., Hertel, J., Hackermüller, J., Hofacker, I. L., Bell, I., Cheung, E., Drenkow, J., Dumais, E., Patel, S., Helt, G., Ganesh, M., Ghosh, S., Piccolboni, A., Sementchenko, V., ... Gingeras, T.R. (2007). RNA maps reveal new RNA classes and a possible function for pervasive transcription. *Science (New York, N.Y.)*, 316(5830), 1484–1488.
<https://doi.org/10.1126/science.1138341>
17. Engreitz, J.M., Haines, J.E., Perez, E.M., Munson, G., Chen, J., Kane, M., McDonel, P.E., Guttman, M., and Lander, E.S. (2016). Local regulation of gene expression by lncRNA promoters, transcription and splicing. *Nature*, 539(7629), 452–455.
<https://doi.org/10.1038/nature20149>
18. Ebisuya, M., Yamamoto, T., Nakajima, M., and Nishida, E. (2008). Ripples from neighbouring transcription. *Nature cell biology*, 10(9), 1106–1113.
<https://doi.org/10.1038/ncb177>
19. Hurst, L. D., Pál, C., and Lercher, M.J. (2004). The evolutionary dynamics of eukaryotic gene order. *Nature reviews. Genetics*, 5(4), 299–310. <https://doi.org/10.1038/nrg1319>
20. Purmann, A., Toedling, J., Schueler, M., Carninci, P., Lehrach, H., Hayashizaki, Y., Huber, W., and Sperling, S. (2007). Genomic organization of transcriptomes in mammals: Coregulation and cofunctionality. *Genomics*, 89(5), 580–587. <https://doi.org/10.1016/j.ygeno.2007.01.010>
21. Sproul, D., Gilbert, N., and Bickmore, W.A. (2005). The role of chromatin structure in regulating the expression of clustered genes. *Nature reviews. Genetics*, 6(10), 775–781.
<https://doi.org/10.1038/nrg1688>

22. Lee, M.P., DeBaun, M.R., Mitsuya, K., Galonek, H.L., Brandenburg, S., Oshimura, M., and Feinberg, A.P. (1999). Loss of imprinting of a paternally expressed transcript, with antisense orientation to KVLQT1, occurs frequently in Beckwith-Wiedemann syndrome and is independent of insulin-like growth factor II imprinting. *Proceedings of the National Academy of Sciences of the United States of America*, 96(9), 5203–5208. <https://doi.org/10.1073/pnas.96.9.5203>
23. Mabb, A.M., Judson, M.C., Zylka, M.J., and Philpot, B.D. (2011). Angelman syndrome: insights into genomic imprinting and neurodevelopmental phenotypes. *Trends in neurosciences*, 34(6), 293–303. <https://doi.org/10.1016/j.tins.2011.04.001>
22. Wevrick, R., Kerns, J.A., and Francke, U. (1994). Identification of a novel paternally expressed gene in the Prader-Willi syndrome region. *Human molecular genetics*, 3(10), 1877–1882. <https://doi.org/10.1093/hmg/3.10.1877>
23. Kanduri, C. (2016). Long noncoding RNAs: Lessons from genomic imprinting. *Biochimica et biophysica acta*, 1859(1), 102–111. <https://doi.org/10.1016/j.bbagr.2015.05.006>
24. Kurzrock, R., Kantarjian, H.M., Druker, B.J., and Talpaz, M. (2003). Philadelphia chromosome-positive leukemias: from basic mechanisms to molecular therapeutics. *Annals of internal medicine*, 138(10), 819–830. <https://doi.org/10.7326/0003-4819-138-10-200305200-00010>
25. Metzeler, K.H., and Bloomfield, C.D. (2017). Clinical Relevance of RUNX1 and CBFβ Alterations in Acute Myeloid Leukemia and Other Hematological Disorders. *Advances in experimental medicine and biology*, 962, 175–199. https://doi.org/10.1007/978-981-10-3233-2_12
26. Yu, N. Y., Hallström, B.M., Fagerberg, L., Ponten, F., Kawaji, H., Carninci, P., Forrest, A.R., Fantom Consortium, Hayashizaki, Y., Uhlén, M., and Daub, C.O. (2015). Complementing tissue characterization by integrating transcriptome profiling from the Human Protein Atlas and from the FANTOM5 consortium. *Nucleic acids research*, 43(14), 6787–6798. <https://doi.org/10.1093/nar/gkv608>
27. Li, L., Zhang, J.A., Dose, M., Kueh, H.Y., Mosadeghi, R., Gounari, F., and Rothenberg, E.V. (2013). A far downstream enhancer for murine Bcl11b controls its T-cell specific expression. *Blood*, 122(6), 902–911. <https://doi.org/10.1182/blood-2012-08-447839>
28. Engreitz, J.M., Ollikainen, N., and Guttman, M. (2016). Long non-coding RNAs: spatial amplifiers that control nuclear structure and gene expression. *Nature reviews. Molecular cell biology*, 17(12), 756–770. <https://doi.org/10.1038/nrm.2016.126>
29. Strehle, M., and 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>

30. Brockdorff, N., Ashworth, A., Kay, G.F., Cooper, P., Smith, S., McCabe, V.M., Norris, D.P., Penny, G.D., Patel, D., and Rastan, S. (1991). Conservation of position and exclusive expression of mouse Xist from the inactive X chromosome. *Nature*, 351(6324), 329–331. <https://doi.org/10.1038/351329a0>
31. Brown, C.J., Ballabio, A., Rupert, J.L., Lafreniere, R.G., Grompe, M., Tonlorenzi, R., and Willard, H.F. (1991). A gene from the region of the human X inactivation centre is expressed exclusively from the inactive X chromosome. *Nature*, 349(6304), 38–44. <https://doi.org/10.1038/349038a0>
32. Penny, G.D., Kay, G.F., Sheardown, S.A., Rastan, S., and Brockdorff, N. (1996). Requirement for Xist in X chromosome inactivation. *Nature*, 379(6561), 131–137. <https://doi.org/10.1038/379131a0>
33. Marahrens, Y., Panning, B., Dausman, J., Strauss, W., and Jaenisch, R. (1997). Xist-deficient mice are defective in dosage compensation but not spermatogenesis. *Genes & development*, 11(2), 156–166. <https://doi.org/10.1101/gad.11.2.156>
34. Wutz, A., and Jaenisch, R. (2000). A shift from reversible to irreversible X inactivation is triggered during ES cell differentiation. *Molecular cell*, 5(4), 695–705. [https://doi.org/10.1016/s1097-2765\(00\)80248-8](https://doi.org/10.1016/s1097-2765(00)80248-8)
35. Wutz, A., Rasmussen, T.P., and Jaenisch, R. (2002). Chromosomal silencing and localization are mediated by different domains of Xist RNA. *Nature genetics*, 30(2), 167–174. <https://doi.org/10.1038/ng820>
36. Minajigi, A., Froberg, J., Wei, C., Sunwoo, H., Kesner, B., Colognori, D., Lessing, D., Payer, B., Boukhali, M., Haas, W., and Lee, J.T. (2015). Chromosomes. 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>
37. Barr, M.L., and Bertram, E.G. (1949). A morphological distinction between neurones of the male and female, and the behaviour of the nucleolar satellite during accelerated nucleoprotein synthesis. *Nature*, 163(4148), 676. <https://doi.org/10.1038/163676a0>
38. Dyer, K.A., Canfield, T.K., and Gartler, S.M. (1989). Molecular cytological differentiation of active from inactive X domains in interphase: implications for X chromosome inactivation. *Cytogenetics and cell genetics*, 50(2-3), 116–120. <https://doi.org/10.1159/000132736>
39. Chen, C.K., Blanco, M., Jackson, C., Aznauryan, E., Ollikainen, N., Surka, C., Chow, A., Cerase, A., McDonel, P., & Guttman, M. (2016). Xist recruits the X chromosome to the nuclear lamina to enable chromosome-wide silencing. *Science (New York, N.Y.)*, 354(6311), 468–472. <https://doi.org/10.1126/science.aae0047>

40. Chaumeil, J., Le Baccon, P., Wutz, A., and Heard, E. (2006). A novel role for Xist RNA in the formation of a repressive nuclear compartment into which genes are recruited when silenced. *Genes & development*, 20(16), 2223–2237. <https://doi.org/10.1101/gad.380906>
41. Nozawa, R.S., Nagao, K., Igami, K.T., Shibata, S., Shirai, N., Nozaki, N., Sado, T., Kimura, H., and Obuse, C. (2013). Human inactive X chromosome is compacted through a PRC2-independent SMCHD1-HbiX1 pathway. *Nature structural & molecular biology*, 20(5), 566–573. <https://doi.org/10.1038/nsmb.2532>
42. Giorgetti, L., Lajoie, B.R., Carter, A.C., Attia, M., Zhan, Y., Xu, J., Chen, C.J., Kaplan, N., Chang, H.Y., Heard, E., and Dekker, J. (2016). Structural organization of the inactive X chromosome in the mouse. *Nature*, 535(7613), 575–579. <https://doi.org/10.1038/nature18589>
43. Gdula, M.R., Nesterova, T.B., Pintacuda, G., Godwin, J., Zhan, Y., Ozadam, H., McClellan, M., Moralli, D., Krueger, F., Green, C.M., Reik, W., Kriaucionis, S., Heard, E., Dekker, J., and Brockdorff, N. (2019). The non-canonical SMC protein SmcHD1 antagonises TAD formation and compartmentalization on the inactive X chromosome. *Nature communications*, 10(1), 30. <https://doi.org/10.1038/s41467-018-07907-2>
44. Wang, C. Y., Jégu, T., Chu, H. P., Oh, H. J., and Lee, J. T. (2018). SMCHD1 Merges Chromosome Compartments and Assists Formation of Super-Structures on the Inactive X. *Cell*, 174(2), 406–421.e25. <https://doi.org/10.1016/j.cell.2018.05.007>
45. Nora, E.P., Lajoie, B.R., Schulz, E.G., Giorgetti, L., Okamoto, I., Servant, N., Piolot, T., van Berkum, N.L., Meisig, J., Sedat, J., Gribnau, J., Barillot, E., Blüthgen, N., Dekker, J., & Heard, E. (2012). Spatial partitioning of the regulatory landscape of the X-inactivation centre. *Nature*, 485(7398), 381–385. <https://doi.org/10.1038/nature11049>
46. Brown, C.J., and Willard, H.F. (1994). The human X-inactivation centre is not required for maintenance of X-chromosome inactivation. *Nature*, 368(6467), 154–156. <https://doi.org/10.1038/368154a0>
47. Su, H., Liu, Y., and Zhao, X. (2015). Split End Family RNA Binding Proteins: Novel Tumor Suppressors Coupling Transcriptional Regulation with RNA Processing. *Cancer Translational Medicine*, 1, 21-25.
48. Maier, D. (2019). The evolution of transcriptional repressors in the Notch signaling pathway: a computational analysis. *Hereditas*, 156, 5. <https://doi.org/10.1186/s41065-019-0081-0>
49. Kuroda, K., Han, H., Tani, S., Tanigaki, K., Tun, T., Furukawa, T., Taniguchi, Y., Kurooka, H., Hamada, Y., Toyokuni, S., and Honjo, T. (2003). Regulation of marginal zone B cell development by MINT, a suppressor of Notch/RBP-J signaling pathway. *Immunity*, 18(2), 301–312. [https://doi.org/10.1016/s1074-7613\(03\)00029-3](https://doi.org/10.1016/s1074-7613(03)00029-3)

50. Dossin, F., Pinheiro, I., Żylicz, J.J., Roensch, J., Collombet, S., Le Saux, A., Chelmicki, T., Attia, M., Kapoor, V., Zhan, Y., Dingli, F., Loew, D., Mercher, T., Dekker, J., and Heard, E. (2020). SPEN integrates transcriptional and epigenetic control of X-inactivation. *Nature*, 578(7795), 455–460. <https://doi.org/10.1038/s41586-020-1974-9>
51. Tamura, K., Taniguchi, Y., Minoguchi, S., Sakai, T., Tun, T., Furukawa, T., and Honjo, T. (1995). Physical interaction between a novel domain of the receptor Notch and the transcription factor RBP-J kappa/Su(H). *Current biology : CB*, 5(12), 1416–1423. [https://doi.org/10.1016/s0960-9822\(95\)00279-x](https://doi.org/10.1016/s0960-9822(95)00279-x)
52. Taniguchi, Y., Furukawa, T., Tun, T., Han, H., and Honjo, T. (1998). LIM protein KyoT2 negatively regulates transcription by association with the RBP-J DNA-binding protein. *Molecular and cellular biology*, 18(1), 644–654. <https://doi.org/10.1128/mcb.18.1.644>
53. Oswald, F., Kostezka, U., Astrahantseff, K., Bourteele, S., Dillinger, K., Zechner, U., Ludwig, L., Wilda, M., Hameister, H., Knöchel, W., Liptay, S., and Schmid, R. M. (2002). SHARP is a novel component of the Notch/RBP-Jkappa signalling pathway. *The EMBO journal*, 21(20), 5417–5426. <https://doi.org/10.1093/emboj/cdf549>
54. Shi, Y., Downes, M., Xie, W., Kao, H.Y., Ordentlich, P., Tsai, C.C., Hon, M., & Evans, R.M. (2001). Sharp, an inducible cofactor that integrates nuclear receptor repression and activation. *Genes & development*, 15(9), 1140–1151. <https://doi.org/10.1101/gad.871201>
55. Lu, Z., Zhang, Q.C., Lee, B., Flynn, R.A., Smith, M.A., Robinson, J.T., Davidovich, C., Gooding, A.R., Goodrich, K.J., Mattick, J.S., Mesirov, J.P., Cech, T.R., and Chang, H.Y. (2016). RNA Duplex Map in Living Cells Reveals Higher-Order Transcriptome Structure. *Cell*, 165(5), 1267–1279. <https://doi.org/10.1016/j.cell.2016.04.028>
56. 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., and 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>
57. Smilinich, N.J., Day, C.D., Fitzpatrick, G.V., Caldwell, G.M., Lossie, A.C., Cooper, P.R., Smallwood, A.C., Joyce, J.A., Schofield, P.N., Reik, W., Nicholls, R.D., Weksberg, R., Driscoll, D.J., Maher, E.R., Shows, T.B., and Higgins, M.J. (1999). A maternally methylated CpG island in KvLQT1 is associated with an antisense paternal transcript and loss of imprinting in Beckwith-Wiedemann syndrome. *Proceedings of the National Academy of Sciences of the United States of America*, 96(14), 8064–8069. <https://doi.org/10.1073/pnas.96.14.8064>
58. Mancini-Dinardo, D., Steele, S.J., Levorse, J.M., Ingram, R.S., and Tilghman, S.M. (2006). Elongation of the Kcnqlot1 transcript is required for genomic imprinting of neighboring genes. *Genes & development*, 20(10), 1268–1282. <https://doi.org/10.1101/gad.1416906>

59. Mohammad, F., Mondal, T., Guseva, N., Pandey, G.K., and Kanduri, C. (2010). Kcnq1ot1 noncoding RNA mediates transcriptional gene silencing by interacting with Dnmt1. *Development (Cambridge, England)*, 137(15), 2493–2499. <https://doi.org/10.1242/dev.048181>
60. Mohammad, F., Pandey, G.K., Mondal, T., Enroth, S., Redrup, L., Gyllensten, U., and Kanduri, C. (2012). Long noncoding RNA-mediated maintenance of DNA methylation and transcriptional gene silencing. *Development (Cambridge, England)*, 139(15), 2792–2803. <https://doi.org/10.1242/dev.079566>
61. Umlauf, D., Goto, Y., Cao, R., Cerqueira, F., Wagschal, A., Zhang, Y., and Feil, R. (2004). Imprinting along the Kcnq1 domain on mouse chromosome 7 involves repressive histone methylation and recruitment of Polycomb group complexes. *Nature genetics*, 36(12), 1296–1300. <https://doi.org/10.1038/ng1467>
62. Fitzpatrick, G.V., Soloway, P.D., and Higgins, M.J. (2002). Regional loss of imprinting and growth deficiency in mice with a targeted deletion of KvDMR1. *Nature genetics*, 32(3), 426–431. <https://doi.org/10.1038/ng988>
63. Weksberg, R., Shuman, C., and Beckwith, J.B. (2010). Beckwith-Wiedemann syndrome. *European journal of human genetics : EJHG*, 18(1), 8–14. <https://doi.org/10.1038/ejhg.2009.106>
64. 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., and Ponting, C.P. (2014). Considerations when investigating lncRNA function in vivo. *eLife*, 3, e03058. <https://doi.org/10.7554/eLife.03058>
65. Zhang, H., Zeitz, M.J., Wang, H., Niu, B., Ge, S., Li, W., Cui, J., Wang, G., Qian, G., Higgins, M.J., Fan, X., Hoffman, A.R., and Hu, J.F. (2014). Long noncoding RNA-mediated intrachromosomal interactions promote imprinting at the Kcnq1 locus. *The Journal of cell biology*, 204(1), 61–75. <https://doi.org/10.1083/jcb.201304152>
66. Thakur, N., Kanduri, M., Holmgren, C., Mukhopadhyay, R., and Kanduri, C. (2003). Bidirectional silencing and DNA methylation-sensitive methylation-spreading properties of the Kcnq1 imprinting control region map to the same regions. *The Journal of biological chemistry*, 278(11), 9514–9519. <https://doi.org/10.1074/jbc.M212203200>
67. Kanduri, C., Thakur, N., and Pandey, R.R. (2006). The length of the transcript encoded from the Kcnq1ot1 antisense promoter determines the degree of silencing. *The EMBO journal*, 25(10), 2096–2106. <https://doi.org/10.1038/sj.emboj.7601090>
68. Pandey, R.R., Mondal, T., Mohammad, F., Enroth, S., Redrup, L., Komorowski, J., Nagano, T., Mancini-Dinardo, D., and 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>

69. Mohammad, F., Pandey, R.R., Nagano, T., Chakalova, L., Mondal, T., Fraser, P., and Kanduri, C. (2008). Kcnq1ot1/Lit1 noncoding RNA mediates transcriptional silencing by targeting to the perinucleolar region. *Molecular and cellular biology*, 28(11), 3713–3728. <https://doi.org/10.1128/MCB.02263-07>
70. Vigushin, D.M., Ali, S., Pace, P.E., Mirsaidi, N., Ito, K., Adcock, I., and Coombes, R.C. (2001). Trichostatin A is a histone deacetylase inhibitor with potent antitumor activity against breast cancer in vivo. *Clinical cancer research : an official journal of the American Association for Cancer Research*, 7(4), 971–976.
71. Cho, S.W., Kim, S., Kim, J.M., and Kim, J.S. (2013). Targeted genome engineering in human cells with the Cas9 RNA-guided endonuclease. *Nature biotechnology*, 31(3), 230–232. <https://doi.org/10.1038/nbt.2507>
72. Lin, Y., Protter, D.S., Rosen, M.K., and Parker, R. (2015). Formation and Maturation of Phase-Separated Liquid Droplets by RNA-Binding Proteins. *Molecular cell*, 60(2), 208–219. <https://doi.org/10.1016/j.molcel.2015.08.018>
73. Banani, S.F., Rice, A.M., Peeples, W.B., Lin, Y., Jain, S., Parker, R., and Rosen, M.K. (2016). Compositional Control of Phase-Separated Cellular Bodies. *Cell*, 166(3), 651–663. <https://doi.org/10.1016/j.cell.2016.06.010>
74. Alberti, S., Saha, S., Woodruff, J.B., Franzmann, T.M., Wang, J., and Hyman, A.A. (2018). A User's Guide for Phase Separation Assays with Purified Proteins. *Journal of molecular biology*, 430(23), 4806–4820. <https://doi.org/10.1016/j.jmb.2018.06.038>
75. Mancini-DiNardo, D., Steele, S.J., Ingram, R.S., and Tilghman, S.M. (2003). A differentially methylated region within the gene Kcnq1 functions as an imprinted promoter and silencer. *Human molecular genetics*, 12(3), 283–294. <https://doi.org/10.1093/hmg/ddg024>
76. Quinodoz, S.A., Ollikainen, N., Tabak, B., Palla, A., Schmidt, J. M., Detmar, E., Lai, M.M., Shishkin, A.A., Bhat, P., Takei, Y., Trinh, V., Aznauryan, E., Russell, P., Cheng, C., Jovanovic, M., Chow, A., Cai, L., McDonel, P., Garber, M., and Guttman, M. (2018). Higher-Order Inter-chromosomal Hubs Shape 3D Genome Organization in the Nucleus. *Cell*, 174(3), 744–757.e24. <https://doi.org/10.1016/j.cell.2018.05.024>
77. Van Nostrand, E.L., Pratt, G.A., Shishkin, A.A., Gelboin-Burkhart, C., Fang, M.Y., Sundararaman, B., Blue, S.M., Nguyen, T.B., Surka, C., Elkins, K., Stanton, R., Rigo, F., Guttman, M., and Yeo, G.W. (2016). Robust transcriptome-wide discovery of RNA-binding protein binding sites with enhanced CLIP (eCLIP). *Nature methods*, 13(6), 508–514. <https://doi.org/10.1038/nmeth.3810>
78. Bolger, A.M., Lohse, M., and Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics (Oxford, England)*, 30(15), 2114–2120. <https://doi.org/10.1093/bioinformatics/btu170>

79. Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., and Gingeras, T. R. (2013). STAR: ultrafast universal RNA-seq aligner. *Bioinformatics (Oxford, England)*, 29(1), 15–21. <https://doi.org/10.1093/bioinformatics/bts635>
80. Quinodoz, S.A., Bhat, P., Ollikainen, N., Jachowicz, J., Banerjee, A.K., Chovanec, P., Blanco, M.R., Chow, A., Markaki, Y., Plath, K. and Guttman, M. (2020). RNA promotes the formation of spatial compartments in the nucleus. bioRxiv 2020.08.25.267435; doi: <https://doi.org/10.1101/2020.08.25.267435>
81. Shishkin, A.A., Giannoukos, G., Kucukural, A., Ciulla, D., Busby, M., Surka, C., Chen, J., Bhattacharyya, R.P., Rudy, R.F., Patel, M.M., Novod, N., Hung, D.T., Gnirke, A., Garber, M., Guttman, M., and Livny, J. (2015). Simultaneous generation of many RNA-seq libraries in a single reaction. *Nature methods*, 12(4), 323–325. <https://doi.org/10.1038/nmeth.3313>
82. Kechin, A., Boyarskikh, U., Kel, A., and Filipenko, M. (2017). cutPrimers: A New Tool for Accurate Cutting of Primers from Reads of Targeted Next Generation Sequencing. *Journal of computational biology : a journal of computational molecular cell biology*, 24(11), 1138–1143. <https://doi.org/10.1089/cmb.2017.0096>
83. Jin, Y., Tam, O.H., Paniagua, E., and Hammell, M. (2015). Tetrascripts: 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>
84. Langmead, B., and Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. *Nature methods*, 9(4), 357–359. <https://doi.org/10.1038/nmeth.1923>
85. Krueger, F., and Andrews, S.R. (2016). SNPsplite: Allele-specific splitting of alignments between genomes with known SNP genotypes. *F1000Research*, 5, 1479. <https://doi.org/10.12688/f1000research.9037.2>
86. Smit, A., Hubley, R., and Grenn, P. RepeatMasker Open-4.0. RepeatMasker Open-4.0.7. (2015).
87. Quinlan, A.R., and 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>
88. Ewels, P., Magnusson, M., Lundin, S., and Källér, M. (2016). MultiQC: summarize analysis results for multiple tools and samples in a single report. *Bioinformatics (Oxford, England)*, 32(19), 3047–3048. <https://doi.org/10.1093/bioinformatics/btw354>
89. Kuehn, M. R., Bradley, A., Robertson, E. J., & Evans, M. J. (1987). A potential animal model for Lesch-Nyhan syndrome through introduction of HPRT mutations into mice. *Nature*, 326(6110), 295–298. <https://doi.org/10.1038/326295a0>

90. Baron-Benhamou, J., Gehring, N.H., Kulozik, A.E., and Hentze, M.W. (2004). Using the lambdaN peptide to tether proteins to RNAs. *Methods in molecular biology (Clifton, N.J.)*, 257, 135–154. <https://doi.org/10.1385/1-59259-750-5:135>
91. Zhao, S., Jiang, E., Chen, S., Gu, Y., Shangguan, A. J., Lv, T., Luo, L., and Yu, Z. (2016). PiggyBac transposon vectors: the tools of the human gene encoding. *Translational lung cancer research*, 5(1), 120–125. <https://doi.org/10.3978/j.issn.2218-6751.2016.01.05>
92. Halpern, M.E., Rhee, J., Goll, M.G., Akitake, C.M., Parsons, M., and Leach, S.D. (2008). Gal4/UAS transgenic tools and their application to zebrafish. *Zebrafish*, 5(2), 97–110. <https://doi.org/10.1089/zeb.2008.0530>

7.0 FIGURES

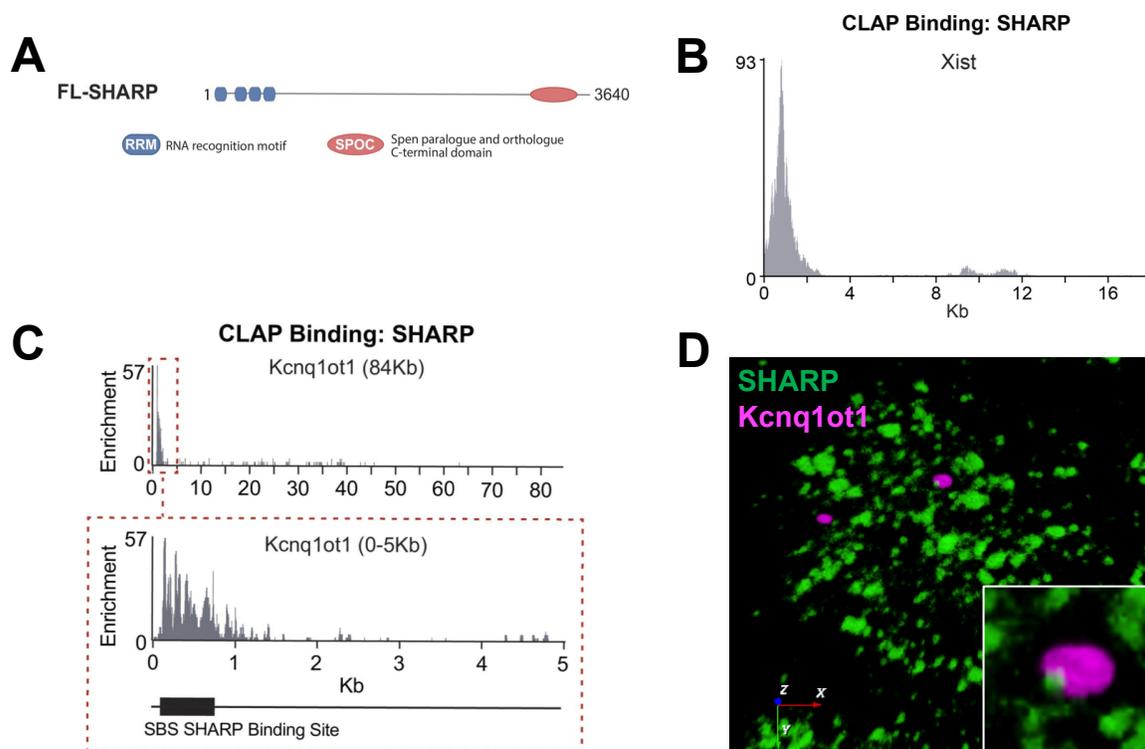
Figure 1

Figure 1: SHARP Covalent Linkage Affinity Purification reveals interaction with lncRNA *Kcnq1ot1*. (A) Diagram of the functional domains in full length (FL) SHARP (also referred to as Spen) protein. SHARP contains four RNA recognition motif (RRM, blue) domains and one Spen paralogue and orthologue C-terminal (SPOC, orange) domain. The region in between the RRM and SPOC domains are referred to as Intrinsically Disordered Domains (IDR). (B) SHARP preferentially binds to the 0-2kb region of the *Xist* lncRNA. (C) SHARP protein binding on the entire 84,000 nucleotide *Kcnq1ot1* lncRNA (top), and a zoom in on the first 5,000 nucleotides of the lncRNA (bottom). We define a region called the SHARP binding site (SBS) shown as a black box. (D) Confocal 3D image reconstruction of mouse embryonic stem cell, visualized with SHARP immunostaining and *Kcnq1ot1* RNA-FISH.

Figure 2

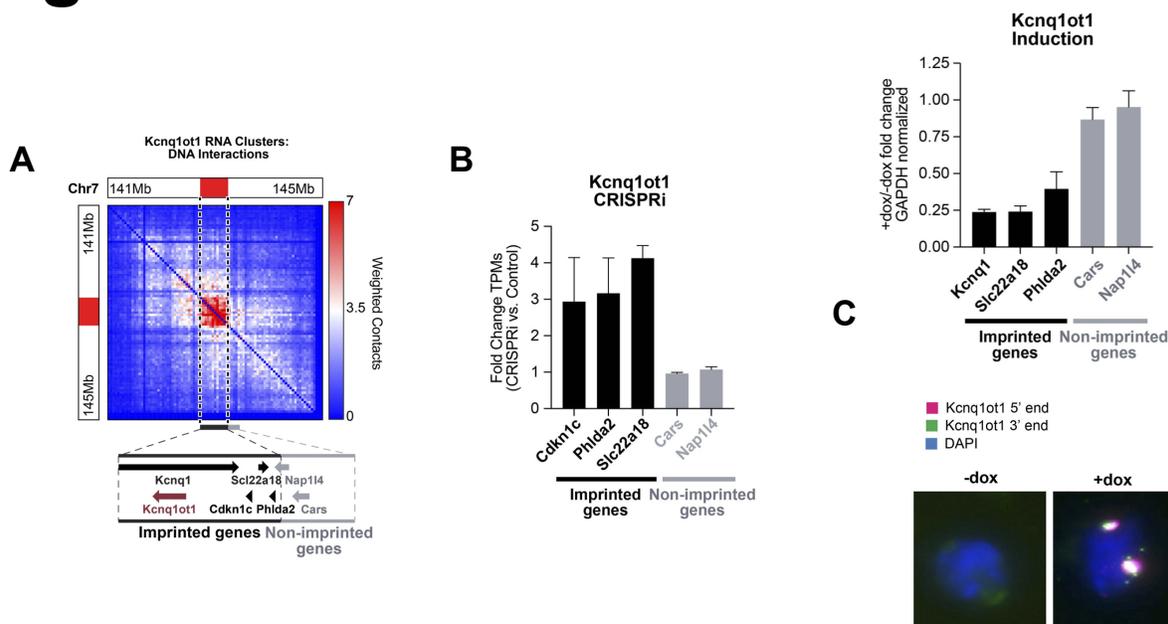


Figure 2: Kcnq1ot1 is necessary and sufficient to seed an RNA-mediated compartment over its imprinted target genes. (A) Weighted DNA-DNA contacts within SPRITE clusters containing the Kcnq1ot1 RNA. Vertical dashed lines indicate the boundaries of the Kcnq1ot1-enriched spatial compartment. (Zoom out) Genomic locations in this domain of the Kcnq1ot1 gene (burgundy) and imprinted target genes Kcnq1, Slc22a18, Cdkn1c, and Phlda2 (black) and non-imprinted neighboring genes Nap114 and Cars (gray). (B) Changes in mean gene expression upon CRISPR inhibition (CRISPRi) of the Kcnq1ot1 lncRNA. Genes contained within the Kcnq1ot1-associated domain (e.g. Cdkn1c, Phlda2, Slc22a18) are shown in black and genes outside the domain (e.g. Cars, Nap114) are shown in gray. Error bars represent standard deviation. (C) Changes in mean gene expression upon doxycycline mediated induction (+Dox) of Kcnq1ot1 relative to cells with no doxycycline (-Dox). Genes contained within the Kcnq1ot1-associated domain (e.g. Cdkn1c, Phlda2, Slc22a18) are shown in black and genes outside the

domain (e.g. Cars, Nap114) are shown in gray. Error bars represent standard deviation. RNA-FISH of inducible line with or without addition of doxycycline; probes aligned against two separate regions of *Kcnq1ot1* lncRNA in two separate colors.

Figure 3

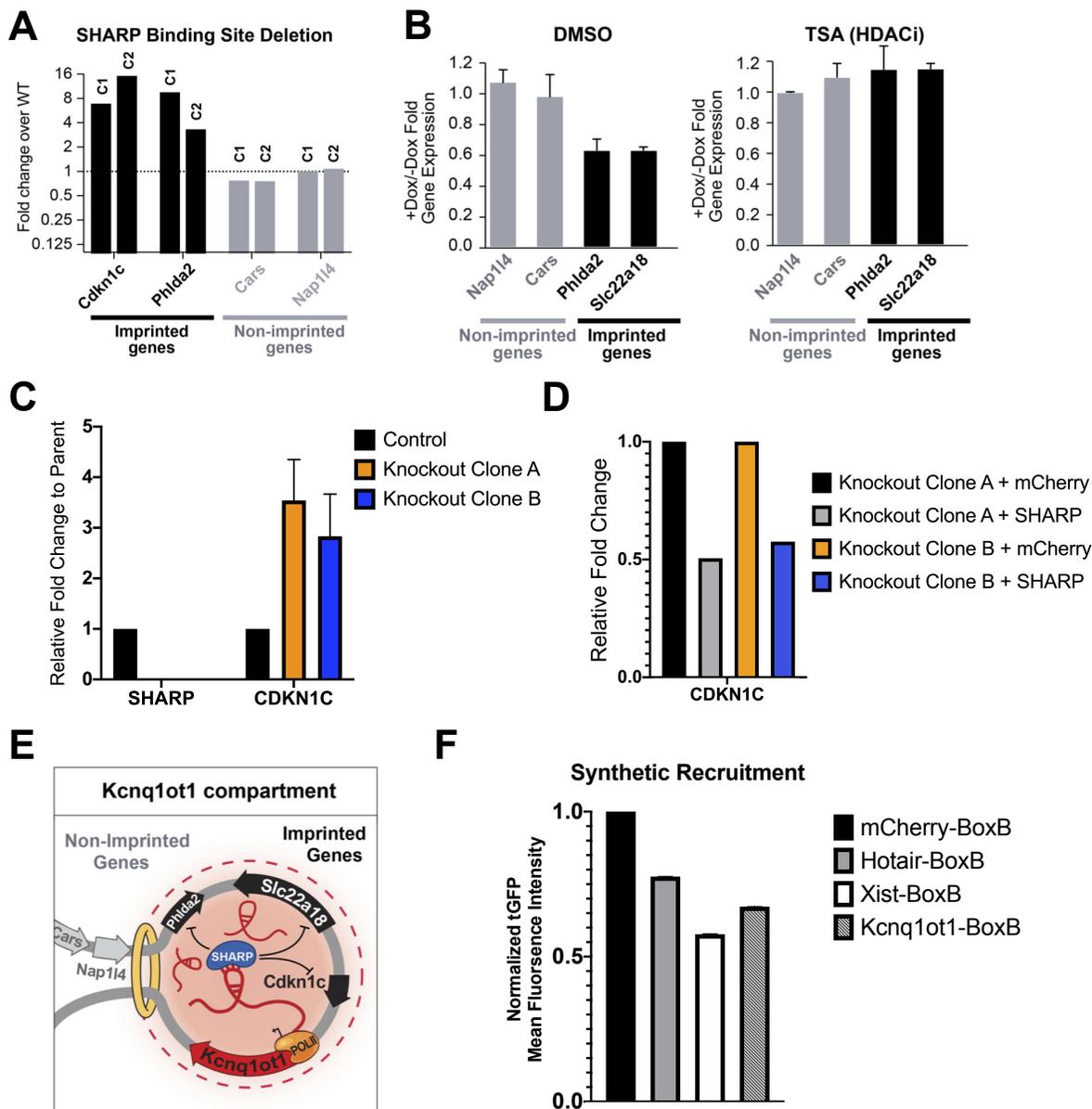


Figure 3: Kcnq1ot1 transcriptionally silences imprint-target genes in a SHARP-dependent manner. (A) RT-qPCR analysis of two lines engineered to have homozygous deletions of the SHARP binding site (Delta SBS) within Kcnq1ot1. (B) Mean gene expression differences of Kcnq1ot1-regulated and Kcnq1ot1-non-regulated genes between induced (+Dox) and non-induced (-Dox) samples treated with DMSO (left) or the HDAC inhibitor, Trichostatin A (TSA)

(right). Error bars represent standard deviation. **(C and D)** CDKN1C is repressed by SHARP in a SHARP-dependent manner. Data presented is GapdH-normalized gene expression to parent control line (**SHARP Deletion**) and relative fold change from sequencing data between control and rescue (**SHARP Rescue**). Primer targeting SHARP was designed against gRNA-targeted region (SHARP Exon 3-5). **(E)** Cartoon of Kcnq1ot1-mediated compartmentalization of SHARP in the Kcnq1 topological associated domain. **(F)** Flow cytometry analysis of synthetic recruitment of select RNAs to PGK-tGFP locus. Data is presented as normalized mean fluorescence intensity, with experimental conditions normalized to mCherry control. Standard deviation is presented as inter-experimental variability between two replicate experiments.

Figure 4

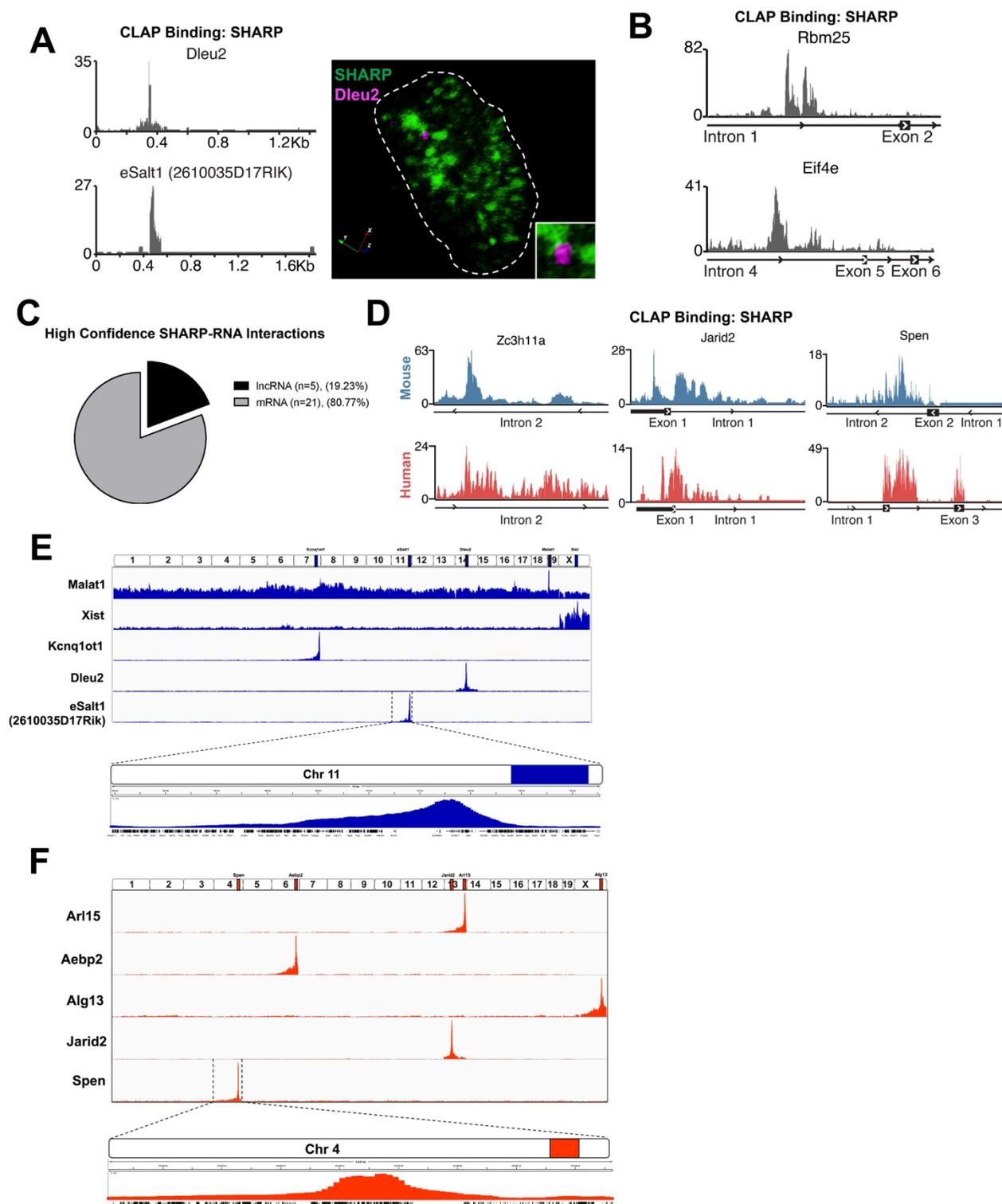


Figure 4: SHARP-RNA interactions are pervasive throughout the nucleus. (A) Input normalized enrichment plot of SHARP CLAP-Seq to Dleu2 and 2610035D17Rik lncRNAs, along with confocal 3D image reconstruction of mouse embryonic stem cell, visualized with SHARP immunostaining and Dleu2 RNA-FISH. **(B)** Input-normalized enrichment plot of SHARP CLAP-Seq to Rbm25 and Eif4e mRNA. **(C)** Representation of SHARP-RNA interactions identified from mouse embryonic stem cell SHARP CLAP-Seq. **(D)** Input-normalized enrichment plot of SHARP CLAP-Seq to Zc3h11a, Jarid2, and Spen (SHARP) mRNAs across mouse (blue) and human (red). **(E and F)** Genome-wide localization of individual **(E)** SHARP-interacting lncRNAs (blue) and **(F)** SHARP-interacting pre-mRNAs (red). Malat1, while not a SHARP-interacting lncRNA, is included to contrast *cis* versus *trans* localization on chromatin.

