Revealing the Mechanism of Xist-mediated Silencing

Thesis by

Chun-Kan Chen

In Partial Fulfillment of the Requirements for

the degree of

Doctor of Philosophy

CALIFORNIA INSTITUTE OF TECHNOLOGY

Pasadena, California

2018

Defended November 1, 2017

© 2017

Chun-Kan Chen
ORCID: 0000-0002-1194-9137

ACKNOWLEDGEMENTS

First of all, I'd like to thank my great mentor, Dr. Mitch Guttman (California Institute of Technology, Pasadena, CA), who led me to become an independent researcher and gave me valuable advice that guided me to accomplish this thesis. He has always been supportive of my future plans and career goals. I really enjoyed every discussion we have had. We often generated some interesting ideas for projects during our discussions. I would also like to send my thanks to my lab mates, Amy Chow, Mario Blanco, and Erik Aznauryan, who helped me with many experiments to move the project forward. I'd like to acknowledge Dr. Kathrin Plath (University of California, Los Angeles, Los Angeles, CA) for the collaboration and his critical comments on this project. Also, I want to thank Jesse Engreitz and Patrick McDonel, who provided helpful comments and suggestions to the project.

I want to thank my parents, brother, and parents-in-law who provided both instrumental and emotional support to assist me in completing my Ph.D. degree. I also want to thank my friends, Lily Chen, Pei-Ying Lin, Tzu-Yao Wang, and Wei Li, for giving me valuable social support during my years in graduate school.

Last but not least, I would like to send my special thanks to my wife, Christine Juang, who has always been supportive. She always encouraged me to pursue my goal, even though sometimes it meant going to the lab with me on weekends and waiting with me doing experiments for several hours. Her support enabled me to overcome the obstacles I have encountered these years, and made me to have what I have achieved today. I am truly grateful for all your support. Thank you.

ABSTRACT

Xist initiates XCI by spreading across the future inactive X-chromosome, excluding RNA polymerase II, recruiting the polycomb repressive complex and its associated repressive chromatin modifications, and repositioning active genes into a transcriptionally silenced nuclear compartment. While much is known about the events that occur during XCI, the mechanism by which Xist carries out these various roles remains unclear. Here we identify ten proteins that directly associate with Xist, and we further show that three of these proteins are required for Xist-mediated transcriptional silencing. One of these proteins, SHARP, which is known to interact with the SMRT co-repressor that activates HDAC3, is not only essential for silencing, but is also required for the exclusion of PolII from the inactive X. We show that both SMRT and HDAC3 are required for Xist-mediated silencing and RNA polymerase II exclusion. Another of these proteins, LBR, is required for repositioning actively transcribed genes into the Xist-silenced compartment. We further show that Xist, through its interaction with LBR, a protein that is anchored in the inner nuclear membrane, would effectively reposition Xist-coated DNA to the nuclear lamina, thereby changing the accessibility of other genes on the X-chromosome to enable Xist to spread to active genes across the entire chromosome to silence chromosome-wide transcription. Together, these results present an integrative picture of how Xist can scaffold multiple proteins to orchestrate the complex functions required for the establishment of the inactive X-chromosome.

PUBLISHED CONTENT AND CONTRIBUTIONS

1. McHugh, Colleen A., et al. "The Xist lncRNA directly interacts with SHARP to silence transcription through HDAC3." Nature 521.7551 (2015): 232.

   CKC participated in the conception of the project, designed, performed, and analyzed Xist functional experiments, and wrote the manuscript with input from all authors.

2. Chen, Chun-Kan, et al. "Xist recruits the X chromosome to the nuclear lamina to enable chromosome-wide silencing." Science 354.6311 (2016): 468-472.

   CKC participated in the conception of the project, designed, performed, and analyzed the data, and wrote the manuscript with input from all authors.

All contents are reprinted with permission from the copyright holder, Springer and American Association for the Advancement of Science, respectively.

TABLE OF CONTENTS

LIST OF ILLUSTRATIONS AND/OR TABLES

Number *Page*

Chapter 1

INTRODUCTION

1.1 X Chromosome Inactivation

In mammals, males and females carry a different number of X chromosomes. Females have two X chromosomes, while males have one X and one Y chromosome(*1*). The X chromosome is larger and carries many genes, while the Y chromosome is smaller and carries much fewer genes(*2, 3*). In this case, females would seem to express twice the amount of the X-linked genes in comparison to males. Therefore, to compensate for the dosage difference between males and females, one of the two X chromosomes in females is transcriptionally inactive during early development, and the inactive state is inherited throughout cell division(*1, 4*). This process, or X-chromosome inactivation (XCI), is essential for proper embryo development of females. Failure of proper XCI will lead to sex-differential embryo lethality(*5*). XCI is found in all mammals, but has not yet been found in other groups of animals, such as *Drosophila* or *C. elegans*, in which they compensate the dosage difference of X chromosome with different mechanisms(*6, 7*).

In female mice, XCI occurs in two waves during early development. The first wave, imprinted XCI, occurs around 4-8 cell stage, where the paternal X chromosome (Xp) is transcriptionally inactive at this stage(*8-10*). The inactive state of Xp is reversed to active at the blastocyst stage, followed by the second wave of XCI, random XCI, where either the maternal X chromosome (Xm) or Xp is transcriptionally silenced(*11, 12*). Once the

inactive state of either Xm or Xp is established, the inactive state of the chromosome is maintained throughout cell division, except for some specific cell types, such as germ cells(*13*). In this thesis, we will mainly focus on the molecular mechanism of the regulation of random XCI.

Chromosome-wide transcriptional silencing on Xi is one of the most significant events occurs during XCI. Immunofluorescence staining reveals that RNA polymerase II (PolII) is depleted (excluded) from the inactive X chromosome (Xi). PolII exclusion is one of the earliest events of XCI, which cause the transcriptional silencing on Xi(*14*). In addition to the transcriptional silencing, many other events also occur exclusively on Xi. One of them is the dramatic change of chromatin structure. From visual observation, Xi forms a dense and distinct nuclear compartment known as the Barr body(*15*). In comparison to the active X chromosome (Xa), Xi undergoes chromosome-wide loss of topological association domains (TADs) and forms two mega-domains(*16, 17*). Xi becomes more compact and are enriched for heterochromatin structure(*18*). In addition, the nuclear organization also changes during XCI. Xi becomes localized at the periphery of the nucleus, and is associated with the nuclear lamina compartment(*14, 19*). Furthermore, XCI results in a cascade of chromatin modification on Xi, such as hypoacetylation of H4, H2A, and H3(*20, 21*), H3K9 methylaion(*22-25*), and macroH2A recruitment(*26*). Interestingly, the regions contain these chromatin modification is associated with the heterochromatin region on Xi(*18*). XCI also leads to the enrichment of some proteins on Xi, such as polycomb repressive complex 1 and 2 (PRC1/2), which mediates H2AK119Ub and H3K27me3 on Xi respectively(*27, 28*), and SMCHD1, a noncanonical member of the SMC family of chromosomal proteins, which

plays an important role in the DNA methylation on Xi(*29*). Finally, at the later stage of XCI, the CpG islands on the promoter region of the silenced genes on Xi are heavily methylated, which is crucial to maintain the silencing state of Xi throughout cell division(*30, 31*).

1.2 Xist/XIST LncRNA is a Key Regulatory Component for the Initiation of XCI

The regulatory element on X chromosome for the initiation of XCI was identified by genetic studies, which was later called the X-inactivation center (XIC)(*32, 33*). The exact location of the XIC on X chromosome was identified by chromosome translocation study, where the autosome was silenced when the X chromosome fragment containing the XIC was translocated and fused with that autosome(*34*). This finding also suggesting that the XIC alone is sufficient to initiate chromosome-wide silencing even in autosomes. By comparing the differential gene expression pattern on Xa and Xi, a gene located within the XIC that was exclusively expressed on Xi, the Xi-specific transcript (Xist, or XIST in human), was identified(*35*). Sequence analysis of Xist/XIST RNA showed no conserved open reading frame, suggesting that it does not encode for any protein(*36*), and was later characterized as a long non-coding RNA (lncRNA). Imaging analysis showed that Xist/XIST RNA is localized in the nucleus and coated the entire Xi(*36, 37*). The expression timing of Xist/XIST RNA coincides with the developmental window of the initiation of XCI, suggesting that Xist/XIST RNA may play a role in the initiation of XCI(*8*).

Mouse Xist transcript is ~17 kb in length (~21 kb for human XIST) and is polyadenylated(*36, 38*). Mouse female embryonic stem cells (ESCs) carry a heterozygous deletion of Xist promoter region or its first exon show skewed allele-specific transcriptional silencing; only the wild-type X chromosome is transcriptional silenced and becomes Xi, while the X chromosome carries the deletion are always actively transcribed and becomes Xa(*39, 40*). In addition, blocking Xist RNA to coat the X chromosome with antisense oligonucleotide analogs also prevents the transcriptional silencing during

XCI(*41*). These results showed that the expression of Xist and its capability of coating the X chromosome is required for the initiation of XCI and the chromosome-wide transcriptional silencing on Xi. Furthermore, inserting the entire Xist/XIST locus into any of the autosomes and driving its expression leads to chromosome-wide transcriptional silencing of the autosome(*42*). In addition, expressing Xist/XIST RNA on the X chromosome in male cells results in transcriptional silencing of the X chromosome, which eventually leads to cell death due to lack of the X-linked genes expression(*43*). These results showed that Xist/XIST a key regulator of XCI which is necessary and sufficient for chromosome-wide transcriptional silencing.

Intensive studies on the function of Xist/XIST reveal a more extensive role that it plays in initiating and regulating XCI. Xist/XIST initiates XCI by spreading across the future inactive X-chromosome, excluding RNA PolII, and repositioning active genes into a transcriptionally silenced nuclear compartment(*14, 44-46*). All of these roles – localization, RNA PolII exclusion, and repositioning – are required for proper silencing of transcription during the initiation of XCI(*46*). Furthermore, Xist/XIST triggers a cascade of events on the inactive X-chromosome including recruitment of the polycomb repressive complex and its associated H3K27me3 repressive chromatin modifications(*27, 47, 48*), loss of active acetylation(*49-52*) and methylation(*22*) chromatin modifications, chromosome-wide compaction(*19, 53*), and repositioning to the nuclear lamina(*54, 55*). Xist/XIST is also required for the change of chromatin structure on Xi. Disrupting proper function of Xist in female mouse ESCs resulted in lack of loss of TADs on Xi and no mega-domain formation, leading to a

chromatin structure similar to Xa(*56*). These results showed that Xist/XIST is a key component that is essential for initiating and regulating various events during XCI.

1.3 The Mechanism of Chromosome-wide Xist/XIST-mediated Transcriptional Silencing

While much is known about the events that occur during XCI, the mechanism by which Xist carries out these various roles remains unclear because we still do not know the protein complexes that interact with Xist to initiate transcriptional silencing, recruit chromatin-modifying proteins, compact chromatin structure, and reposition the inactive X-chromosome(*57-59*). Over the last two decades, numerous attempts have been made to define the protein complexes that interact with Xist(*57, 58*) and that are required for its various roles in XCI[24,25 26,27 28,29]. Most studies have used prior knowledge of the molecular events that occur on the X-chromosome, as well as their timing during the initiation of XCI, to define potential Xist-interacting proteins(*27, 57, 58, 60*). While individual proteins have been identified that associate with Xist(*60-65*), these proteins cannot explain the various functional roles mediated by Xist. For example, we still do not know how Xist initiates transcriptional silencing: indeed, perturbations of the proteins identified so far, including components of the PRC2 complex, have no impact on Xist-mediated transcriptional silencing(*61, 62, 66, 67*).

The main challenge in deciphering the mechanisms by which Xist, or other lncRNAs, function is that there are currently no methods to comprehensively define the proteins that interact with a lncRNA in the cell. Currently, the two classes of methods for studying lncRNA-protein interactions are immunoprecipitation of a specific protein(*60, 68-70*) and *in vitro* association between a labeled RNA and cellular lysates(*71-77*). Both of these approaches are limited in their ability to define lncRNA-protein complexes that occur in cells

because the immunoprecipitation methods require selecting specific candidate interacting proteins to study, and *in vitro* association methods fail to distinguish between interactions that occur in the cell from those that occur in solution (*25, 68, 77*). Accordingly, defining the mechanism of Xist and other lncRNAs requires new approaches that can specifically identify the direct lncRNA-interacting proteins *in vivo* without prior knowledge of their identity. Therefore, we decided to develop a method that allows us to identify the direct-interacting proteome of a lncRNA of interest. With the method. We will be able to reveal the Xist-interacting proteins and give us some insights of the mechanism of Xist-mediated silencing.

In Chapter II, it is demonstrated that a newly developed method, RNA Antisense Purification (RAP), can identify the direct interacting proteins of Xist *in vivo*, and a screening system that identifies three proteins which are essential for Xist-mediated silencing.

In Chapter III, it is demonstrated that one of these essential proteins, SHARP, is required for Xist-mediated silencing through excluding PolII from Xist-coated territory.

In Chapter IV, it is demonstrated that another essential protein, LBR, is required for Xist-mediated silencing through recruiting Xist-associated DNA to nuclear lamina that enables Xist to spread across the entire chromosome.

Chapter V is the discussion of a proposed model of the mechanism of Xist-mediated silencing based on the findings of this thesis, and some unanswered questions about Xist and XCI that require more studies in the future.

References

1. J. A. Graves, Evolution of vertebrate sex chromosomes and dosage compensation. *Nat Rev Genet* **17**, 33-46 (2016).
2. Y. Q. Soh *et al.*, Sequencing the mouse Y chromosome reveals convergent gene acquisition and amplification on both sex chromosomes. *Cell* **159**, 800-813 (2014).
3. J. L. Mueller *et al.*, Independent specialization of the human and mouse X chromosomes for the male germ line. *Nat Genet* **45**, 1083-1087 (2013).
4. M. F. Lyon, Sex chromatin and gene action in the mammalian X-chromosome. *Am J Hum Genet* **14**, 135-148 (1962).
5. S. M. Gartler, A. D. Riggs, Mammalian X-chromosome inactivation. *Annu Rev Genet* **17**, 155-190 (1983).
6. J. C. Lucchesi, Gene dosage compensation and the evolution of sex chromosomes. *Science* **202**, 711-716 (1978).
7. J. C. Lucchesi, Dosage compensation in flies and worms: the ups and downs of X-chromosome regulation. *Current opinion in genetics & development* **8**, 179-184 (1998).
8. G. F. Kay *et al.*, Expression of Xist during mouse development suggests a role in the initiation of X chromosome inactivation. *Cell* **72**, 171-182 (1993).
9. N. Takagi, N. Wake, M. Sasaki, Cytologic evidence for preferential inactivation of the paternally derived X chromosome in XX mouse blastocysts. *Cytogenet Cell Genet* **20**, 240-248 (1978).
10. I. Okamoto, A. P. Otte, C. D. Allis, D. Reinberg, E. Heard, Epigenetic dynamics of imprinted X inactivation during early mouse development. *Science* **303**, 644-649 (2004).

11. E. G. Schulz, E. Heard, Role and control of X chromosome dosage in mammalian development. *Curr Opin Genet Dev* **23**, 109-115 (2013).

12. I. Okamoto, D. Arnaud, P. Le Baccon, A. P. Otte, Evidence for de novo imprinted X-chromosome inactivation independent of meiotic inactivation in mice. *Nature* **438**, 369 (2005).

13. P. Tam, S. X. Zhou, S.-S. Tan, X-chromosome activity of the mouse primordial germ cells revealed by the expression of an X-linked lacZ transgene. *Development* **120**, 2925-2932 (1994).

14. J. Chaumeil, P. Le Baccon, A. Wutz, E. Heard, A novel role for Xist RNA in the formation of a repressive nuclear compartment into which genes are recruited when silenced. *Genes Dev* **20**, 2223-2237 (2006).

15. E. G. Barr Ml Fau - Bertram, E. G. Bertram, A morphological distinction between neurones of the male and female, and the behaviour of the nucleolar satellite during accelerated nucleoprotein synthesis.

16. E. Splinter *et al.*, The inactive X chromosome adopts a unique three-dimensional conformation that is dependent on Xist RNA. *Genes Dev* **25**, 1371-1383 (2011).

17. X. Deng *et al.*, Bipartite structure of the inactive mouse X chromosome. *Genome biology* **16**, 152 (2015).

18. E. J. Richards, S. C. Elgin, Epigenetic codes for heterochromatin formation and silencing: rounding up the usual suspects.

19. D. Smeets *et al.*, Three-dimensional super-resolution microscopy of the inactive X chromosome territory reveals a collapse of its active nuclear compartment harboring distinct Xist RNA foci. *Epigenetics & chromatin* **7**, 8 (2014).

20. N. D. Belyaev, A. M. Keohane, B. M. Turner, Differential underacetylation of histones H2A, H3 and H4 on the inactive X chromosome in human female cells. *Human genetics* **97**, 573-578 (1996).

21. P. Jeppesen, B. M. Turner, The inactive X chromosome in female mammals is distinguished by a lack of histone H4 acetylation, a cytogenetic marker for gene expression. *Cell* **74**, 281-289 (1993).

22. B. A. Boggs *et al.*, Differentially methylated forms of histone H3 show unique association patterns with inactive human X chromosomes. *Nat Genet* **30**, 73-76 (2002).

23. E. Heard *et al.*, Methylation of histone H3 at Lys-9 is an early mark on the X chromosome during X inactivation. *Cell* **107**, 727-738 (2001).

24. J. E. Mermoud, B. Popova, A. H. Peters, T. Jenuwein, N. Brockdorff, Histone H3 lysine 9 methylation occurs rapidly at the onset of random X chromosome inactivation. *Current biology* **12**, 247-251 (2002).

25. S. Mili, J. A. Steitz, Evidence for reassociation of RNA-binding proteins after cell lysis: implications for the interpretation of immunoprecipitation analyses. *RNA* **10**, 1692-1694 (2004).

26. C. Costanzi, J. R. Pehrson, Histone macroH2A1 is concentrated in the inactive X chromosome of female mammals. *Nature* **393**, 599 (1998).

27. K. Plath *et al.*, Role of histone H3 lysine 27 methylation in X inactivation. *Science* **300**, 131-135 (2003).

28. J. Fang, T. Chen, B. Chadwick, E. Li, Y. Zhang, Ring1b-mediated H2A ubiquitination associates with inactive X chromosomes and is involved in initiation of X inactivation. *Journal of Biological Chemistry* **279**, 52812-52815 (2004).

29. A.-V. Gendrel *et al.*, Smchd1-dependent and-independent pathways determine developmental dynamics of CpG island methylation on the inactive X chromosome. *Dev Cell* **23**, 265-279 (2012).

30. L. F. Lock, N. Takagi, G. R. Martin, Methylation of the Hprt gene on the inactive X occurs after chromosome inactivation. *Cell* **48**, 39-46 (1987).

31. D. P. Norris, N. Brockdorff, S. Rastan, Methylation status of CpG-rich islands on active and inactive mouse X chromosomes. *Mammalian Genome* **1**, 78-83 (1991).

32. L. B. Russell, Mammalian X-chromosome action: inactivation limited in spread and in region of origin. *Science* **140**, 976-978 (1963).

33. E. Therman, G. E. Sarto, K. Patau, Center for Barr body condensation on the proximal part of the human Xq: a hypothesis. *Chromosoma* **44**, 361-366 (1974).

34. E. Heard, P. Clerc, P. Avner, X-chromosome inactivation in mammals. *Annu Rev Genet* **31**, 571-610 (1997).

35. G. Borsani, R. Tonlorenzi, Characterization of a murine gene expressed from the inactive X chromosome. *Nature* **351**, 325 (1991).

36. N. Brockdorff *et al.*, The product of the mouse Xist gene is a 15 kb inactive X-specific transcript containing no conserved ORF and located in the nucleus. *Cell* **71**, 515-526 (1992).

37. G. D. Penny, G. F. Kay, S. A. Sheardown, S. Rastan, N. Brockdorff, Requirement for Xist in X chromosome inactivation. *Nature* **379**, 131 (1996).

38. Y.-K. Hong, S. D. Ontiveros, C. Chen, W. M. Strauss, A new structure for the murine Xist gene and its relationship to chromosome choice/counting during X-chromosome inactivation. *Proceedings of the National Academy of Sciences* **96**, 6829-6834 (1999).

39. Y. Marahrens, J. Loring, R. Jaenisch, Role of the Xist gene in X chromosome choosing. *Cell* **92**, 657-664 (1998).

40. Y. Marahrens, B. Panning, J. Dausman, W. Strauss, R. Jaenisch, Xist-deficient mice are defective in dosage compensation but not spermatogenesis. *Genes Dev* **11**, 156-166 (1997).

41. A. Beletskii, Y.-K. Hong, J. Pehrson, M. Egholm, W. M. Strauss, PNA interference mapping demonstrates functional domains in the noncoding RNA Xist. *Proceedings of the National Academy of Sciences* **98**, 9215-9220 (2001).

42. A. D. Kelsey *et al.*, Impact of flanking chromosomal sequences on localization and silencing by the human non-coding RNA XIST. *Genome biology* **16**, 208 (2015).

43. A. Wutz, T. P. Rasmussen, R. Jaenisch, Chromosomal silencing and localization are mediated by different domains of Xist RNA. *Nat Genet* **30**, 167-174 (2002).

44. C. M. Clemson, J. A. McNeil, H. F. Willard, J. B. Lawrence, XIST RNA paints the inactive X chromosome at interphase: evidence for a novel RNA involved in nuclear/chromosome structure. *J Cell Biol* **132**, 259-275 (1996).

45. J. M. Engreitz *et al.*, The Xist lncRNA exploits three-dimensional genome architecture to spread across the X chromosome. *Science* **341**, 1237973 (2013).

46. A. Wutz, Gene silencing in X-chromosome inactivation: advances in understanding facultative heterochromatin formation. *Nat Rev Genet* **12**, 542-553 (2011).

47. J. Silva *et al.*, Establishment of histone h3 methylation on the inactive X chromosome requires transient recruitment of Eed-Enx1 polycomb group complexes. *Dev Cell* **4**, 481-495 (2003).

48. M. Almeida *et al.*, PCGF3/5-PRC1 initiates Polycomb recruitment in X chromosome inactivation. *Science* **356**, 1081-1084 (2017).

49. A. M. Keohane, J. S. Lavender, L. P. O'Neill, B. M. Turner, Histone acetylation and X inactivation. *Developmental genetics* **22**, 65-73 (1998).

50. A. M. Keohane, P. O'Neill L, N. D. Belyaev, J. S. Lavender, B. M. Turner, X-Inactivation and histone H4 acetylation in embryonic stem cells. *Developmental biology* **180**, 618-630 (1996).

51. M. J. Wakefield, A. M. Keohane, B. M. Turner, J. A. Graves, Histone underacetylation is an ancient component of mammalian X chromosome inactivation. *Proc Natl Acad Sci U S A* **94**, 9665-9668 (1997).

52. G. Csankovszki, A. Nagy, R. Jaenisch, Synergism of Xist RNA, DNA methylation, and histone hypoacetylation in maintaining X chromosome inactivation. *The Journal of cell biology* **153**, 773-784 (2001).

53. C. Naughton, D. Sproul, C. Hamilton, N. Gilbert, Analysis of active and inactive X chromosome architecture reveals the independent organization of 30 nm and large-scale chromatin structures. *Mol Cell* **40**, 397-409 (2010).

54. A. S. Belmont, F. Bignone, P. O. Ts'o, The relative intranuclear positions of Barr bodies in XXX non-transformed human fibroblasts. *Experimental cell research* **165**, 165-179 (1986).

55. C. L. Walker, C. B. Cargile, K. M. Floy, M. Delannoy, B. R. Migeon, The Barr body is a looped X chromosome formed by telomere association. *Proc Natl Acad Sci U S A* **88**, 6191-6195 (1991).

56. E. Heard, J. Dekker, Structural organization of the inactive X chromosome in the mouse. *Nature*, (2016).

57. N. Brockdorff, X-chromosome inactivation: closing in on proteins that bind Xist RNA. *Trends in genetics : TIG* **18**, 352-358 (2002).

58. O. Masui, E. Heard, RNA and protein actors in X-chromosome inactivation. *Cold Spring Harbor symposia on quantitative biology* **71**, 419-428 (2006).

59. N. Brockdorff, Noncoding RNA and Polycomb recruitment. *RNA* **19**, 429-442 (2013).

60. J. Zhao, B. K. Sun, J. A. Erwin, J. J. Song, J. T. Lee, Polycomb proteins targeted by a short repeat RNA to the mouse X chromosome. *Science* **322**, 750-756 (2008).

61. Y. Hasegawa, N. Brockdorff, S. Kawano, K. Tsutui, S. Nakagawa, The matrix protein hnRNP U is required for chromosomal localization of Xist RNA. *Dev Cell* **19**, 469-476 (2010).

62. Y. Jeon, J. T. Lee, YY1 Tethers Xist RNA to the Inactive X Nucleation Center. *Cell* **146**, 119-133 (2011).

63. C. J. Brown, S. E. Baldry, Evidence that heteronuclear proteins interact with XIST RNA in vitro. *Somatic cell and molecular genetics* **22**, 403-417 (1996).