Figure 5

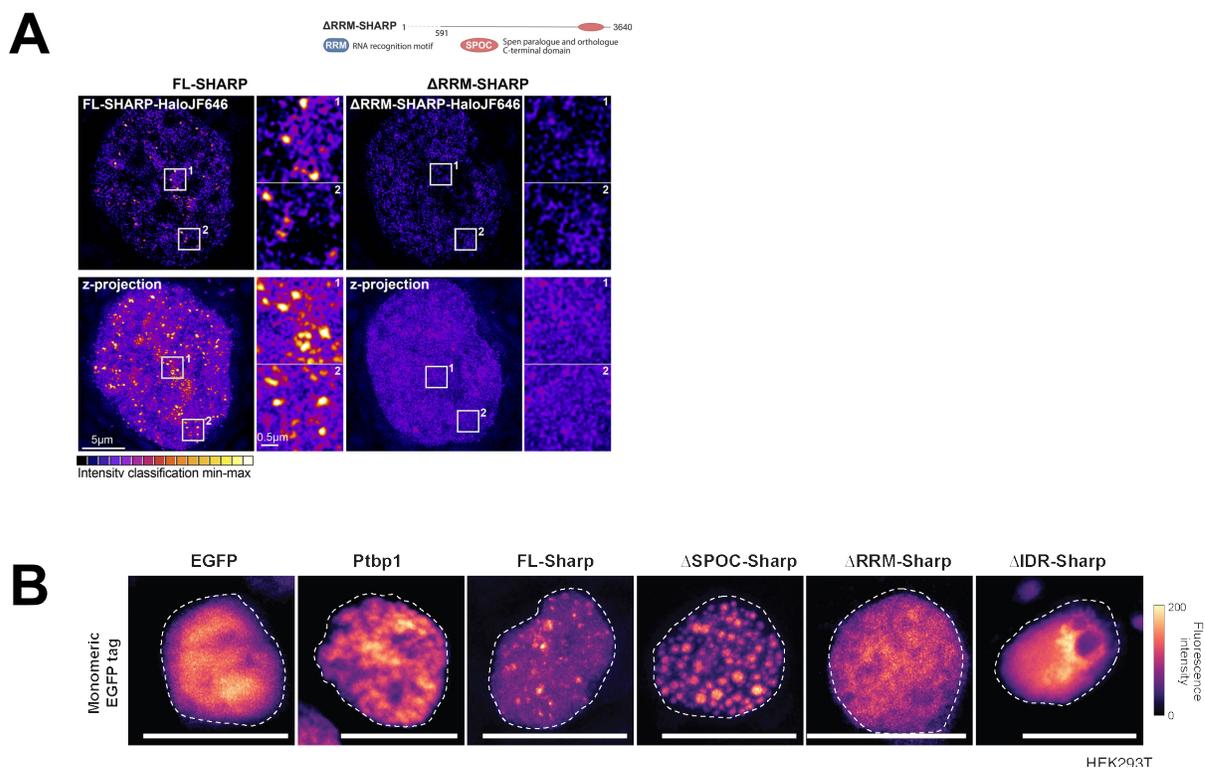


Figure 5: SHARP aggregates in a RRM- and IDR-dependent Manner. (A) (Top) 3D-Structural Illumination Microscopy 125 nm optical sections of FL-SHARP (left) and Delta RRM-SHARP (right) and z-projections (bottom) Halo-tagged FL- and Delta RRM-SHARPJF646. FL-SHARP localizes in foci throughout the nucleus (zoom in panels 1-2), while Delta RRM-SHARP leads to diffusive localization. Bar: 5 μ m, insets: 0.5 μ m (magnifications). Intensities are depicted in 16-color grading from black (minimum) to white (maximum). Diagram of the functional domains of the Delta RRM SHARP mutant, generated by deletion of the first 591 amino acids is shown above. **B)** Images representing localization of meGFP-tagged FL-SHARP, Delta RRM-SHARP, Delta IDR-SHARP, as well as controls, in transiently transfected HEK293T cells. Images shown as max projections; scale bars show 10 μ m.

Table 1: List of High Confidence SHARP-Interacting RNAs

RNA Class	Gene Symbol	Name
mRNA	Aebp2	AE binding Protein 2
mRNA	Alg13	Asparagine-Linked Glycosylation 13
mRNA	Arl15	ADP-Ribosylation Factor-like 15
mRNA	Eif4e	Eukaryotic Translation Initiation Factor 4E
mRNA	Fkbp3	FK506 Binding Protein 3
mRNA	HnrnpC	Heterogeneous Nuclear Ribonucleoprotein C
mRNA	Jarid2	Jumonji, AT- Rich Interactive Domain 2
mRNA	Luc7l2	Luc7-like 2
mRNA	Mbtd1	MBT Domain Containing 1
mRNA	Psme4	Proteasome Activator Subunit 4
mRNA	Rbm25	RNA Binding Motif Protein 25
mRNA	Serbp1	Serpine1 mRNA Binding Protein 1
mRNA	Smg1	SMG1 Homolog, Phosphatidylinositol 3-kinase related kinase
mRNA	Spen	Split Ends Transcriptional Repressor (SHARP)
mRNA	Sumo2	Small Ubiquitin-like Modifier 2
mRNA	U2surp	U2 snRNP-Associated SURP Domain Containing
mRNA	Ube2d3	Ubiquitin-conjugating enzyme E2D
mRNA	Ubxn2a	UBX Domain Protein 2A
mRNA	Zc3h11a	Zinc Finger CCCH Type Containing 11A
mRNA	Zfp42	Zinc Finger Protein 42, Rex-1
lncRNA	2610035D17Rik	<i>Embryonic stem cell SHARP-Associated Long Coding Transcript 1, eSalt1</i>
lncRNA	1810026B05Rik	Chaserr (CHD2 adjacent suppressive regulatory RNA)
lncRNA	Dleu2	Deleted In Lymphocytic Leukemia 2
lncRNA	Kcnq1ot1	KCNQ1 overlapping transcript 1
lncRNA	Pvt1	Pvt1 oncogene
lncRNA	Xist	X inactive specific transcript

*Chapter 4*THE SMRT/HDAC1-ASSOCIATED REPRESSOR PROTEIN POSITIVELY BUFFERS
NOTCH SIGNALING IN EARLY T-CELL DEVELOPMENT

A.K. Banerjee, X. Wang, H. Hosokawa, J. Wangen, M. Guttman, and E.V. Rothenberg

1. ABSTRACT

During early T-cell development, multipotent hematopoietic progenitor cells systematically lose alternative cell fate potential and acquire components of the T-cell developmental program.

Notch signaling is critical for early T-cell development and lineage commitment, however in spite of this essential role, the transcriptional regulation and downstream targets of Notch signaling remain unknown. One important regulator of Notch signaling is the SMRT/HDAC1 Associated Repressor Protein (SHARP), which has been demonstrated to act as a transcriptional repressor and negative regulator of the pathway in other contexts. In contrast to Notch pathway gain of function (which results in ectopic T-cell differentiation and tumorigenesis), SHARP deficiency results in stage-specific developmental delay. This suggests that SHARP's relationship to Notch activity may be more nuanced in early T-cell development. Here we demonstrate that SHARP acts to buffer a Notch-mediated Incoherent Feedforward Loop. By repressing a key negative regulator of Notch activity, Notch-Regulated Ankyrin Repeat Protein (Nrarp), SHARP exerts an overall positive influence on the circuit. SHARP is also a well-defined RNA-binding protein. We demonstrate that SHARP can recognize its own pre-mRNA and participate in RNA-mediated autofeedback, suggesting that RNA-binding may add an additional layer of regulatory complexity to developmentally sensitive circuits and processes.

2.0 INTRODUCTION

2.1 NOTCH SIGNAL TRANSDUCTION AND EARLY T-CELL DEVELOPMENT

Early T-cell development is a highly stereotyped biological process, through which multipotent hematopoietic progenitor cells systematically lose alternative cell fate potential and acquire

components of the T-cell developmental program¹. Notch signaling is critical for early T-cell development and lineage commitment, and dysregulation of the pathway has been linked to a variety of hematologic and solid malignancies, including T-Acute lymphoblastic leukemia and pancreatic cancer etc..¹⁻⁴

At its core, Notch signaling relies on the interactions between Notch-family receptors and their cognate, transmembrane-bound Delta-like/Jagged family ligands. Upon ligand recognition and binding, Notch receptors undergo a series of proteolytic cleavage events. This results in the release and translocation of Notch intracellular domain (Notch^{ICD}) into the cell nucleus. Within the nucleus, Notch^{ICD} can then interact with its DNA-binding partner, Recombination Signal Binding Protein-Jk (RBPJk, also known as CSL or SuH), to activate expression of target genes (**Figure 1**)³⁻⁶. Despite the essential role for Notch signaling in early T-cell development, its transcriptional regulation and downstream targets remain critical mysteries in the field.

2.2 THE RELATIONSHIP BETWEEN SHARP AND NOTCH PATHWAY REGULATION IS COMPLEX

One important regulator of Notch signaling is SMRT/HDAC1 Associated Repressor Protein (SHARP), also known as Split Ends Transcriptional Repressor (Spen) or Msx2-Interacting Nuclear Target Protein (Mint)⁷⁻¹⁰. SHARP is a member of the Split Ends protein family, which consists of SHARP, RBM15, and RBM15B¹¹. It is an approximately 3600 amino acid protein and contains four N-terminal RNA-Recognition Motifs and a C-terminal Spen Paralog and Ortholog C-terminal (SPOC) domain, demonstrated to be responsible for recruitment of SMRT (Silencing Mediator for Retinoid or Thyroid-Hormone Receptor, also known as Nuclear Receptor Co-Repressor 2 or NCOR2) and ultimately transcriptional silencing¹²⁻¹³.

SHARP has been shown to compete with Notch^{ICD} for RBPJk binding and suppress Notch-mediated gene activation; it is thought to act as a transcriptional repressor, recruiting co-repressive complexes to remodel chromatin and repress genomic loci^{9,14-16,12}. Moreover, SHARP has been demonstrated to regulate several Notch-mediated processes, including marginal vs. follicular B-cell choice¹⁶. Despite evidence supporting SHARP's role as a negative regulator of Notch in other developmental contexts, its relationship to Notch signaling is unclear in early T-cell development. In contrast to Notch pathway gain of function which results in ectopic T-cell differentiation and tumorigenesis, SHARP deficiency results in stage-specific developmental delay^{15,17}. What is the relationship between SHARP and Notch signaling in early T-cell development, and what are possible impacts of SHARP's ability to recognize RNA?

3.0 RESULTS

3.1 RESULTS: SHARP DEFICIENCY DOES NOT STRICTLY PHENOCOPY NOTCH GAIN OF FUNCTION

We first aimed to recapitulate the developmental phenotype associated with SHARP deficiency¹⁵. Cas9-expressing E13.5 fetal liver progenitors were retrovirally transduced with pools of guide RNA retroviruses targeting either *Luciferase* (SgControl), *IL2r α* (CD25, SgIL2r α), or *Spen*, the gene body for SHARP (SgSpen). Transduced cells were sorted and later differentiated on OP9-D11 stromal layer with supportive cytokines for 4 and 7 days prior to analysis via flow cytometry (see Methods). We specifically looked for differences in the proportions of early pro-T cell progenitors (hereby referred to as **DN1**: CD44+CD25-; **DN2**: CD44+CD25+; **DN3**: CD44-CD25+; and **DN4**: CD44-CD25-)¹⁸⁻¹⁹. We observed an enriched DN1 compartment in SgSpen samples compared to SgControl samples, recapitulating the

development delay observed in Tsuji et al 2007 (**Figure 2A**)¹⁵. This trend continued on day 7 of analysis (**Figure 2B**).

To complement our primary cell experiments, we repeated CRISPR/Cas9-based targeting in SCID.adh.2c2 cells, a DN3-like T-cell leukemia line originally isolated and subcloned from a Severe Combined Immunodeficiency T-cell lymphoma²⁰. In comparison to *ex vivo* differentiated primary hematopoietic progenitors, SCID.adh.2c2 cells exhibit gamma-secretase-dependent, stromal-layer independent activation of Notch signaling. SCID.adh.2c2 cells are much more robust than primary cells and are more amenable to downstream genetic, pharmacologic, and biochemical manipulation.

After transducing SCID.adh.2c2 cells with retroviruses encoding Cas9 and gRNAs as before, we performed flow cytometry against Il2r α (CD25), an established Notch responsive cell-surface marker²¹⁻²². Surprisingly, we did not initially observe a change in CD25 cell-surface expression upon SHARP knockout. To rule out the possibility that phenotypes associated with SHARP deficiency were outside the sensitivity range of our system, we repeated CRISPR/Cas9 targeting, but additionally cultured cells in media supplemented with either DMSO vehicle or titrated amounts of Gamma Secretase Inhibitor (GSI) prior to flow cytometry analysis. We observed a modest decrease in CD25 signal in SHARP knockout cells compared to control, which became more pronounced with increasing concentration of GSI (**Figure 2C**). As with previous studies of Notch signaling and SHARP in early T-cell development, this GSI Hypersensitivity phenotype was in stark contrast to those observed upon over-expression of dominant active Notch (Notch^{ICD}) (**Figure 2D**)^{17,15}.

3.2 SHARP NEGATIVELY REGULATES NRARP

We wanted to better understand the genetic circuitry underlying SHARP deficiency and the observed GSI Hypersensitivity phenotype. To address this, we performed RNA-Sequencing in SHARP knockout and control cells, cultured with various concentrations of GSI. Among the most statistically significant, differentially expressed genes was Notch-Regulated Ankyrin Protein (Nrarp), a well-characterized negative regulator of Notch-mediated transactivation (**Figure 2A**)²³. Nrarp has been shown to form a ternary complex with Notch^{ICD} and SuH, and it has been postulated to reduce Notch^{ICD} protein levels²³⁻²⁴. Interestingly, Tsuji et al also observed Nrarp up-regulation in SHARP knockout DN1 cells in their 2007 study, and Yun et al. demonstrated that Nrarp is sufficient to induce a DN1-DN2 developmental delay in early pro-T-cells, similar to those observed by Tsuji et al. and by our group^{15,25}. Given the strict requirement for Notch pathway signaling in early T-cell development and the regulatory relationship between Nrarp and Notch, this is an example of a Notch-mediated Incoherent Feedforward Loop, with SHARP-repression of Nrarp exerting an overall positive input on Notch signaling and developmental progression. In addition to Nrarp, several other SHARP-regulated genes appear to be Notch pathway-regulated, including Cpa3 (Carboxypeptidase A3, Mast Cell) and Sla (Src-like Adaptor) (**Table 1 and 2**). It should be noted that there are only 13 differentially expressed, statistically significant genes upon SHARP deficiency, which may reflect incomplete knockout or severe fitness deficits (**Figure 2A**). Therefore, the full expression overlap between SHARP deficiency and Notch-regulation remains to be determined.

3.3 NRARP IS SUFFICIENT FOR GSI HYPERSENSITIVITY

Since Yun et al. demonstrated that *Nrarp* is sufficient to phenocopy the developmental delay observed with SHARP deficiency, we wanted to test if this were also true with respect to the GSI Hypersensitivity phenotypes observed²⁵. To do this, we retrovirally transduced SCID.adh.2c2 cells with either Empty Vector Control or an *Nrarp* over-expression vector and subjected cells to the GSI hypersensitivity assay. We observed downregulation of CD25 upon *Nrarp* over-expression, mimicking the deficits observed upon SHARP knockout and demonstrating that *Nrarp* is sufficient to induce GSI Hypersensitivity (**Figure 3B**).

3.4 SHARP ACTS IN AN NRARP-DEPENDENT MANNER

To test if SHARP acts in an *Nrarp*-dependent manner, we isolated Cas9- and gRNA-expressing subclones (targeting either *Luciferase* or *Spen*) and additionally transduced subclones with retroviruses encoding gRNAs targeting *Luciferase* or *Nrarp*, along with an NGFR cell-surface receptor for gating. We subsequently repeated the GSI Hypersensitivity Assay in these cells and were able to effectively compare single knockout (**Spen**: SgSpen + SgControl ; **Nrarp**: SgControl + SgNrarp), double knockout (SgSpen + SgNrarp), and no knockout (SgControl + SgControl) conditions. Although we encountered slight baseline differences in Notch pathway activation between SgControl and SgSpen samples in the DMSO vehicle condition, we reasoned that this was due to genetic drift from subcloning. We focused instead on the 0.1 μ M GSI condition, since it had enabled us to look at more subtle differences in Notch-mediated gene expression previously (**Figures 2C and 3B**). We observed reversion of GSI Hypersensitivity in the SHARP/*Nrarp* double knockout condition, demonstrating that *Nrarp* is downstream of

SHARP in this circuit and ultimately that SHARP acts to positively buffer Notch signaling in an *Nrarp*-dependent manner (**Figure 3C**).

3.5 SHARP PARTICIPATES IN RNA-MEDIATED AUTOFEEDBACK

SHARP is a well-characterized RNA-binding protein that has been studied extensively within the context of nuclear hormone signaling and X-chromosome inactivation^{12,26-29}. Many RNA-binding proteins exhibit autoregulation, including alternative splicing-coupled nonsense-mediated decay, poison cassette exon inclusion, RNA destabilization, and control of 3' end processing, among other mechanisms³⁰⁻³⁵. Given its interaction with RNA in other contexts, we were interested in characterizing SHARP-RNA interactions within the context of early T-cell development.

To do this, we performed Covalent Linkage Affinity Purification in SCID.adh.2c2 lines engineered to express Halo-epitope tagged SHARP and identified interactions between the protein and its own mRNA transcript (*Spen*) within intron 2. This binding interaction was conserved among cell types (mouse embryonic stem cell and SCID.adh.2c2) and species (mouse and human), leading us to hypothesize that SHARP negatively regulates its own production in early T-cell development (**Figure 4A, Chapter 3 – Figure 4D**). To test the significance of this interaction, we generated homozygous SHARP Binding Site deletion lines in mouse embryonic stem cells and measured *Spen* mRNA using RT-qPCR. We observed an increase in steady state *Spen* transcript levels compared to parent line control, demonstrating that SHARP participates in autofeedback (**Figure 4B**). We independently confirmed this finding by analyzing recently published RNA-Seq datasets from Dossin et al, who had generated mouse embryonic stem cells with SHARP tagged with an auxin-inducible degron¹³ (**Figure 4D**).

Based on SHARP's interactions with SMRT and HDAC3 in Xist-mediated silencing, we reasoned that HDAC inhibition would also result in an increase of steady state Spen mRNA. To test this prediction, we exposed SCID.adh.2c2 cells to several histone de-acetylase inhibitors, including Suberoylanilide Hydroxamic Acid (SAHA), RGFP966 (an HDAC3-specific Inhibitor), and Trichostatin A (TSA), and observed an increase in steady state Spen levels compared to the DMSO vehicle (**Figure 4C**)³⁶⁻³⁸. While the effects of deleting the SHARP Binding Site within Spen in SCID.adh.2c2 cells or hematopoietic progenitors remain to be explicitly determined, these data support the claim that SHARP participates in RNA-mediated autofeedback and match previous reports of SHARP's stable expression profile across T-cell development⁸⁴.

4.0 DISCUSSION: IMPLICATIONS FOR GENE REGULATORY NETWORKS

Taken together, our data supports a gene circuit architecture whereby SHARP simultaneously buffers a Notch-mediated Incoherent Feedforward Loop and its own production (**Figure 5**). Incoherent Feedforward Loops are named because they contain two branches with opposing actions on a particular process. They are a commonly used regulatory architecture, documented in bacteria, mouse embryonic stem cells, hematopoietic stem cells, innate immune regulation, fly eye development, and downstream of Notch signaling, among other contexts³⁹⁻⁴⁹. Incoherent Feedforward Loops, like the circuit discussed above, can be used for several different purposes, including improving stability and dynamics of a conjoined negative feedback loop, increasing gene-regulatory network response time and expression pulsatility, biphasic expression responses, fold-change detection, and adaptive tuning of gene expression^{39,42,49-60}. With respect to Notch signaling during early T-cell development, SHARP establishes a threshold for the repressor Nrarp, which in turn allows sustained Notch signaling¹⁵ (**Figure 3A**). In the absence of SHARP-

mediated inhibition, upregulated (Notch-induced) Nrarp causes loss of sustained Notch signal intensity, resulting in robust developmental delay^{15,25} (**Figure 2A and 2B**).

RNA has already been implicated in Incoherent Feedforward Loops. Several studies examining the use of small inhibitory RNAs point towards their utility in establishing homeostatic robustness in response to change (rather than generation of expression pulses^{61-72,49}). Through the ability to detect their respective mRNA, RNA-binding chromatin regulatory proteins can offer an additional layer of regulatory complexity to developmentally sensitive circuits and processes. As an example, Notch signal transduction can have variable transcriptional after-effects, based on degree of pulsatility⁷³. Circuits stabilized by RNA-binding proteins could off-set environmentally variable inputs (as potentially observed with Delta-like ligand availability during progenitor migration within the thymic microenvironment) or establish a band-pass filtering for sustained effective signal over time. The specific consequences of SHARP's interaction with its own mRNA for regulation of Notch signaling remain to be determined, as does the extent to which such a regulatory mechanism may occur in broader development.

5.0 MATERIALS AND METHODS

Experimental Animal Lines and Protocols

Mice: C57BL/6 and B6.Gt(ROSA)26Sor^{tm1.1}(CAG-cas9*,- EGFP)Fezh/J (Cas9) mice were purchased from the Jackson Laboratory. All animals were bred and maintained in the California Institute of Technology Laboratory Animal Facility under specific pathogen free conditions. The protocol supporting animal breeding for this work was reviewed and approved by the Institute

Animal Care and Use Committee of the California Institute of Technology (Protocol #1445, entitled “Conventional and Immunodeficient Mouse Breeding. Developmental Control of Transcription Factors”).

For primary cell CRISPR knockout experiments, we used F1 progeny of B6 and Cas9 mice. Fetal mice were collected from pregnant females at embryonic day 13.5 for fetal liver cell collection. Fetal livers were dissected from E13.5 (day of plug, E0.5) C57BL/6 animals. Pooled suspensions of FL cells were stained for lineage markers using biotin-conjugated lineage antibodies (CD11c, Gr1, TER-119, NK1.1, CD19, F4/80), incubated with streptavidin-coated magnetic beads (Miltenyi Biotec), and passed through a magnetic column (Miltenyi Biotec) prior to freezedown in media containing 40% Serum and 10% DMSO; male and female embryos were pooled together and used in experiments⁷⁴.

Cell Culture Conditions

SCID.adh.2c2: Scid.adh.2c2 cells were cultured in RPMI1640 with 10% fetal bovine serum (Sigma-Aldrich), sodium pyruvate, non-essential amino acids, Pen-Strep-Glutamine and 50 μ M β -mercaptoethanol as previously described⁷⁴.

OP9-D11: OP9-D11 were cultured as previously described⁷⁵.

Mouse Embryonic Stem Cells (mESC): All mouse ES cell lines were cultured in serum-free 2i/LIF medium as previously described⁸³. SHARP-Binding Site targeting was performed in F1 2-1 parent line, a F1 hybrid wild-type mouse ES cell line derived from a 129 \times *castaneous* cross (provided by K. Plath).

Retroviral Infection: Non-TC treated plates were coated overnight at 4C with 500 μ l of 50ug/ml Retronectin (Takara) in 1X Sterile Filtered PBS per well. On the following day, wells were briefly rinsed with 1ml of sterile-filtered 1X PBS, before being exposed to MCSV-based retroviral supernatants. Retroviral coated plates were centrifuged at 2000G X 2 hrs X 32C and excess unbound virus was rinsed away with 1ml of sterile-filtered 1X PBS. 250-500K SCID.adh.2c2 cells were then added to each respective well, centrifuged at 800 RPM X 20 minutes X 32C, and then left to incubate with virus for 24hrs at 37C prior to further downstream experiments. For primary cell experiments, Lin⁻ Fetal Liver Precursors (FLPs) were cultured in OP9 media supplemented with 5ng/ml Flt3, Il7, and SCF for 6-12 hours prior to retroviral infection. Cells were subsequently transferred to retrovirus-coated wells as previously described, centrifuged at 800 RPM X 20 minutes X 32C, and left to incubate with retrovirus for 48hrs at 37C prior to fluorescence activated cell sorting (FACS).

OP9-D11 Co-Culture and Pro-T Cell Ex vivo Differentiation: Following FACS for Lin⁻, mTurquoise2⁺, CD27⁺, cKIT⁺ precursors, cells were seeded onto prepared OP9-D11 stromal layers, with media supplemented with 5ng/ml Flt3 and Il7. Cells were analyzed on Day 4 of co-culture and re-seeded to new OP9-D11 stromal layers with media supplemented with 1ng/ml Il7. Cells were subsequently analyzed on Day 7 of co-culture. Co-culture experiments were performed in replicate, starting from retroviral infection and sorting through Day 7 co-culture flow cytometry. Data presented in Figure 1 is representative of both co-culture series.

Gamma Secretase Inhibition (GSI): GSI (InSolution γ -Secretase Inhibitor X - EMD Millipore) was administered to SCID.adh.2c2 cells as previously documented in Del Real, 2013⁷⁵. GSI dosages were titrated by diluting 2000X stock solutions with additional Dimethyl-sulfoxide vehicle prior to 2000-fold dilution in media to working concentrations. Retrovirally transduced

SCID.adh.2c2 cells were resuspended and plated in SCID.adh.2c2 media containing working concentrations of GSI and allowed to culture for 48hrs prior to flow cytometry analysis or RNA harvesting post-FACS.

Histone De-Acetylase Inhibitor Administration: RGFP966, SAHA, and TSA were diluted in DMSO solvent to generate stock concentrations and further diluted in SCID.adh.2c2 media for final working concentrations of 0.1 μ M SAHA, 5 μ M RGFP966, and 10nM TSA. Cells were grown in HDAC Inhibitor media for 48hrs prior to RNA harvesting for RT-qPCR. GapDH-normalized gene expression was fold-normalized to DMSO Vehicle and averaged across three separate Spen-targeting primer pairs.

Generation of Halo-tagged SHARP SCID.adh.2c2 Line: To generate an N-terminal Halo-tagged SHARP mammalian expression construct, a LR-compatible pB-Halo-DEST-IRES-eGFP destination vector was first generated. Generation of this destination vector made use of the pB-TAG-ERN backbone (a gift from Knut Woltjen; Addgene plasmid # 80476; <http://n2t.net/addgene:80476> ; RRID:Addgene_80476) and the coding sequence for the Halo epitope (gift from Ward Walkup IV)⁷⁶. Resulting expression vectors drive protein expression by a dox-inducible promoter, contain the rtTA needed for dox induction, and produce an N-terminally-tagged Halo fusion protein. LR recombination was subsequently performed using the Full Length SHARP entry clone (Sp22, gift from Alexander Shiskin) to generate the pB-Halo-SHARP,IRES-eGFP expression vector. SCID.adh.2c2 cells were transfected using the pB-Halo-SHARP,IRES-eGFP, along with piggybac transposase (gift from Xun Wang), using a previously published Lonza/Amaza Nucleofection (Solution-V, program D-19) method⁷⁷. Cells with stable integration of the construct were selected using repeated cycles of 2 μ g/ml doxycycline administration, eGFP FACS, and doxycycline removal to generate polyclonal Halo-SHARP

SCID.adh.2c2 lines.

Deletion of SHARP Binding Site in Spen: F1 2-1 line were CRISPR-targeted with gRNAs targeting the SHARP-Binding Site identified previously via Covalent Linkage Affinity Purification Sequencing (SHARP Binding Site Coordinates: mm10 - chr4:141,520,614-141,521,890; gRNA sequences were TCTACCGGGAGATCACATGG and ACTGTGCCAGGTATCCACTG). Following selection using 1 μ g/ml of puromycin in standard 2i culture conditions, transfected cells were allowed to recover in standard 2i media prior to sub-cloning. Clones were subsequently screened by genomic DNA PCR using primers flanking the deletion region (GGAGACGGAAGTTATTTCTATCC and GTTTCTCACTTAACCTGGAGC respectively). In contrast to the wild type 1833bp band, successfully targeted alleles would produce a 215bp band. Sub-clones homozygous for the targeted allele were subject to RT-qPCR and GAPDH-normalized gene expression was further normalized to the F1 parent line.

Flow Cytometry: Flow cytometry was performed as previously discussed⁷⁵. More specifically for Day 4 and Day 7 *Ex Vivo* Differentiation studies, cells were stained against cKit, CD45, CD44, and CD25 cell-surface markers, as well as 7-AAD to assess viability. For GSI Hypersensitivity assays, SCID.adh.2c2 samples were stained against CD25 and NGFR cell-surface marker and viability was assessed using 7-AAD.

Molecular Biology Protocols

Retroviral Cloning: The NRARP open reading frame was cloned into the pMX-IRES-eGFP (MIG, gift from X. Wang and H. Hosokawa) backbone (pMX-NRARP). Notch^{ICD} expression

vector was also a gift from H. Hosokawa. All retroviruses were packaged as previously described⁸¹.

gRNA Retroviral Vector Cloning: gRNA sequences were cloned and packaged as discussed previously, with the following modifications⁷⁴. To enable selection of SHARP and NRARP double knockouts, E42 dTet-CFP was modified to replace the CFP selection marker with Truncated Human Nerve Growth Factor Receptor (NFGR) (coding sequence from Addgene plasmid #27489, a gift from Warren Pear; <http://n2t.net/addgene:27489> ; RRID:Addgene_27489) to generate E42 dTet-NGFR⁷⁸. gRNA sequences targeting the *Spn* gene body were cloned into E42 dTet-CFP (GGAGACCGAGACCTCCGCA, TATGAGCGGAGACTCGATG, and GACTGGGAGAACTAACACA). Control gRNA sequence, as well as gRNA sequences targeting *Nrarp* gene body were cloned into E42 dTet-NGFR. *Nrarp*-targeting gRNAs were TACACCAGTCAGTCATCGA, GTGCGCAAGGGCAACACGC, and GTCTGTGGCGCCGAGCAGG).

RT-qPCR: Total RNA was extracted from cultured SCID.adh.2c2 or mESCs with Silane beads (Sigma) according to manufacturer's conditions and treated with Turbo DNase (Life Technologies) for 15min at 37C to remove genomic DNA. RT reactions were performed according to the Superscript II protocol (Thermo Fisher Scientific #18064022) with random 9mers. qPCRs were performed in technical replicates using a Roche Lightcycler. Plots were generated using GraphPad software. ddCt values were calculated by normalizing to GAPDH and to samples transfected with control treatment or parent line to compare fold gene expression differences between samples⁸².

mRNA and RNA-seq library preparation: RNA and RNA-Seq libraries were processed as discussed previously⁷⁴. Total RNA was isolated using an RNAeasy MicroKit (Qiagen) according

to manufacturer's recommendations. Libraries were constructed using NEBNext Ultra RNA Library Prep Kit for Illumina (NEB #E7530) from ~1 µg of total RNA following manufacturer's instructions. Libraries were sequenced on an Illumina HiSeq2500 in single read mode with the read length of 50 nt according to manufacturer's instructions⁷⁴.

SHARP Covalent Linkage Affinity Purification (CLAP): CLAP and associated library-preparation were performed as previously described (Chapters 2 and 3), with the following modifications for suspension SCID.adh.2c2 cells: Halo-SHARP SCID.adh.2c2 cells were resuspended in media supplemented with 2µg/ml doxycycline for 48 hours prior to harvesting. Cells were washed with 1X PBS, transferred to a 15cm plate and crosslinked on ice using 0.25 J cm⁻² (UV2.5k) of UV at 254 nm in a Spectrolinker UV Crosslinker. Following UV-cross linking, cells were transferred from a culture dish to a 50ml conical vial, washed once with PBS, pelleted by centrifugation at 1,500g for 4 min and flash-frozen in liquid nitrogen for storage at -80C.

Quantification and Statistical Analysis

Flow cytometry Analysis: When available, median fluorescence intensities were averaged between replicate wells within a given experiment. Mean fluorescence intensities were then internally normalized to a reference control for fold normalized intensity (please see legend for respective reference control used for each experiment). Normalized intensity was then averaged between experimental series replicates, with standard deviation representing inter-experimental variability.

Differential Gene Expression: Calls were made according to previously documented pipelines and software packages⁷⁴. Use of the DeSeq2 package is denoted in respective figure legends, with false discovery rate set to 0.05. High Confidence Notch target genes were determined by comparing between two separate Notch pathway perturbations: GSI inhibition treatment (0.1 μ M GSI: CD25+ versus CD25- subsets) and over-expression of Notch^{ICD} (CD25+ subsets from Notch^{ICD} versus Empty Vector Control). Notch-dependent genes (GSI-repressed and NotchICD-enhanced) and Notch-repressed genes (GSI-enhanced and NotchICD-repressed) are presented in Tables 1 and 2. Calls were made using DeSeq2 pipeline, with FDR set to 0.05.

RNA-seq data used in **Figure 4D** was adapted from Dossin et al., where mouse-ES cells were engineered to express SHARP tagged with an auxin-inducible degron¹³. Fastq files were trimmed using skewer and pseudoaligned using Salmon to mm10 using the latest transcripts from Gencode (GRCm38_p6)⁸⁵⁻⁸⁶. TPM's were calculated using Salmon for untreated and 28-hour auxin treated samples (3 replicates for each condition)⁸⁵. TPMs were further averaged across aligned transcript isoforms and presented in Figure 4D.

For analysis performed for **Chapter 5 - Figure 1**, comparing SHARP-mediated differential gene expression (log-2 fold change after auxin treatment against untreated control) versus Spen RNA localization on chromatin from RNA-DNA SPRITE, log-2 fold changes for each gene with a 1 Mb window around SHARP, excluding genes that had a mean TPM of zero in either condition were plotted on the left y-axis. Weighted SHARP-RNA to DNA contacts from the RNA-DNA SPRITE data were counted in 50kb bins around the SHARP locus and plotted on the right y-axis.

CLAP Peaks: Calls and input-normalized enrichments were made using CLIPCLAPEnrichment javascript pipeline. Enrichment figures were made with alignments to mm10.

6.0 REFERENCES

1. Yui, M.A., and Rothenberg, E.V. (2014). Developmental gene networks: a triathlon on the course to T cell identity. *Nature reviews. Immunology*, 14(8), 529–545. <https://doi.org/10.1038/nri3702>
2. Radtke, F., Wilson, A., Stark, G., Bauer, M., van Meerwijk, J., MacDonald, H.R., and Aguet, M. (1999). Deficient T cell fate specification in mice with an induced inactivation of Notch1. *Immunity*, 10(5), 547–558. [https://doi.org/10.1016/s1074-7613\(00\)80054-0](https://doi.org/10.1016/s1074-7613(00)80054-0)
3. Previs, R.A., Coleman, R.L., Harris, A.L., and Sood, A.K. (2015). Molecular pathways: translational and therapeutic implications of the Notch signaling pathway in cancer. *Clinical cancer research : an official journal of the American Association for Cancer Research*, 21(5), 955–961. <https://doi.org/10.1158/1078-0432.CCR-14-0809>
4. Ntziachristos, P., Lim, J.S., Sage, J., and Aifantis, I. (2014). From fly wings to targeted cancer therapies: a centennial for notch signaling. *Cancer cell*, 25(3), 318–334. <https://doi.org/10.1016/j.ccr.2014.02.018>
5. Dexter, J.S. (1914). The analysis of continuous variation in *Drosophila* by a study of its linkage relations. *Am Nat* 48, 712-758.
6. Tanigaki, K., and Honjo, T. (2007). Regulation of lymphocyte development by Notch signaling. *Nature immunology*, 8(5), 451–456. <https://doi.org/10.1038/ni1453>
7. Tamura, K., Taniguchi, Y., Minoguchi, S., Sakai, T., Tun, T., Furukawa, T., and Honjo, T. (1995). Physical interaction between a novel domain of the receptor Notch and the transcription factor RBP-J kappa/Su(H). *Current biology : CB*, 5(12), 1416–1423. [https://doi.org/10.1016/s0960-9822\(95\)00279-x](https://doi.org/10.1016/s0960-9822(95)00279-x)
8. Taniguchi, Y., Furukawa, T., Tun, T., Han, H., and Honjo, T. (1998). LIM protein KyoT2 negatively regulates transcription by association with the RBP-J DNA-binding protein. *Molecular and cellular biology*, 18(1), 644–654. <https://doi.org/10.1128/mcb.18.1.644>
9. Oswald, F., Kostezka, U., Astrahantseff, K., Bourteele, S., Dillinger, K., Zechner, U., Ludwig, L., Wilda, M., Hameister, H., Knöchel, W., Liptay, S., and Schmid, R.M. (2002). SHARP is a novel component of the Notch/RBP-Jkappa signalling pathway. *The EMBO journal*, 21(20), 5417–5426. <https://doi.org/10.1093/emboj/cdf549>
10. Newberry, E.P., Latifi, T., and Towler, D.A. (1999). The RRM domain of MINT, a novel Msx2 binding protein, recognizes and regulates the rat osteocalcin promoter. *Biochemistry*, 38(33), 10678–10690. <https://doi.org/10.1021/bi990967j>
11. Su, H., Liu, Y., and Zhao, X. (2015). Split End Family RNA Binding Proteins: Novel Tumor Suppressors Coupling Transcriptional Regulation with RNA Processing. *Cancer Translational Medicine*, 1, 21-25.