64. M. E. Royce-Tolland *et al.*, The A-repeat links ASF/SF2-dependent Xist RNA processing with random choice during X inactivation. *Nat Struct Mol Biol* **17**, 948-954 (2010).

65. R. Agrelo *et al.*, SATB1 defines the developmental context for gene silencing by Xist in lymphoma and embryonic cells. *Dev Cell* **16**, 507-516 (2009).

66. S. Schoeftner *et al.*, Recruitment of PRC1 function at the initiation of X inactivation independent of PRC2 and silencing. *Embo J* **25**, 3110-3122 (2006).

67. S. Kalantry, T. Magnuson, The Polycomb group protein EED is dispensable for the initiation of random X-chromosome inactivation. *PLoS Genet* **2**, e66 (2006).

68. R. B. Darnell, HITS-CLIP: panoramic views of protein-RNA regulation in living cells. *Wiley interdisciplinary reviews. RNA* **1**, 266-286 (2010).

69. J. Zhao *et al.*, Genome-wide identification of polycomb-associated RNAs by RIP-seq. *Mol Cell* **40**, 939-953 (2010).

70. A. M. Khalil *et al.*, Many human large intergenic noncoding RNAs associate with chromatin-modifying complexes and affect gene expression. *Proc Natl Acad Sci U S A* **106**, 11667-11672 (2009).

71. K. Hartmuth *et al.*, Protein composition of human prespliceosomes isolated by a tobramycin affinity-selection method. *Proc Natl Acad Sci U S A* **99**, 16719-16724 (2002).

72. J. R. Hogg, K. Collins, RNA-based affinity purification reveals 7SK RNPs with distinct composition and regulation. *RNA* **13**, 868-880 (2007).

73. F. Butter, M. Scheibe, M. Morl, M. Mann, Unbiased RNA-protein interaction screen by quantitative proteomics. *Proc Natl Acad Sci U S A* **106**, 10626-10631 (2009).

74. M. Scheibe *et al.*, Quantitative interaction screen of telomeric repeat-containing RNA reveals novel TERRA regulators. *Genome Res* **23**, 2149-2157 (2013).

75. N. Lin *et al.*, An evolutionarily conserved long noncoding RNA TUNA controls pluripotency and neural lineage commitment. *Mol Cell* **53**, 1005-1019 (2014).

76. M. Huarte *et al.*, A large intergenic noncoding RNA induced by p53 mediates global gene repression in the p53 response. *Cell* **142**, 409-419 (2010).

77. C. A. McHugh, P. Russell, M. Guttman, Methods for comprehensive experimental identification of RNA-protein interactions. *Genome Biol* **15**, 203 (2014).

Chapter 2

RAP-MS IDENTIFIES DIRECT XIST-INTERACTING PROTEINS REQUIRED FOR XIST-MEDIATED SILENCING

The work was first published as:

McHugh, Colleen A., et al. "The Xist lncRNA directly interacts with SHARP to silence transcription through HDAC3." Nature 521.7551 (2015): 232.

2.1 RAP-MS: A Method to Identify the Proteins that Interact with lncRNAs *in vivo*

In order to characterize the mechanism of Xist-mediated silencing and define the functions of other lncRNAs, a new method that can identify the direct-interacting proteins of a lncRNA of interest is essential. To develop a method for identifying the proteins that directly interact with a specific lncRNA *in vivo*, we adapted our RNA Antisense Purification (RAP) method(*1*) to purify a lncRNA complex and identify the interacting proteins by quantitative mass spectrometry (RAP-MS) (**Methods, Figure 2.1a**). Specifically, we used UV crosslinking to create covalent linkages between directly interacting RNA and protein(*2*), then purified lncRNAs in denaturing conditions to disrupt non-covalent interactions(*3, 4*). This approach, utilized by methods such as CLIP(*4*), is known to separate interactions that occur in the cell from those that merely associate in solution(*3, 5*).

Adapting this UV-crosslinking and denaturing approach to enable purification of a specific lncRNA is challenging for several reasons: **(i)** In order to purify lncRNA complexes in denaturing conditions, we need an RNA capture method that can withstand harsh denaturing conditions. **(ii)** In order to detect the proteins associated with a given lncRNA, we need to achieve high purification yields of a lncRNA complex because, unlike nucleic acids, we cannot amplify proteins prior to detection. **(iii)** Because any individual RNA is likely to be present at a very low percentage of the total cellular RNA, we need to achieve high levels of enrichment to identify specific interacting proteins. **(iv)** Because the number of background proteins will be high, even after enrichment, we need accurate and sensitive methods for protein quantification to detect specific lncRNA interacting proteins.

The RAP-MS method addresses these challenges because **(i)** RAP uses long biotinylated antisense probes, which form very stable RNA-DNA hybrids, and therefore can be used to purify lncRNA complexes in denaturing conditions (4M urea and 500 mM lithium chloride at 67˚C). Also, **(ii)** we optimized the RAP method to achieve high yields of endogenous RNA complexes. In our original protocol(*1*), we achieved <2% yield of the endogenous RNA complex; by optimizing hybridization, washing, and elution conditions, we were able to reproducibly achieve ~70% yield **(Figure 2.1b)**. **(iii)** Using our optimized conditions, we increased the enrichment levels for the target lncRNA complex (~5,000-fold) relative to our already high levels of enrichment achieved previously (~100-fold)(*1*) **(Figure 2.1c)**. **(iv)** To achieve sensitive quantification and to distinguish between specific proteins and background proteins, we used Stable Isotope Labeling by Amino acids in Culture (SILAC)(*6*) to label

proteins (**Methods**), which enables quantitative comparisons of purified proteins by mass spectrometry(*7-9*).

We validated the RAP-MS approach by defining the proteins that interact with two well-characterized non-coding RNAs: (i) U1 small nuclear RNA, a core component of the spliceosome(*10*) and (ii) 18S ribosomal RNA, a component of the small ribosomal subunit(*11*). In the U1 purifications, we identified 9 enriched proteins, all of which are known to interact with the U1 snRNA. The list includes 7 of the 10 proteins that made up the core U1 snRNP complex (U1-A, U1-C, U1-70K, Sm-B, Sm-D2, Sm-D3, Sm-E)(*12*) as well as the Gemin5 processing factor involved in U1 snRNP biogenesis(*13*) **(Figure 2.1d)**. The ninth enriched protein, SF3a1, had not previously been identified as a U1-interacting protein but was recently shown to interact directly with the U1 snRNA *in vivo*(*14*). In the 18S purification, we identified 105 enriched proteins; 98 of these (93%) were previously characterized as ribosomal proteins, ribosomal processing and assembly factors, translational regulators, or other known ribosome interactors **(Figure 2.1d)**. In particular, we identified 21 of the 31 known small ribosomal subunit proteins. The U1 and 18S proteins that were missed appear to fall predominately into two categories: (i) proteins that make few direct contacts with the RNA, and (ii) small proteins that contain only a single possible peptide that could be detected by mass spectrometry.

Together, these results demonstrate that the RAP-MS method identifies the majority of known RNA interacting proteins, and that the proteins identified by RAP-MS are highly specific for the purified lncRNA complex.

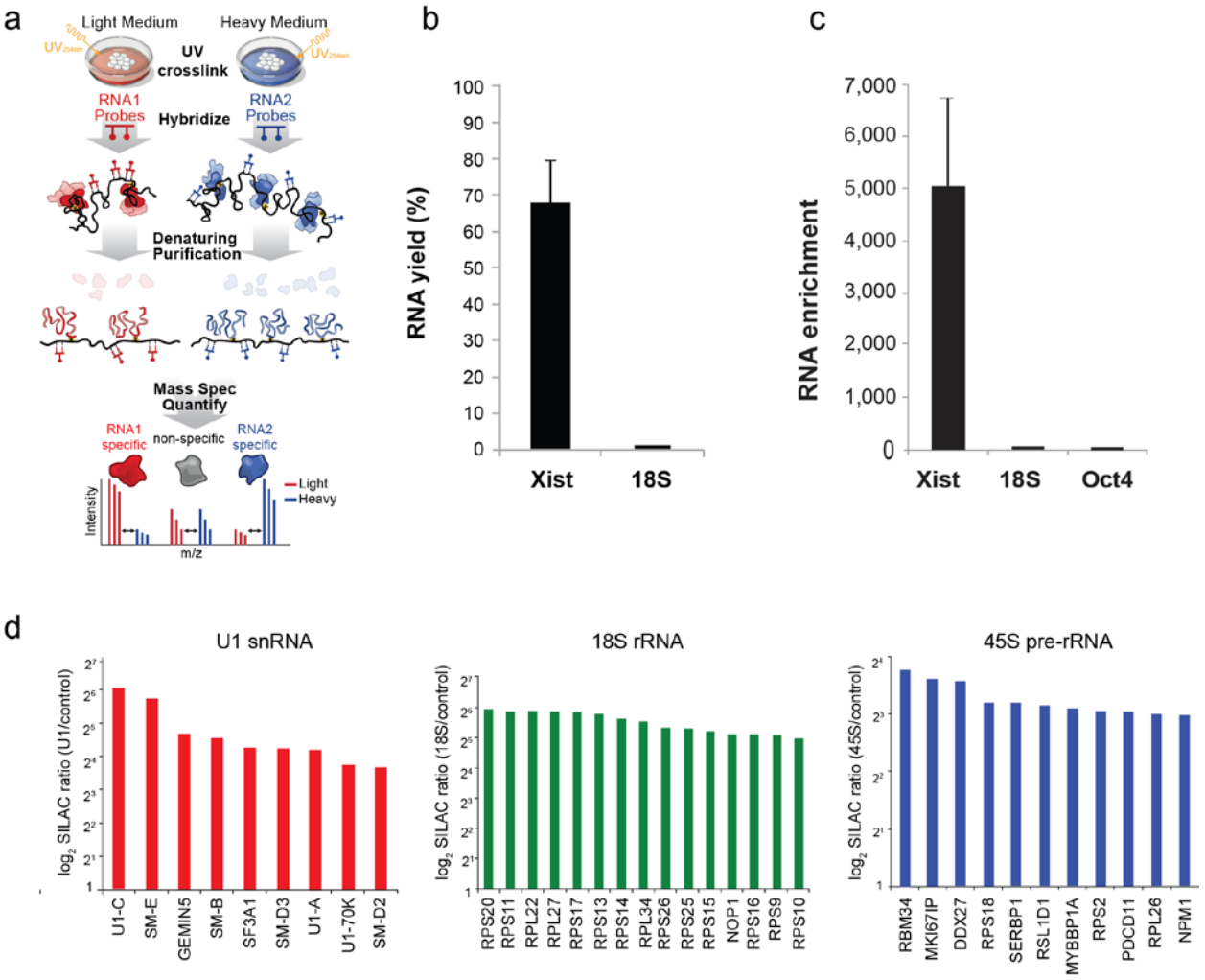

Figure 2.1. RAP-MS identifies proteins that are known to directly interact with specific ncRNAs. (a) A schematic overview of the RAP-MS method. (b) RT-qPCR measuring the percentage of the total cellular Xist or 18S recovered after RAP-MS of Xist. Values are computed as the amount of each RNA in the elution divided by the amount of RNA in the starting ("input") lysate material. Error bars represent the standard error of the mean from 5 biological replicates. (c) Enrichment of Xist after RAP-MS captures from pSM33 cells as measured by qPCR. Bars indicate RNA levels of Xist, 18S, and Oct4 after purification of Xist, normalized to RNA in input sample. Each bar represents the RNA

levels of Xist, 18S, and Oct4 after purification of Xist, normalized to RNA in input sample, from 3 biological replicates. (d) SILAC ratios of top proteins enriched in the RAP-MS U1 snRNA, 18S rRNA, and 45S pre-rRNA experiments.

2.2 RAP-MS Identifies Direct Xist-interacting Proteins

To define the proteins that interact with the Xist lncRNA during the initiation of XCI, we performed RAP-MS in a male mouse embryonic stem (ES) cell line containing a doxycycline-inducible Xist after six hours of induction(*1*) (**Methods**). This system is known to represent a well-synchronized and accurate model for the initiation of XCI(*1, 15-17*). We performed RAP in nuclear extracts from UV-crosslinked SILAC-labeled cells using probes that are antisense to the Xist RNA and achieved a ~5,000-fold enrichment of the Xist RNA relative to its level in total nuclear RNA (**Methods, Figure 2.1b**). To control for background proteins or non-specific proteins that might interact with any nuclear RNA, we separately purified the abundant U1 snRNA, which is not expected to interact with the same proteins as Xist. We identified the proteins in each sample using liquid chromatography-mass spectrometry and calculated a SILAC ratio for each protein based on the intensity of all heavy or light peptides originating from the Xist or U1 purification.

We identified 10 proteins that were specifically enriched in the Xist purification compared to the U1 purification (SILAC ratio >8-fold, **Methods, Figure 2.2a**). Nearly all of these proteins were reproducibly enriched in multiple Xist purifications from independent biological samples; the sole exception was missed only because its enrichment level (4-fold) fell below our stringent enrichment cutoff (8-fold) in some replicate samples (**Methods**). Consistent with the notion that these proteins are direct Xist RNA-interacting proteins, 9 of these proteins contain a well-characterized RNA binding domain (**Figure 2.2b**).

The identified Xist-interacting proteins are SHARP, RBM15, MYEF2, CELF1, hnRNPC, LBR, SAF-A, RALY, hnRNPM, and PTBP1. Interestingly, SAF-A (also called hnRNPU) was previously shown to interact directly with Xist by UV-based CLIP experiments(*18*) and is required for tethering Xist to the inactive X-chromosome in differentiated cells(*18-20*). In addition, 5 of these proteins have been previously implicated in transcriptional repression, chromatin regulation, and nuclear organization (**Figure 2.2b**). These include the SMRT and HDAC Associated Repressor Protein (SHARP, also called SPEN), a member of the SPEN family of transcriptional repressors that directly interacts with the SMRT (also called NCoR-2) component of the nuclear co-repressor complex(*21, 22*) to activate HDAC3 deacetylation activity on chromatin(*23*). Interestingly, we also identified RBM15, another member of the SPEN family of transcriptional repressors, which shares the same domain structure as SHARP(*24*), but appears to have a distinct functional role during development(*25, 26*). Myef2 has been shown to function as a negative regulator of transcription in multiple cell types, although its mechanism of regulation is still unknown(*27, 28*). hnRNP-M is a paralog of Myef2. Finally, we identified the Lamin B receptor (LBR), a protein that contains 8 transmembrane domains that are anchored in the inner nuclear membrane(*29-31*), a domain that interacts with repressive chromatin regulatory proteins(*29, 32, 33*), and an independent domain that interacts with Lamin B(*29, 30*).

To confirm that these proteins reflect specific interactions with Xist, and are not due to non-specific purification of other RNAs, we performed RAP using the same Xist probes in uninduced cells in which Xist is not expressed. Furthermore, to confirm that the identified interactions represent proteins that are crosslinked to Xist *in vivo* rather than interactions that

form in solution, we purified Xist from cells that were not crosslinked (no UV light). In both cases, we identified none of these 10 Xist-interacting proteins, nor any other specifically enriched proteins in either of these control samples (**Methods**), demonstrating that the identified proteins are covalently crosslinked to Xist in cells.

Finally, we tested whether we could enrich Xist RNA using UV-crosslinking and immunoprecipitation of the identified proteins (**Methods**). We tested 5 of the 10 identified proteins, for which we could obtain high-quality IP-grade antibodies (Ptbp1, hnRNPC, RBM15, LBR, and RALY), along with several negative controls (IGG and Pum1). In all 5 cases, we observed a strong enrichment for Xist RNA relative to total input RNA levels (**Figure 2.2c, d**). Notably, we did not observe any enrichment for Xist in the negative controls, nor did we observe enrichment of mRNAs (i.e. Oct4 or Nanog) or other nuclear lncRNAs (i.e. Neat1) (**Figure 2.2e, f**).

Together, these results identify a set of highly specific and reproducible proteins that directly interact with Xist during the initiation of XCI. Given the generality of the RAP-MS approach, we expect that it will be broadly applicable for defining the protein complexes that interact with other lncRNAs.

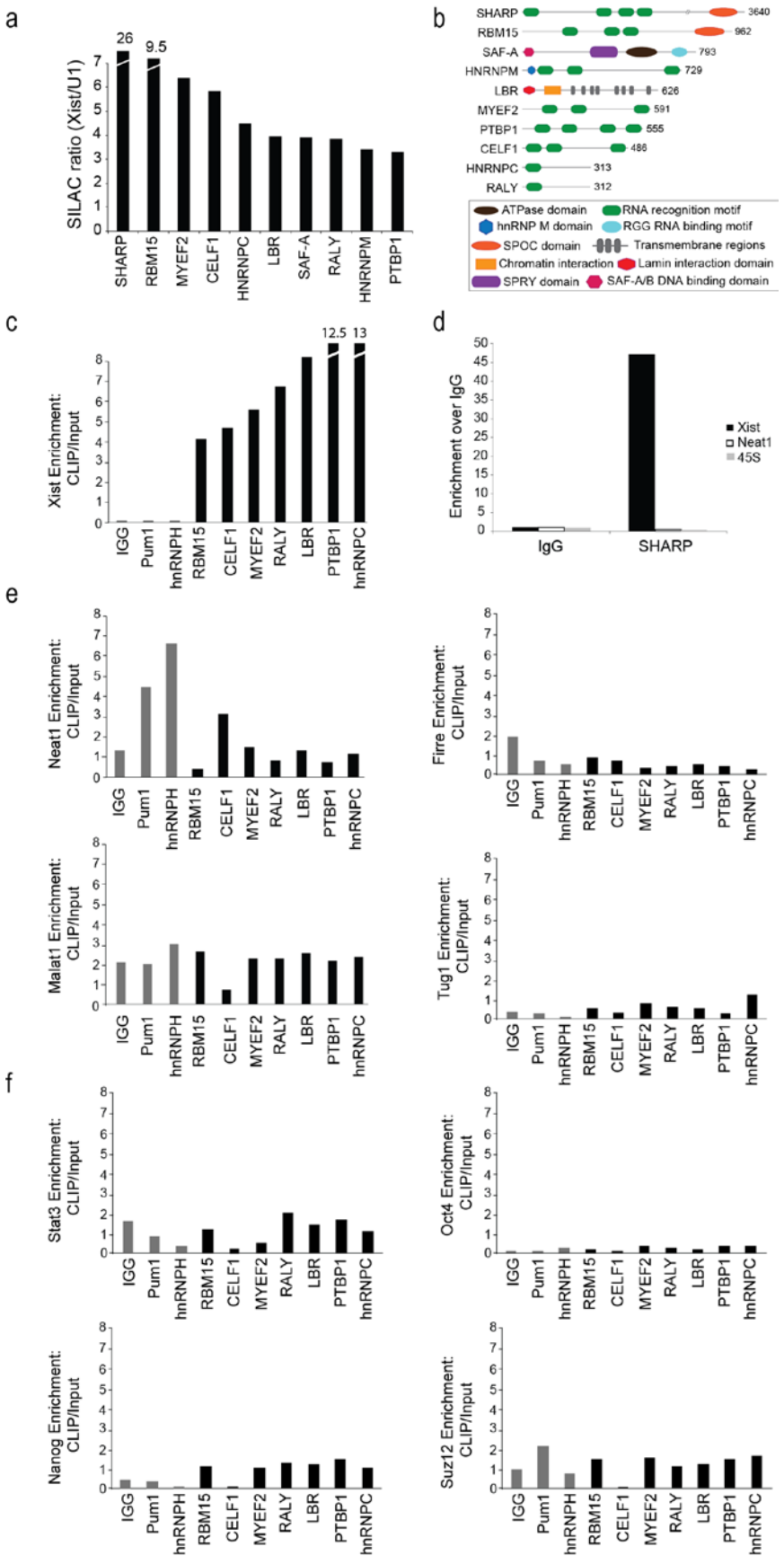
a
SILAC ratio (Xist/U1)
26
9.5
SHARP
RBM15
MYEF2
CELF1
HNRNPC
LBR
SAF-A
RALY
HNRNPM
PTBP1
b
SHARP 3640
RBM15 962
SAF-A 793
HNRNPM 729
LBR 626
MYEF2 591
PTBP1 555
CELF1 486
HNRNPC 313
RALY 312
ATPase domain
RNA recognition motif
hnRNP M domain
RGG RNA binding motif
SPOC domain
Transmembrane regions
Chromatin interaction
Lamin interaction domain
SPRY domain
SAF-A/B DNA binding domain
c
Xist Enrichment: CLIP/Input
12.5
13
IGG
Pum1
hnRNPH
RBM15
CELF1
MYEF2
RALY
LBR
PTBP1
hnRNPC
d
Enrichment over IgG
Xist
Neat1
45S
IgG
SHARP
e
Neat1 Enrichment: CLIP/Input
Firre Enrichment: CLIP/Input
Malat1 Enrichment: CLIP/Input
Tug1 Enrichment: CLIP/Input
f
Stat3 Enrichment: CLIP/Input
Oct4 Enrichment: CLIP/Input
Nanog Enrichment: CLIP/Input
Suz12 Enrichment: CLIP/Input

Figure 2.2. RAP-MS identifies direct Xist-interacting proteins. (a) The SILAC ratio (Xist/U1) for each Xist-enriched protein identified by RAP-MS for one representative sample of four biological replicates. For SHARP and RBM15, the enrichment values are indicated above their bars. (b) Each Xist-interacting protein is shown (scaled to protein length). The locations of functional domains are shown. (c) Enrichment of the Xist lncRNA after immunoprecipitation from a sample of pSM33 male cells. (d) Immunoprecipitation of SHARP was performed from a sample of UV-crosslinked females ES cells that were treated with retinoic acid for 24 hours. The levels of recovered Xist lncRNA (black bars), Neat1 lncRNA (white bars), and 45S pre-ribosomal RNA (gray bars) were measured by RT-qPCR. Enrichment of each RNA after capture with anti-SHARP antibody was calculated relative to the level of RNA captured with IgG control antibody. (e) The enrichment of various lncRNAs after immunoprecipitation in pSM33 male cells – including Neat1, Malat1, Firre, and Tug1 – are shown. (f) The enrichment of various mRNA controls after immunoprecipitation in pSM33 male cells – including Oct4, Nanog, Stat3, and Suz12 – are shown.

2.3 SHARP, LBR, and SAF-A are required for Xist-mediated silencing

Having defined the direct Xist-interacting proteins, we sought to determine which Xist-interacting proteins are required for Xist-mediated transcriptional silencing on the inactive X-chromosome. To do this, we knocked down each of the identified proteins prior to Xist induction and assayed for the failure to silence gene expression on the X-chromosome upon induction of Xist expression (**Figure 2.3a**).

Specifically, we selected two X-linked genes, Gpc4 and Atrx, that are well expressed in the absence of Xist expression, but are normally silenced by 16 hours of Xist induction in our doxycycline-inducible system in male cells (**Figure 2.3b, c**). This male-inducible system is more sensitive for identifying proteins that affect silencing compared to a female system because Xist-mediated silencing in males will lead to loss of 100% of X-chromosome transcripts rather than only 50% in a female system, which still retains one active X. We used siRNAs to knock down the mRNA levels of each of the proteins identified by RAP-MS along with several negative controls (**Methods**). We ensured that each cell examined showed both successful depletion of the siRNA-targeted mRNA as well as induction of Xist expression using single molecule RNA FISH(*34*) (**Methods**). We observed no difference in the percentage of cells that induce Xist expression in any of the siRNA conditions relative to untreated cells. Within each of these cells, we quantified the mRNA level of each of the two X-linked genes prior to Xist induction (-dox) and after Xist induction (+dox).

As a control, we transfected several non-targeting siRNAs (**Methods**). In these negative controls, we observed the expected silencing of the X-linked genes studied (Gpc4 expression

decreased from an average of 20 copies (-dox) to 2 copies (+dox) per cell and Atrx expression decreased from 22 to 3 copies per cell; **Figure 2.3d**). Similarly, knock down of Rbm15, Myef2, Ptbp1, Celf1, hnRNPC, Raly, or hnRNPM did not alter gene silencing on the X-chromosome (**Figure 2.3b, e**). Consistent with previous observations, we also observed no effect on X-chromosome gene silencing upon knock down of EED[42,43], a component of PRC2 that is required for its localization and activity on chromatin(*35, 36*) (**Figure 2.3b, e**).

In contrast, knock down of SHARP, Lamin B Receptor (LBR), or SAF-A largely abolished the silencing of X-chromosome genes following Xist induction (**Figure 2.3b, e**). Indeed, the expression levels of the X-chromosome genes studied did not significantly change following Xist expression (**Figure 2.3b, c**), indicating that SHARP, LBR, and SAF-A are required for Xist to initiate transcriptional silencing on the X-chromosome.