12. Shi, Y., Downes, M., Xie, W., Kao, H.Y., Ordentlich, P., Tsai, C.C., Hon, M., and Evans, R.M. (2001). Sharp, an inducible cofactor that integrates nuclear receptor repression and activation. *Genes & development*, 15(9), 1140–1151. <https://doi.org/10.1101/gad.871201>
13. Dossin, F., Pinheiro, I., Żylicz, J.J., Roensch, J., Collombet, S., Le Saux, A., Chelmicki, T., Attia, M., Kapoor, V., Zhan, Y., Dingli, F., Loew, D., Mercher, T., Dekker, J., and Heard, E. (2020). SPEN integrates transcriptional and epigenetic control of X-inactivation. *Nature*, 578(7795), 455–460. <https://doi.org/10.1038/s41586-020-1974-9>
14. Osborne, B.A., and Minter, L.M. (2007). Notch signalling during peripheral T-cell activation and differentiation. *Nature reviews. Immunology*, 7(1), 64–75. <https://doi.org/10.1038/nri1998>
15. Tsuji, M., Shinkura, R., Kuroda, K., Yabe, D., and Honjo, T. (2007). Msx2-interacting nuclear target protein (Mint) deficiency reveals negative regulation of early thymocyte differentiation by Notch/RBP-J signaling. *Proceedings of the National Academy of Sciences of the United States of America*, 104(5), 1610–1615. <https://doi.org/10.1073/pnas.0610520104>
16. Kuroda, K., Han, H., Tani, S., Tanigaki, K., Tun, T., Furukawa, T., Taniguchi, Y., Kurooka, H., Hamada, Y., Toyokuni, S., and Honjo, T. (2003). Regulation of marginal zone B cell development by MINT, a suppressor of Notch/RBP-J signaling pathway. *Immunity*, 18(2), 301–312. [https://doi.org/10.1016/s1074-7613\(03\)00029-3](https://doi.org/10.1016/s1074-7613(03)00029-3)
17. Pui, J.C., Allman, D., Xu, L., DeRocco, S., Karnell, F.G., Bakkour, S., Lee, J.Y., Kadesch, T., Hardy, R.R., Aster, J.C., and Pear, W.S. (1999). Notch1 expression in early lymphopoiesis influences B versus T lineage determination. *Immunity*, 11(3), 299–308. [https://doi.org/10.1016/s1074-7613\(00\)80105-3](https://doi.org/10.1016/s1074-7613(00)80105-3)
18. Holmes, R., and Zúñiga-Pflücker, J.C. (2009). The OP9-DL1 system: generation of T-lymphocytes from embryonic or hematopoietic stem cells in vitro. *Cold Spring Harbor protocols*, 2009(2), pdb.prot5156. <https://doi.org/10.1101/pdb.prot5156>
19. Godfrey, D.I., and Zlotnik, A. (1993). Control points in early T-cell development. *Immunology today*, 14(11), 547–553. [https://doi.org/10.1016/0167-5699\(93\)90186-O](https://doi.org/10.1016/0167-5699(93)90186-O)
20. Dionne, C.J., Tse, K.Y., Weiss, A.H., Franco, C.B., Wiest, D.L., Anderson, M.K., and Rothenberg, E.V. (2005). Subversion of T lineage commitment by PU.1 in a clonal cell line system. *Developmental biology*, 280(2), 448–466. <https://doi.org/10.1016/j.ydbio.2005.01.027>
21. Adler, S.H., Chiffolleau, E., Xu, L., Dalton, N.M., Burg, J.M., Wells, A.D., Wolfe, M.S., Turka, L.A., and Pear, W.S. (2003). Notch signaling augments T cell responsiveness by enhancing CD25 expression. *Journal of immunology (Baltimore, Md. : 1950)*, 171(6), 2896–2903. <https://doi.org/10.4049/jimmunol.171.6.2896>

22. Bellavia, D., Campese, A.F., Alesse, E., Vacca, A., Felli, M.P., Balestri, A., Stoppacciaro, A., Tiveron, C., Tatangelo, L., Giovarelli, M., Gaetano, C., Ruco, L., Hoffman, E. S., Hayday, A.C., Lendahl, U., Frati, L., Gulino, A., and Screpanti, I. (2000). Constitutive activation of NF-kappaB and T-cell leukemia/lymphoma in Notch3 transgenic mice. *The EMBO journal*, *19*(13), 3337–3348. <https://doi.org/10.1093/emboj/19.13.3337>
23. Lamar, E., Deblandre, G., Wettstein, D., Gawantka, V., Pollet, N., Niehrs, C., & Kintner, C. (2001). Nrarp is a novel intracellular component of the Notch signaling pathway. *Genes & development*, *15*(15), 1885–1899. <https://doi.org/10.1101/gad.908101>
24. Jarrett, S.M., Seegar, T., Andrews, M., Adelmant, G., Marto, J.A., Aster, J.C., and Blacklow, S.C. (2019). Extension of the Notch intracellular domain ankyrin repeat stack by NRARP promotes feedback inhibition of Notch signaling. *Science signaling*, *12*(606), eaay2369. <https://doi.org/10.1126/scisignal.aay2369>
25. Yun, T.J., and Bevan, M.J. (2003). Notch-regulated ankyrin-repeat protein inhibits Notch1 signaling: multiple Notch1 signaling pathways involved in T cell development. *Journal of immunology (Baltimore, Md. : 1950)*, *170*(12), 5834–5841. <https://doi.org/10.4049/jimmunol.170.12.5834>
26. 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., and 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>
27. Chu, C., Zhang, Q.C., da Rocha, S. T., Flynn, R.A., Bharadwaj, M., Calabrese, J.M., Magnuson, T., Heard, E., and 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>
28. Moindrot, B., Cerase, A., Coker, H., Masui, O., Grijzenhout, A., Pintacuda, G., Schermelleh, L., Nesterova, T.B., and Brockdorff, N. (2015). A Pooled shRNA Screen Identifies Rbm15, Spen, and Wtap as Factors Required for Xist RNA-Mediated Silencing. *Cell reports*, *12*(4), 562–572. <https://doi.org/10.1016/j.celrep.2015.06.053>
29. Monfort, A., Di Minin, G., Postlmayr, A., Freimann, R., Arieti, F., Thore, S., and Wutz, A. (2015). Identification of Spen as a Crucial Factor for Xist Function through Forward Genetic Screening in Haploid Embryonic Stem Cells. *Cell reports*, *12*(4), 554–561. <https://doi.org/10.1016/j.celrep.2015.06.067>
30. Wollerton, M.C., Gooding, C., Wagner, E.J., Garcia-Blanco, M.A., & Smith, C.W. (2004). Autoregulation of polypyrimidine tract binding protein by alternative splicing leading to nonsense-mediated decay. *Molecular cell*, *13*(1), 91–100. [https://doi.org/10.1016/s1097-2765\(03\)00502-1](https://doi.org/10.1016/s1097-2765(03)00502-1)

31. Jumaa, H., and Nielsen, P.J. (2000). Regulation of SRp20 exon 4 splicing. *Biochimica et biophysica acta*, 1494(1-2), 137–143. [https://doi.org/10.1016/s0167-4781\(00\)00233-5](https://doi.org/10.1016/s0167-4781(00)00233-5)
32. Ottens, F., and Gehring, N.H. (2016). Physiological and pathophysiological role of nonsense-mediated mRNA decay. *Pflugers Archiv : European journal of physiology*, 468(6), 1013–1028. <https://doi.org/10.1007/s00424-016-1826-5>
33. Rossbach, O., Hung, L.H., Schreiner, S., Grishina, I., Heiner, M., Hui, J., and Bindereif, A. (2009). Auto- and cross-regulation of the hnRNP L proteins by alternative splicing. *Molecular and cellular biology*, 29(6), 1442–1451. <https://doi.org/10.1128/MCB.01689-08>
34. Boelens, W.C., Jansen, E.J., van Venrooij, W.J., Stripecke, R., Mattaj, I.W., and Gunderson, S.I. (1993). The human U1 snRNP-specific U1A protein inhibits polyadenylation of its own pre-mRNA. *Cell*, 72(6), 881–892. [https://doi.org/10.1016/0092-8674\(93\)90577-d](https://doi.org/10.1016/0092-8674(93)90577-d)
35. Müller-McNicoll, M., Rossbach, O., Hui, J., and Medenbach, J. (2019). Auto-regulatory feedback by RNA-binding proteins. *Journal of molecular cell biology*, 11(10), 930–939. <https://doi.org/10.1093/jmcb/mjz043>
36. Malvaez, M., McQuown, S. C., Rogge, G.A., Astarabadi, M., Jacques, V., Carreiro, S., Rusche, J.R., and Wood, M.A. (2013). HDAC3-selective inhibitor enhances extinction of cocaine-seeking behavior in a persistent manner. *Proceedings of the National Academy of Sciences of the United States of America*, 110(7), 2647–2652. <https://doi.org/10.1073/pnas.1213364110>
37. Shi, X.Y., Ding, W., Li, T. Q., Zhang, Y.X., and Zhao, S.C. (2017). Histone Deacetylase (HDAC) Inhibitor, Suberoylanilide Hydroxamic Acid (SAHA), Induces Apoptosis in Prostate Cancer Cell Lines via the Akt/FOXO3a Signaling Pathway. *Medical science monitor : international medical journal of experimental and clinical research*, 23, 5793–5802. <https://doi.org/10.12659/msm.904597>
38. Vigushin, D.M., Ali, S., Pace, P.E., Mirsaidi, N., Ito, K., Adcock, I., and Coombes, R.C. (2001). Trichostatin A is a histone deacetylase inhibitor with potent antitumor activity against breast cancer in vivo. *Clinical cancer research : an official journal of the American Association for Cancer Research*, 7(4), 971–976.
39. Reeves, G.T. (2019). The engineering principles of combining a transcriptional incoherent feedforward loop with negative feedback. *Journal of biological engineering*, 13, 62. <https://doi.org/10.1186/s13036-019-0190-3>
40. Milo, R., Shen-Orr, S., Itzkovitz, S., Kashtan, N., Chklovskii, D., and Alon, U. (2002). Network motifs: simple building blocks of complex networks. *Science (New York, N.Y.)*, 298(5594), 824–827. <https://doi.org/10.1126/science.298.5594.824>

41. Eichenberger, P., Fujita, M., Jensen, S.T., Conlon, E.M., Rudner, D.Z., Wang, S.T., Ferguson, C., Haga, K., Sato, T., Liu, J.S., and Losick, R. (2004). The program of gene transcription for a single differentiating cell type during sporulation in *Bacillus subtilis*. *PLoS biology*, 2(10), e328. <https://doi.org/10.1371/journal.pbio.0020328>
42. Mangan, S., Itzkovitz, S., Zaslaver, A., and Alon, U. (2006). The incoherent feed-forward loop accelerates the response-time of the gal system of *Escherichia coli*. *Journal of molecular biology*, 356(5), 1073–1081. <https://doi.org/10.1016/j.jmb.2005.12.003>
43. Lee, T.I., Rinaldi, N.J., Robert, F., Odom, D.T., Bar-Joseph, Z., Gerber, G.K., Hannett, N. M., Harbison, C.T., Thompson, C.M., Simon, I., Zeitlinger, J., Jennings, E.G., Murray, H.L., Gordon, D.B., Ren, B., Wyrick, J.J., Tagne, J.B., Volkert, T.L., Fraenkel, E., Gifford, D.K., ... Young, R.A. (2002). Transcriptional regulatory networks in *Saccharomyces cerevisiae*. *Science (New York, N.Y.)*, 298(5594), 799–804. <https://doi.org/10.1126/science.1075090>
44. Boyer, L.A., Lee, T.I., Cole, M.F., Johnstone, S.E., Levine, S.S., Zucker, J.P., Guenther, M.G., Kumar, R.M., Murray, H.L., Jenner, R.G., Gifford, D.K., Melton, D.A., Jaenisch, R., and Young, R.A. (2005). Core transcriptional regulatory circuitry in human embryonic stem cells. *Cell*, 122(6), 947–956. <https://doi.org/10.1016/j.cell.2005.08.020>
45. Swiers, G., Patient, R., and Loose, M. (2006). Genetic regulatory networks programming hematopoietic stem cells and erythroid lineage specification. *Developmental biology*, 294(2), 525–540. <https://doi.org/10.1016/j.ydbio.2006.02.051>
46. Chevrier, N., Mertins, P., Artyomov, M.N., Shalek, A.K., Iannaccone, M., Ciaccio, M.F., Gat-Viks, I., Tonti, E., DeGrace, M.M., Clauser, K.R., Garber, M., Eisenhaure, T.M., Yosef, N., Robinson, J., Sutton, A., Andersen, M.S., Root, D.E., von Andrian, U., Jones, R.B., Park, H., ... Hacohen, N. (2011). Systematic discovery of TLR signaling components delineates viral-sensing circuits. *Cell*, 147(4), 853–867. <https://doi.org/10.1016/j.cell.2011.10.022>
47. Johnston, R.J., Jr, Otake, Y., Sood, P., Vogt, N., Behnia, R., Vasiliauskas, D., McDonald, E., Xie, B., Koenig, S., Wolf, R., Cook, T., Gebelein, B., Kussell, E., Nakagoshi, H., and Desplan, C. (2011). Interlocked feedforward loops control cell-type-specific Rhodopsin expression in the *Drosophila* eye. *Cell*, 145(6), 956–968. <https://doi.org/10.1016/j.cell.2011.05.003>
48. Krejci, A., Bernard, F., Housden, B.E., Collins, S., and Bray, S.J. (2009). Direct response to Notch activation: signaling crosstalk and incoherent logic. *Science signaling*, 2(55), ra1. <https://doi.org/10.1126/scisignal.2000140>
49. Hart, Y., and Alon, U. (2013). The utility of paradoxical components in biological circuits. *Molecular cell*, 49(2), 213–221. <https://doi.org/10.1016/j.molcel.2013.01.004>
50. Basu, S., Mehreja, R., Thiberge, S., Chen, M.T., and Weiss, R. (2004). Spatiotemporal control of gene expression with pulse-generating networks. *Proceedings of the National Academy of Sciences of the United States of America*, 101(17), 6355–6360. <https://doi.org/10.1073/pnas.0307571101>

51. Mangan, S., and Alon, U. (2003). Structure and function of the feed-forward loop network motif. *Proceedings of the National Academy of Sciences of the United States of America*, 100(21), 11980–11985. <https://doi.org/10.1073/pnas.2133841100>
52. Alon, U. (2007). Network motifs: theory and experimental approaches. *Nature reviews. Genetics*, 8(6), 450–461. <https://doi.org/10.1038/nrg2102>
53. Shoval, O., and Alon, U. (2010). SnapShot: network motifs. *Cell*, 143(2), 326–e1. <https://doi.org/10.1016/j.cell.2010.09.050>
54. Macía, J., Widder, S., and Solé, R. (2009). Specialized or flexible feed-forward loop motifs: a question of topology. *BMC systems biology*, 3, 84. <https://doi.org/10.1186/1752-0509-3-84>
55. Entus, R., Aufderheide, B., and Sauro, H.M. (2007). Design and implementation of three incoherent feed-forward motif based biological concentration sensors. *Systems and synthetic biology*, 1(3), 119–128. <https://doi.org/10.1007/s11693-007-9008-6>
56. Kaplan, S., Bren, A., Dekel, E., and Alon, U. (2008). The incoherent feed-forward loop can generate non-monotonic input functions for genes. *Molecular systems biology*, 4, 203. <https://doi.org/10.1038/msb.2008.43>
57. Kim, D., Kwon, Y.K., and Cho, K.H. (2008). The biphasic behavior of incoherent feed-forward loops in biomolecular regulatory networks. *BioEssays : news and reviews in molecular, cellular and developmental biology*, 30(11-12), 1204–1211. <https://doi.org/10.1002/bies.20839>
58. Goentoro, L., Shoval, O., Kirschner, M.W., and Alon, U. (2009). The incoherent feedforward loop can provide fold-change detection in gene regulation. *Molecular cell*, 36(5), 894–899. <https://doi.org/10.1016/j.molcel.2009.11.018>
59. Shoval, O., Goentoro, L., Hart, Y., Mayo, A., Sontag, E., and Alon, U. (2010). Fold-change detection and scalar symmetry of sensory input fields. *Proceedings of the National Academy of Sciences of the United States of America*, 107(36), 15995–16000. <https://doi.org/10.1073/pnas.1002352107>
60. Hong, J., Brandt, N., Abdul-Rahman, F., Yang, A., Hughes, T., and Gresham, D. (2018). An incoherent feedforward loop facilitates adaptive tuning of gene expression. *eLife*, 7, e32323. <https://doi.org/10.7554/eLife.32323>
61. Bleris, L., Xie, Z., Glass, D., Adadey, A., Sontag, E., and Benenson, Y. (2011). Synthetic incoherent feedforward circuits show adaptation to the amount of their genetic template. *Molecular systems biology*, 7, 519. <https://doi.org/10.1038/msb.2011.49>
62. Ebert, M.S., and Sharp, P.A. (2012). Roles for microRNAs in conferring robustness to biological processes. *Cell*, 149(3), 515–524. <https://doi.org/10.1016/j.cell.2012.04.005>

63. Elowitz, M.B., Levine, A.J., Siggia, E.D., and Swain, P.S. (2002). Stochastic gene expression in a single cell. *Science (New York, N.Y.)*, 297(5584), 1183–1186. <https://doi.org/10.1126/science.1070919>
64. Blake, W.J., KAern, M., Cantor, C.R., and Collins, J.J. (2003). Noise in eukaryotic gene expression. *Nature*, 422(6932), 633–637. <https://doi.org/10.1038/nature01546>
65. Blake, W.J., Balázsi, G., Kohanski, M.A., Isaacs, F.J., Murphy, K.F., Kuang, Y., Cantor, C.R., Walt, D.R., and Collins, J.J. (2006). Phenotypic consequences of promoter-mediated transcriptional noise. *Molecular cell*, 24(6), 853–865. <https://doi.org/10.1016/j.molcel.2006.11.003>
66. Paulsson, J. (2004). Summing up the noise in gene networks. *Nature*, 427(6973), 415–418. <https://doi.org/10.1038/nature02257>
67. Raser, J.M., and O'Shea, E.K. (2005). Noise in gene expression: origins, consequences, and control. *Science (New York, N.Y.)*, 309(5743), 2010–2013. <https://doi.org/10.1126/science.1105891>
68. Newman, J.R., Ghaemmaghami, S., Ihmels, J., Breslow, D.K., Noble, M., DeRisi, J.L., and Weissman, J.S. (2006). Single-cell proteomic analysis of *S. cerevisiae* reveals the architecture of biological noise. *Nature*, 441(7095), 840–846. <https://doi.org/10.1038/nature04785>
69. Sigal, A., Milo, R., Cohen, A., Geva-Zatorsky, N., Klein, Y., Liron, Y., Rosenfeld, N., Danon, T., Perzov, N., and Alon, U. (2006). Variability and memory of protein levels in human cells. *Nature*, 444(7119), 643–646. <https://doi.org/10.1038/nature05316>
70. Osella, M., Bosia, C., Corá, D., and Caselle, M. (2011). The role of incoherent microRNA-mediated feedforward loops in noise buffering. *PLoS computational biology*, 7(3), e1001101. <https://doi.org/10.1371/journal.pcbi.1001101>
71. Li, X., Cassidy, J.J., Reinke, C.A., Fischboeck, S., and Carthew, R.W. (2009). A microRNA imparts robustness against environmental fluctuation during development. *Cell*, 137(2), 273–282. <https://doi.org/10.1016/j.cell.2009.01.058>
72. Hornstein, E., and Shomron, N. (2006). Canalization of development by microRNAs. *Nature genetics*, 38 Suppl, S20–S24. <https://doi.org/10.1038/ng1803>
73. Nandagopal, N., Santat, L.A., LeBon, L., Sprinzak, D., Bronner, M.E., and Elowitz, M.B. (2018). Dynamic Ligand Discrimination in the Notch Signaling Pathway. *Cell*, 172(4), 869–880.e19. <https://doi.org/10.1016/j.cell.2018.01.002>
74. Hosokawa, H., Ungerbäck, J., Wang, X., Matsumoto, M., Nakayama, K. I., Cohen, S.M., Tanaka, T., and Rothenberg, E.V. (2018). Transcription Factor PU.1 Represses and Activates Gene Expression in Early T Cells by Redirecting Partner Transcription Factor Binding. *Immunity*, 48(6), 1119–1134.e7. <https://doi.org/10.1016/j.immuni.2018.04.024>