Since previous studies have shown that Xist can no longer initiate transcriptional silencing after a certain critical window during differentiation(*16*), we wanted to ensure that the loss of Xist silencing upon knock down of SHARP, LBR, and SAF-A was not merely due to cellular differentiation. To address this, we performed single molecule FISH for Gpc4 mRNA along with immunofluorescence for Nanog, a marker of the pluripotent state that is rapidly lost upon differentiation(*37*). We confirmed that knockdown of SHARP, LBR, or SAF-A also abolished gene silencing on the X-chromosome in Nanog-positive cells (**Figure 2.3f**).

To ensure that these silencing defects were not specific to our inducible male ES cell system, we knocked down SHARP, LBR, or SAF-A along with several controls in wild-type female ES cells and induced Xist expression through retinoic acid-mediated differentiation (**Methods**). In the negative controls, we observed silencing of X-chromosome genes upon induction of Xist (30 to 14 copies of Gpc4). Consistent with our previous observations, knock down of SHARP, LBR, or SAF-A led to a loss of gene silencing on the X-chromosome, whereas knock down of EED had no effect on gene silencing within cells that induce Xist expression (**Figure 2.3g**).

Since SAF-A was previously shown to be required for Xist localization to chromatin in differentiated cells(*18*), we hypothesized that the observed SAF-A silencing defect might be because it is required for Xist localization to the X-chromosome during initiation of XCI. To test this, we looked at the Xist distribution in the nucleus in all of these perturbations. Consistent with previous observation, we observed a diffuse Xist localization pattern in the nucleus upon knock down of SAF-A, but not in any other control or protein knock down (**Figure 2.3f**). This suggests that SAF-A is required for localizing Xist and its silencing proteins to the X-chromosome.

Together, these results demonstrate that SHARP, LBR, and SAF-A are required for the initiation of Xist-mediated transcriptional silencing of the X-chromosome. Although the remaining seven Xist-interacting proteins showed no effect on X-chromosome gene silencing, they may still be important for Xist function: (i) some may have redundant functions (e.g. Myef2 and hnRNP-M, which are known paralogs), (ii) in some of these cases, the small amount of protein remaining after knock down may still be sufficient for Xist

function, or (iii) some of these proteins may be important for alternative Xist-mediated roles, such as the maintenance of XCI, which would not be captured by this silencing assay.

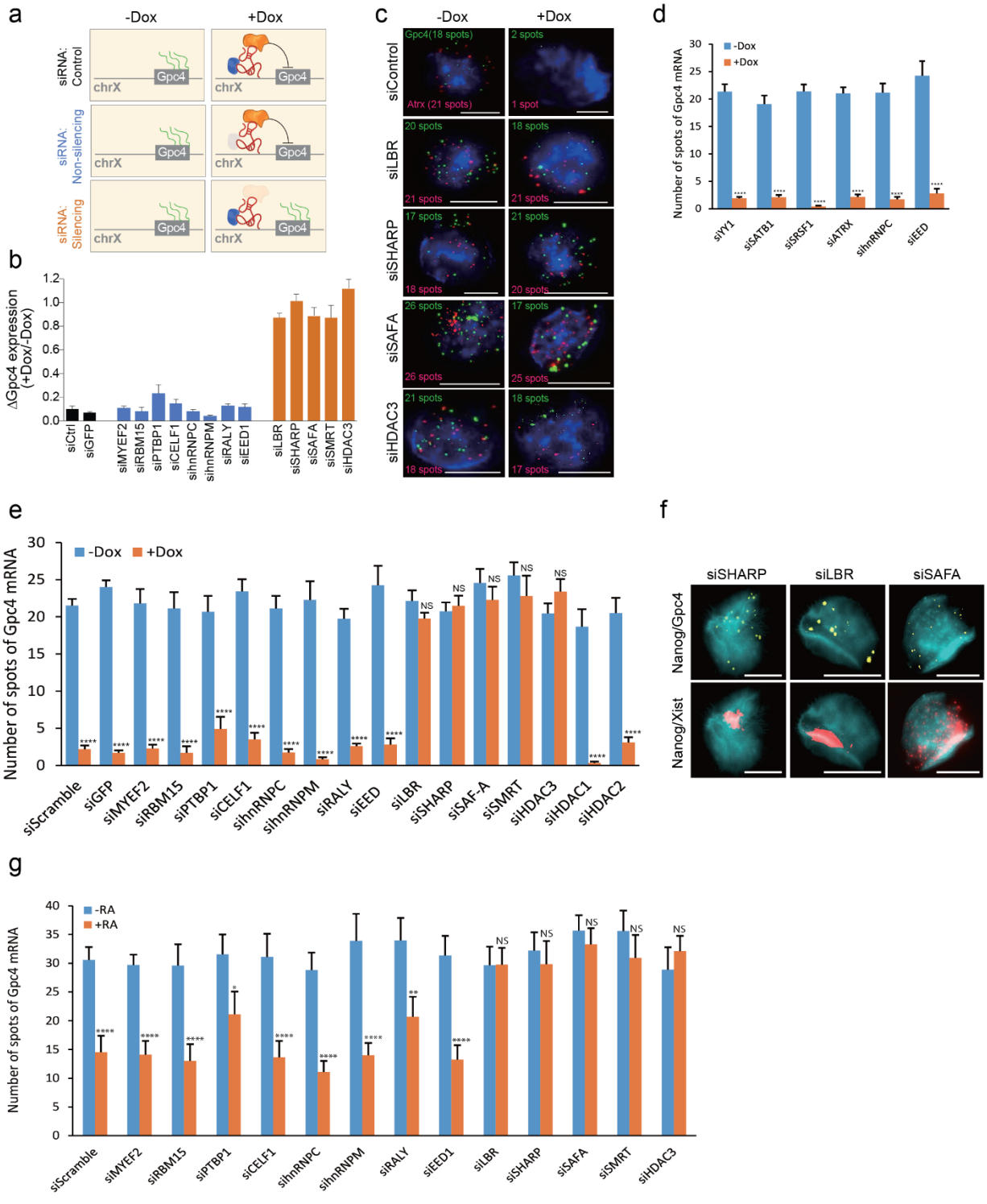
a
-Dox
+Dox
siRNA: Control
siRNA: Non-silencing
siRNA: Silencing
chrX
Gpc4
b
ΔGpc4 expression (+Dox/-Dox)
siCtrl
siGFP
siMYEF2
siRBM15
siPTBP1
siCELF1
sihnRNPC
sihnRNPM
siRALY
siEED1
siLBR
siSHARP
siSAFA
siSMRT
siHDAC3
c
-Dox
+Dox
siControl
siLBR
siSHARP
siSAFA
siHDAC3
Gpc4(18 spots)
Atrx (21 spots)
2 spots
1 spot
20 spots
21 spots
18 spots
17 spots
26 spots
25 spots
d
Number of spots of Gpc4 mRNA
-Dox
+Dox
siYY1
siSATB1
siSRSF1
siATRX
sihnRNPC
siEED
e
Number of spots of Gpc4 mRNA
-Dox
+Dox
siScramble
siGFP
siMYEF2
siRBM15
siPTBP1
siCELF1
sihnRNPC
sihnRNPM
siRALY
siEED
siLBR
siSHARP
siSAF-A
siSMRT
siHDAC3
siHDAC1
siHDAC2
NS
f
siSHARP
siLBR
siSAFA
Nanog/Gpc4
Nanog/Xist
g
Number of spots of Gpc4 mRNA
-RA
+RA
siScramble
siMYEF2
siRBM15
siPTBP1
siCELF1
sihnRNPC
sihnRNPM
siRALY
siEED1
siLBR
siSHARP
siSAFA
siSMRT
siHDAC3
NS

Figure 2.3. SHARP, LBR, and SAF-A are required for Xist-mediated gene silencing.

(a) Screen for Xist-mediated gene silencing for knockdown of control (top), non-silencing proteins (middle), or silencing proteins (bottom). (b) Gpc4 mRNA levels after induction of Xist (+dox) normalized to Gpc4 levels before Xist induction (-dox). Error bars: standard error of the mean across 50 cells from one experiment. siCtrl: scrambled siRNA control. (c) Images of individual cells for two X-linked mRNAs, Gpc4 (green) and Atrx (red), and DAPI (blue) after treatment with different siRNAs (rows). The number of identified mRNAs is shown. (d) Quantification of the copy number of Gpc4 before and after Xist induction upon treatment with different siRNAs. (e) Quantification of the copy number of Gpc4 in –Dox and +Dox cells after knockdown with siRNAs targeting different mRNAs. (f) Knockdown of SHARP, LBR, or SAF-A abrogates Xist-mediated gene silencing without causing pluripotency defects. Representative images showing staining of Nanog (cyan), Xist (red), and Gpc4 (green) upon knockdown of SHARP, LBR or SAF-A after 16 hours of Xist induction with doxycycline. (g) Quantification of the copy number of Gpc4 for –RA and +RA cells upon transfection with different siRNAs. Error bars represent the standard error across 50 individual cells from one experiment. Error bars represent the standard error of the mean across 50 individual cells from one experiment. NS: not significantly different, **** represents values with a p-value<0.001, ** represents values with a p-value<0.01, and * represents values with a p-value<0.05 between +Dox and –Dox, or +RA and –RA, cells based on an unpaired two-sample t-test. Scale bars on the images represent 5 μm.

Methods

Mouse ES cell culture

All mouse ES cell lines were cultured in serum-free 2i/LIF medium as previously described(*1*). We used the following cell lines: (i) Wild-type male ES cells (*V6.5 line*); (ii) Male ES cells expressing Xist from the endogenous locus under control of a tet-inducible promoter (*pSM33* ES cell line) as previously described(*1*). (iii) Male ES cells carrying a cDNA Xist transgene without the A-repeat integrated into the Hprt locus under control of the tet-inducible promoter (*A-repeat deletion*: kindly provided by A. Wutz)(*15*). (iv) Female ES cells (*F1 2-1 line*). This wild-type female mouse ES cell line is derived from a 129 x castaneous F1 mouse cross as previously described(*1*).

Xist induction

For Dox inducible cells (pSM33 and A-repeat deletion), we induced Xist expression by treating cells with 2 μg/ml doxycycline (Sigma) for 6 hours, 16 hours, or 24 hours based on the application. For female ES cells (F1 2-1 line), we induced Xist expression by inducing differentiation; 2i was replaced with MEF media (DMEM, 10% Gemini Benchmark FBS, 1× L-glutamine, 1× NEAA, 1× Pen/Strep; Life Technologies unless otherwise indicated) for 24 hours followed by treatment with 1 μM retinoic acid (RA) (Sigma) for an additional 24 hours.

We measured the amount of Xist RNA in both the doxycycline-inducible cells (6 hours induction) and differentiating female ES cells (24 hour induction) by qRT-PCR. We normalized this level to various RNA housekeeping controls, 18S, 28S, and U6, in both cell

populations and calculated the fold expression difference between male and female cells using the comparative Ct method. We observed a range of expression, with the male inducible system expressing from 5-20 fold (12-fold average) more Xist than the female cells**.** We note that this estimate likely represents an upper limit of the actual differences because the female ES cell system is known to be heterogeneous in Xist-induction, such that not every cell will induce Xist to the same level after 24 hours of retinoic acid treatment. Accordingly, we expect that the actual differences between the male inducible system and differentiating female ES cells are actually significantly lower. While the precise levels are hard to compare by single molecule FISH, the size and intensity of each Xist RNA cloud is similar in both systems at the time points used.

The male-inducible system is more sensitive for identifying proteins that affect silencing compared to a female system because Xist-mediated silencing in males will lead to loss of 100% of X-chromosome transcripts rather than only 50% in a female system, which still retains one active X.

UV crosslinking

Cells were washed once with PBS and then crosslinked on ice using 0.8 Joules/cm^2 (UV8k) 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 1500 × g for 4 minutes, and flash frozen in liquid nitrogen for storage at – 80 ˚C.

SILAC ES cell culture

For SILAC experiments, we adapted our ES cell culture procedures to incorporate either light or heavy lysine and arginine amino acids. The 2i/LIF SILAC medium was composed as follows: custom DMEM/F-12 without lysine or arginine (Dundee Cell Products) was supplemented with either 0.398 mM heavy Arg10 (Sigma) or unlabeled arginine (Sigma) and either 0.798 mM heavy Lys8 (Cambridge Isotope Labs) or unlabeled lysine (Sigma), 0.5× B-27 (Gibco), 2 mg/mL bovine insulin (Sigma), 1.37 μg/mL progesterone (Sigma), 5 mg/mL BSA Fraction V (Gibco), 0.1 mM 2-mercaptoethanol (Sigma), 5 ng/mL murine LIF (GlobalStem), 0.1 μM PD0325901 (SelleckChem) and 0.3 μM CHIR99021 (SelleckChem). Cells in both heavy and light 2i/LIF SILAC medium were also supplemented with 0.2 mg/mL of unlabeled proline (Sigma) to prevent conversion of labeled arginine to proline. 2i inhibitors were added fresh with each medium change.

Adapting cells to SILAC conditions

Prior to mass spectrometry, ES cells were adapted to SILAC conditions over three passages. The heavy or light culture medium was replaced every 24-48 hours depending on cell density, and cells were passaged every 72 hours using 0.025% trypsin (Gibco), rinsing dissociated cells from the plates with DMEM/F12 containing 0.038% BSA Fraction V (Gibco). Cells were grown in two different types of medium: (i) 2i/LIF SILAC medium with light (unlabeled) lysine and arginine, or (ii) 2i/LIF SILAC medium with heavy isotope labeled lysine and arginine.

Measuring SILAC incorporation

To examine the efficiency of SILAC labeling in pSM33 cells, we tested for the incorporation of labeled amino acids after 10 days of growth (3 cell passages) in heavy 2i/LIF SILAC medium. Pellets of 2 million cells were boiled for 10 minutes in LDS Sample Loading Buffer (Invitrogen) and then proteins were separated by SDS-PAGE on a 4-12% Tris-Glycine polyacrylamide gel (Invitrogen). Total protein was stained with Colloidal Coomassie (Invitrogen) and gel slices were excised with a clean scalpel and transferred to microcentrifuge tubes for in-gel tryptic digest. Protein disulfide bonds were reduced with DTT and then alkylated with iodoacetamide. Proteins were digested with trypsin overnight and then extracted using successive washes with 1% formic acid/2% acetonitrile, 1:1 acetonitrile/water, and 1% formic acid in acetonitrile. Peptides were collected, lyophilized, and then resuspended in 1% formic acid for mass spectrometry analysis (described below in ***Mass Spectrum Measurements)***. Peptides were identified from mass spectra using MaxQuant (described below in ***MS data analysis***). The incorporation rate of labeled amino acids was calculated based on the ratio of the intensity of heavy and light versions of each peptide identified. In cells used for subsequent assays, we confirmed that over 95% of peptides from cellular proteins showed >95% incorporation of labeled amino acids (**Extended Data 1b**).

RNA Affinity Purification-Mass Spectrometry (RAP-MS)

Probe design and generation

To create the probes used to capture target RNAs, we designed and synthesized 90-mer DNA oligonucleotides (Eurofins Operon) that spanned the entire length of the target RNA. The

sequence of each DNA oligonucleotide probes was antisense to the complementary target RNA sequence. Each DNA oligonucleotide probe was also modified with a 5' biotin in order to enable capture of DNA-RNA hybrids on streptavidin coated magnetic beads (described below). While we had previously used 120-mer probes, we found that 90-mer probes provided comparable stringency and yield in the conditions used. For Xist, we used 142 probes that covered the entire mature RNA sequence, with the exception of regions that match to other transcripts or genomic regions as previously described(*1*, *38*).

Total cell lysate preparation

For the 18S and U1 experiments we used total cellular lysates prepared in the following manner. We lysed batches of 20 million cells by completely resuspending frozen cell pellets in ice cold detergent-based Cell Lysis Buffer (10 mM Tris pH 7.5, 500 mM LiCl, 0.5% dodecyl maltoside (DDM, Sigma), 0.2% sodium dodecyl sulfate (SDS, Ambion), 0.1% sodium deoxycholate (Sigma)). Next, 1× Protease Inhibitor Cocktail (Set III, EDTA-free, Calbiochem) and 920 U of Murine RNase Inhibitor (New England Biolabs) were added and the sample was incubated for 10 minutes on ice to allow lysis to proceed. During this incubation period, the cell sample was passed 3-5× through a 26-gauge needle attached to a 1 mL syringe in order to disrupt the pellet and shear genomic DNA. Each sample was then sonicated using a Branson Digital Sonifier with a microtip set at 5 watts power for a total of 30 seconds in intermittent pulses (0.7 seconds on, 1.3 seconds off). During sonication the samples were chilled to prevent overheating of the lysate. The samples were then treated for 10 minutes at 37 ˚C with 2.5 mM $MgCl_2$, 0.5 mM $CaCl_2$, and 20 U of TURBO DNase (Ambion) to digest DNA. Samples were returned to ice and the reaction was immediately

terminated by the addition of 10 mM EDTA and 5 mM EGTA. Disulfide bonds were reduced by addition of 2.5 mM Tris-(2-carboxyethyl) phosphine (TCEP) and samples were then mixed with twice the lysate volume of 1.5× LiCl/Urea Buffer (the final1× Buffer contains 10 mM Tris pH 7.5, 5 mM EDTA, 500 mM LiCl, 0.5% DDM, 0.2% SDS, 0.1% deoxycholate, 4M urea, 2.5 mM TCEP). Lysates were incubated on ice for 10 minutes and then cleared by centrifugation in an Eppendorf 5424R centrifuge for 10 minutes at 16,000 × *g*. Supernatants were pooled and flash frozen in liquid nitrogen for storage at -80 ˚C.

Nuclear lysate preparation

For the Xist versus U1 and 45S versus U1 comparisons, we used nuclear lysates prepared in the following manner. We lysed batches of 50 million cells by resuspending frozen pellets in 1 mL Lysis Buffer 1 (10 mM HEPES pH7.2, 20 mM KCl, 1.5 mM $MgCl_2$, 0.5 mM EDTA, 1 mM Tris(2-carboxyethyl)phosphine (TCEP), 0.5 mM PMSF). Then the samples were centrifuged at 3,300 × *g* for 10 minutes to pellet cells. The cell pellets were resuspended in 1 mL Lysis Buffer 1 with 0.1% dodecyl maltoside (DDM) and dounced 20 times using a glass dounce homogenizer with the small clearance pestle (Kontes). Nuclei released from the cells after douncing were pelleted by centrifugation at 3,300 × *g* then resuspended in 550 µl Lysis Buffer 2 (20 mM Tris pH 7.5, 50 mM KCl, 1.5 mM $MgCl_2$, 2 mM TCEP, 0.5 mM PMSF, 0.4% sodium deoxycholate, 1% DDM, and 0.1% N-lauroylsarcosine (NLS)). Samples were incubated on ice for 10 minutes, and then each sample was sonicated using a Branson Sonifier at 5 watts power for a total of 1 minute in intermittent pulses (0.7 seconds on, 3.3 seconds off) to lyse nuclei and solubilize chromatin. During sonication the samples were chilled to prevent overheating of the nuclear lysate. Samples were then treated with 2.5 mM

$MgCl_2$, 0.5 mM $CaCl_2$, and 330 U TURBO DNase (Ambion) for 12 minutes at 37 ˚C to further solubilize chromatin. After DNase treatment, lysates were mixed with equal volume of 2× Hybridization Buffer (the final 1× Buffer contains 10 mM Tris pH 7.5, 5 mM EDTA, 500 mM LiCl, 0.5% DDM, 0.2% SDS, 0.1% deoxycholate, 4M urea, 2.5 mM TCEP). Finally, lysates were cleared by centrifugation for 10 minutes at 16,000 × *g* in an Eppendorf 5424R centrifuge and the resulting supernatants were pooled and flash frozen in liquid nitrogen for storage at -80 ˚C.

RNA affinity purification of crosslinked complexes

Lysates from 200 million or 800 million cells were used for each capture. For 200 million cells the following protocol was used, and scaled appropriately for larger cell numbers. For each capture, a sample of heavy or light SILAC labeled frozen lysate was warmed to 37 ˚C. For each sample, 1.2 mL of Streptavidin Dynabeads MyOne C1 magnetic beads (Invitrogen) were washed 6 times with equal volume of hybridization buffer (10 mM Tris pH 7.5, 5 mM EDTA, 500 mM LiCl, 0.5% DDM, 0.2% SDS, 0.1% deoxycholate, 4M urea, 2.5 mM TCEP). Lysate samples were pre-cleared by incubation with the washed Streptavidin C1 magnetic beads at 37 ˚C for 30 minutes with intermittent shaking at 1100 rpm on a Eppendorf Thermomixer C (30 seconds mixing, 30 seconds off). Streptavidin beads were then magnetically separated from lysate samples using a Dynamag magnet (Life Technologies). The beads used for pre-clearing lysate were discarded and the lysate sample was transferred to fresh tubes twice to remove all traces of magnetic beads. Biotinylated 90-mer DNA oligonucleotide probes specific for the RNA target of interest (20 μg per sample, in water) were heat-denatured at 85 ˚C for 3 minutes and then snap-cooled on ice. Probes and pre-

cleared lysate were mixed and incubated at 67 ˚C using an Eppendorf thermomixer with intermittent shaking (30 seconds shaking, 30 seconds off) for 2 hours to hybridize probes to the capture target RNA. Hybrids of biotinylated DNA probes and target RNA were then bound to streptavidin beads by incubating each sample with 1.2 mL of washed Streptavidin coated magnetic beads at 67 ˚C for 30 minutes on an Eppendorf Thermomixer C with intermittent shaking as above. Beads with captured hybrids were washed 6 times with LiCl/Urea Hybridization Buffer at 67 ˚C for 5 minutes to remove non-specifically associated proteins. Between 0.5 – 1% of the total beads were removed and transferred to a fresh tube after the final wash to examine RNA captures by qPCR (see "Elution and analysis of RNA samples"). The remaining beads were resuspended in Benzonase Elution Buffer (20 mM Tris pH 8.0, 2 mM $MgCl_2$, 0.05% NLS, 0.5 mM TCEP) for subsequent processing of the protein samples.

Elution of protein samples

Elution of captured proteins from streptavidin beads was achieved by digesting all nucleic acids (both RNA and DNA, double-stranded and single-stranded) using 125 U of Benzonase nonspecific RNA/DNA nuclease for 2 hours at 37 ˚C (Millipore, #71206-3). Beads were then magnetically separated from the sample using a DynaMag magnet (Life Technologies) and the supernatant containing eluted Xist-specific proteins was precipitated overnight at 4 ˚C with 10% trichloroacetic acid (TCA). TCA treated protein elution samples were pelleted by centrifugation for 30 minutes at >20,000 × *g*, and then washed with 1 mL cold acetone and recentrifuged. Final protein elution pellets were air dried to remove acetone and stored at -20 ˚C until processing for mass spectrometry.

Elution and analysis of RNA samples

Beads with hybrids were magnetically separated using a 96-well DynaMag (Life Technologies) and the supernatant was discarded. Beads were then resuspended by pipetting in 20 μL NLS RNA Elution Buffer (20 mM Tris pH 8.0, 10 mM EDTA, 2% NLS, 2.5 mM TCEP). To release the target RNA, beads were heated for 2 minutes at 95 ˚C in an Eppendorf Thermomixer C. Beads were then magnetically separated using a 96-well DynaMag (Life Technologies) and the supernatants containing eluted target RNA were digested by the addition of 1 mg/mL Proteinase K for 1 hour at 55 ˚C to remove all proteins. The remaining nucleic acids were then purified by ethanol precipitation onto SILANE beads (Invitrogen) as previously described(*1, 38*). DNA probes were removed by digestion with TURBO DNase (Ambion). To quantify RNA yield and enrichment, qPCR was performed as previously described(*1*).

Mass Spectrometry Analysis

Preparation of proteins for mass spectrometry

Proteins from RAP-MS captures were resuspended in fresh 8 M urea dissolved in 40 μL of 100 mM Tris-HCl pH 8.5. Disulfide bonds were reduced by incubation with 3 mM TCEP for 20 minutes at room temperature, followed by alkylation with 11 mM iodoacetamide for 15 minutes at room temperature in the dark. Samples were then digested with 0.1 μg endoproteinase Lys-C for 4 hours at room temperature. After Lys-C digestion the samples were diluted to a final concentration of 2M urea by the addition of 100 mM Tris-HCl pH 8.5, and $CaCl_2$ was added to a final concentration of 1 mM. Tryptic peptides were generated by

treatment with 0.1 to 0.5 μg of trypsin overnight at room temperature. Contaminating detergents were removed from peptides using HiPPR detergent removal columns (Thermo), and peptides were protonated by the addition of 5% formic acid before desalting on a Microm Bioresources C8 peptide MicroTrap column. Peptide fractions were collected and lyophilized, and dried peptides were resuspended in 0.2% formic acid with 5% acetonitrile.

Mass spectrum measurements

Liquid chromatography-mass spectrometry and data analyses of the digested samples were carried out as previously described(*39*) with the following modifications. All experiments were performed on a nanoflow LC system, EASY-nLC 1000 coupled to a hybrid linear ion trap Orbitrap Elite mass spectrometer (Thermo Fisher Scientific, Bremen, Germany) equipped with a nanoelectrospray ion source (Thermo Fisher Scientific). For the EASY-nLC II system, solvent A consisted of 97.8% H_2O, 2% ACN, and 0.2% formic acid and solvent B consisted of 19.8% H_2O, 80% ACN, and 0.2% formic acid. For the LC-MS/MS experiments, 200 ng of digested peptides were directly loaded at a flow rate of 500 nL/min onto a 16-cm analytical HPLC column (75 μm ID) packed in-house with ReproSil-Pur C_{18}AQ 3 μm resin (120 Å pore size, Dr. Maisch, Ammerbuch, Germany). The column was enclosed in a column heater operating at 30°C. After 30 min of loading time, the peptides were separated with a 75 min gradient at a flow rate of 350 nL/min. The gradient was as follows: 0–2% Solvent B (5 min), 2–30% B (60 min), and 100% B (10 min). The Elite was operated in data-dependent acquisition mode to automatically alternate between a full scan (m/z=400–1600) in the Orbitrap and subsequent rapid 20 CID MS/MS scans in the linear ion trap. CID was

performed with helium as collision gas at a normalized collision energy of 35% and 10 ms of activation time.

MS data analysis

Thermo RAW files were searched with MaxQuant (v 1.5.0.30)(*40, 41*). Spectra were searched against all UniProt mouse entries (43,565 entries, downloaded 02 Oct 14) and MaxQuant contaminant database (245 entries). Decoy sequences (reversed peptide sequences) were generated in MaxQuant to estimate the false discovery rate(*42*). Search parameters included multiplicity of 2 with heavy Arg (+10.0083) and heavy Lys (+8.0142) as heavy peptide modifications. Variable modifications included oxidation of Met (+15.9949) and protein N-terminal acetylation (+42.0106). Carboxyamidomethylation of Cys (+57.0215) was specified as a fixed modification. Protein and peptide false discovery rates were thresholded at 1%. Precursor mass tolerance was 7 ppm (or less for individual peptides). Fragment mass tolerance was 0.5 Da. Requantify and match between runs were both enabled. Trypsin was specified as the digestion enzyme with up to 2 missed cleavages.

Identification of RNA interacting proteins

Proteins of interest from RAP-MS captures were identified based on several criteria. First, proteins were considered identified only if 2 or more unique peptides were found in the mass spectrum. Then proteins of interest were selected based on the SILAC ratio of capture versus control samples. SILAC ratios for each peptide were calculated based on the intensity ratios of heavy and light SILAC pairs. The protein ratio is the median of all calculated peptide ratios, with a minimum of two SILAC pairs required for a SILAC ratio to be assigned to a

given protein. A SILAC ratio cutoff of ≥3.0 (fold enrichment over control sample) was used as a cutoff for further analysis. We excluded known contaminants, including human keratin and proteins introduced during the sample purification and preparation process (such as streptavidin, Benzonase, and trypsin), as well as naturally biotinylated proteins that contaminate the preparation by binding to streptavidin beads.

RAP-MS experiments and controls

18S rRNA versus U1 snRNA

To validate the RAP-MS method and identify proteins specifically interacting with 18S ribosomal RNA or U1 snRNA, we performed captures of each target RNA in parallel samples from heavy and light labeled lysates from wild-type V6.5 ES cells. The total protein quantity in elution samples from each RAP-MS capture was measured by comparing the median intensity of peptides identified in a single quantitation MS run for each sample. The heavy and light label swapped samples were then mixed equally based on total protein quantity and analyzed by mass spectrometry to identify the SILAC enrichment ratio of proteins originating from 18S rRNA or U1 snRNA captures. The experiment was performed twice and each experimental set contained two biological replicates of 18S and U1 captures (heavy and light labeling states).

Xist lncRNA versus U1 snRNA captures

To identify proteins specifically interacting with Xist lncRNA, we performed captures as described above with either 200M cells or 800M pSM33 cells treated with doxycycline for

6 hours. The total protein quantity in elution samples from each RAP-MS capture was measured by a single quantitation MS run for each sample. Heavy and light label swapped samples were mixed equally based on total protein quantity, and analyzed by mass spectrometry. SILAC ratios of Xist enriched proteins versus U1 enriched proteins were calculated and used to identify Xist-specific interacting proteins for further analysis. The experiment was performed twice and each experimental set contained two biological replicates of Xist and U1 captures, from heavy and light labeled samples. Proteins replicated well between samples, with a sole exception (LBR) that was missed only because its enrichment level (2-fold) fell below our enrichment cutoff (3-fold) in some replicate samples.

Xist lncRNA capture from non-crosslinked cells

As a control to ensure that purified proteins are not non-specifically associated or binding *in vitro* with target RNAs during capture, we performed RAP-MS captures of Xist from non-crosslinked cells otherwise treated in the same manner (i.e. doxycycline treated for 6 hours).

Xist lncRNA capture from cells where Xist is not expressed

To confirm that the identified proteins are not resulting from background proteins or probe association with other RNAs or proteins in the pSM33 cells, we performed RAP captures of Xist from pSM33 cells that were not treated with doxycycline, but which were otherwise treated identically.

45S pre-rRNA capture versus U1 capture

To ensure that the proteins enriched in Xist captures using RAP-MS are not simply due to increased protein capture as a consequence of long target RNA transcripts, we additionally performed captures of the 13,000 nucleotide long 45S pre-ribosomal RNA as a control. To ensure specific capture only of the 45S, and not the mature 18S and 28S, we designed probes that specifically targeted the internal transcribed spacer regions (ITS1 and ITS2) that are only present in the 45S pre-ribosomal RNA. The experiment was performed in the same manner and with the same conditions as the Xist lncRNA captures described above. To compare Xist protein enrichment to 45S protein enrichment, we used a SILAC approach based on direct comparison of two samples that share a common denominator (called spike-in SILAC(*43*)). Specifically, we calculated an overall Xist/45S SILAC ratio by multiplying the Xist/U1 ratio by the U1/45S ratio for each identified protein.

Protein domain classification

We defined the conserved domain structures of proteins using the Protein Families database (Pfam(*44*)).

RNA Immunoprecipitation in UV-crosslinked cells

We crosslinked pSM33 cells after 6 hours of doxycycline-treatment with 0.4 Joules/cm^2 of UV_{254}. Cells were lysed and RNA was digested with RNase I to achieve a size range of 100-500 nucleotides in length. Lysate preparations were precleared by mixing with Protein G beads for 1 hour at 4 ˚C. For each sample, target proteins were immunoprecipitated from 20 million cells with 10 μg of antibody (**Supplementary Table 1**) and 60 μl of Protein G magnetic beads (Invitrogen). The antibodies were pre-coupled to the beads for 1 hour at room

temperature with mixing before incubating the precleared lysate to the antibody-bead complexes for 2 hours at 4 ˚C. After the immunoprecipitation, the beads were rinsed with a wash buffer of 1× PBS with detergents. After a dephosphorylation treatment, the RNA in each sample was ligated to a mixture of barcoded adapters in which each adapter had a unique barcode identifier. After ligation, beads were rinsed with 1× PBS and detergents and then 5× PBS (750 mM NaCl) and detergents prior to pooling 3-4 antibodies in a new tube. The proteins and RNA were then eluted from the Protein G beads with 6 M urea and 40 mM DTT at 60 °C. Protein-RNA complexes were separated away from free RNA by covalently coupling proteins to NHS-magnetic beads (Pierce) and washing 3 times in 6 M GuSCN (Qiagen RLT buffer) and heating in 1% NLS at 98 °C for 10 minutes. The proteins were then digested with Proteinase K and RNA was purified for subsequent analysis. From the barcoded RNA in each pool, we generated Illumina sequencing libraries as previously described(*38*). We saved a small percentage (~1%) of starting material prior to immunoprecipitation and processed and sequenced this sample in parallel.

Analysis of crosslinked RNA Immunoprecipitation Data

We computed the enrichment for any RNA upon immunoprecipitation with a specific protein relative to its total levels in the cell. To do this, we counted the total number of reads overlapping the RNA in either the immunoprecipitation sample or the input control. To account for differences in read coverage between samples, each of these numbers was normalized to the total number of reads within the same experiment. This generates a normalized score, per RNA, within each sample. We then computed an enrichment metric by taking the ratio of these normalized values (IP/Input). We then compared these enrichment

levels across different proteins and controls (i.e. IgG). To enable direct comparison across proteins for a given gene, we need to account for differences in the protein specific background level, which may occur to differences in IP efficiency or non-specific binding of each antibody. To do this, we computed a normalized enrichment ratio by dividing the ratio for each gene by the average ratio across all genes for a given protein, as previously described(*45*).

To exclude the possibility of promiscuous binding to all RNAs, we considered various mRNA controls, which are not expected to bind to these proteins, including Oct4, Nanog, Stat3, and Suz12. These mRNAs were selected as examples because they are expressed in ES cells, although many mRNAs show similar results. To account for the possibility that the Xist RNA non-specifically binds to any RBP, we evaluated Xist with other RBPs that we did not identify as interacting with Xist by RAP-MS (Pum1 and hnRNP-H). To ensure that a negative result (i.e. no enrichment for Xist) is meaningful and does not reflect a failed immunoprecipitation experiment, we evaluated Neat1-1, which we previously found immunoprecipitates with hnRNPH(*45*). To further evaluate the level of enrichment on other lncRNAs, we considered several lncRNAs including Malat1, Firre, and Tug1. These lncRNAs were selected as examples because they are well-known and expressed in ES cells, although many ES lncRNAs show similar results.

Immunoprecipitation and RT-qPCR

Female ES cells were differentiated then crosslinked with UV4k as described above. Pellets of 20M cells were lysed and treated with TURBO DNase (Ambion) to destroy DNA by

incubation for 10 minutes at 37 ˚C in an Eppendorf Thermomixer C. The lysate was pre-cleared by incubation with 180 μL of Dynabeads Protein G magnetic beads (Life Technologies). Meanwhile, 10 μg of antibody for immunoprecipitation (SHARP antibody, Novus NBP1-82952 or IgG antibody, Cell Signaling 2729S) was coupled to 60 μl Protein G magnetic beads. After pre-clearing was completed, the lysate was then mixed with the appropriate antibody-coupled Protein G magnetic beads and incubated for 2 hours at 4 ˚C on a Hulamixer sample mixer (Life Technologies) for protein capture. After immunoprecipitation, beads were washed with a wash buffer of 1× PBS with detergents and then captured nucleic acids were eluted by digesting all proteins with 5.6 U proteinase K (New England Biolabs). Eluted RNA was purified using the RNA Clean and Concentrator-5 Kit (Zima Research) and RT-qPCR was performed as described previously(*1*) to evaluate RNA enrichment.

V5-epitope tagged protein expression

For V5-tagged protein expression and immunoprecipitation, mouse ES cells were electroporated using the Neon transfection system (Invitrogen) with an episomally-replicating vector (pCAG-GW-V5-Hygro) encoding expression of a C-terminal V5 tagged ORF driven by a CAG promoter. ORFs were obtained from the DNASU plasmid repository as Gateway entry clones and inserted into pCAG-GW-V5-Hygro using an LR recombination reaction (Invitrogen). Transfected cells were selected on 125ug/mL Hygromycin B (Invitrogen) to generate stably expressing lines.

siRNA Transfections

For siRNA knockdown experiments, 20 nM siRNAs were transfected using the Neon transfection system (settings: 1200V, 40ms width, 1 pulse). For each transfection, two 10 μL transfections with the same siRNA were carried out in succession using 100,000 cells each, mixed, and plated equally between two poly-L-lysine or poly-D-lysine (Sigma) and 0.2% gelatin (Sigma)-coated #1.5 coverslips placed into wells of a 24-well plate containing 2i media. After 48 hours, 2i media was replaced and cells on one coverslip of each pair were treated with 2 μg/mL doxycycline (Sigma) for 16hr to induce Xist expression. Coverslips were then fixed in Histochoice (Sigma) for 5 min, washed thoroughly in PBS, and dehydrated in ethanol for storage until FISH staining.

For all proteins we used siRNAs pool from Dharmacon (ON-TARGETplus SMARTpool siRNAs). For each of these, we tested whether the siRNA successfully reduced the targeted mRNA expression by >70%. For SAF-A and SMRT, the siRNAs failed to achieve this level of mRNA reduction, so we purchased additional siRNAs (and their associated controls) for SAF-A and SMRT from Qiagen and Ambion respectively, and selected siRNAs that successfully reduced on-target mRNA levels. siRNA against GFP was purchased from Qiagen. For additional independent siRNAs, the siRNAs were purchased as a pool from Dharmacon, Qiagen, and Ambion, or as each individual siRNA deconvoluted from the pool from Dharmacon and Qiagen (**Supplementary Table 2**).

siRNA experiments in female ES cells

Female ES F1 2-1 cells were similarly transfected. To initiate differentiation and Xist expression for these cells, 2i was replaced with MEF media (DMEM, 10% Gemini

Benchmark FBS, 1x L-glutamine, 1x NEAA, 1x Pen/Strep; Life Technologies unless otherwise indicated) at 12 hours post-transfection. Forty-eight hours after transfection, 1uM retinoic acid (Sigma) was administered for 24 hours and cells were fixed as described above. For cells not undergoing differentiation, 2i was replaced 12hr and 48hr after transfection.

Single molecule RNA FISH

Single molecule RNA Fluorescence *in situ* hybridization (FISH) experiments were done using QuantiGene ViewRNA ISH Cell Assay (Affymetrix) and QuantiGene ViewRNA ISH Cell 740 Module (Affymetrix) according to manufacturer's protocol. Cells fixed on coverslips were first permeabilized with Detergent Solution QC at room temperature for 5 min, and then incubated with desired mixture of probe set (Affymetrix) in Probe Set Diluent QF at 40°C for 3 h, followed by incubated with PreAmplifier Mix at 40°C for 30 min, Amplifier Mix at 40°C for 30 min, and Label Probe Mix at 40°C for 30 min sequentially. For DAPI staining, coverslips were incubated in 30 nM DAPI in PBS at room temperature for 15-20 min. Probe set and conjugated fluorophore for FISH were TYPE 1-XIST (550 nm), TYPE 4-GPC4, RBMX, SMC1A, MECP2 (488 nm), TYPE 10-ATRX (740 nm), and TYPE 6-EED1, SHARP, LBR, SAFA, RBM15, MYEF2, PTBP1, HNRNPC, HNRNPM, CELF1, RALY, HDAC3, NCOR2, MID1, PIR (650 nm).

Immunofluorescence and RNA FISH

For immunofluorescence (IF), cells were fixed on coverslips and permeabilized with 0.1% Triton-X in PBS at room temperature for 10 min, and blocked with 5% normal goat serum in PBS at room temperature for 10 min. Cells were then incubated with primary antibodies

at room temperature for 1 h, followed by incubating with secondary antibodies at room temperature for 1 h. The samples were then processed using the RNA FISH protocol, as described above. Primary antibodies and the dilution used for IF were anti-RNA polymerase II CTD repeat YSPTSPS (phospho S2) (Abcam; ab5095) (1:100), anti-Nanog (Abcam; ab80892) (1:100), and anti-EZH2 (Active Motif; 39933) (1:100). Secondary antibodies and the dilution used for IF were Alexa Fluor® 405 goat anti-rabbit IgG (H+L) (Life Technology; 1575534) (1:100) and Alexa Fluor® 488 $F(ab')_2$ fragment of goat anti-rabbit IgG (H+L) (Life Technology; 1618692) (1:100).

Microscopic Imaging

FISH and IF/FISH samples were imaged using a Leica DMI 6000 Deconvolution Microscope with the Leica HC PL APO 63x/1.30 GLYC CORR CS2 objective. Samples stained with TYPE 10-ATRX (740 nm) were imaged using Nikon Ti Eclipse with the Nikon CFI Plan Apochromat λ DM 60x/1.40 oil objective. Images were projected with maximum projection (3 μm; step size, 0.5 μm).

X-chromosome Silencing Assay

Cells were stained for Xist RNA, Gpc4 mRNA, Atrx mRNA and siRNA-targeted mRNA by FISH and imaged. In addition, in some siScramble and siSHARP samples, we used probes against Rbm15, Mecp2, Smc1a, Mid1 or Pir mRNA. Images were then analyzed using Matlab R2013b (described below). Cells were selected if the copy number of the targeted mRNA was less than 30% of the level of the no siRNA treated cells and if they induced Xist expression. Within these cells, the copy number of Gpc4 mRNA and Atrx mRNA were

quantified using a peak finding method (described below) and compared across conditions. We quantified mRNA levels for 50 individual cells. We also evaluated Xist expression in siRNA-treated cells, and observed no difference in the percentage of cells that induced Xist expression in any of the siRNA conditions relative to untreated cells.

Quantifying mRNAs by single molecule FISH

All image analysis was carried out using Matlab (version R2013b) utilizing built-in functions from the Image Processing toolbox. Images were first filtered using a two-dimensional median filter to remove background. Cell boundaries were outlined manually, guided by DAPI staining, to create a binary mask and applied to the various channels from the same field of view. Top-hat morphological filtering, a background subtraction method that enhances the individual focal spots, was applied to the images(*46*). The spots were then identified using a 2D peak finding algorithm that identifies local maximal signals within the cell. Once regional maxima were identified, the number of spots was counted for each cell.

References

1. J. M. Engreitz *et al.*, The Xist lncRNA exploits three-dimensional genome architecture to spread across the X chromosome. *Science* **341**, 1237973 (2013).
2. R. Brimacombe, W. Stiege, A. Kyriatsoulis, P. Maly, Intra-RNA and RNA-protein cross-linking techniques in Escherichia coli ribosomes. *Methods in enzymology* **164**, 287-309 (1988).
3. R. B. Darnell, HITS-CLIP: panoramic views of protein-RNA regulation in living cells. *Wiley interdisciplinary reviews. RNA* **1**, 266-286 (2010).
4. J. Ule *et al.*, CLIP identifies Nova-regulated RNA networks in the brain. *Science* **302**, 1212-1215 (2003).
5. S. Mili, J. A. Steitz, Evidence for reassociation of RNA-binding proteins after cell lysis: implications for the interpretation of immunoprecipitation analyses. *RNA* **10**, 1692-1694 (2004).
6. S. E. Ong, M. Mann, A practical recipe for stable isotope labeling by amino acids in cell culture (SILAC). *Nature protocols* **1**, 2650-2660 (2006).
7. M. Pardo, J. S. Choudhary, Assignment of protein interactions from affinity purification/mass spectrometry data. *Journal of proteome research* **11**, 1462-1474 (2012).
8. R. M. Kaake, X. Wang, L. Huang, Profiling of protein interaction networks of protein complexes using affinity purification and quantitative mass spectrometry. *Molecular & cellular proteomics : MCP* **9**, 1650-1665 (2010).

9. L. Trinkle-Mulcahy *et al.*, Identifying specific protein interaction partners using quantitative mass spectrometry and bead proteomes. *The Journal of cell biology* **183**, 223-239 (2008).

10. S. M. Mount, I. Pettersson, M. Hinterberger, A. Karmas, J. A. Steitz, The U1 small nuclear RNA-protein complex selectively binds a 5' splice site in vitro. *Cell* **33**, 509-518 (1983).

11. T. A. Steitz, A structural understanding of the dynamic ribosome machine. *Nat Rev Mol Cell Biol* **9**, 242-253 (2008).

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

13. J. Yong, M. Kasim, J. L. Bachorik, L. Wan, G. Dreyfuss, Gemin5 delivers snRNA precursors to the SMN complex for snRNP biogenesis. *Mol Cell* **38**, 551-562 (2010).

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

15. A. Wutz, T. P. Rasmussen, R. Jaenisch, Chromosomal silencing and localization are mediated by different domains of Xist RNA. *Nat Genet* **30**, 167-174 (2002).

16. A. Wutz, R. Jaenisch, A shift from reversible to irreversible X inactivation is triggered during ES cell differentiation. *Mol Cell* **5**, 695-705 (2000).

17. J. Chaumeil, P. Le Baccon, A. Wutz, E. Heard, A novel role for Xist RNA in the formation of a repressive nuclear compartment into which genes are recruited when silenced. *Genes Dev* **20**, 2223-2237 (2006).

18. Y. Hasegawa, N. Brockdorff, S. Kawano, K. Tsutui, S. Nakagawa, The matrix protein hnRNP U is required for chromosomal localization of Xist RNA. *Dev Cell* **19**, 469-476 (2010).

19. R. Helbig, F. O. Fackelmayer, Scaffold attachment factor A (SAF-A) is concentrated in inactive X chromosome territories through its RGG domain. *Chromosoma* **112**, 173-182 (2003).

20. D. Pullirsch *et al.*, The Trithorax group protein Ash2l and Saf-A are recruited to the inactive X chromosome at the onset of stable X inactivation. *Development* **137**, 935-943 (2010).

21. M. Ariyoshi, J. W. Schwabe, A conserved structural motif reveals the essential transcriptional repression function of Spen proteins and their role in developmental signaling. *Genes Dev* **17**, 1909-1920 (2003).

22. S. Mikami *et al.*, Structural insights into the recruitment of SMRT by the corepressor SHARP under phosphorylative regulation. *Structure* **22**, 35-46 (2014).

23. V. Perissi, K. Jepsen, C. K. Glass, M. G. Rosenfeld, Deconstructing repression: evolving models of co-repressor action. *Nat Rev Genet* **11**, 109-123 (2010).

24. T. Mercher *et al.*, Involvement of a human gene related to the Drosophila spen gene in the recurrent t(1;22) translocation of acute megakaryocytic leukemia. *Proc Natl Acad Sci U S A* **98**, 5776-5779 (2001).

25. G. D. Raffel *et al.*, Ott1 (Rbm15) is essential for placental vascular branching morphogenesis and embryonic development of the heart and spleen. *Molecular and cellular biology* **29**, 333-341 (2009).

26. G. D. Raffel *et al.*, Ott1(Rbm15) has pleiotropic roles in hematopoietic development. *Proc Natl Acad Sci U S A* **104**, 6001-6006 (2007).

27. S. Haas, A. Steplewski, L. D. Siracusa, S. Amini, K. Khalili, Identification of a sequence-specific single-stranded DNA binding protein that suppresses transcription of the mouse myelin basic protein gene. *The Journal of biological chemistry* **270**, 12503-12510 (1995).

28. B. van Riel *et al.*, A novel complex, RUNX1-MYEF2, represses hematopoietic genes in erythroid cells. *Molecular and cellular biology* **32**, 3814-3822 (2012).

29. A. Chu, R. Rassadi, U. Stochaj, Velcro in the nuclear envelope: LBR and LAPs. *FEBS letters* **441**, 165-169 (1998).

30. A. L. Olins, G. Rhodes, D. B. Welch, M. Zwerger, D. E. Olins, Lamin B receptor: multi-tasking at the nuclear envelope. *Nucleus* **1**, 53-70 (2010).

31. H. J. Worman, J. Yuan, G. Blobel, S. D. Georgatos, A lamin B receptor in the nuclear envelope. *Proc Natl Acad Sci U S A* **85**, 8531-8534 (1988).

32. C. Maison *et al.*, Higher-order structure in pericentric heterochromatin involves a distinct pattern of histone modification and an RNA component. *Nat Genet* **30**, 329-334 (2002).

33. Q. Ye, I. Callebaut, A. Pezhman, J. C. Courvalin, H. J. Worman, Domain-specific interactions of human HP1-type chromodomain proteins and inner nuclear

membrane protein LBR. *The Journal of biological chemistry* **272**, 14983-14989 (1997).

34. A. Raj, P. van den Bogaard, S. A. Rifkin, A. van Oudenaarden, S. Tyagi, Imaging individual mRNA molecules using multiple singly labeled probes. *Nat Methods* **5**, 877-879 (2008).

35. R. Margueron, D. Reinberg, The Polycomb complex PRC2 and its mark in life. *Nature* **469**, 343-349 (2011).

36. J. A. Simon, R. E. Kingston, Mechanisms of polycomb gene silencing: knowns and unknowns. *Nat Rev Mol Cell Biol* **10**, 697-708 (2009).

37. K. Mitsui *et al.*, The homeoprotein Nanog is required for maintenance of pluripotency in mouse epiblast and ES cells. *Cell* **113**, 631-642 (2003).

38. J. M. Engreitz *et al.*, RNA-RNA interactions enable specific targeting of noncoding RNAs to nascent Pre-mRNAs and chromatin sites. *Cell* **159**, 188-199 (2014).

39. A. Kalli, S. Hess, Effect of mass spectrometric parameters on peptide and protein identification rates for shotgun proteomic experiments on an LTQ-orbitrap mass analyzer. *Proteomics* **12**, 21-31 (2012).

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

41. J. Cox *et al.*, Andromeda: a peptide search engine integrated into the MaxQuant environment. *Journal of proteome research* **10**, 1794-1805 (2011).

42. J. E. Elias, S. P. Gygi, Target-decoy search strategy for mass spectrometry-based proteomics. *Methods in molecular biology* **604**, 55-71 (2010).