75. Del Real, M.M., and Rothenberg, E.V. (2013). Architecture of a lymphomyeloid developmental switch controlled by PU.1, Notch and Gata3. *Development (Cambridge, England)*, 140(6), 1207–1219. <https://doi.org/10.1242/dev.088559>
76. Kim, S.I., Oceguera-Yanez, F., Sakurai, C., Nakagawa, M., Yamanaka, S., and Woltjen, K. (2016). Inducible Transgene Expression in Human iPS Cells Using Versatile All-in-One piggyBac Transposons. *Methods in molecular biology (Clifton, N.J.)*, 1357, 111–131. https://doi.org/10.1007/7651_2015_251
77. Zarnegar, M.A., Chen, J., and Rothenberg, E.V. (2010). Cell-type-specific activation and repression of PU.1 by a complex of discrete, functionally specialized cis-regulatory elements. *Molecular and cellular biology*, 30(20), 4922–4939. <https://doi.org/10.1128/MCB.00354-10>
78. Izon, D.J., Punt, J.A., Xu, L., Karnell, F.G., Allman, D., Myung, P.S., Boerth, N.J., Pui, J.C., Koretzky, G.A., and Pear, W.S. (2001). Notch1 regulates maturation of CD4+ and CD8+ thymocytes by modulating TCR signal strength. *Immunity*, 14(3), 253–264. [https://doi.org/10.1016/s1074-7613\(01\)00107-8](https://doi.org/10.1016/s1074-7613(01)00107-8)
79. Lai, E.C. (2004). Notch signaling: control of cell communication and cell fate. *Development (Cambridge, England)*, 131(5), 965–973. <https://doi.org/10.1242/dev.01074>
80. Amsen, D., Antov, A., and Flavell, R.A. (2009). The different faces of Notch in T-helper-cell differentiation. *Nature reviews. Immunology*, 9(2), 116–124. <https://doi.org/10.1038/nri2488>
81. Ng, K.K., Yui, M.A., Mehta, A., Siu, S., Irwin, B., Pease, S., Hirose, S., Elowitz, M.B., Rothenberg, E.V., and Kueh, H.Y. (2018). A stochastic epigenetic switch controls the dynamics of T-cell lineage commitment. *eLife*, 7, e37851. <https://doi.org/10.7554/eLife.37851>
82. Quinodoz, S.A., Bhat, P., Ollikainen, N., Jachowicz, J., Banerjee, A.K., Chovanec, P., Blanco, M.R., Chow, A., Markaki, Y., Plath, K. and Guttman, M. (2020). RNA promotes the formation of spatial compartments in the nucleus. bioRxiv 2020.08.25.267435; doi: <https://doi.org/10.1101/2020.08.25.267435>
83. Quinodoz, S.A., Ollikainen, N., Tabak, B., Palla, A., Schmidt, J. M., Detmar, E., Lai, M.M., Shishkin, A.A., Bhat, P., Takei, Y., Trinh, V., Aznauryan, E., Russell, P., Cheng, C., Jovanovic, M., Chow, A., Cai, L., McDonel, P., Garber, M., and Guttman, M. (2018). Higher-Order Inter-chromosomal Hubs Shape 3D Genome Organization in the Nucleus. *Cell*, 174(3), 744–757.e24. <https://doi.org/10.1016/j.cell.2018.05.024>
84. Heng, T.S., Painter, M.W., and Immunological Genome Project Consortium (2008). The Immunological Genome Project: networks of gene expression in immune cells. *Nature immunology*, 9(10), 1091–1094. <https://doi.org/10.1038/ni1008-1091>
85. Patro, R., Duggal, G., Love, M.I., Irizarry, R.A., and Kingsford, C. (2017). Salmon provides fast and bias-aware quantification of transcript expression. *Nature methods*, 14(4), 417–419. <https://doi.org/10.1038/nmeth.4197>

86. Jiang, H., Lei, R., Ding, S.W., and Zhu, S. (2014). Skewer: a fast and accurate adapter trimmer for next-generation sequencing paired-end reads. *BMC bioinformatics*, 15, 182. <https://doi.org/10.1186/1471-2105-15-182>

7.0 FIGURES

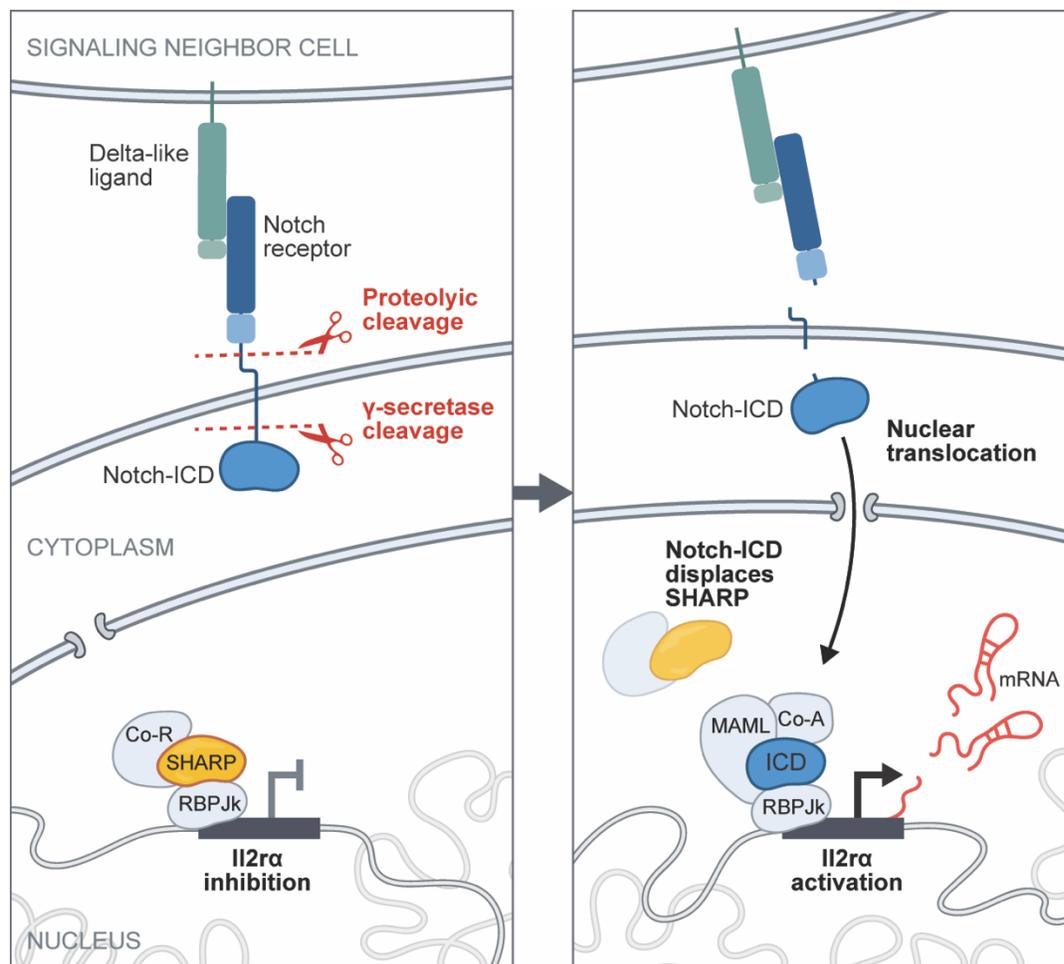
Figure 1

Figure 1: Classic model of Notch signaling. Upon ligand-interaction between Notch receptor and its cognate cell-surface bound receptors, Notch Receptor undergoes proteolytic cleavage, releasing its intracellular domain to translocate to the nucleus and convert RBP-Jk from transcriptional repressor to activator. In this model, SHARP negatively regulates Notch-mediated gene expression^{6,80-81}.

Figure 2

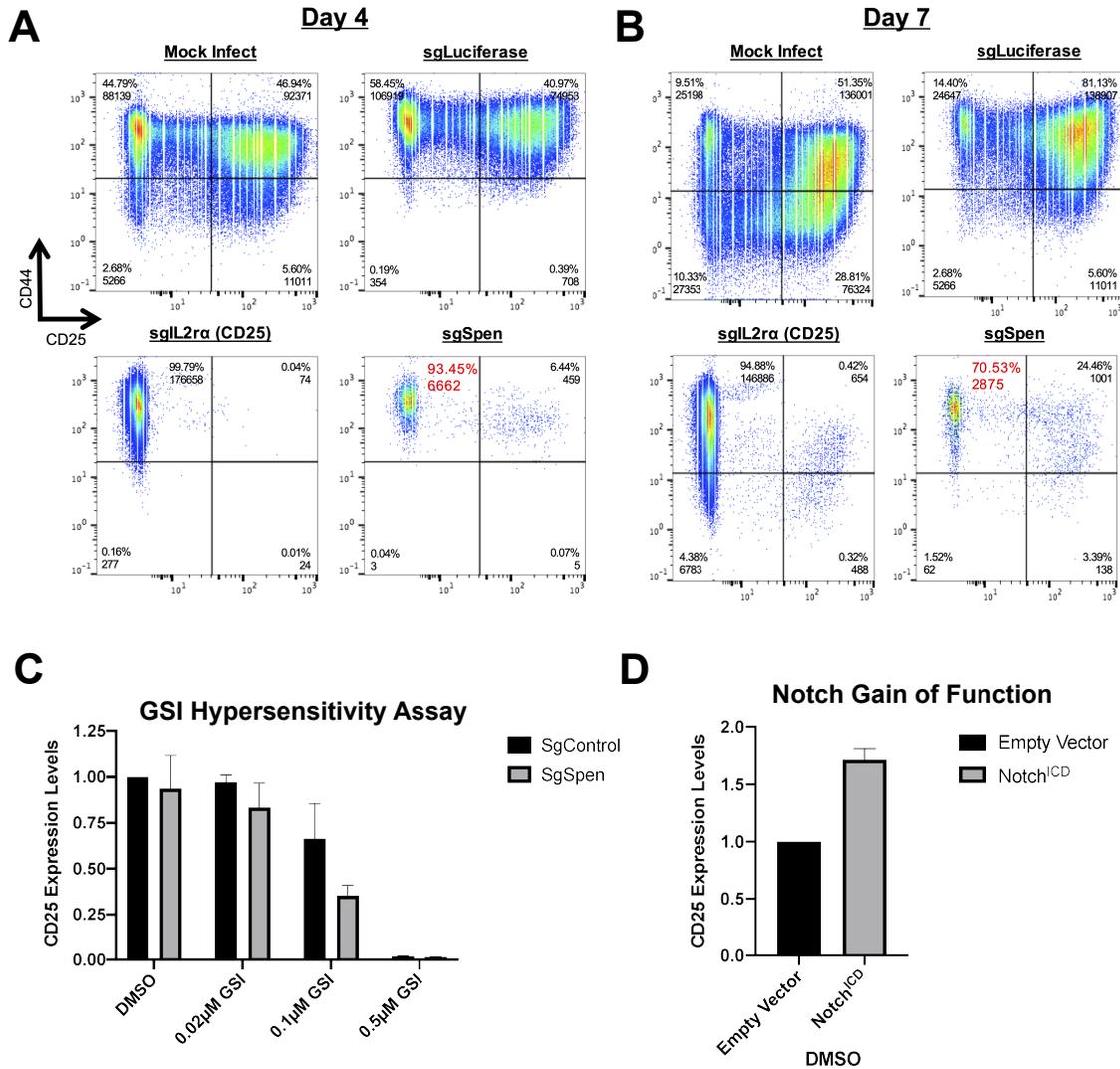


Figure 2: SHARP deficiency does not functionally phenocopy Notch gain of function. A and B) Flow cytometric analysis of *ex vivo* differentiated Cas9-targeted fetal liver progenitors (FLP). Cas9-eGFP expressing E13.5 FLPs were retrovirally transduced with pools of guide RNA retroviruses targeting either Luciferase (Control), IL2 α , and SpEn. Transduced cells were differentiated on OP9-D11 stromal layer with supportive cytokines for (A) 4 and (B) 7 days prior to flow cytometry analysis. (C) Quantification of flow cytometry against IL2 α (CD25) from

Gamma Secretase Inhibitor Hypersensitivity Assay from SgControl and SgSpen SCID.adh.2c2 experimental conditions. Internally normalized to SgControl DMSO vehicle condition (see methods) and average of two biological experimental series. **(D)** Quantification of CD25 flow cytometry in SCID.adh.2c2 cells retrovirally infected with Empty Vector Control or Dominant Active Notch (Notch^{ICD}). Internally normalized to Empty Vector Control DMSO vehicle condition and average of two biological experimental series.

Figure 3

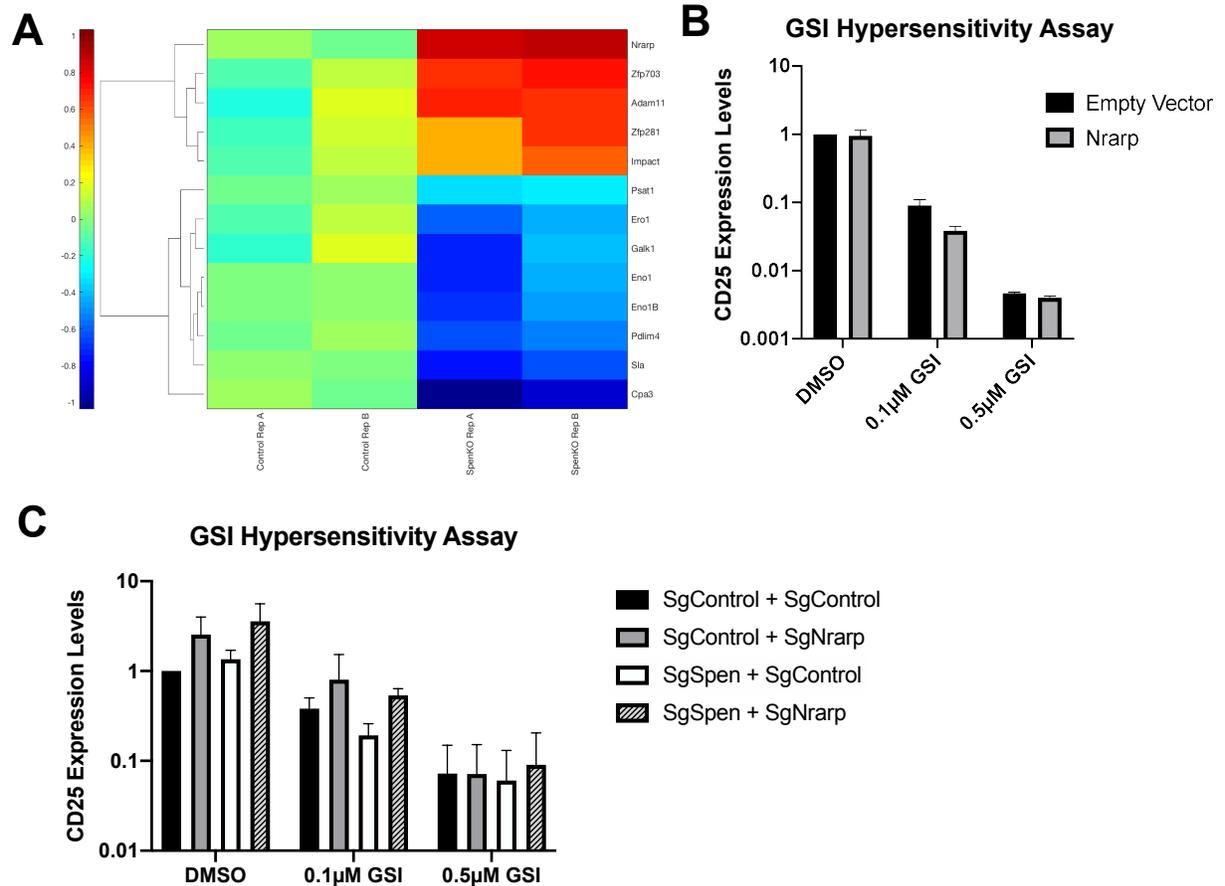


Figure 3: SHARP regulates Notch-mediated gene expression in a NRARP-dependent manner. (A) Heat map representation of RNA-Sequencing data comparing DMSO exposed CD25⁺ subsets from SgControl and SgSpen transduced SCID.adh.2c2 cells. Average of two biological experimental series. Data is represented in log₂ fold change. (B) Quantification of CD25 flow cytometry from Gamma Secretase Inhibitor Hypersensitivity Assay from Empty Vector Control or Nrarp Over-expression vector. Internally normalized to Empty Vector Control DMSO vehicle condition (see methods) and average of two biological experimental series. (C) Quantification of CD25 flow cytometry from Gamma Secretase Inhibitor Hypersensitivity Assay from subcloned SgControl and SgSpen lines, additionally transduced with SgControl,

NGFR or SgNrarp, NGFR. Internally normalized to SgControl + SgControl, DMSO vehicle condition (see methods) and average of three biological experimental series.

Figure 4

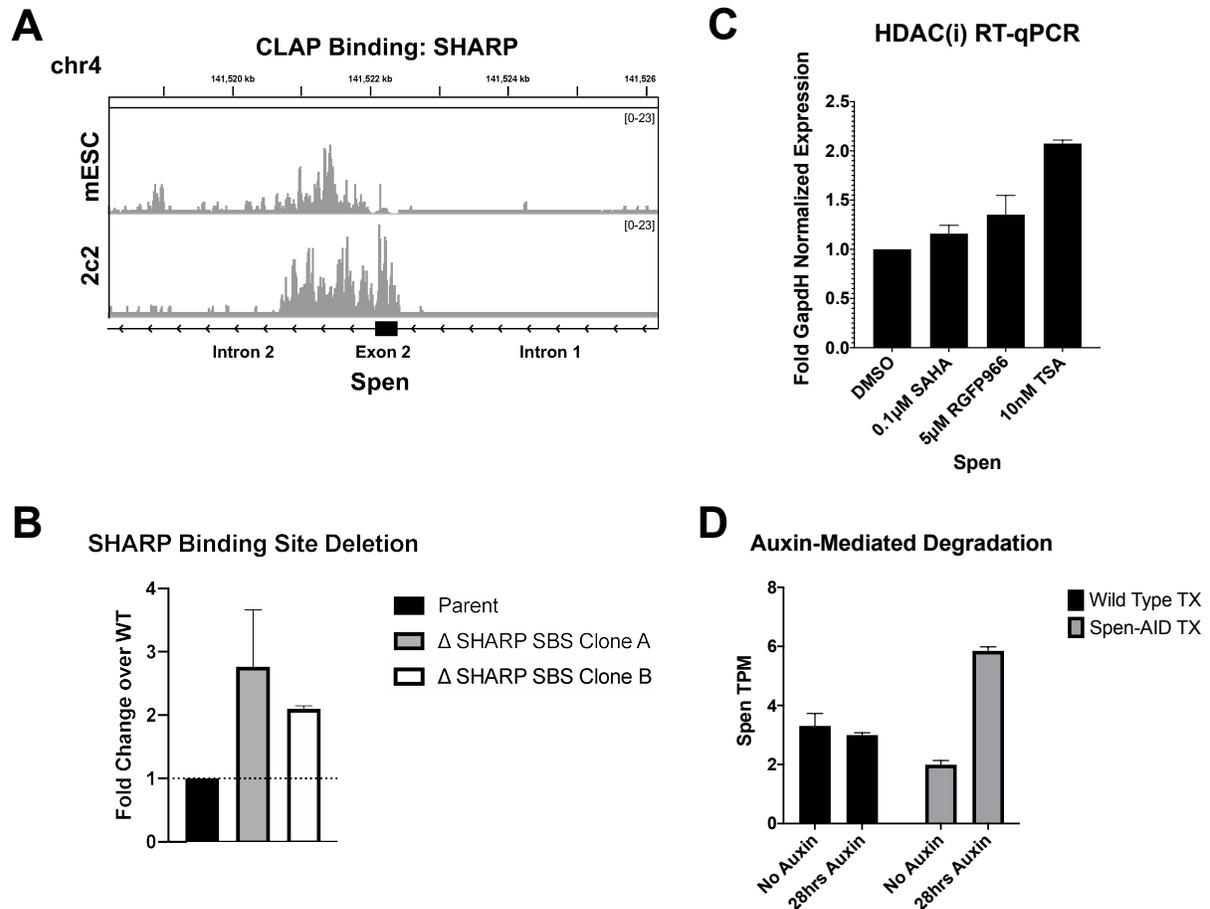


Figure 4: SHARP exhibits negative auto-regulation. (A) Input normalized enrichment plots of SHARP CLAP-Seq of Spen pre-mRNA transcript, performed in both mouse embryonic stem cells and in Halo-SHARP SCID.adh.2c2 lines. Comparison between mouse embryonic stem cell and Human Embryonic Kidney 293T SHARP CLAP-Seq is included in Figure 4D of Chapter 3 of this dissertation. (B) RT-qPCR analysis of two different mouse embryonic stem cell clones, harboring homozygous deletion of the SHARP binding site (Delta SBS) within Spen. Expression normalized to WT, average of 3 separate primer pairs. (C) RT-qPCR analysis of SCID.adh.2c2 cells treated with histone de-acetylase inhibitor treatment for 48 hours. Normalized to DMSO

vehicle and average of two biological experimental series. **(D)** RNA-Seq analysis of published dataset from Dossin et al. 2020¹³. Authors generated mouse embryonic stem cell harboring auxin-degron tagged-SHARP and compared differential gene expression between parent and auxin lines, with or without Auxin exposure. Presented are isoform-averaged TPM values of Spen transcript, averaged between biological triplicate.