43. T. Geiger *et al.*, Use of stable isotope labeling by amino acids in cell culture as a spike-in standard in quantitative proteomics. *Nature protocols* **6**, 147-157 (2011).

44. R. D. Finn *et al.*, Pfam: the protein families database. *Nucleic Acids Res* **42**, D222-230 (2014).

45. M. Guttman *et al.*, lincRNAs act in the circuitry controlling pluripotency and differentiation. *Nature* **477**, 295-300 (2011).

46. Z. Theodosiou *et al.*, Automated analysis of FISH and immunohistochemistry images: a review. *Cytometry A* **71**, 439-450 (2007).

Chapter 3

SHARP IS REQUIRED FOR XIST-MEDIATED EXCLUSION OF RNA POLYMERASE II

The work was first published as:

McHugh, Colleen A., et al. "The Xist lncRNA directly interacts with SHARP to silence transcription through HDAC3." Nature 521.7551 (2015): 232.

3.1 SHARP/SMRT/HDAC3 Complex is Required for Xist-mediated Exclusion of RNA Polymerase II

Xist-mediated transcriptional silencing involves several distinct roles, including (i) initial localization to sites on the X-chromosome(*1*, *2*), (ii) establishment of the initial silenced compartment – an Xist-coated nuclear domain from which RNA PolII is excluded(*3*), and (iii) repositioning of actively transcribed genes into the RNA PolII-excluded nuclear compartment(*3*, *4*). To determine the proteins that are directly responsible for establishing the initial silenced compartment on the X-chromosome, we explored whether SHARP or LBR might be required for the exclusion of RNA PolII from the Xist-coated region. We excluded SAF-A because we observed a diffuse Xist localization pattern in the nucleus upon knock down of SAF-A, but not in any other sample, consistent with the previous observation that SAF-A is required for Xist localization to chromatin(*5*).

To identify the proteins required for RNA PolII exclusion, we measured co-localization in single cells using RNA FISH for Xist and immunofluorescence for PolII upon Xist induction. In wild-type cells after 16 hours of Xist induction, we observed a depletion of RNA PolII over the Xist-coated territory (**Methods, Figure 3.1a**). We observed a similar depletion in the negative controls and upon knock down of EED (**Figure 3.1a, b**). Similarly, knock down of LBR did not alter the exclusion of PolII from the Xist-coated region (**Figure 3.1b**). In contrast, upon knock down of SHARP, we observed higher levels of PolII over the Xist-coated territory relative to the control samples (**Figure 3.1b**). We note that the effects on transcriptional silencing and PolII exclusion are robust, being seen with two independent siRNA pools (**Figure 3.1d**).

Together, these results demonstrate that SHARP is required to exclude RNA PolII on the inactive X-chromosome and may be required for creating the initial silenced compartment upon Xist localization(*4*). While LBR is not required for initial PolII exclusion on the X-chromosome, it is likely to play an alternative role during the initiation of Xist-mediated transcriptional silencing, such as repositioning genes into this PolII excluded compartment(*2, 3*).

Having identified SHARP as the direct Xist-interacting protein that is required for excluding RNA PolII on the X-chromosome, we sought to determine how it might carry out this role. SHARP is a direct RNA binding protein(*6, 7*) that was identified based on its interaction with the SMRT co-repressor complex(*7, 8*), which is required for activating the deacetylation and transcriptional silencing activity of HDAC3 *in vivo* (*9, 10*). Based on these previous

observations, we hypothesized that Xist-mediated transcriptional silencing through SHARP would occur through SMRT and the silencing function of HDAC3.

To test this hypothesis, we knocked down either SMRT or HDAC3 and measured the expression of X-linked genes before and after Xist induction. Knock down of SMRT or HDAC3 abrogated silencing of X-linked genes upon induction of Xist expression (**Figure 2.2b, g**). To determine whether this effect is similar to the effect produced by knock down of SHARP or a distinct defect in transcriptional silencing, we tested whether HDAC3, the silencing protein in this complex(*10, 11*), is required for the exclusion of RNA PolII from the Xist-coated territory. We found that knock down of HDAC3 eliminated the exclusion of RNA PolII from the Xist-coated compartment to a similar degree to that seen for knock down of SHARP (**Figure 3.1a-d**).

Together, these results suggest that SHARP silences transcription through the HDAC3 silencing protein. This role for HDAC3 in Xist-mediated silencing would explain the long-standing observation of global hypoacetylation on the entire X-chromosome as one of the very first events that occur upon initiation of XCI(*4, 12*).

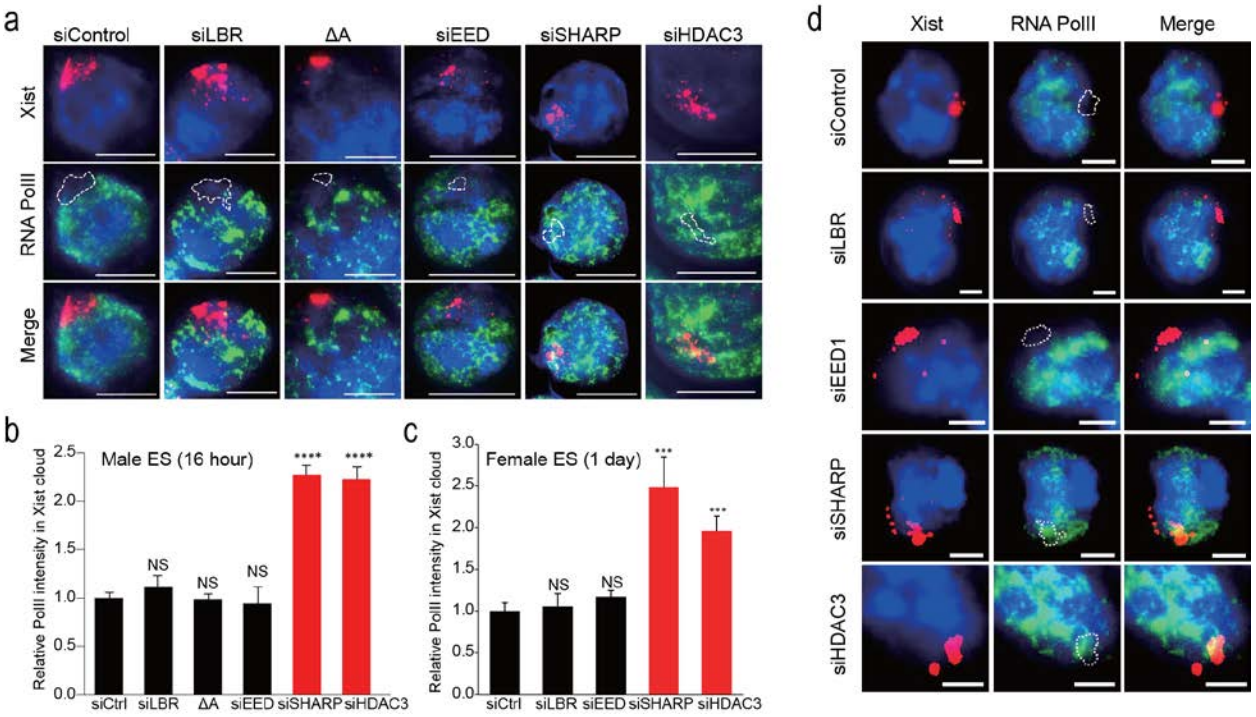

Figure 3.1 SHARP is required for exclusion of PolII from the Xist-coated territory. (a) Individual male ES cells labelled with Xist (red), PolII (green), and DAPI (blue) across different siRNA conditions (rows). Quantification of fluorescence intensity of PolII within Xist territory normalized to control siRNA levels for (b) male ES cells after 16 hours of doxycycline treatment and (c) female ES cells after 1 day of retinoic acid induced differentiation. (d) Images of individual cells that are labeled with Xist (red), RNA Polymerase II (green), and DAPI (blue) across different siRNA conditions (rows) in female ES cells after 24 hours of retinoic acid treatment. The dashed white region represents the outlined Xist coated territory. Error bars: standard error across 50 cells from one experiment. NS: not significant, **** p-value<0.001 relative to siControl by unpaired two-sample t-test. ΔA: Genetic deletion of A-repeat. Scale bars, 5 micrometers.

3.2 SHARP is Required for Xist-mediated Recruitment of PRC2

One of the features of the initiation of X-chromosome inactivation is the recruitment of the PRC2 complex across the X-chromosome in an Xist-dependent manner(*13-15*). While PRC2 is not required for the initiation of XCI(*16, 17*) (**Figure 2.2b, g**), it or its associated H3K27me3 repressive chromatin modifications may be involved in establishing an epigenetically silenced state(*18, 19*). Yet how Xist recruits the PRC2 complex across the X-chromosome is unknown.

Since we failed to identify any PRC2 components in our RAP-MS data, we reasoned that PRC2 recruitment across the X-chromosome might occur through one of the other identified direct interacting proteins. Specifically, we hypothesized that SHARP might be required to recruit PRC2 to the X-chromosome for three main reasons: (i) SHARP is required for RNA PolII exclusion, which has been shown to be sufficient to trigger PRC2 recruitment in other contexts(*20*); (ii) Previous studies have shown that the PRC2 complex can interact with various HDAC complexes(*21*) or may be recruited through HDAC mediated compaction of chromatin(*22*); (iii) SHARP has been shown to interact *in vitro* with RbAp48(*7*), a component of several chromatin regulatory complexes including the PRC2(*23, 24*) and HDAC3 complexes(*25, 26*).

To test this hypothesis, we looked at PRC2 recruitment to the Xist-coated territory. In wild-type cells, we observe a strong enrichment of EZH2, a component of PRC2, over the Xist-coated territory after 16 hours of induction (**Figure 3.2a**). Upon knock down of EED, a distinct component of the PRC2 complex that is required for its proper localization to

chromatin(*23, 24*), we observe no enrichment of EZH2 over the Xist cloud at this same time point (**Figure 3.2a**). Similarly, upon knock down of SHARP, we identified a loss of EZH2 over the Xist coated territory, of comparable magnitude to that observed in the absence of EED (**Figure 3.2a**). Conversely, upon knock down of LBR, we observed a strong enrichment of EZH2 over the Xist coated territory, of comparable magnitude to the levels of recruitment in wild-type conditions (**Figure 3.2b**). To determine whether HDAC3 is required for PRC2 recruitment, we knocked down HDAC3 and observed a loss of PRC2 recruitment (**Figure 3.2a**), of comparable magnitude to that observed upon loss of SHARP (**Figure 3.2b**).

Together, these results argue that PRC2 recruitment across the X-chromosome is dependent on the Xist interaction with SHARP and the activity of HDAC3. Whether this occurs through an interaction with HDAC3 (direct recruitment) or due to the HDAC3-induced silenced transcription state, chromatin modifications, or compact chromatin structure (indirect recruitment) remains unclear. We note that our results are in contrast to previous reports that PRC2 is recruited to the X-chromosome through a direct interaction between Ezh2 and the A-repeat of Xist(*15*). Instead, our results are consistent with reports that the deletion of the A-repeat, unlike knock down of SHARP or HDAC3, has no significant effect on PRC2 recruitment(*13, 27*) (**Figure 3.2b**).

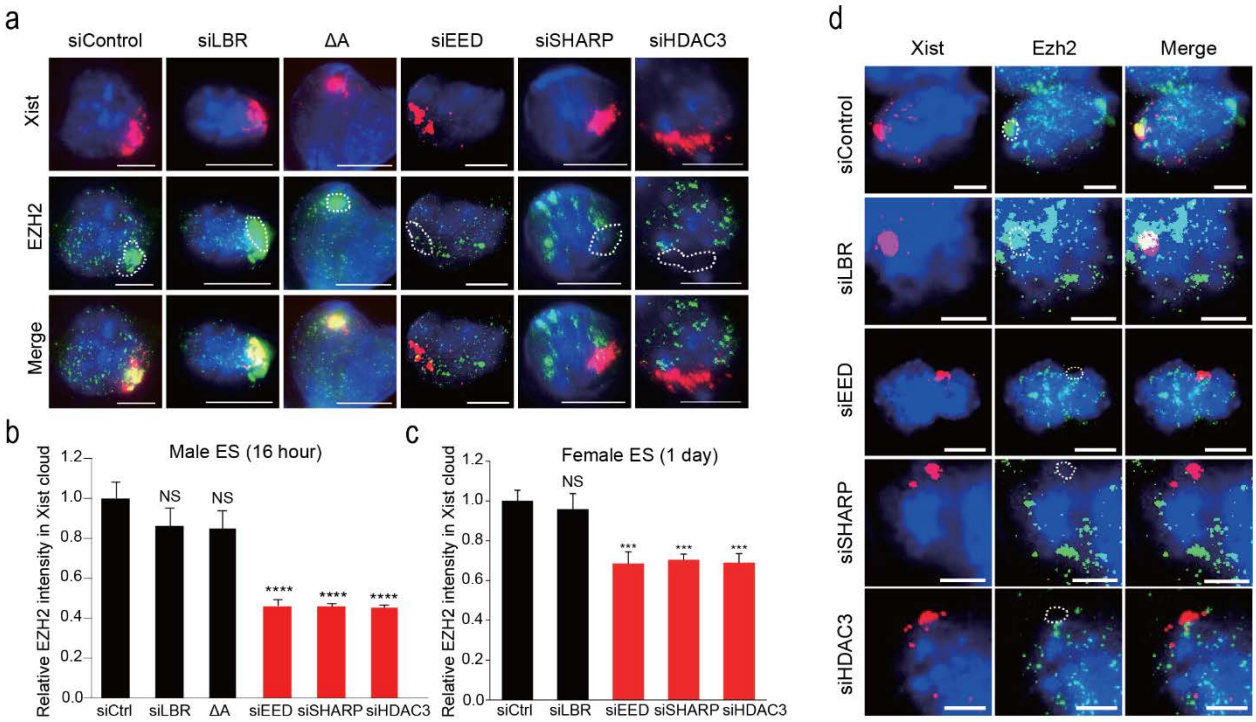

Figure 3.2. SHARP is required for PRC2 recruitment across the Xist-coated territory. (a) Images of individual male ES cells that are labeled with Xist (red), Ezh2 (green) and DAPI (blue) across different siRNA conditions (rows). Quantification of the level of Ezh2 within the defined Xist territory normalized to the levels in the control siRNA sample for (b) male ES cells and (c) differentiating female ES cells. (d) **.** Images of individual cells that are labeled with Xist (red), Ezh2 (green), and DAPI (blue) across different siRNA conditions (rows) in female ES cells after 24 hours of differentiation. Error bars: standard error of the mean across 50 cells from one experiment. NS: not significant, ***, *p*-value<0.005, **** *p*-value<0.001 relative to siControl by an unpaired two-sample t-test. Scale bars, 5 micrometers. The dashed white region represents the outlined Xist coated territory.

Methods

Only the methods that have not yet been described in previous Chapters are listed

Ezh2 Recruitment and PolII Exclusion

Cells were stained for Xist RNA and the siRNA-targeted mRNA (FISH) along with Ezh2 or PolII (IF) as described above. For image acquisition, the exposure time for each individual channel was kept the same across all samples. Images were then analyzed and selected for XIST-induced and cells showing knock down of the target mRNA, as described above. Specifically, the nuclei of individual cells were identified manually using the DAPI staining. We identified the Xist cloud by using an intensity-based threshold to partition the image within the nucleus and find contiguous 2-dimensional regions of high intensity. The threshold was determined based on Otsu method as previously described[50], which splits the image into 2 bins – high and low – and identifies a threshold that minimizes the variance within the partition. This creates a binary mask on the image. We visually confirmed that this binary mask accurately reflected the Xist cloud. We then applied this binary mask to all other images in that field of view (PolII or Ezh2) for all images. We then quantified the intensity of fluorescence signal by taking the average intensity of all the pixels within the region (i.e. PolII or Ezh2, respectively). We computed this average intensity (1 number per cell) across all conditions and compared them using a 2-same unpaired t-test relative to the scramble sample across 50 single cells.

References

1. M. D. Simon *et al.*, High-resolution Xist binding maps reveal two-step spreading during X-chromosome inactivation. *Nature* **504**, 465-469 (2013).
2. J. M. Engreitz *et al.*, The Xist lncRNA exploits three-dimensional genome architecture to spread across the X chromosome. *Science* **341**, 1237973 (2013).
3. J. Chaumeil, P. Le Baccon, A. Wutz, E. Heard, A novel role for Xist RNA in the formation of a repressive nuclear compartment into which genes are recruited when silenced. *Genes Dev* **20**, 2223-2237 (2006).
4. A. Wutz, Gene silencing in X-chromosome inactivation: advances in understanding facultative heterochromatin formation. *Nat Rev Genet* **12**, 542-553 (2011).
5. Y. Hasegawa, N. Brockdorff, S. Kawano, K. Tsutui, S. Nakagawa, The matrix protein hnRNP U is required for chromosomal localization of Xist RNA. *Dev Cell* **19**, 469-476 (2010).
6. F. Arieti *et al.*, The crystal structure of the Split End protein SHARP adds a new layer of complexity to proteins containing RNA recognition motifs. *Nucleic acids research* **42**, 6742-6752 (2014).
7. Y. Shi *et al.*, Sharp, an inducible cofactor that integrates nuclear receptor repression and activation. *Genes Dev* **15**, 1140-1151 (2001).
8. M. Ariyoshi, J. W. Schwabe, A conserved structural motif reveals the essential transcriptional repression function of Spen proteins and their role in developmental signaling. *Genes Dev* **17**, 1909-1920 (2003).

9. M. G. Guenther, O. Barak, M. A. Lazar, The SMRT and N-CoR corepressors are activating cofactors for histone deacetylase 3. *Molecular and cellular biology* **21**, 6091-6101 (2001).

10. S. H. You *et al.*, Nuclear receptor co-repressors are required for the histone-deacetylase activity of HDAC3 in vivo. *Nature structural & molecular biology* **20**, 182-187 (2013).

11. J. Li, Q. Lin, W. Wang, P. Wade, J. Wong, Specific targeting and constitutive association of histone deacetylase complexes during transcriptional repression. *Genes Dev* **16**, 687-692 (2002).

12. K. Plath, S. Mlynarczyk-Evans, D. A. Nusinow, B. Panning, Xist RNA and the mechanism of X chromosome inactivation. *Annu Rev Genet* **36**, 233-278 (2002).

13. K. Plath *et al.*, Role of histone H3 lysine 27 methylation in X inactivation. *Science* **300**, 131-135 (2003).

14. J. Silva *et al.*, Establishment of histone h3 methylation on the inactive X chromosome requires transient recruitment of Eed-Enx1 polycomb group complexes. *Dev Cell* **4**, 481-495 (2003).

15. J. Zhao, B. K. Sun, J. A. Erwin, J. J. Song, J. T. Lee, Polycomb proteins targeted by a short repeat RNA to the mouse X chromosome. *Science* **322**, 750-756 (2008).

16. S. Kalantry, T. Magnuson, The Polycomb group protein EED is dispensable for the initiation of random X-chromosome inactivation. *PLoS Genet* **2**, e66 (2006).

17. S. Schoeftner *et al.*, Recruitment of PRC1 function at the initiation of X inactivation independent of PRC2 and silencing. *Embo J* **25**, 3110-3122 (2006).

18. S. Kalantry *et al.*, The Polycomb group protein Eed protects the inactive X-chromosome from differentiation-induced reactivation. *Nature cell biology* **8**, 195-202 (2006).

19. A. Kohlmaier *et al.*, A chromosomal memory triggered by Xist regulates histone methylation in X inactivation. *PLoS Biol* **2**, E171 (2004).

20. E. M. Riising *et al.*, Gene silencing triggers polycomb repressive complex 2 recruitment to CpG islands genome wide. *Mol Cell* **55**, 347-360 (2014).

21. J. van der Vlag, A. P. Otte, Transcriptional repression mediated by the human polycomb-group protein EED involves histone deacetylation. *Nat Genet* **23**, 474-478 (1999).

22. W. Yuan *et al.*, Dense chromatin activates Polycomb repressive complex 2 to regulate H3 lysine 27 methylation. *Science* **337**, 971-975 (2012).

23. R. Margueron, D. Reinberg, The Polycomb complex PRC2 and its mark in life. *Nature* **469**, 343-349 (2011).

24. J. A. Simon, R. E. Kingston, Mechanisms of polycomb gene silencing: knowns and unknowns. *Nat Rev Mol Cell Biol* **10**, 697-708 (2009).

25. E. Nicolas, S. Ait-Si-Ali, D. Trouche, The histone deacetylase HDAC3 targets RbAp48 to the retinoblastoma protein. *Nucleic acids research* **29**, 3131-3136 (2001).

26. Y. Zhang *et al.*, Analysis of the NuRD subunits reveals a histone deacetylase core complex and a connection with DNA methylation. *Genes Dev* **13**, 1924-1935 (1999).

27. S. T. da Rocha *et al.*, Jarid2 Is Implicated in the Initial Xist-Induced Targeting of PRC2 to the Inactive X Chromosome. *Mol Cell* **53**, 301-316 (2014).

Chapter 4

LBR IS REQUIRED FOR XIST SPREADING TO ACTIVELY TRANSCRIBED GENES ACROSS THE X-CHROMOSOME

The work was first published as:

Chen, Chun-Kan, et al. "Xist recruits the X chromosome to the nuclear lamina to enable chromosome-wide silencing." Science 354.6311 (2016): 468-472.

4.1 LBR Requires its Arginine-Serine (RS) Motif to Interact with Xist

Although the 3-dimensional structure of the nucleus is dynamically organized in different cellular conditions (*1-3*), it is generally unclear whether these structural changes reflect distinct regulatory states or whether they lead to changes in gene regulation (*2, 4, 5*). X chromosome inactivation (XCI) represents an ideal model to study the relationship between dynamic 3-dimensional nuclear organization and gene regulation because initiation of XCI entails chromosome-wide transcriptional silencing (*6, 7*) and large scale remodeling of the 3-dimensional structure of the X chromosome (*8-12*).

The Xist lncRNA initiates XCI by spreading across the future inactive X chromosome and excluding RNA Polymerase II (PolII) to silence chromosome-wide transcription (*7, 13-15*). Xist initially localizes to genomic DNA regions on the X chromosome that are not actively transcribed (*16-19*) before spreading to actively transcribed genes (*17*). Deletion of a highly conserved region of Xist that is required for transcriptional silencing, called the A-repeat

region (*20*), leads to a defect in Xist spreading to actively transcribed genes on the X chromosome (*17*) and spatial exclusion of active genes from the Xist-coated nuclear compartment (*21*). While these results suggest a role for Xist in remodeling the 3-dimensional structure of the X chromosome, whether these structural changes are required for, or merely a consequence of, transcriptional silencing mediated by the A-repeat of Xist remains unclear (*17, 22*).

Addressing this question requires understanding the molecular components involved in Xist-mediated transcriptional silencing. Recently, we and others identified the proteins that interact with Xist using mass spectrometry (*23-25*). One of these proteins is the Lamin B Receptor (LBR) (*23, 25*), a transmembrane protein that is anchored in the inner nuclear membrane (*26*), interacts with Lamin B (*27, 28*), and is required for anchoring chromatin to the nuclear lamina (*27*) – a nuclear compartment that helps shape the 3-dimensional structure of DNA (*2, 3, 29*) and is enriched for silencing proteins (*2, 3, 30, 31*). Based on these observations, along with the observation that induction of XCI leads to recruitment of the inactive X-chromosome to the nuclear lamina (*11, 12*), we hypothesized that the Xist-LBR interaction might be important both for shaping nuclear organization and regulating gene expression during XCI.

To determine whether LBR-mediated silencing is due to its interaction with Xist, we sought to disrupt its RNA binding region. However, among the 10 Xist-interacting proteins that we previously identified (11), LBR is the only protein that does not contain a canonical RNA binding domain. We hypothesized that the Arginine-Serine (RS) motif of LBR might be required for interacting with Xist because the RS motif is present within a class of mRNA

binding proteins involved in splicing (SR proteins) (31–33), is overrepresented in RNA binding proteins that lack canonical RNA binding domains (34), and the RS motif of LBR was previously shown to interact with RNA in vitro (35). To test this, we generated a truncated LBR protein containing a deletion of the RS motif (ΔRS-LBR, **Figure 4.1a**, **Methods**). As a control, we also deleted seven of the eight transmembrane domains in LBR (ΔTM-LBR, **Figure 4.1a**, **Methods**). Consistent with previous observations (*32*), we find that both ΔRS-LBR and ΔTM-LBR localize properly in the nuclear envelope (**Figure 4.1b**). Importantly, ΔRS-LBR did not interact with Xist (~97% reduction relative to wild-type levels, **Figure 4.1c**, **Methods**) and failed to rescue the silencing defect upon knock down of LBR (**Figure 4.1d**). In contrast, ΔTM-LBR did not impact Xist binding (**Figure 4.1c**) and was able to rescue the silencing defect upon knock down of LBR (**Figure 4.1d**).

To ensure that ΔRS-LBR fails to silence X chromosome genes because of its RNA binding ability and specifically its interaction with Xist, we tested whether artificially tethering ΔRS-LBR to the Xist RNA can re-establish Xist-mediated silencing. To do this, we fused 3 copies of the viral BoxB RNA aptamer, which binds tightly to the viral λN coat protein (*33-35*), to the 3' end of the endogenous Xist RNA (Xist-BoxB, **Figure 4.1e**, **Methods**) and ensured that Xist-BoxB still silences X chromosome genes (**Figure 4.1f**). Expression of ΔRS-LBR-λN in Xist-BoxB cells rescued the silencing defect observed upon LBR knock down (**Figure 4.1g**). Together, these results demonstrate that the Xist-LBR interaction is required for Xist-mediated transcriptional silencing (**Figure 4.1h**).

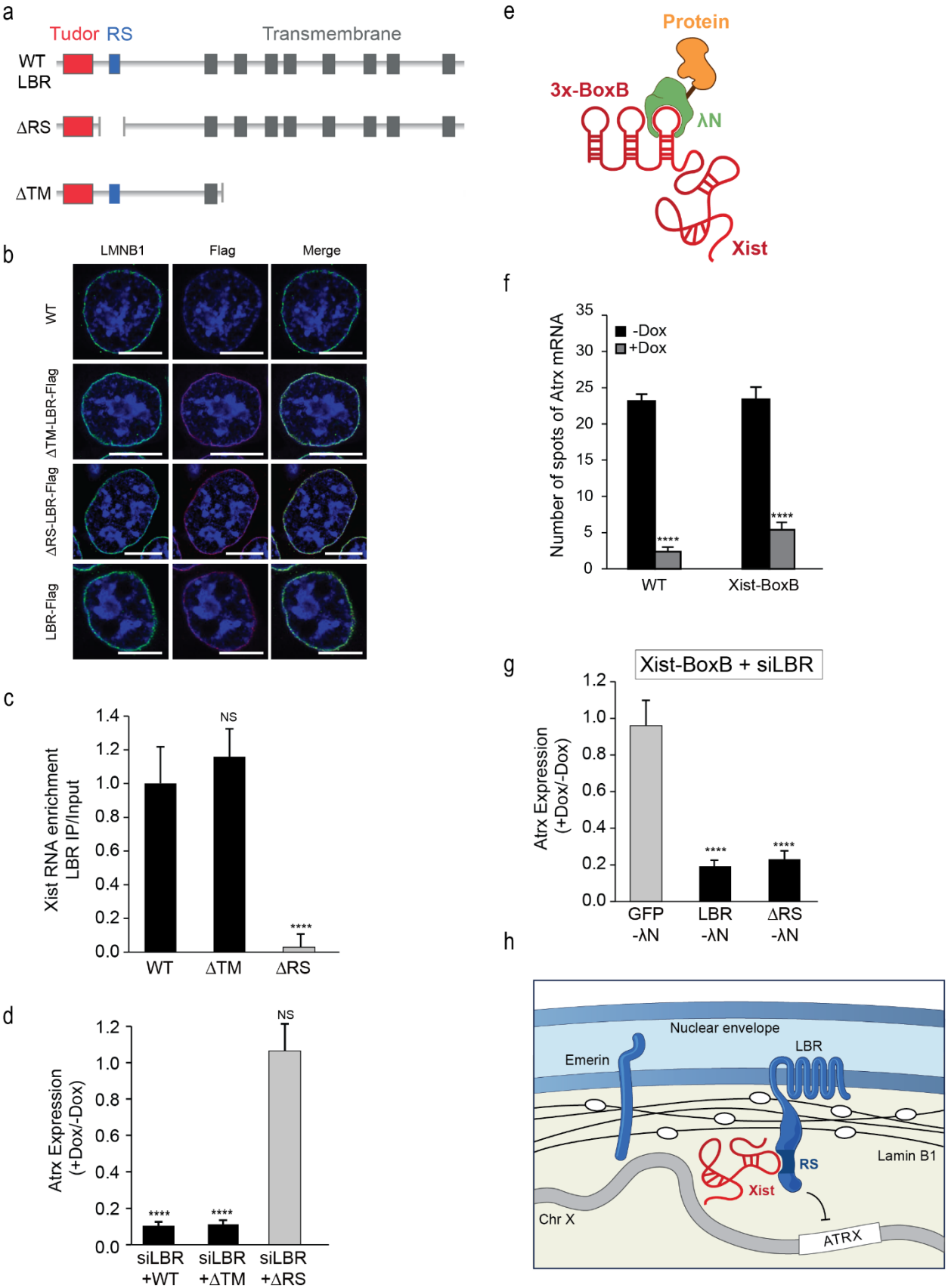
a
Tudor RS
Transmembrane
WT LBR
ΔRS
ΔTM
b
LMNB1
Flag
Merge
WT
ΔTM-LBR-Flag
ΔRS-LBR-Flag
LBR-Flag
c
Xist RNA enrichment LBR IP/Input
NS
****
WT
ΔTM
ΔRS
d
Atrx Expression (+Dox/-Dox)
NS
****
siLBR +WT
siLBR +ΔTM
siLBR +ΔRS
e
Protein
3x-BoxB
λN
Xist
f
Number of spots of Atrx mRNA
-Dox
+Dox
****
WT
Xist-BoxB
g
Xist-BoxB + siLBR
Atrx Expression (+Dox/-Dox)
****
GFP -λN
LBR -λN
ΔRS -λN
h
Nuclear envelope
Emerin
LBR
Lamin B1
RS
Xist
Chr X
ATRX

Figure 4.1. LBR requires its RS motif to interact with Xist and silence transcription.

(a) The domain structure of the LBR protein highlighting the Tudor domain (red), RS motif (blue), and 8 transmembrane domains (gray). The regions deleted in ΔRS-LBR (amino acids 71-90) and ΔTM-LBR (amino acids 237-615). (b) Images of individual cell stained with endogenous LMNB1 (green) and FLAG-tag (red) in cells transfected with empty vector (WT) or λN-3xFLAG tagged ΔTM-LBR, ΔRS-LBR, or full-length LBR. The λN-3xFLAG tagged proteins colocalize with endogenous LMNB1 demonstrating that these mutants localize properly in the nuclear envelope. Scale bars: 5 micrometers. (c) Xist RNA levels measured by RT-qPCR after immunoprecipitation of a 3x-FLAG tagged LBR mutants relative to the input in cells expressing full-length LBR (WT), ΔRS-LBR, or ΔTM-LBR. Error bars represent the standard error from three independent IP experiments. NS: not significantly different from the input; **** represents values significantly different from the input with a p-value<0.001 based on an unpaired two-sample t-test. (d) Relative Atrx mRNA expression upon knockdown of the endogenous LBR and expression of full length LBR (WT), ΔTM-LBR, or ΔRS-LBR. (e) A schematic of the interaction between λN-fusion protein and Xist containing 3 copies of the BoxB RNA aptamer. (f) Quantification of the copy number of Atrx mRNA prior to Xist-BoxB induction (-Dox) and after Xist-BoxB induction (+Dox). (g) Relative Atrx mRNA expression in Xist-BoxB cells after knockdown of the endogenous LBR and expression of GFP-λN (control), LBR-λN, or ΔRS-LBR-λN. (h) A schematic of the nuclear lamina and the interaction between LBR and Xist. Error bars: SEM across 50 individual cells. NS: not significant, **** p-value<0.001 relative to wild type cells (d, f), or cells transfected with GFP-λN (g) by an unpaired two-sample t-test.

4.2 LBR Binds to Precise Regions of the Xist RNA that are Required for Silencing

To determine where LBR binds to Xist, we UV-crosslinked RNA-protein complexes in cells, digested RNA into short fragments, immunoprecipitated LBR, gel extracted crosslinked RNA-protein complexes, and sequenced the Xist RNA (CLIP (*41–43*), **Methods**). We identified 3 discrete LBR binding sites (LBS) that are spread across >10,000 nucleotides of the Xist RNA (**Figure 4.2a**). These LBR binding sites are distinct from the binding sites of other Xist interacting proteins, including SHARP and PTBP1 (**Figure 4.2a**). Interestingly, one of these LBR binding sites (LBS-1) is present within the ~900 nucleotide region of Xist that was previously shown to be required for Xist-mediated silencing (ΔA-repeat region) (*20*) (**Figure 4.2a**).

We tested LBR binding within a previously generated ΔA-repeat cell line (*20*) and found that LBR binding is lost across the entire Xist RNA (**Figure 4.2b**). Because SHARP also binds within the ΔA-repeat region (*24*, *44*) (**Figure 4.2a)** and its binding is also disrupted in ΔA-Xist (*24*) (**Figure 4.2b**), we generated a mutant Xist that precisely deletes the LBR binding site, but not the SHARP binding site, that overlaps the ΔA-repeat region (ΔLBS, **Figure 4.2a**). In ΔLBS-Xist, LBR binding was lost across the entire Xist RNA without impacting SHARP binding (**Figure 4.2b**). Importantly, ΔLBS-Xist fails to silence X chromosome transcription to a similar level as observed in the ΔA-Xist (**Figure 4.2c**).

To ensure that the observed silencing defect in ΔLBS-Xist cells is due to LBR binding alone and not due to disruption of another unknown protein interaction, we tested whether we could rescue the observed silencing defect by re-establishing the ΔLBS-LBR interaction. To do

this, we generated an endogenous ΔLBS-BoxB Xist RNA (**Methods**) and confirmed that expression of LBR-λN fusion protein, but not LBR fused to a different RNA binding domain (MS2-coat protein(*45*)), was able to rescue the silencing defect observed in ΔLBS-BoxB cells (**Figure 4.2d, Methods**). In contrast, expression of other silencing proteins fused to λN, such as SHARP and EED, did not rescue the observed silencing defect (**Figure 4.2d**). Together, these results demonstrate that the LBR binding site that overlaps the ΔA-repeat region of Xist is required for silencing.

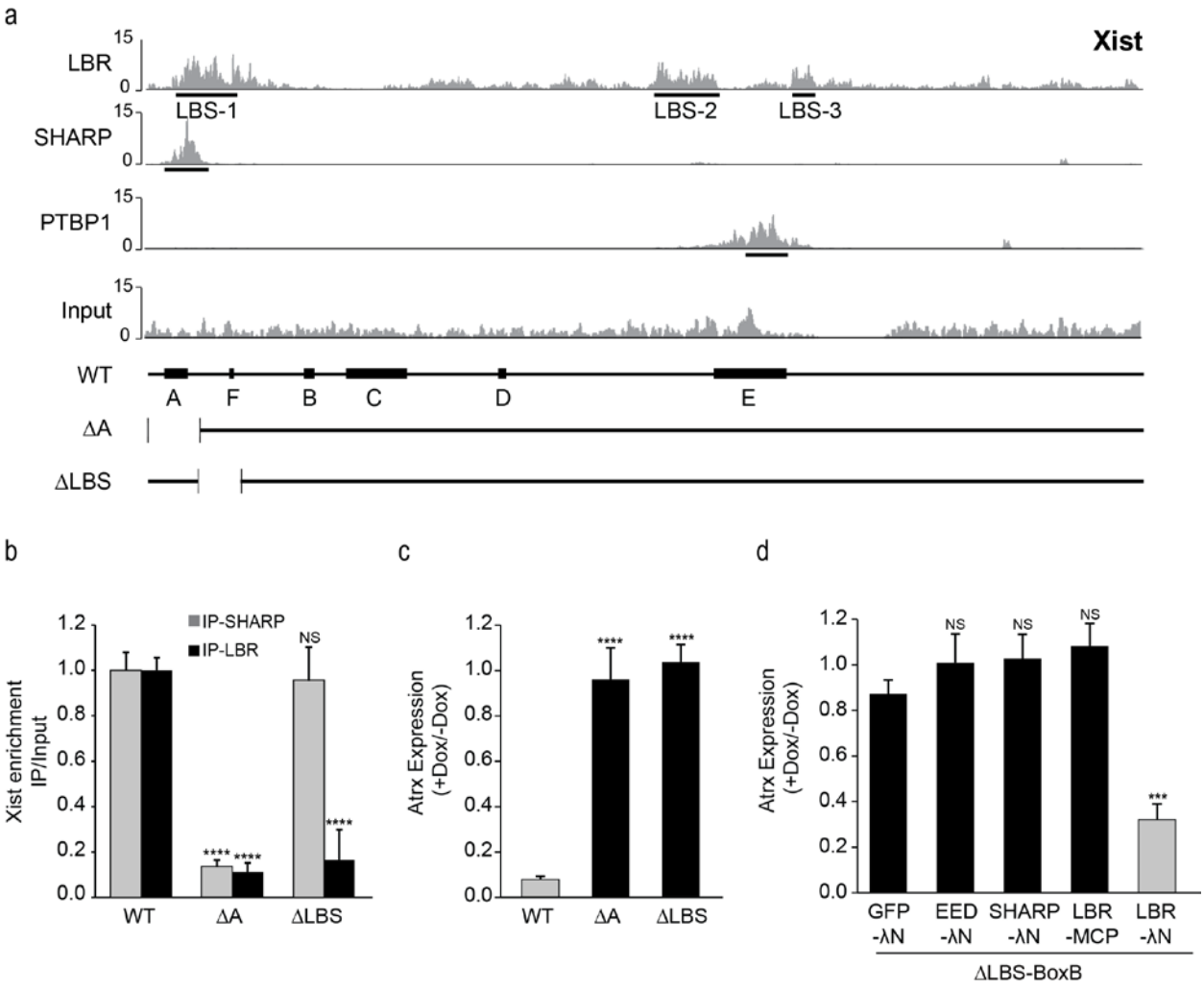

Figure 4.2. LBR binds to precise regions of the Xist RNA that are required for silencing.

(a) CLIP data plotted across the Xist RNA for LBR, SHARP, and PTBP1 proteins. The values represent fold-enrichment at each position on Xist normalized to a size-matched input RNA control. Input represents the total RNA control for the LBR sample. Bottom: A schematic of the annotated repeat regions on the Xist RNA (WT) and the locations of the deleted regions in ΔA (nucleotides 1-937) and ΔLBS (nucleotides 998-1782). (b) Xist RNA enrichment level measured by RT-qPCR after immunoprecipitation of endogenous LBR or SHARP in wild-type, ΔA, or ΔLBS cells. Error bars: SEM from four independent IP

experiments. (c) Relative Atrx mRNA expression in wild-type, ΔA, or ΔLBS-Xist cells. (d) Expression of ΔLBS-Xist with a 3x-BoxB fusion (ΔLBS-BoxB) along with expression of GFP-λN (control), EED-λN, SHARP-λN, or LBR-λN. As an additional control, we expressed LBR fused with the bacteriophage MS2 Coat Protein (LBR-MCP). Error bars: SEM across 50 individual cells. NS: not significant, *** p-value<0.005, **** p-value<0.001 relative to wild type cells (b, c), or cells transfected with GFP-λN (d) by an unpaired two-sample t-test.

4.3 LBR is Required for Recruitment of the Xist-coated Compartment to Nuclear Lamina

Because induction of XCI is known to lead to recruitment of the inactive X chromosome to the nuclear lamina (*11*, *12*, *46*), we hypothesized that the Xist-LBR interaction might be required for mediating these structural changes. To test this, we measured the distance between the Xist-coated nuclear compartment and Lamin B1 in the nucleus using RNA FISH and immunofluorescence (**Figure 4.3a, Methods**). Specifically, we computed the minimum distance between the edge of the Xist compartment and Lamin B1. To account for differences in nuclear size, we normalized the measured distance by the nuclear radius.

Upon Xist induction in wild-type cells, we find that the Xist compartment overlaps Lamin B1 signal in the vast majority of wild-type cells (~88%, **Figure 4.3**). In contrast, in ΔLBS-Xist cells, which disrupt the Xist-LBR interaction, the vast majority of cells (~85%) displayed a clear separation between the Xist-coated compartment and Lamin B1 (**Figure 4.3**), demonstrating a ~17-fold increase in the distance between the Xist compartment and Lamin B1 relative to the distances observed for wild-type Xist. In ΔA-Xist cells, which also ablate the Xist-LBR interaction, we observe a comparable distance distribution between Xist and Lamin B1 to that observed in the ΔLBS-Xist (**Figure 4.3**). These results demonstrate that recruitment of the inactive X chromosome to the nuclear lamina is an active process that is directly mediated by the Xist RNA through a direct interaction with LBR.

Having demonstrated that the Xist-LBR interaction is important for recruitment of the X chromosome to the nuclear lamina, we sought to explore why LBR is important for Xist-

mediated transcriptional silencing. One possibility is that recruitment of the X chromosome to the nuclear lamina, a nuclear territory enriched for silenced DNA and repressive chromatin regulators(*36*), acts to directly silence transcription. Consistent with this notion, recruitment to the lamina has been shown in some cases to be sufficient to silence transcription (*37, 38*). To test this hypothesis, we explored the nuclear lamina association upon knock down of SHARP, which we previously showed is required for Xist-mediated transcriptional silencing (*23, 24, 39, 40*). We found that in the absence of SHARP, the Xist-coated compartment is still localized at the nuclear lamina, demonstrating a comparable distance distribution between Xist and Lamin B1 to that observed for wild-type Xist (**Figure 4.3**). These results demonstrate that Xist-mediated recruitment of the X chromosome to the lamina does not directly lead to transcriptional silencing.

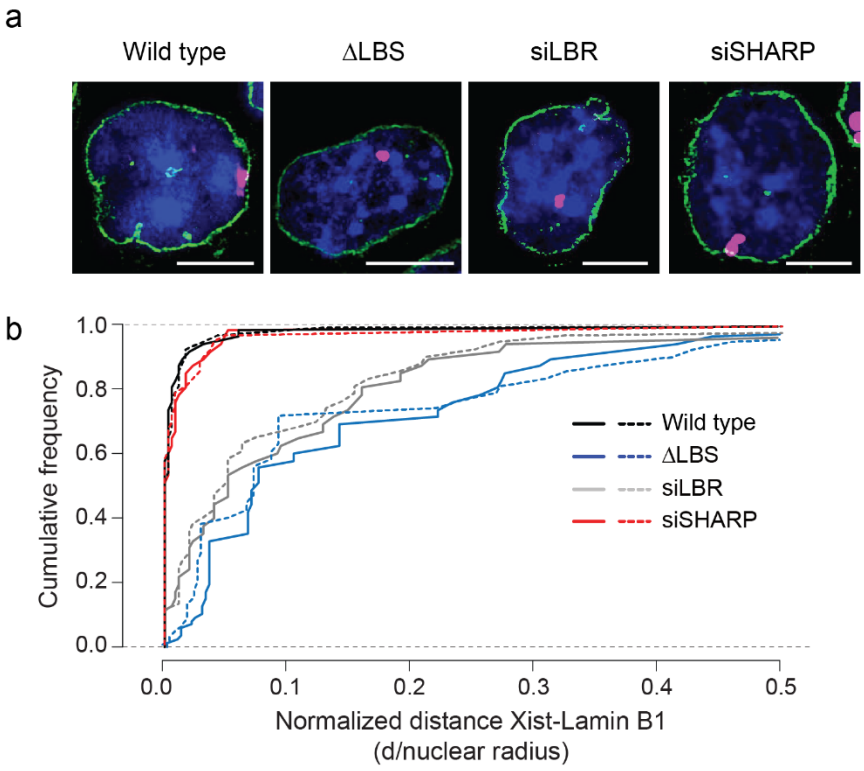

Figure 4.3. Xist-mediated recruitment of DNA to the nuclear lamina is required for transcriptional silencing. (a) Images of Xist (red), Lamin B1 (green) and DAPI (blue) across different conditions. Scale bars: 5 micrometers. (b) The cumulative frequency distribution of normalized distances between Xist and Lamin B1 across 40 individual cells across different conditions. Dashed lines represents a second independent experiment.

4.4 LBR is Required for Xist Spreading to Actively Transcribed Genes across the X-Chromosome

The A-repeat region of Xist was previously shown to lead to an Xist spreading defect such that genes that are actively transcribed prior to initiation of XCI are no longer repositioned into the Xist compartment (*17, 21*). Because LBR binds within the previously deleted A-repeat region of Xist (ΔA), we hypothesized that LBR-mediated recruitment of the X chromosome to the nuclear lamina might act to reposition actively transcribed genes into the Xist-silenced compartment.

To test this hypothesis, we explored the localization of Xist across the X chromosome using RAP-DNA. RAP-DNA enables comprehensive mapping of Xist to all genomic DNA sites – including active and inactive genes as well as intergenic regions. We confirmed that Xist RNA localization is depleted over large regions that contain actively transcribed genes, but not inactive genes, in ΔA-Xist expressing cells(*17*). In ΔLBS-Xist cells or upon knock down of LBR, we observed a comparable Xist localization defect as in ΔA-Xist, with strong depletion of Xist observed across all actively transcribed genes(*17*) (**Figure 4.4a, b**). To ensure that this result is specific to LBR, we knocked down SHARP, a protein that is also required for silencing and that also interact with the A-repeat region of Xist. Upon knock down of SHARP, we did not observe an Xist RNA localization defect; instead Xist localization was comparable to that observed in wild-type conditions (**Figure 4.4b**).

To determine whether this spreading defect is due to a failure to reposition actively transcribed genes into the Xist-coated compartment, we measured the position of active

genes relative to the Xist coated compartment using RNA FISH (**Figure 4.4c**, **Methods**). We labeled the genomic locus of an actively transcribed X chromosome gene (Gpc4) using probes against its intronic region and calculated the distance to the Xist compartment (**Methods**). As a control, we measured the distance between the genomic locus of an autosomal gene (Notch2), which should always be excluded from the Xist compartment. In ΔA cells, the distance between the Xist compartment and the Gpc4, Mecp2, or Pgk1 loci was comparable to the distance between Xist and the autosomal locus (**Figure 4.4d, e**). Similarly, in ΔLBS cells or upon knock down of LBR, we observed a comparably large distance between the Gpc4, Mecp2, or Pgk1 loci and the Xist compartment (**Figure 4.4d, e**). In contrast, upon knockdown of SHARP, we found that the Gpc4 locus overlapped with the Xist compartment in the vast majority (~80%) of all cells (**Figure 4.4d, e**) displaying a comparable frequency of overlap to that observed for the Xist genomic locus itself.

Together, these results demonstrate that the Xist-LBR interaction is required for Xist spreading to actively transcribed DNA loci. Because Xist can still spread to active genes upon knock down of SHARP, this argues that spreading to active genes occurs independently of RNA PolII exclusion and that both roles – spreading and RNA PolII exclusion – are required for silencing. As both roles are mediated by the A-repeat, this would explain why the A-repeat mutant fails to display an RNA PolII exclusion defect, as observed upon knock down of SHARP, because in the absence of Xist spreading to active genes, SHARP cannot localize to and exclude PolII from these active genes.