Figure 5

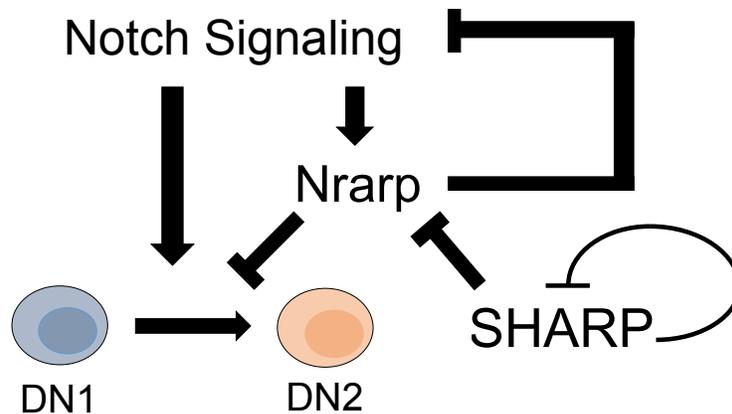


Figure 5: SHARP positively buffers a Notch-mediated Incoherent Feedforward Loop. In early T-cell development, Notch signaling promotes the DN1-DN2 developmental transition, as well as production of Nrarp, which simultaneously inhibits this developmental process. SHARP negatively regulates Nrarp production, exerting a net positive effect. SHARP negatively auto-regulates itself, additionally buffering the transcriptional circuit.

Table 1: List of High Confidence Notch-Dependent Target Genes in SCID.adh.2c2

Gene Symbol	Name
2700054A10Rik	Putative Non-Coding RNA
Afdn	Adherens Junction Formation Factor
Agfg1	Arf-GAP Domain and FG Repeats-Containing Protein
Agtr1a	Angiotensin II Receptor, Type 1A
Aldh1b1	Aldehyde Dehydrogenase 1 Family, member B1
Alp1	Alkaline Phosphatase, Intestinal
Arhgef10l	Rho Guanine Nucleotide Exchange Factor 10-like
Arsi	Arylsulfatase i
Capn5	Calpain 5
Colq	Collagen-like Tail Subunit (single strand of homotrimer) of asymmetric acetylcholinesterase
Deptor	DEP Domain Containing MTOR-interacting Protein
Dlg1	Discs Large MAGUK Scaffold Protein 1
Dsc2	Desmocollin 2
Dtx1	Deltex 1, E3 Ubiquitin Ligase
Fkbp5	FK506 Binding Protein 5
Frmd4a	FERM Domain Containing 4A
Il2ra	Interleukin 2 Receptor, Alpha Chain (CD25)
Itga9	Integrin Alpha 9
Nav2	Neuron Navigator 2
Notch1	Notch 1
Nrarp	Notch-regulated Ankyrin Repeat Protein
Shb	SRC Homology 2 Domain-Containing Transforming Protein B
Slc16a10	Solute Carrier Family 16 (monocarboxylic acid transporters), member 10
Svep1	Sushi, von Willebrand factor Type A, EGF, and Pentraxin Domain Containing 1
Tcof1	Treacle Ribosome Biogenesis Factor I
Tnfrsf8	Tumor Necrosis Factor Receptor Superfamily, member 8

Table 2: List of High Confidence Notch-Repressed Target Genes in SCID.adh.2c2

Gene Symbol	Name
5830411N06Rik	SCART 2
Btg2	B cell translocation gene 2, anti-proliferative
Cpa3	Carboxypeptidase A3, Mast Cell
Ctsl	Cathepsin L
Dok2	Docking Protein 2
Fam214a	Family with Sequence Similarity 214, Member A
Glcc1	Glucocorticoid Induced Transcript 1
Il21r	Interleukin 21 Receptor
Lztfl1	Leucine Zipper Transcription Factor-like 1
Minpp1	Multiple Inositol Polyphosphate Histidine Phosphatase 1
Nr3c1	Nuclear Receptor Subfamily 3, Group C, Member 1
Pdcd4	Programmed Cell Death 4
Pik3r3	Phosphoinositide-3-kinase Regulatory Subunit 3
Ptpre	Protein Tyrosine Phosphatase, Receptor Type, C
Rora	RAR-Related Orphan Receptor Alpha
Sdc4	Syndecan 4
Sla	Src-like Adaptor
Slc22a3	Solute Carrier Family 22 (organic cation transporter), member 3
Synpo2l	Synaptopodin 2 like
Sys1	SYS1 Golgi-Localized Integral Member Protein Homology (S. cerevisiae)

Chapter 5

CONCLUSION

A.K. Banerjee and M. Guttman

Modified versions of this conclusion were published as “Banerjee, A. K., Blanco, M. R., Bruce, E. A., Honson, D. D., Chen, L. M., Chow, A., Bhat, P., Ollikainen, N., Quinodoz, S. A., Loney, C., Thai, J., Miller, Z. D., Lin, A. E., Schmidt, M. M., Stewart, D. G., Goldfarb, D., De Lorenzo, G., Rihn, S. J., Voorhees, R. M., Botten, J. W., ... Guttman, M. (2020). SARS-CoV-2 Disrupts Splicing, Translation, and Protein Trafficking to Suppress Host Defenses. *Cell*, S0092-8674(20)31310-6. Advance online publication. <https://doi.org/10.1016/j.cell.2020.10.004>” ;

“Quinodoz, S.A., Bhat, P., Ollikainen, N., Jachowicz, J., Banerjee, A.K., Chovanec, P., Blanco, M.R., Chow, A., Markaki, Y., Plath, K. and Guttman, M. (2020). RNA promotes the formation of spatial compartments in the nucleus. *bioRxiv* 2020.08.25.267435; doi: <https://doi.org/10.1101/2020.08.25.267435>” ; and

“Blanco, M.R., Walkup IV, W.G., Bonesteele, G., Banerjee, A.K., Peyda, P., Amaya, E., Guo, J., Chow, A., Trinh, V., and Guttman, M. (in submission/review). Denaturing purifications demonstrate that PRC2 and other chromatin proteins do not bind directly to RNA *in vivo*.”

1.0 DISCUSSIONS AND IMPLICATIONS

RNA-binding interactions are critical to core cellular function. It is important then to be able to accurately identify *bona fide* RNA-protein interactions. In Chapter 2, we demonstrated that the current gold standard approach, Crosslinking Immunoprecipitation (CLIP), does not fully exclude interactions that occur in solution (*in vitro* associations). These artifacts can overshadow true interactions or lead to further inaccurate studies. In contrast, we showed that Covalent Linkage Affinity Purification (CLAP) can simultaneously identify true interactions while resolving *in vitro* association artifacts and provides a critical new framework for studying *bona fide* RNA-protein interactions *in vivo*.

In response to the current COVID-19 pandemic, we used CLAP to ask if RNA is a central target of host cell takeover by SARS-CoV-2 and observed that the virus uses at least three RNA-mediated mechanisms to blunt the innate immune system's ability to engage the interferon pathway. More specifically, we identified interactions between SARS-CoV-2 NSP16 with the U1 and U2 snRNAs to inhibit pre-mRNA splicing, NSP1 with the 18S ribosomal RNA to inhibit mRNA translation, and NSP8/9 with 7SL RNA to inhibit protein trafficking. We expect that mechanistic insights gained from these three examples, as well as the comprehensive RNA binding maps of the SARS-CoV-2 proteins presented in Chapter 2, will be a critical resource for the larger scientific community working to develop therapeutic strategies and dissect mechanisms of viral protection. More generally, global mapping of RNA binding by viral proteins via CLAP could be a generally applicable strategy for rapidly characterizing mechanisms and potential therapeutic strategies for newly emerging pathogenic RNA viruses, as well as defining new targets for well-established viruses.

In Chapter 3, we examined the role of RNA-binding interactions in gene regulation. More specifically, we leveraged the CLAP method to study a key, RNA-binding, chromatin regulatory protein called SHARP. We demonstrated that the paternally imprinted long non-coding RNA (lncRNA) *Kcnq1ot1* acts to transcriptionally repress targets in a SHARP-dependent manner, similar to the activity of *Xist* lncRNA during X-chromosome inactivation. This work supports a broader model of RNA-mediated gene regulation in the nucleus, where RNA can recruit and concentrate proteins to defined chromatin territories to spatially amplify local regulatory signals *in cis*¹⁻². *Kcnq1ot1* recruits SHARP to a defined gene cluster to silence target genes. Through its avidity to RNA and its intrinsically disordered domains, SHARP can then aggregate and compartmentalize over this chromatin territory. Control over the entire gene cluster and RNA-mediated compartment can therefore be coordinated from a single *cis* regulatory DNA element: the *Kcnq1ot1* promoter (also known as the *Kcnq1* Imprint Control Region)³⁻⁴.

What was most surprising to us from this dataset was the identification of SHARP/pre-mRNA interactions, which actually outnumber the SHARP/lncRNA interactions observed. Many of these pre-mRNAs localize broadly to chromatin in a similar manner to lncRNAs. These data suggest that RNA-mediated compartmentalization of effector proteins may act more pervasively to coordinate gene regulation across the nucleus. While the reasons underlying SHARP's interaction with chromatin-retained pre-mRNAs and possible gene regulation *in cis* are currently unknown, it is interesting to note that several pre-mRNA targets are documented to be haploinsufficient or have pathologies sensitive to dosage⁵⁻⁸. If these SHARP/pre-mRNA interactions are capable of seeding functional regulatory compartments within the nucleus (as suggested by preliminary data from SHARP's interaction with *Spen* mRNA on expression of neighboring genes adjacent to *Spen*) it questions a fundamental dichotomy within RNA biology

(that is regulatory RNA and coding RNA are mutually exclusive) and expands the field's definition of functional regulatory RNAs (**Figure 1**).

In Chapter 4, we dissected a SHARP-mediated transcriptional circuit in early T-cell development. Given previous evidence demonstrating SHARP to be a negative regulator of Notch signaling in other hematopoietic contexts, we were surprised to observe that SHARP loss of function does not functionally phenocopy Notch gain of function in this system⁹⁻¹⁰. Through a series of targeted perturbations, we demonstrated that SHARP negatively regulates a negative regulator of Notch (*Nrarp*) and in extension, positively buffers a Notch-mediated Incoherent Feedforward Loop¹¹. We also identified an interaction between SHARP and its mRNA using CLAP and showed that SHARP autoregulates its own transcription.

Incoherent Feedforward Loops, like the circuit discussed above, are commonly used regulatory circuits that can be used for several different purposes, including improving stability and dynamics of a conjoined negative feedback loop, increasing gene-regulatory network response time and expression pulsatility, biphasic expression responses, fold-change detection, and adaptive tuning of gene expression¹²⁻²⁵. Although the specific consequences of SHARP's interaction with its own mRNA on Notch signaling remains to be experimentally determined, circuits stabilized by RNA-binding proteins could off-set environmentally variable inputs (as potentially observed with Delta-like ligand availability for progenitors as they migrate within the thymic microenvironment) or establish a band-pass filtering for sustained effective signal over time.

2.0 FUTURE DIRECTIONS

As stated very eloquently by Alan Chalmers regarding the progress of scientific understanding, “[w]e start off confused and end up confused on a higher level”²⁶. While the three presented studies illustrate the diversity of roles played by RNA-interactions in development, cellular homeostasis, and disease, our data poses many additional questions for discussion and further investigation.

To begin, what is the implication of an interaction between RNA and an RNA-binding protein? Important clues for tackling such questions include the function of the specific RNA or RNA-binding protein in question (if available), localization of these components within the cell (if available), nucleotide resolution localization of the RNA-binding protein on its interacting RNA, and context of identification. For some interactions, interpretation may be clear. For example, in Chapter 2 we observed specific interactions between NSP16 and the U1 and U2 snRNAs, specifically at the splice site recognition sequence of U1 and adjacent to the branch point recognition sequence of U2; given the importance of these snRNAs in pre-mRNA splicing, it was straightforward to hypothesize the implications of such a viral protein/host RNA interaction²⁷.

While it is tempting to apply a reductionist framework to viral protein/host RNA interactions and assume that all interactions must promote viral fitness, such hypothesis may not be justified. For example, we do not understand the reasons underlying interactions between NSP12 and JUN mRNA, NSP9 and COPS5 mRNA, or those between the Nucleocapsid protein and a specific subset of 83 host mRNAs. Further study is required to determine the role, if any, that these interactions play in viral propagation.

There are many cases where cellular context is not as clear (**Figure 2**). For example, Lamin B Receptor (LBR) protein plays a key role in tethering the inactive X-chromosome to the nuclear lamina and also has annotated sterol reductase enzymatic activity²⁸⁻²⁹. Given the interaction between LBR and Dsg2 (Desmoglein 2) pre-mRNA or Ttc28 (Tetratricopeptide Repeat Domain 28) pre-mRNA, it may or may not be reasonable to assume all LBR interactions act in a similar manner. While the CLAP method has enabled genome-wide, high-specificity identification of RNA-binding interactions, it is now equally (if not more) important to develop massively multiplexed, hypothesis-agnostic, experimental perturbation systems to dissect their significance without reliance on *a priori* knowledge.

Historically, much has been learned by bootstrapping upon previously characterized biological mechanisms and using this approach to understand SHARP is no exception. The interactions of SHARP with RNA, as well as its role in gene regulatory networks and chromatin regulation are well known; SHARP is therefore amenable to carefully designed experiments with clear, logical and testable predictions^{9,30-37}. Its functional interactions with the Kcnq1ot1 lncRNA, along with the identification of several more targets, point towards pervasive mechanisms of RNA-mediated gene regulation and compartment formation within the nucleus. However, this raises several new questions to be explored.

First given the similarities between Xist- and Kcnq1ot1-mediated silencing, it remains to be determined why Xist is capable of establishing durable transcriptional memory, while Kcnq1ot1 is not³⁸⁻⁴⁰. Second as discussed in Chapter 4 within the context of Incoherent Feedforward Loops, transcriptional circuits stabilized by RNA can promote robustness in response to change, potentially off-setting environmentally variable inputs or establishing band-pass filtering^{25,41-52}.

SHARP autoregulates its transcription. Therefore, the role of SHARP autoregulation in compartment robustness and crosstalk remains to be explored (**Figure 3**).

We began discussing the advantages of RNA as a flexible biomolecule with exclusive capabilities based on its unique attributes. We showed that RNA plays a key role in core cellular machinery, such as pre-mRNA splicing and translation, and as such, is the target of host-cell takeover by SARS-CoV-2. We showed that RNA can spatially amplify regulatory signals within the nucleus to coordinate gene regulation *in cis* and speculate that it has yet uncharacterized roles in nuclear compartment stabilization and crosstalk. It is therefore appropriate to conclude with a simple yet open-ended question: How deep does the RNA rabbit hole go?

3.0 REFERENCES

1. Engreitz, J.M., Ollikainen, N., and Guttman, M. (2016). Long non-coding RNAs: spatial amplifiers that control nuclear structure and gene expression. *Nature reviews. Molecular cell biology*, 17(12), 756–770. <https://doi.org/10.1038/nrm.2016.126>
2. Strehle, M., and 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>
3. Smilnich, N.J., Day, C.D., Fitzpatrick, G.V., Caldwell, G.M., Lossie, A.C., Cooper, P.R., Smallwood, A.C., Joyce, J.A., Schofield, P.N., Reik, W., Nicholls, R.D., Weksberg, R., Driscoll, D.J., Maher, E.R., Shows, T.B., and Higgins, M.J. (1999). A maternally methylated CpG island in KvLQT1 is associated with an antisense paternal transcript and loss of imprinting in Beckwith-Wiedemann syndrome. *Proceedings of the National Academy of Sciences of the United States of America*, 96(14), 8064–8069. <https://doi.org/10.1073/pnas.96.14.8064>
4. Mancini-DiNardo, D., Steele, S.J., Ingram, R.S., and Tilghman, S.M. (2003). A differentially methylated region within the gene *Kcnq1* functions as an imprinted promoter and silencer. *Human molecular genetics*, 12(3), 283–294. <https://doi.org/10.1093/hmg/ddg024>

5. Verberne, E.A., Goh, S., England, J., van Ginkel, M., Rafael-Croes, L., Maas, S., Polstra, A., Zarate, Y.A., Bosanko, K.A., Pechter, K.B., Bedoukian, E., Izumi, K., Chaudhry, A., Robin, N.H., Boothe, M., Lippa, N.C., Aggarwal, V., De Vivo, D.C., Lehman, A., Study, C., ... Campeau, P.M. (2020). JARID2 haploinsufficiency is associated with a clinically distinct neurodevelopmental syndrome. *Genetics in medicine : official journal of the American College of Medical Genetics*, 10.1038/s41436-020-00992-z. Advance online publication. <https://doi.org/10.1038/s41436-020-00992-z>
6. Ge, Y., Schuster, M.B., Pundhir, S., Rapin, N., Bagger, F.O., Sidiropoulos, N., Hashem, N., and Porse, B.T. (2019). The splicing factor RBM25 controls MYC activity in acute myeloid leukemia. *Nature communications*, 10(1), 172. <https://doi.org/10.1038/s41467-018-08076-y>
7. Kim, H., Kang, K., Ekram, M.B., Roh, T.Y., and Kim, J. (2011). Aebp2 as an epigenetic regulator for neural crest cells. *PloS one*, 6(9), e25174. <https://doi.org/10.1371/journal.pone.0025174>
8. Müller-McNicoll, M., Rossbach, O., Hui, J., and Medenbach, J. (2019). Auto-regulatory feedback by RNA-binding proteins. *Journal of molecular cell biology*, 11(10), 930–939. <https://doi.org/10.1093/jmcb/mjz043>
9. Kuroda, K., Han, H., Tani, S., Tanigaki, K., Tun, T., Furukawa, T., Taniguchi, Y., Kurooka, H., Hamada, Y., Toyokuni, S., and Honjo, T. (2003). Regulation of marginal zone B cell development by MINT, a suppressor of Notch/RBP-J signaling pathway. *Immunity*, 18(2), 301–312. [https://doi.org/10.1016/s1074-7613\(03\)00029-3](https://doi.org/10.1016/s1074-7613(03)00029-3)
10. Yabe, D., Fukuda, H., Aoki, M., Yamada, S., Takebayashi, S., Shinkura, R., Yamamoto, N., & Honjo, T. (2007). Generation of a conditional knockout allele for mammalian Spen protein Mint/SHARP. *Genesis (New York, N.Y. : 2000)*, 45(5), 300–306. <https://doi.org/10.1002/dvg.20296>
11. Lamar, E., Deblandre, G., Wettstein, D., Gawantka, V., Pollet, N., Niehrs, C., & Kintner, C. (2001). Nrarp is a novel intracellular component of the Notch signaling pathway. *Genes & development*, 15(15), 1885–1899. <https://doi.org/10.1101/gad.908101>
12. Reeves, G.T. (2019). The engineering principles of combining a transcriptional incoherent feedforward loop with negative feedback. *Journal of biological engineering*, 13, 62. <https://doi.org/10.1186/s13036-019-0190-3>
13. Basu, S., Mehreja, R., Thiberge, S., Chen, M.T., and Weiss, R. (2004). Spatiotemporal control of gene expression with pulse-generating networks. *Proceedings of the National Academy of Sciences of the United States of America*, 101(17), 6355–6360. <https://doi.org/10.1073/pnas.0307571101>
14. Mangan, S., and Alon, U. (2003). Structure and function of the feed-forward loop network motif. *Proceedings of the National Academy of Sciences of the United States of America*, 100(21), 11980–11985. <https://doi.org/10.1073/pnas.2133841100>

15. Mangan, S., Itzkovitz, S., Zaslaver, A., and Alon, U. (2006). The incoherent feed-forward loop accelerates the response-time of the gal system of *Escherichia coli*. *Journal of molecular biology*, 356(5), 1073–1081. <https://doi.org/10.1016/j.jmb.2005.12.003>
16. Alon, U. (2007). Network motifs: theory and experimental approaches. *Nature reviews. Genetics*, 8(6), 450–461. <https://doi.org/10.1038/nrg2102>
17. Shoval, O., and Alon, U. (2010). SnapShot: network motifs. *Cell*, 143(2), 326–e1. <https://doi.org/10.1016/j.cell.2010.09.050>
18. Macía, J., Widder, S., and Solé, R. (2009). Specialized or flexible feed-forward loop motifs: a question of topology. *BMC systems biology*, 3, 84. <https://doi.org/10.1186/1752-0509-3-84>
19. Entus, R., Aufderheide, B., and Sauro, H.M. (2007). Design and implementation of three incoherent feed-forward motif based biological concentration sensors. *Systems and synthetic biology*, 1(3), 119–128. <https://doi.org/10.1007/s11693-007-9008-6>
20. Kaplan, S., Bren, A., Dekel, E., and Alon, U. (2008). The incoherent feed-forward loop can generate non-monotonic input functions for genes. *Molecular systems biology*, 4, 203. <https://doi.org/10.1038/msb.2008.43>
21. Kim, D., Kwon, Y.K., and Cho, K.H. (2008). The biphasic behavior of incoherent feed-forward loops in biomolecular regulatory networks. *BioEssays : news and reviews in molecular, cellular and developmental biology*, 30(11-12), 1204–1211. <https://doi.org/10.1002/bies.20839>
22. Goentoro, L., Shoval, O., Kirschner, M.W., and Alon, U. (2009). The incoherent feedforward loop can provide fold-change detection in gene regulation. *Molecular cell*, 36(5), 894–899. <https://doi.org/10.1016/j.molcel.2009.11.018>
23. Shoval, O., Goentoro, L., Hart, Y., Mayo, A., Sontag, E., and Alon, U. (2010). Fold-change detection and scalar symmetry of sensory input fields. *Proceedings of the National Academy of Sciences of the United States of America*, 107(36), 15995–16000. <https://doi.org/10.1073/pnas.1002352107>
24. Hong, J., Brandt, N., Abdul-Rahman, F., Yang, A., Hughes, T., and Gresham, D. (2018). An incoherent feedforward loop facilitates adaptive tuning of gene expression. *eLife*, 7, e32323. <https://doi.org/10.7554/eLife.32323>
25. Hart, Y., and Alon, U. (2013). The utility of paradoxical components in biological circuits. *Molecular cell*, 49(2), 213–221. <https://doi.org/10.1016/j.molcel.2013.01.004>
26. Chalmers, A. (1982). *What is This Thing Called Science?* Open University Press. Introduction, xix.

27. Séraphin, B., Kretzner, L., and Rosbash, M. (1988). A U1 snRNA:pre-mRNA base pairing interaction is required early in yeast spliceosome assembly but does not uniquely define the 5' cleavage site. *The EMBO journal*, 7(8), 2533–2538.
28. Tsai, P.L., Zhao, C., Turner, E., and Schlieker, C. (2016). The Lamin B receptor is essential for cholesterol synthesis and perturbed by disease-causing mutations. *eLife*, 5, e16011. <https://doi.org/10.7554/eLife.16011>
29. Chen, C.K., Blanco, M., Jackson, C., Aznauryan, E., Ollikainen, N., Surka, C., Chow, A., Cerase, A., McDonel, P., and Guttman, M. (2016). Xist recruits the X chromosome to the nuclear lamina to enable chromosome-wide silencing. *Science (New York, N.Y.)*, 354(6311), 468–472. <https://doi.org/10.1126/science.aae0047>
30. 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., and 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>
31. 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>
32. Moindrot, B., Cerase, A., Coker, H., Masui, O., Griizenhout, A., Pintacuda, G., Schermelleh, L., Nesterova, T.B., and Brockdorff, N. (2015). A Pooled shRNA Screen Identifies Rbm15, Spen, and Wtap as Factors Required for Xist RNA-Mediated Silencing. *Cell reports*, 12(4), 562–572. <https://doi.org/10.1016/j.celrep.2015.06.053>
33. Monfort, A., Di Minin, G., Postlmayr, A., Freimann, R., Arieti, F., Thore, S., and Wutz, A. (2015). Identification of Spen as a Crucial Factor for Xist Function through Forward Genetic Screening in Haploid Embryonic Stem Cells. *Cell reports*, 12(4), 554–561. <https://doi.org/10.1016/j.celrep.2015.06.067>
34. Shi, Y., Downes, M., Xie, W., Kao, H.Y., Ordentlich, P., Tsai, C.C., Hon, M., and Evans, R.M. (2001). Sharp, an inducible cofactor that integrates nuclear receptor repression and activation. *Genes & development*, 15(9), 1140–1151. <https://doi.org/10.1101/gad.871201>
35. Oswald, F., Kostezka, U., Astrahantseff, K., Bourteele, S., Dillinger, K., Zechner, U., Ludwig, L., Wilda, M., Hameister, H., Knöchel, W., Liptay, S., and Schmid, R.M. (2002). SHARP is a novel component of the Notch/RBP-Jkappa signalling pathway. *The EMBO journal*, 21(20), 5417–5426. <https://doi.org/10.1093/emboj/cdf549>
36. Tsuji, M., Shinkura, R., Kuroda, K., Yabe, D., and Honjo, T. (2007). Msx2-interacting nuclear target protein (Mint) deficiency reveals negative regulation of early thymocyte differentiation by Notch/RBP-J signaling. *Proceedings of the National Academy of Sciences of the United States of America*, 104(5), 1610–1615. <https://doi.org/10.1073/pnas.0610520104>

37. Dossin, F., Pinheiro, I., Żylicz, J.J., Roensch, J., Collombet, S., Le Saux, A., Chelmicki, T., Attia, M., Kapoor, V., Zhan, Y., Dingli, F., Loew, D., Mercher, T., Dekker, J., and Heard, E. (2020). SPEN integrates transcriptional and epigenetic control of X-inactivation. *Nature*, 578(7795), 455–460. <https://doi.org/10.1038/s41586-020-1974-9>
38. Brown, C.J., and Willard, H.F. (1994). The human X-inactivation centre is not required for maintenance of X-chromosome inactivation. *Nature*, 368(6467), 154–156. <https://doi.org/10.1038/368154a0>
39. Wutz, A., and Jaenisch, R. (2000). A shift from reversible to irreversible X inactivation is triggered during ES cell differentiation. *Molecular cell*, 5(4), 695–705. [https://doi.org/10.1016/s1097-2765\(00\)80248-8](https://doi.org/10.1016/s1097-2765(00)80248-8)
40. Zhang, H., Zeitz, M.J., Wang, H., Niu, B., Ge, S., Li, W., Cui, J., Wang, G., Qian, G., Higgins, M.J., Fan, X., Hoffman, A.R., and Hu, J.F. (2014). Long noncoding RNA-mediated intrachromosomal interactions promote imprinting at the Kcnq1 locus. *The Journal of cell biology*, 204(1), 61–75. <https://doi.org/10.1083/jcb.201304152>
41. Bleris, L., Xie, Z., Glass, D., Adadey, A., Sontag, E., and Benenson, Y. (2011). Synthetic incoherent feedforward circuits show adaptation to the amount of their genetic template. *Molecular systems biology*, 7, 519. <https://doi.org/10.1038/msb.2011.49>
42. Ebert, M.S., and Sharp, P.A. (2012). Roles for microRNAs in conferring robustness to biological processes. *Cell*, 149(3), 515–524. <https://doi.org/10.1016/j.cell.2012.04.005>
43. Elowitz, M.B., Levine, A.J., Siggia, E.D., and Swain, P.S. (2002). Stochastic gene expression in a single cell. *Science (New York, N.Y.)*, 297(5584), 1183–1186. <https://doi.org/10.1126/science.1070919>
44. Blake, W.J., KAern, M., Cantor, C.R., and Collins, J.J. (2003). Noise in eukaryotic gene expression. *Nature*, 422(6932), 633–637. <https://doi.org/10.1038/nature01546>
45. Blake, W.J., Balázsi, G., Kohanski, M.A., Isaacs, F.J., Murphy, K.F., Kuang, Y., Cantor, C.R., Walt, D.R., and Collins, J.J. (2006). Phenotypic consequences of promoter-mediated transcriptional noise. *Molecular cell*, 24(6), 853–865. <https://doi.org/10.1016/j.molcel.2006.11.003>
46. Paulsson, J. (2004). Summing up the noise in gene networks. *Nature*, 427(6973), 415–418. <https://doi.org/10.1038/nature02257>
47. Raser, J.M., and O'Shea, E.K. (2005). Noise in gene expression: origins, consequences, and control. *Science (New York, N.Y.)*, 309(5743), 2010–2013. <https://doi.org/10.1126/science.1105891>

48. Newman, J.R., Ghaemmaghami, S., Ihmels, J., Breslow, D.K., Noble, M., DeRisi, J.L., and Weissman, J.S. (2006). Single-cell proteomic analysis of *S. cerevisiae* reveals the architecture of biological noise. *Nature*, 441(7095), 840–846. <https://doi.org/10.1038/nature04785>
49. Sigal, A., Milo, R., Cohen, A., Geva-Zatorsky, N., Klein, Y., Liron, Y., Rosenfeld, N., Danon, T., Perzov, N., and Alon, U. (2006). Variability and memory of protein levels in human cells. *Nature*, 444(7119), 643–646. <https://doi.org/10.1038/nature05316>
50. Osella, M., Bosia, C., Corá, D., and Caselle, M. (2011). The role of incoherent microRNA-mediated feedforward loops in noise buffering. *PLoS computational biology*, 7(3), e1001101. <https://doi.org/10.1371/journal.pcbi.1001101>
51. Li, X., Cassidy, J.J., Reinke, C.A., Fischboeck, S., and Carthew, R.W. (2009). A microRNA imparts robustness against environmental fluctuation during development. *Cell*, 137(2), 273–282. <https://doi.org/10.1016/j.cell.2009.01.058>
52. Hornstein, E., and Shomron, N. (2006). Canalization of development by microRNAs. *Nature genetics*, 38 Suppl, S20–S24. <https://doi.org/10.1038/ng1803>
53. Quinodoz, S.A., Bhat, P., Ollikainen, N., Jachowicz, J., Banerjee, A.K., Chovanec, P., Blanco, M.R., Chow, A., Markaki, Y., Plath, K. and Guttman, M. (2020). RNA promotes the formation of spatial compartments in the nucleus. bioRxiv 2020.08.25.267435; <https://doi.org/10.1101/2020.08.25.267435>

4.0 FIGURES

Figure 1

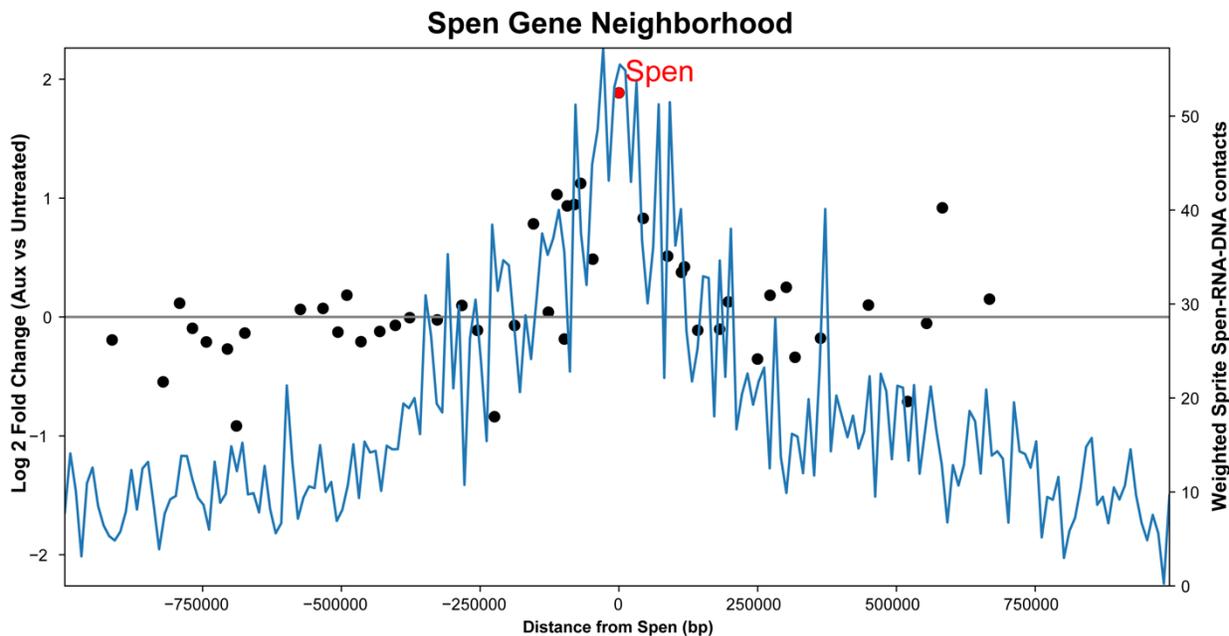


Figure 1: Preliminary evidence of *cis* gene regulation by SHARP/pre-mRNAs. Comparison of SHARP-mediated differential gene expression, published in Dossin et al., with localization of Spen RNA on chromatin from RNA-DNA SPRITE (both datasets aligned to mm10)^{37,53}. Data is presented along linear chromatin distance, with genes centered around *Spen* gene body (displayed in red). Differential gene expression for individual genes is plotted in black, while localization of Spen RNA on chromatin is plotted in blue. For quantification methods used, please refer to Chapter 4 Materials and Methods section.

Figure 2

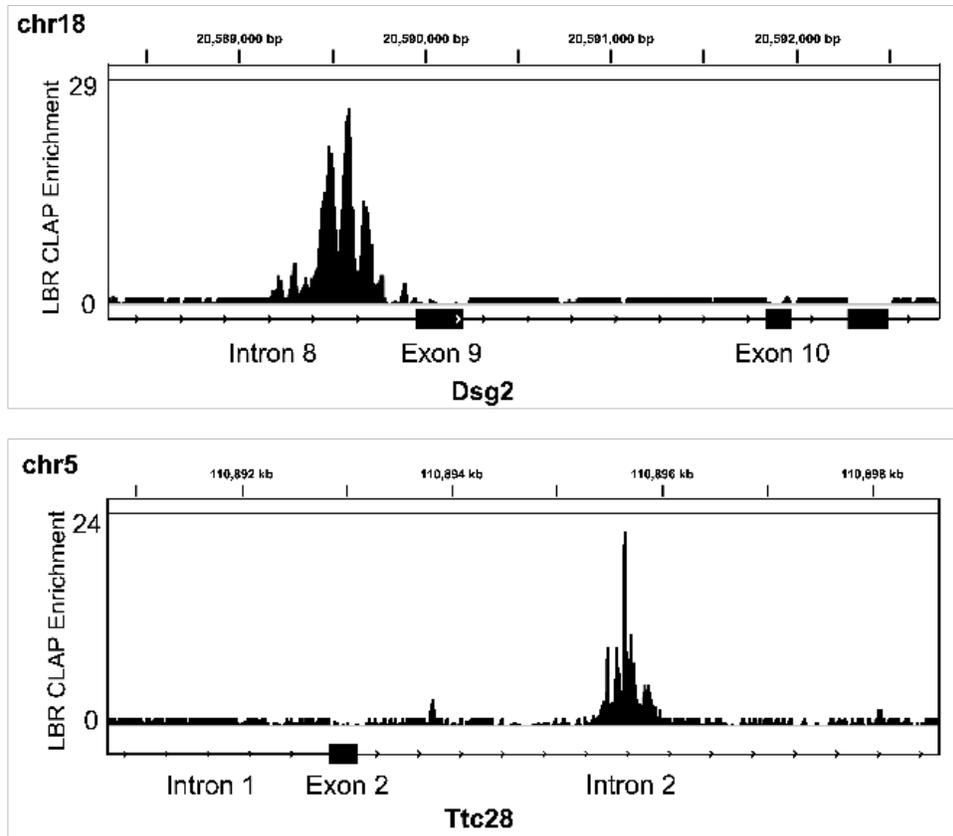


Figure 2: CLAP reveals a growing number of uncharacterized RNA-protein interactions.

Input normalized enrichment plot of LBR CLAP-Seq to Dsg2 and Ttc28 pre-mRNAs, performed using mouse embryonic stem cell lines engineered with endogenous integrations of Spy-V5 tags into the *LBR* gene body.

Figure 3

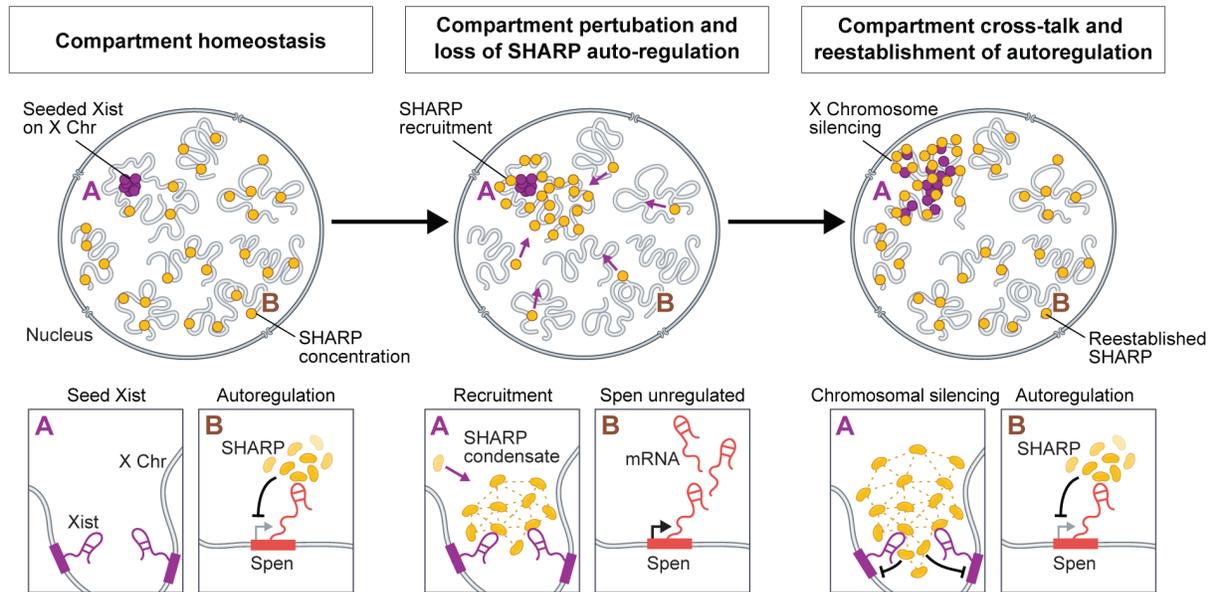


Figure 3: Shared effector proteins enable compartment cross-talk. Upon Xist-mediated silencing, SHARP is dynamically recruited and concentrated to the inactive X-chromosome. Recruitment of SHARP away from its own locus would lead to de-repression and lack of autoregulation, resulting increased Spen transcription and SHARP protein production to re-establish autoregulation.