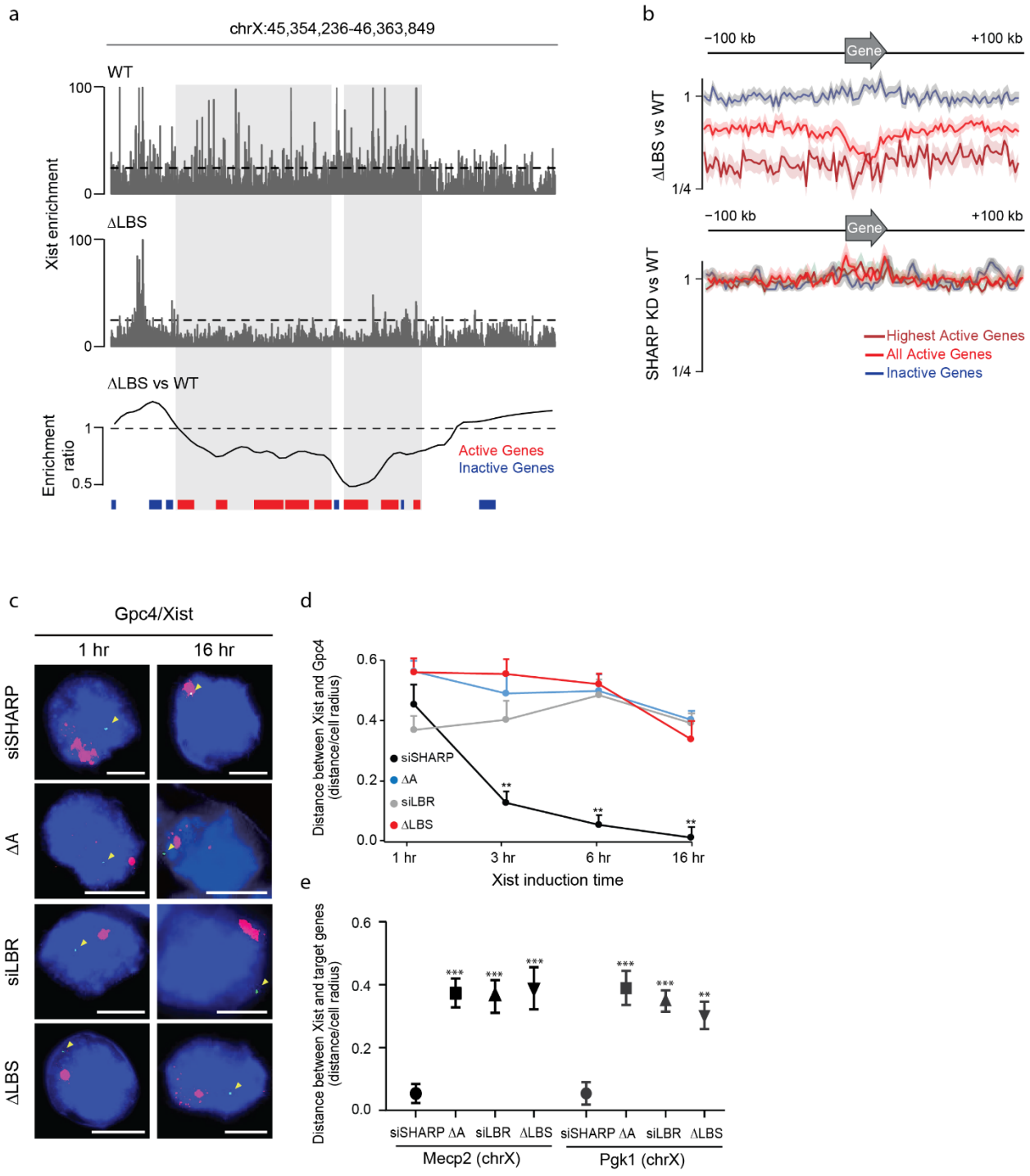

Fig. 4.4. LBR is Required for Xist Spreading to Actively Transcribed Genes across the X-Chromosome. (a) Xist RNA localization as measured by RAP-DNA for wild type (top),

ΔLBS-Xist (middle), and the smoothed fold change (bottom) across a region of the X chromosome containing active (red) and inactive (blue) genes. Dashed line: average Xist enrichment in wild type cells. (b) Aggregate Xist enrichment relative to the genomic locations of highly active genes (dark red, RPKM > 5), all active genes (red, RPKM >1), and inactive genes (blue) on the X-chromosome for ΔLBS and SHARP knockdown cells compared to wild type cells. Shaded areas represent 95% confidence interval. (c) Images of Xist (red), Gpc4 locus (green) and DAPI (blue) across different cell lines (rows) after Xist induction for 1 or 16 hours. Scale bars: 5 micrometers. (d) The median distance from Gpc4 locus to the Xist-compartment after Xist induction for 1, 3, 6, and 16 hours. Error bars represent the standard error of the median across 50 individual cells. ** p-value<0.01 relative to 1-hour induction by an unpaired two-sample t-test. (e) The median distance from the Mecp2 and Pgk1 locus to the Xist compartment after Xist induction for 16 hours across different conditions. Error bars represent the standard error of the median across 50 individual cells. ** p-value<0.01, *** p-value<0.005 relative to siSHARP by an unpaired two-sample t-test. (F) A model for how Xist-mediated recruitment to the nuclear lamina enables spreading to active genes and transcriptional silencing on the X chromosome.

4.5 Recruitment to the Nuclear Lamina Enables Xist Spreading and Silencing of Active Genes

Having established that LBR is required for Xist-mediated transcriptional silencing, recruitment of the Xist-coated compartment to the nuclear lamina, and for Xist RNA spreading to actively transcribed genes, we hypothesized that Xist-mediated recruitment of DNA to the nuclear lamina enables Xist to spread to active genes and silence transcription. We reasoned that if we recruit Xist to the nuclear lamina, independent of LBR, then Xist will spread to active genes and rescue Xist-mediated transcriptional silencing **(Figure 4.5a)**. To test this, we used our ΔLBS-BoxB Xist, which fails to interact with LBR, to create an interaction between Xist and other components of the nuclear lamina that on their own do not impact X chromosome silencing in normal conditions. Specifically, we created a fusion between Lamin B1 and λN and confirmed that expression of this fusion protein in ΔLBS-BoxB Xist cells led to recruitment of the Xist-compartment to the nuclear lamina using RNA FISH and immunofluorescence of Lamin B1 (**Figure 4.5b-d**).

Having synthetically recruited Xist to the nuclear lamina, we tested whether Xist can spread to actively transcribed genes using RAP-DNA. Indeed, tethering Xist to the nuclear lamina is sufficient to enable Xist to spread to active genes, with Xist localizing to active genes and inactive genes at comparable levels, similar to that observed in wild-type conditions (**Figure 4.5e**). Consistent with a functional requirement for Xist spreading across active genes to enable X chromosome silencing, we find that tethering Xist to the nuclear lamina also rescues the Xist silencing defect observed in ΔLBS cells to the same extent as that observed after rescuing directly with LBR-λN (**Figure 4.5f**).

Together, these results demonstrate that Xist-mediated recruitment of the X chromosome to the nuclear lamina leads to Xist spreading to active genes, and through this enables Xist-mediated transcriptional silencing.

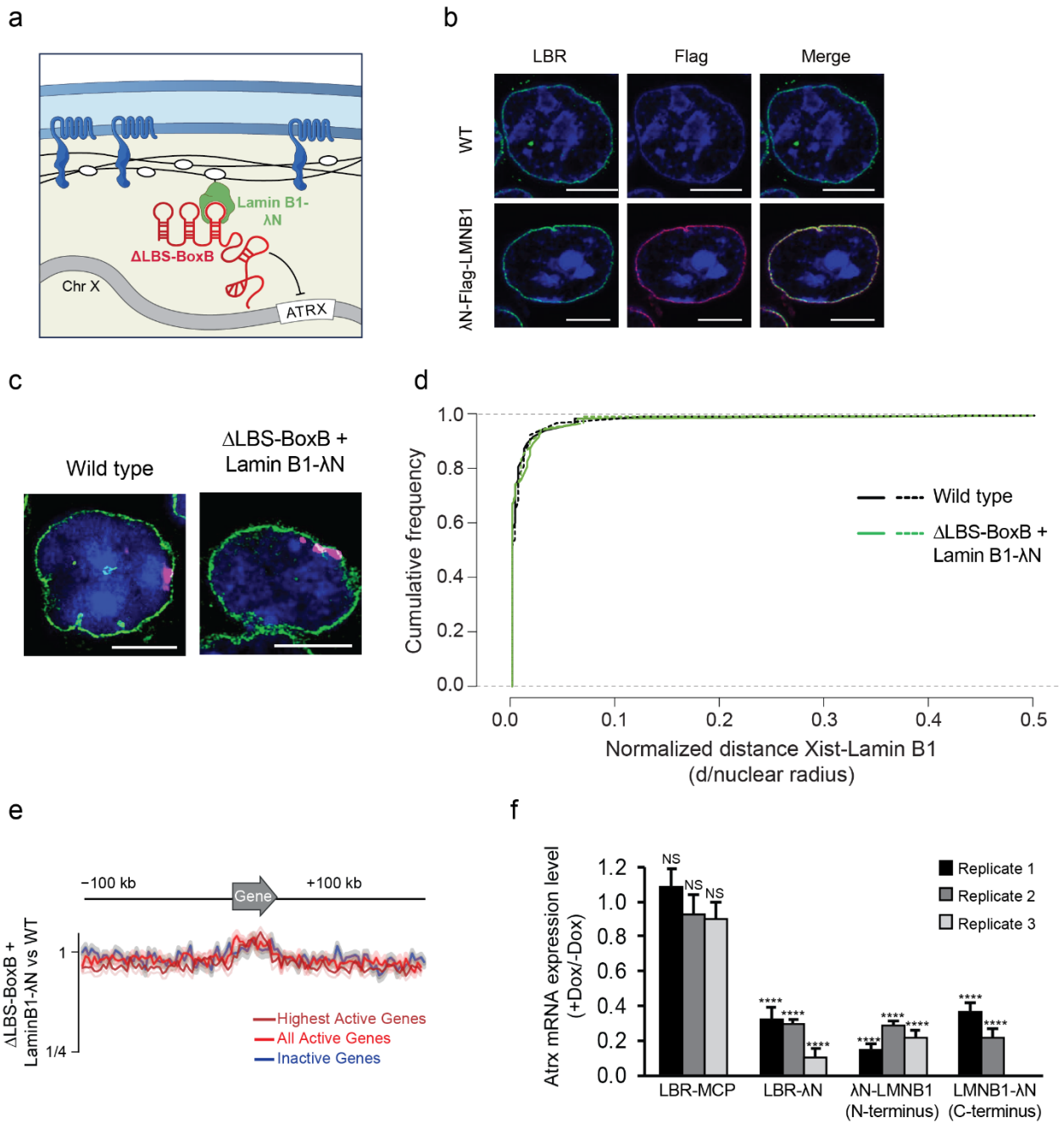

Figure 4.5. Recruitment to the nuclear lamina enables Xist spreading and silencing of active genes. (a) A schematic illustrating the tethering of ΔLBS-BoxB to the nuclear lamina using the LaminB1-λN fusion protein. (b) Images of individual cell stained with endogenous LBR (green) and FLAG-tag (red) in cells transfected with empty vector (WT) or λN-

3xFLAG tagged LMNB1. The λN-3xFLAG tagged LMNB1 localizes properly in the nuclear envelope and colocalizes with endogenous LBR. Scale bars: 5 micrometers. (c) Images of Xist (red), Lamin B1 (green) and DAPI (blue) across different conditions. Scale bars: 5 micrometers. (d) The cumulative frequency distribution of normalized distances between Xist and Lamin B1 across 40 individual cells across different conditions. Dashed lines represent a second independent experiment. (e) Aggregate Xist enrichment relative to the genomic locations of highly active genes (dark red, RPKM > 5), all active genes (red, RPKM >1), and inactive genes (blue) on the X-chromosome ΔLBS-BoxB + LMNB1-λN cells compared to wild type cells. Shaded areas represent 95% confidence interval. (f) Three independent replicates (two for LMNB1-λN) showing the copy number of Atrx mRNA molecules after induction of Xist normalized to the number of the mRNA molecules prior to Xist induction in cells expressing ΔLBS-BoxB Xist transfecting with LBR-MCP, LBR-λN, λN-LMNB1 or LMNB1-λN fusion protein. Error bars represent the standard error across 50 individual cells. NS: not significantly different from 1; **** represents values significantly different from 1 with a p-value<0.0001 based on a one-sample t-test.

Methods

Only the methods that have not yet been described in previous Chapters are listed

LBR knockout female ES cells

We generated an LBR knockout in female ES cells using the CRISPR-Cas9 system. Specifically, we co-transfected a construct expressing Cas9 driven by a pCAG promoter and a pool of sgRNAs targeting an upstream and downstream region of LBR locus to delete the entire genomic locus (upstream sgRNA sequences: GGCGATGATTCAAAAGGTCG, AGCGCCGGCGATGATTCAAA, GGGCTCCGGCCTGGGCCTGC, TGAAATAAGAGAATGTTATA; downstream sgRNA sequences: TTTAACCTGTTTTTAGGTCT, AGGCTGTCTGGTCAGAATCC, CGAAGAAACCTCCCAGTCAC, CATTTTTGGTTTATTCATGG). We then picked single colonies from transfected cells and verified LBR knockout using PCR and Sanger sequenced successful homozygous knockout lines with primers flanking deletion sites. We confirmed that this deletion leads to a complete loss of function of LBR using RNA FISH and immunofluorescence of the protein.

Integrating the BoxB aptamer sequence into Xist

We knocked in 3 copies of the viral BoxB RNA aptamer, which binds tightly to the viral λN coat protein *(38–40),* into nucleotide 16,523 of the endogenous Xist RNA in *pSM33* ES cell line using CRISPR-mediated homologous recombination. Specifically, we co-transfected a construct expressing Cas9 driven by a pCAG promoter, a sgRNA targeting

the 3’ region of the Xist locus (sgRNA sequence: CCTCATCCTCATGTCTTCTC), and a single strand DNA ultramer (IDT) containing a 3x-BoxB sequence (GGGCCCTGAAGAAGGGCCCATGGGCCCTGAAGAAGGGCCCATAGGGCCCTGAAGAAGGGCCC; underlined nucleotides denote the BoxB sequence) flanked with 70 nucleotides of upstream and downstream homologous sequence of the insertion site. We then picked single colonies from transfected cells and verified BoxB integration using PCR and Sanger sequenced successful integration lines with primers flanking the integration site and confirmed correct insertion. We ensured that the Xist-BoxB was still able to silence the X chromosome by expressing it and measuring transcriptional silencing of Atrx.

UV crosslinking

Cells were washed once with PBS and then crosslinked on ice using 0.4 Joules/cm^2 (UV4k) 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 1500 × g for 4 minutes, and flash frozen in liquid nitrogen for storage at – 80 ˚C.

Immunoprecipitation and RT-qPCR

Mouse ES cells were induced then crosslinked with UV4k as described above. Pellets of 20M cells were lysed and treated with TURBO DNase (Ambion) and incubated for 10 minutes at 37 ˚C in an Eppendorf Thermomixer C to digest genomic DNA. The lysate was pre-cleared by incubation with 180 μL of Dynabeads Protein G magnetic beads (Life Technologies). Meanwhile, 10 μg of antibody for immunoprecipitation was coupled to 75 μl Protein G magnetic beads. After pre-clearing was completed, the lysate was then mixed with

the appropriate antibody-coupled Protein G magnetic beads and incubated overnight at 4 ˚C on a Hulamixer sample mixer (Life Technologies) for protein capture. After immunoprecipitation, beads were washed with a wash buffer of 1× PBS with detergents and then captured nucleic acids were eluted by digesting all proteins with 5.6 U proteinase K (New England Biolabs). Eluted RNA was purified using the RNA Clean and Concentrator-5 Kit (Zymo Research) and RT-qPCR was performed as described previously (*7*) to evaluate RNA enrichment. The antibodies used for immunoprecipitation were anti-FLAG® M2 (Sigma-Aldrich; F1804) (for ΔTM- and ΔRS-LBR transfected cells), anti-SHARP (Bethyl; A301-119A), and customized LBR antibody from GenScript (LBR #4; 540774-1).

Crosslinking and Immunoprecipitation (CLIP) analysis

We crosslinked 6 hour doxycycline-induced pSM33 mouse male ES cells with 0.4 J/cm^2 of UV254. Cells were lysed and RNA was digested with RNase I to achieve a size range of 100-500 nucleotides in length. Lysate preparations were precleared by mixing with Protein G beads for 1 hr at 4C. For each CLIP sample, target proteins were immunoprecipitated from 20 million cells with 10 ug of antibody and 75 ul of Protein G beads. The antibodies were pre-coupled to the beads for 1 hr at room temperature with mixing before incubating the precleared lysate to the beads-antibody overnight at 4C. After the immunoprecipitation, the beads were washed four times with High salt wash buffer (50 mM Tris-HCl 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). RNAs were then eluted with NLS elution buffer (20 mM Tris-HCl pH 7.5, 10 mM EDTA, 2% N-lauroylsacrosine, 2.5 mM TCEP) with 100 mM DTT. Samples were then run through a

standard SDS-PAGE gel and transferred to a nitrocellulose membrane, and a region 75 kDa above the molecular size of the protein of interest was isolated and treated with Proteinase K (NEB) followed by phenol/chloroform/isoamyl alcohol (pH 6.5) extraction to isolate the RNAs. Extracted RNAs were then purified with RNA Clean & Concentrator™-5 (Zymo). After a dephosphorylation treatment, the RNA in each sample was ligated to a mixture of barcoded adapters in which each adapter had a unique barcode identifier according to our Massively Multiplexed RNA Sequencing method (*25*). After ligation, beads were rinsed with 1x PBS and detergents and then 5x PBS and detergents prior to pooling 3-4 IPs per new tube. The proteins and RNA were then eluted from the Protein G beads with 6M urea and 40 mM DTT at 60C. Protein-RNA complexes were separated away from free RNA and the proteins were then digested with Proteinase K. From the barcoded RNA in each pool, we generated Illumina sequencing libraries as previously described(*26*).

Input samples: As a control, we sequenced an "input" RNA control for each immunoprecipitated protein. To do this, we saved 10% of the total cellular lysate prior to the immunoprecipitation step. These samples were then run through an SDS-PAGE gel alongside the immunoprecipitated sample and gel extracted from the identical region as the protein analyzed. We then made sequencing libraries from these samples as described above.

Analysis of CLIP Data

We computed and visualized the enrichment for any RNA region by normalizing the number of reads upon immunoprecipitation with a specific protein relative to the number of reads in its size-matched input control (input sample). Specifically, we counted the total number of

reads overlapping the RNA region in either the immunoprecipitation sample or the input control. To account for differences in read coverage between samples, each of these numbers was normalized to the total number of reads within the same experiment. This generates a normalized score, per region, within each sample. We then computed an enrichment metric by taking the ratio of these normalized values (IP/Input).

We identified protein binding sites on the Xist RNA by identifying regions that were enriched relative to the same region in the input control ("differential enrichment") and also was enriched relative to all other regions on the remainder of the Xist RNA ("local enrichment"). The differential enrichment accounts for biases in the size-selected input sample that would lead to a pile up of reads in specific regions of the RNA, but that do not reflect true protein binding sites. In contrast, the local enrichment accounts for cases where a given RNA might have higher overall levels of protein binding relative to the input. To compute significant enrichment, we computed the differential enrichment as defined above (IP/Input) for each window (window size=100 nucleotides). We computed the local enrichment for each region by taking the normalized number of reads for each region (IP) and dividing it by the normalized number of reads over the entire Xist RNA. To make these rates comparable, we divided each number by their respective region length prior to taking the ratio. We then generated 1,000 random permutations of the reads in the IP samples and paired input samples across the Xist RNA. For each permutation, we computed the differential and local enrichments and generated an empirical distribution of the maximum value observed for each permutation. We assigned a multiple-testing corrected p-value to each region by comparing

the observed differential and local enrichment values to these permutation distributions. We identified significant windows that had a differential p-value <0.01 and a local p-value<0.01.

We identified three distinct LBR binding sites from 535-1608 nucleotides (LBS-1), 9506-10245 nucleotides (LBS-2), and 11732-11956 nucleotides (LBS-3). We also identified a SHARP binding site from 317-1056 nucleotides and PTBP1 binding site from 10859-11344 nucleotides on Xist.

Generating ΔLBS and ΔLBS-BoxB Xist

We generated ΔLBS and ΔLBS-BoxB using CRISPR-mediated knock out. To generate ΔLBS and ΔLBS-BoxB cells, mouse *pSM33* ES cells and Xist-BoxB cells were transfected with two sgRNAs flanking the LBS-1 region of Xist (sgRNA sequence: CACCGAGGAGCACAGCGGAC and TAAGGACGTGAGTTTCGCTT) and co-transfected along with the Cas9 construct described above to create a deletion of LBS-1 by non-homologous end joining. We isolated single colonies from the transfected cells for both cell lines and verified that the LBS-1 region was deleted from the genome using PCR and Sanger sequencing with primers flanking the A-repeat region of Xist. Inversed PCR using the primer inside of deleted region and Sanger sequencing showed that the deleted region had been inserted in Chr12. We confirmed that the ΔLBS Xist transcript lacked of the deleted region and the deleted region was not expressed from its inserted locus by RNA-seq. We ensured that the ΔLBS affected binding of the LBR protein using IP-qPCR and CLIP sequencing across the entire Xist RNA. We also ensured that there was no impact on SHARP

binding using IP-qPCR. The ΔLBS RAP experiment is done in ΔLBS-BoxB cells as a better control for LMNB1-λN rescuing experiment.

dCas9-KRAB silencing

To generate stable LBR and SHARP knock down cells, we co-transfected a puromycin resistant construct expressing dCas9-KRAB driven by an Ef1a promoter and a guide RNA with scaffolding structure targeting the region near the transcription start site of LBR (sgRNA sequence: CGGGACTCCGCCGCGTG) or SHARP (sgRNA sequence: CGGTGGCGTCGGCAGCGG). Transfected cells were selected on 1 μg/mL puromycin (Sigma-Aldrich) for four days to enrich for cells that contain the dCas9-KRAB. We used FISH to verify that >90% of these puro-resistant cells had no detectable amount of mRNA after four days of puromycin selection.

LBR Protein mutagenesis

A human cDNA containing the full-length ORF of LBR was obtained from the DNASU plasmid repository as a Gateway entry clone and inserted into the pCAG-GW-λN-3xFLAG-BSD vector using an LR recombination reaction (Invitrogen). To generate ΔRS-LBR and ΔTM-LBR, λN-3xFLAG tagged full-length LBR construct was truncated using PCR-mediated deletion with primers flanking the deletion region.

Expression of cDNA rescue constructs

Mouse ES cells were electroporated using the Neon transfection system (Invitrogen) with mammalian expression vector (pCAG-GW-λN-3xFLAG-BSD vector) expressing human

ΔRS-LBR, ΔTM-LBR, or full-length LBR construct from above. We knocked down endogenous LBR by treating cells with siRNAs pool from Dharmacon (ON-TARGETplus SMARTpool siRNAs) targeting only mouse LBR, but not human LBR. We ensured that the siRNAs targeted the mouse LBR specifically by ensuring that the human full-length LBR construct could rescue cells with knock down of endogenous LBR.

Generation of λN-3xFLAG epitope tagged proteins

For λN-3xFLAG-tagged protein expression and immunoprecipitation, mouse ES cells were electroporated using the Neon transfection system (Invitrogen) with mammalian expression vector (pCAG-GW-λN-3xFLAG-BSD) encoding expression of a C-terminal λN-3xFLAG tagged ORF driven by CAG. Human ORFs of GFP, LBR, EED1 and LMNB1 were obtained from the DNASU plasmid repository as Gateway entry clones and inserted into pCAG-GW-λN-3xFLAG-BSD vector using an LR recombination reaction (Invitrogen). For λN-3xFLAG-LMNB1, the LMNB1 ORF was inserted into the vector described above but with a N-terminal λN-3xFLAG tag instead (pCAG-λN-3xFLAG-GW-BSD) using an LR recombination reaction (Invitrogen). Mouse SHARP ORF was obtained by RT-PCR from *pSM33* total RNA using SHARP specific primers. The SHARP ORF was then cloned into a pENTR™/D-TOPO Gateway entry clone (Invitrogen) and inserted into the pCAG-GW-λN-3xFLAG-BSD vector as described above. Transfected cells were selected on 4 ug/mL Blasticidin (InvivoGen) to enrich for cells expressing tagged proteins. For LBR-MCP, the LBR ORF was inserted into Ef1a-GW-MCP-V5-Neo vector using an LR recombination reaction (Invitrogen) and selected with 200 ug/mL Geneticin/G418 (Invitrogen). For

analysis, we used immunofluorescence staining with antibodies against 3xFLAG or V5 epitope (described below) to select for cells expressing tagged proteins.

We verified that λN-3xFLAG tagged proteins were still functional by ensuring that they could rescue knock down of the endogenous protein.

Western blotting

pSM33 cells were lysed in buffer containing 50 mM Tris-HCl (pH 7.5), 100 mM NaCl, 1% NP-40, 0.5% sodium deoxycholate and protease inhibitor cocktail (CalBiochem; 539134) and sonicated using a Branson Sonifier at 25 watts for 20 seconds (0.7 seconds on, 3.3 seconds off) on ice. ~30 μg of total protein was separated by SDS–PAGE and transferred to nitrocellulose membranes followed by blocking with Odyssey Blocking Buffer (Licor, 927-40000). Primary antibodies were diluted in blocking buffer as follows: anti-FLAG® M2 (Sigma-Aldrich; F1804) (1:1000), anti-lamin B receptor antibody (Abcam; ab122919) (1:1500), anti-V5 tag antibody (Abcam; ab27671) (1:1000) and anti-actin antibody (Abcam; ab3280) (1:1500). Secondary antibodies were diluted in 0.1% Tween-20 diluted in PBS as following: IRDye® 680RD Goat anti-Mouse IgG (H + L) (LI-COR; 926-68070) (1:10000) and IRDye® 800CW Goat anti-Rabbit IgG (H + L) (LI-COR; 926-32611) (1:10000). Blots were imaged with Odyssey® CLx Imager (LI-COR Biosciences) and the intensity of each band was quantified using ImageJ.

Single molecule RNA FISH

Single molecule RNA Fluorescence *in situ* hybridization (FISH) experiments were done using QuantiGene ViewRNA ISH Cell Assay (Affymetrix) and QuantiGene ViewRNA ISH Cell 740 Module (Affymetrix) according to manufacturer's protocol. Specifically, cells fixed on coverslips were first permeabilized with Detergent Solution QC at room temperature for 5 min, and then incubated with desired mixture of probe set (Affymetrix) in Probe Set Diluent QF at 40°C for 3 h, followed by incubation with PreAmplifier Mix at 40°C for 30 min, Amplifier Mix at 40°C for 30 min, and Label Probe Mix at 40°C for 30 min sequentially. For DAPI staining, coverslips were incubated in 30 nM DAPI in PBS at room temperature for 15-20 min. Probe sets and conjugated fluorophores (excitation wavelengths) for FISH were TYPE 1-XIST (550 nm), TYPE 4-GPC4 (488 nm), TYPE 10-ATRX (740 nm), and TYPE 6-SHARP, LBR, LMNB1, EMD (650 nm).

Immunofluorescence and RNA FISH

For immunofluorescence (IF), cells were fixed on coverslips and permeabilized with 0.1% Triton-X in PBS at room temperature for 10 min, and blocked with 1X blocking buffer (Abcam; ab126587) or 5% normal goat serum in 0.1% Triton-X in PBS at room temperature for 30 min. Cells were then incubated with primary antibodies at room temperature for 1 h, followed by washes with 0.1% Triton-X in PBS and incubation with secondary antibodies at room temperature for 1 h. The samples were then processed using the RNA FISH protocol, as described above. Primary antibodies and the dilution used for IF were anti-Lamin B1 (Abcam; ab16048) (1:50), and anti-FLAG® M2 (Sigma-Aldrich; F3165) (1:50), and anti-Lamin B Receptor antibody (Abcam; ab122919) (1:100). Secondary antibodies and the dilution used for IF were Alexa Fluor® 488 F(ab')$_2$ fragment of goat anti-rabbit IgG (H+L)

(Life Technology; 1618692) (1:100), highly x-ads DyLight® 650 goat anti-Rabbit IgG (H&L) (Bethyl; A120-201D5) (1:200), DyLight® 650 goat anti-Mouse IgG (H&L) (Bethyl; A120-201D3) (1:200), DyLight® 550 goat anti-Rabbit IgG (H&L) (Bethyl; A90-516D5) (1:200) and DyLight® 650 goat anti-Mouse IgG (H&L) (Bethyl; A90-516D3) (1:200).

Microscopic Imaging

FISH, IF/FISH and X-chromosome paint samples were imaged using a Leica DMI 6000 Deconvolution Microscope with the Leica HC PL APO 63x/1.30 GLYC CORR CS2 objective. Samples stained with TYPE 10-ATRX (740 nm) were imaged using Nikon Ti Eclipse with the Nikon CFI Plan Apochromat λ DM 60x/1.40 oil objective. Images were projected with maximum projection (3 μm; step size, 0.2 μm). Samples for 3D deconvolution was imaged using Leica DMI 6000 Deconvolution Microscope with the Leica HC PL APO 63x/1.30 GLYC CORR CS2 objective (15 μm; 0.02 μm step size). 3D deconvolution images were processed using Huygens Professional (SVI; v15.05) with the built-in theoretical point spread function and the classic maximum likelihood estimation method for restoration. A manually defined signal to noise ratio value was applied for each fluorescent channel respectively. Samples stained with LMNB1 or LBR were imaged using a Zeiss LSM 880 Laser scanning confocal system with the Airyscan super-resolution module, mounted on an upright Axio Examiner Z1 microscope. We used a Plan-Apochromat 63x/1.40 NA Oil DIC f/ELYRA objective and the Airyscan module to collect super-resolution images. Single Z-section was used for these images.

X-chromosome Silencing Assay

Cells were stained for Xist RNA, Gpc4 mRNA, Atrx mRNA and siRNA-targeted mRNA by FISH and imaged. Images were then analyzed using Matlab R2013b (described below). Cells were selected if the copy number of the targeted mRNA was less than 30% of the level of the no siRNA treated cells and if they induced Xist expression. Within these cells, the copy number of Gpc4 mRNA and Atrx mRNA were quantified using a peak finding method (described below) and compared across conditions. We quantified mRNA levels for 50 individual cells. We also evaluated Xist expression in siRNA-treated cells, and observed no difference in the percentage of cells that induced Xist expression in any of the siRNA conditions relative to untreated cells.

The mean and the variance of the ratio (+Dox/-Dox) were calculated using the standard Taylor approximation method for estimating the significance of ratios. Accordingly, we calculated the average, standard deviation, and standard error of the mean as follows.

The average (μ) is defined as:

$$\mu\left(\frac{+Dox}{-Dox}\right) = \frac{\mu(+Dox)}{\mu(-Dox)} + \frac{\mu(+Dox)}{\mu(-Dox)^3}\sigma^2(-Dox)$$

The standard deviation (σ^2) is defined as:

$$\sigma^2\left(\frac{+Dox}{-Dox}\right) = \frac{\sigma^2(+Dox)}{\mu(-Dox)^2} + \frac{\mu(+Dox)^2}{\mu(-Dox)^4}\sigma^2(-Dox)$$

and the standard error of the mean is defined as:

$$SEM\left(\frac{+Dox}{-Dox}\right) = \sqrt[2]{\frac{\sigma^2\left(\frac{+Dox}{-Dox}\right)}{50}}$$

Quantifying mRNAs by single molecule FISH

All image analysis was carried out using Matlab (version R2013b) utilizing built-in functions from the Image Processing toolbox. Images were first filtered using a two-dimensional median filter to remove background. Cell boundaries were outlined manually, guided by DAPI staining, to create a binary mask and applied to the various channels from the same field of view. Top-hat morphological filtering, a background subtraction method that enhances the individual focal spots, was applied to the images (*27*). The spots were then identified using a 2D peak finding algorithm that identifies local maximal signals within the cell. Once regional maxima were identified, the number of spots was counted for each cell. For better visualization of spots of mRNAs, we enhanced the spot size of the images using Fiji (ImageJ v1.51d) Maximum Filter plugin with radius of 1.0 pixel for Gpc4 and/or Atrx channel.

Calculating distance between Xist RNA compartment and Lamin B1

The nuclei of individual cells were identified manually using the DAPI staining. We identified the Xist compartment by either staining for Xist RNA (FISH) or X chromosome DNA (chromosome paint) along with immunofluorescence for Lamin B1 protein. We defined the compartment by identifying a region in the nucleus using an intensity-based threshold to partition the image within the nucleus and find contiguous 2-dimensional regions of high intensity. The threshold was determined based on Otsu method as previously

described (*28*), which splits the image into 2 bins – high and low – and identifies a threshold that minimizes the variance within the partition. This creates a binary mask on the image. We visually confirmed that this binary mask accurately reflected the Xist compartment, X chromosome, or Lamin B1 region. The distance between the Xist compartment and Lamin B1 was determined by calculating the distance of each pixel between Xist or the X chromosome and Lamin B1 and finding the minimum value with a customized Fiji macro script. The area of the nucleus (Area) was measured using Fiji, and the radius of the nucleus (r) was calculated using $r = \sqrt{Area/\pi}$. We set the distance as zero if the Lamin B1 fluorescence signal overlapped with the fluorescence signal detect for Xist or the X chromosome (respectively).

Calculating distance between the Xist compartment and genomic loci

The nuclear area and Xist compartment were identified using the method described above. Genomic loci were determined by RNA FISH with probes against the intronic region of the genes using smFISH (*29*) as described above (TYPE 4-GPC4 (Intron1) (488 nm)). We then identified the spot with Analyze Particle function in Fiji and selecting the spot with highest fluorescent intensity within the nucleus. We discarded the small number of images that contained more than one spot. For the GPC4 locus in male cells, the distance between the Xist compartment and the locus was determined by finding the minimum distance between Xist compartment and the locus with a customized Fiji macro script described above. For GPC4 in female ES cell, the distance between the Xist compartment and the loci was determined by calculating the minimum distance between Xist compartment and either one

of the two loci. Median values and the standard error of the median were plotted in the figures. We calculated the standard error of the median as standard error of the median = 1.2533(standard error of the mean)

We identified loci as inside the Xist compartment if the fluorescence signal of the locus overlapped with these fluorescence signal for Xist (for XIST and male ES cell GPC4).

RNA antisense purification (RAP) coupled with DNA sequencing

10 million mouse ES cells were induced with doxycycline for 6 , lysates were prepared, and Xist RNA was captured and purified as previously described (*7*). For Xist RNA capture, we used antisense 5' biotinylated 90-mer DNA oligonucleotides (Eurofins Operon) that spanned the entire length of the Xist RNA as previously described (*11*). To elute captured DNA, we incubated the beads with 15 U RNase H in 20 uL RNase H buffer (NEB Biolabs) at 37°C for 1 hour. The RNase H digested samples were then transferred to a new tube. To reverse crosslinks, we added 25 uL Hybridization Buffer (20 mM Tris-HCl (pH 7.5), 7 mM EDTA, 3 mM EGTA, 150 mM LiCl, 1% NP-40, 0.2% N-lauroylsarcosine, 0.125% Na-Deoxycholate, 3M Guanidinium Thiocyanate, 2.5mM TCEP), 125 uL NLS Elution Buffer (20 mM Tris-HCl (pH 7.5), 10 mM EDTA, 2% N-lauroylsarcosine, 2.5mM TCEP), 500 mM NaCl and 4 U Protease K (NEB Biolabs, Molecular Biology Grade) and incubated at 60°C overnight. Eluted DNA was sequenced, aligned and analyzed as previously described (*7*, *26*).

Aggregate gene analysis

We calculated the metaplot by first scaling the number of reads in each sample to obtain the same total number of reads for all the samples in 1 kb windows. We then normalized each sample to its own input followed by a second normalization to the wild-type Xist sample to obtain the relative ratio of each window. We then plotted the log-ratio of these values 100 Kb upstream and downstream of each gene on X-chromosome along with the gene body region, which was scaled across genes to represent the same overall area. To avoid overcounting, when we extended a given gene, we only included those extensions in our aggregation set if they were not already included in the left or right extensions from a previous gene. Genes within 5 Mb of the Xist transcription locus were excluded from the analysis because they represent outliers in terms of average Xist enrichment. The plots were generated and visualized using DeepTools and Gviz. The “active” and “inactive” genes were defined as previously described (*7*). Expression levels were split based on RPKM levels computed from chromatin RNA-Seq levels as previously described (*26*). We only considered genes with RPKM expression >1. Genes with RPKM expression >5 are grouped as highly actively transcribed genes.

The regional normalization curve was obtained by calculating a smoothed running average across a 10-kilobase window on the chromosome. Accordingly, it was included to demonstrate the overall pattern of the data, which can be more easily seen in a smoothed aggregate representation relative to overlay of each individual data point. Each number in the plot (Wi) was calculated using the simple mean of the 1-kilobase pair windows shown in the top panel, such that:

$$W_i = \frac{W_i + W_{i+1} + \cdots + W_{i+9}}{10}$$

References

1. J. H. Gibcus, J. Dekker, The hierarchy of the 3D genome. *Mol Cell* **49**, 773-782 (2013).
2. W. A. Bickmore, B. van Steensel, Genome architecture: domain organization of interphase chromosomes. *Cell* **152**, 1270-1284 (2013).
3. A. Pombo, N. Dillon, Three-dimensional genome architecture: players and mechanisms.
4. M. R. Hubner, M. A. Eckersley-Maslin, D. L. Spector, Chromatin organization and transcriptional regulation. *Curr Opin Genet Dev* **23**, 89-95 (2013).
5. W. A. Bickmore, The spatial organization of the human genome.
6. S. Augui, E. Nora Ep Fau - Heard, E. Heard, Regulation of X-chromosome inactivation by the X-inactivation centre.
7. A. Wutz, Gene silencing in X-chromosome inactivation: advances in understanding facultative heterochromatin formation. *Nat Rev Genet* **12**, 542-553 (2011).
8. E. Splinter *et al.*, The inactive X chromosome adopts a unique three-dimensional conformation that is dependent on Xist RNA.
9. X. Deng *et al.*, Bipartite structure of the inactive mouse X chromosome. *Genome Biol* **16**, 152 (2015).
10. S. S. Rao *et al.*, A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping.

11. A. S. Belmont, F. Bignone, P. O. Ts'o, The relative intranuclear positions of Barr bodies in XXX non-transformed human fibroblasts. *Experimental cell research* **165**, 165-179 (1986).

12. A. Rego, W. Sinclair Pb Fau - Tao, I. Tao W Fau - Kireev, A. S. Kireev I Fau - Belmont, A. S. Belmont, The facultative heterochromatin of the inactive X chromosome has a distinctive condensed ultrastructure.

13. K. Plath, S. Mlynarczyk-Evans, D. A. Nusinow, B. Panning, Xist RNA and the mechanism of X chromosome inactivation. *Annu Rev Genet* **36**, 233-278 (2002).

14. R. Galupa, E. Heard, X-chromosome inactivation: new insights into cis and trans regulation.

15. A. Cerase, G. Pintacuda, A. Tattermusch, P. Avner, Xist localization and function: new insights from multiple levels.

16. J. C. Chow *et al.*, LINE-1 activity in facultative heterochromatin formation during X chromosome inactivation.

17. J. M. Engreitz *et al.*, The Xist lncRNA exploits three-dimensional genome architecture to spread across the X chromosome. *Science* **341**, 1237973 (2013).

18. M. D. Simon *et al.*, High-resolution Xist binding maps reveal two-step spreading during X-chromosome inactivation. *Nature* **504**, 465-469 (2013).

19. C. M. Clemson, M. Hall Ll Fau - Byron, J. Byron M Fau - McNeil, J. B. McNeil J Fau - Lawrence, J. B. Lawrence, The X chromosome is organized into a gene-rich outer rim and an internal core containing silenced nongenic sequences.

20. A. Wutz, T. P. Rasmussen, R. Jaenisch, Chromosomal silencing and localization are mediated by different domains of Xist RNA. *Nat Genet* **30**, 167-174 (2002).

21. J. Chaumeil, P. Le Baccon, A. Wutz, E. Heard, A novel role for Xist RNA in the formation of a repressive nuclear compartment into which genes are recruited when silenced. *Genes Dev* **20**, 2223-2237 (2006).

22. S. Quinodoz, M. Guttman, Long non-coding RNAs: An emerging link between gene regulation and nuclear organization. *Trends in Cell Biology* ***in press***, (2014).

23. C. A. McHugh *et al.*, The Xist lncRNA interacts directly with SHARP to silence transcription through HDAC3.

24. C. Chu *et al.*, Systematic discovery of Xist RNA binding proteins. *Cell* **161**, 404-416 (2015).

25. A. Minajigi *et al.*, Chromosomes. A comprehensive Xist interactome reveals cohesin repulsion and an RNA-directed chromosome conformation. *Science* **349**, (2015).

26. H. J. Worman, J. Yuan, G. Blobel, S. D. Georgatos, A lamin B receptor in the nuclear envelope. *Proc Natl Acad Sci U S A* **85**, 8531-8534 (1988).

27. Y. Gruenbaum, R. D. Margalit A Fau - Goldman, D. K. Goldman Rd Fau - Shumaker, K. L. Shumaker Dk Fau - Wilson, K. L. Wilson, The nuclear lamina comes of age.

28. B. Burke, C. L. Stewart, The nuclear lamins: flexibility in function.

29. D. Hochstrasser M Fau - Mathog, Y. Mathog D Fau - Gruenbaum, H. Gruenbaum Y Fau - Saumweber, J. W. Saumweber H Fau - Sedat, J. W. Sedat, Spatial organization of chromosomes in the salivary gland nuclei of Drosophila melanogaster.

30. M. Amendola, B. van Steensel, Mechanisms and dynamics of nuclear lamina-genome interactions. *Curr Opin Cell Biol* **28**, 61-68 (2014).

31. B. D. Towbin, S. M. Meister P Fau - Gasser, S. M. Gasser, The nuclear envelope--a scaffold for silencing?

32. S. Smith, G. Blobel, The first membrane spanning region of the lamin B receptor is sufficient for sorting to the inner nuclear membrane.

33. J. Baron-Benhamou, A. E. Gehring Nh Fau - Kulozik, M. W. Kulozik Ae Fau - Hentze, M. W. Hentze, Using the lambdaN peptide to tether proteins to RNAs.

34. C. Keryer-Bibens, H. B. Barreau C Fau - Osborne, H. B. Osborne, Tethering of proteins to RNAs by bacteriophage proteins.

35. N. Daigle, J. Ellenberg, LambdaN-GFP: an RNA reporter system for live-cell imaging.

36. J. Kind, B. van Steensel, Genome-nuclear lamina interactions and gene regulation. *Curr Opin Cell Biol* **22**, 320-325 (2010).

37. L. E. Finlan *et al.*, Recruitment to the nuclear periphery can alter expression of genes in human cells. *PLoS Genet* **4**, e1000039 (2008).

38. K. L. Reddy, J. M. Zullo, E. Bertolino, H. Singh, Transcriptional repression mediated by repositioning of genes to the nuclear lamina. *Nature* **452**, 243-247 (2008).

39. A. Monfort *et al.*, Identification of Spen as a Crucial Factor for Xist Function through Forward Genetic Screening in Haploid Embryonic Stem Cells.

40. B. Moindrot *et al.*, A Pooled shRNA Screen Identifies Rbm15, Spen, and Wtap as Factors Required for Xist RNA-Mediated Silencing.

Chapter 5

CONCLUSION AND FUTURE DIRECTIONS

5.1 Conclusion

With RAP-MS, we are able to identify three Xist direct-interacting proteins, SAF-A, SHARP and LBR, that are required for Xist-mediated chromosome-wide silencing. While previous studies have shown that SAF-A is required for Xist to localize on the chromatin(*1*), by characterizing the roles SHARP and LBR play in the process, we suggest a model for how the interaction between Xist and these three proteins can orchestrate chromosome-wide transcriptional silencing on Xi **(Figure 5.1)**.

Upon initiation of Xist expression **(Figure 5.1; left panel)**, Xist spreads to regions **(Figure 5.1; black regions)** that are closest to the Xist transcription locus **(Figure 5.1; red arrow)** by binding to the SAF-A protein on chromatin(*1, 7, 8*). Xist recruits the SHARP protein and its associated SMRT complex(*2-4*) to these initial sites. This Xist-SHARP complex can then act to either directly recruit HDAC3 to the X-chromosome or exploit HDAC3 that may already be present at active genes across the X-chromosome(*5, 6*) and induce its enzymatic activity by bringing it into proximity with its required SMRT co-repressor(*7, 8*) at Xist target sites across the X-chromosome. Through the activity of HDAC3, Xist can direct the removal of activating histone acetylation marks on chromatin thereby compacting chromatin(*9-11*) and silencing transcription(*6, 12, 13*). Then, these initial Xist-coated DNA regions **(Figure 5.1; black regions)** sample different locations of the nucleus(*44*, *45*, *23*, *46, 47*) and when

they come into close proximity of the nuclear lamina, are sequestered at the nuclear lamina through an interaction between Xist and LBR **(Figure 5.1; middle panel)**. Because DNA that interacts with the nuclear lamina undergoes more constrained mobility(*23, 24*), this recruitment changes the 3-dimensional organization of X-chromosome(*15, 48, 49*) and repositions active genes **(Figure 5.1; green regions)** closer the Xist transcription locus enabling Xist, and its SHARP/SMRT/HDAC3 silencing complex(*11, 18, 22*), to spread to these new sites by 3-dimensional proximity transfer. These sites are then recruited to the nuclear lamina, effectively bringing another set of active genes **(Figure 5.1; yellow regions)** into closer contact with the Xist transcription locus **(Figure 5.1; right panel)**. Because the Xist transcription locus escapes Xist coating and silencing, it is positioned away from the nuclear lamina(*7, 8, 5, 2, 1*) and therefore will be close to sites that have not yet been coated and silenced by Xist. This iterative process would enable Xist to spread to, and silence, actively transcribed genes across the entire X-chromosome.

Xist has long represented a mechanistic paradigm for understanding other lncRNAs, but this is largely because we lacked the tools required to probe any specific lncRNA. Accordingly, most work has focused on the role of lncRNAs in the regulation of gene expression through the recruitment of chromatin regulatory proteins – primarily PRC2. Our results highlight the importance of identifying direct lncRNA-interacting proteins for deciphering lncRNA mechanisms of action. Even for Xist, where we know a tremendous amount about its molecular functions, we uncovered a novel mechanism for its ability to silence transcription. There are likely to be many additional mechanisms of action for lncRNAs, including roles in the nucleoplasm and cytoplasm. Identifying the proteins that interact with any given

lncRNA will be an important step towards deciphering these mechanisms. Importantly, RAP-MS provides a critical tool for achieving this goal and will accelerate the discovery of novel lncRNA mechanisms that have thus far proved elusive.

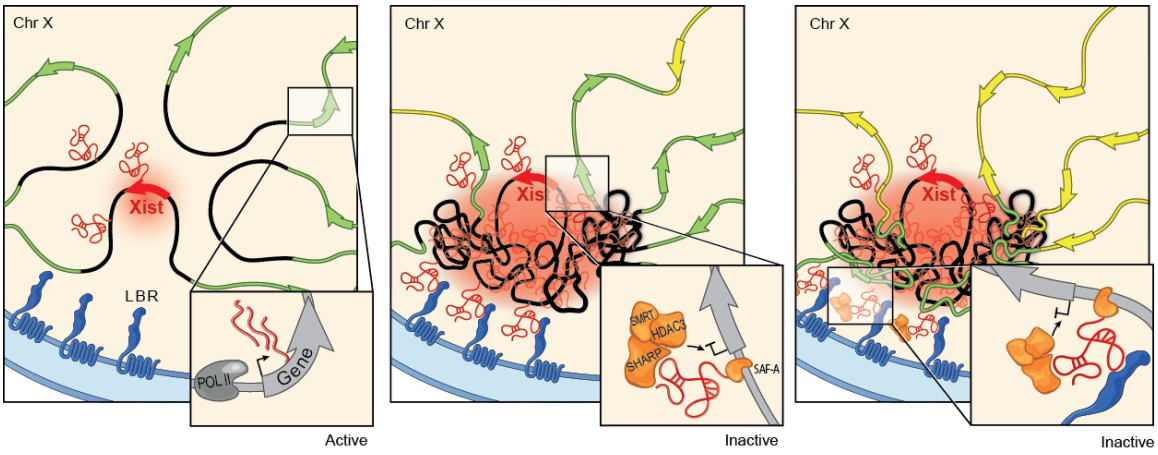

Figure 5.1. A model for how Xist-mediated recruitment to the nuclear lamina enables spreading to active genes and transcriptional silencing on the X chromosome. Xist initially localizes to the core of the X chromosome territory by localizing at DNA sites that are in close 3D proximity to its transcriptional locus. These initial Xist localization sites are generally inactive prior to Xist induction. The Xist-coated DNA, like other chromosomal DNA regions, will dynamically sample different nuclear locations and, because Xist binds LBR, will become tethered at the nuclear lamina when it comes into spatial proximity. This lamina association is known to constrain chromosomal mobility and by doing so would position the Xist-coated DNA away from the actively transcribed Xist transcription locus. This would enable other DNA regions on the X chromosome, which are physically linked to these tethered regions, to be brought into closer spatial proximity to the Xist transcription locus. In this way, Xist and its silencing factors can spread to these newly accessible DNA regions on the X chromosome.

5.2 Future Directions

Although we have identified the direct Xist-interacting proteins and demonstrated a model for the mechanism of Xist-mediated silencing, some questions remain unclear during the process of XCI. For example, although Xist can spread across the entire X chromosome, some genes on Xi, including Xist itself, can escape from Xist-mediated silencing and remain actively transcribed(*14*). Genome-wide analysis of Xist associated DNA on Xi reveals that Xist is depleted from the regions of these escaping genes and their promoters(*15*). The result suggests that the escapees escape from Xist-mediated silencing by preventing Xist binding to the regions. Furthermore, the depleted Xist association is unlikely result from the features of the DNA sequences of these escapees, since no common sequences have been identified among escapees. Also, the depleted Xist association could not be simply explained by the interaction between Xist and SAF-A, since SAF-A showed homogenous chromatin association across the entire X chromosome. It is possible that the depleted Xist association is due to the spatial separation between Xist and the escapees, which makes Xist unable to access to the regions by 3D proximity search. This hypothesis is supported by a recent study showing that changing the chromosome organization of Xi can lead to lower expression level of certain escapees(*16*). However, further studies are required to identify the components involved in the process, which may help us to reveal the mechanism of escaping Xist-mediated silencing. Although escapee only contributes a small fraction among the genes on X chromosome (~20-80 in mouse depending on different cell types)(*14*, *16*), studies have suggested that some escapees may be related to some clinical features of polyX karyotypes in humans(*17*). Therefore, studying the

mechanism of escaping may also provide us some clinical insights of how to compensate the dosage difference of certain escapees for polyX karyotypes patients.

Another interesting question waiting to be answered is that how the maintenance state of Xist-mediated silencing is established. At the late stage of XCI, the CpG island of the promoter regions of silenced genes on Xi is heavily methylated(*18, 19*). One possible model is that once the DNA methylation has been established *de novo* by DNMT3A/B, the silencing state can be maintained by DNMT1-mediated self-propagation of DNA methylation pattern(*20*). But how is DNMT1 recruited to the target sites across the entire Xi at the first place? Since Xist is capable of spreading across the entire Xi, it is likely that DNMT1 is recruited to Xi through an Xist-mediated mechanism. However, since DNMT1 is not identified as one of the Xist-interacting proteins, either directly or indirectly(*21-23*), it is possible that DNMT1 is recruited to the target sites through a secondary downstream event of Xist spreading and silencing. Some studies suggested that SMCHD1, a protein interacting with Xist indirectly, plays a role in DNMT1 recruitment(*24, 25*). However, the CpG island methylation still occurs in some silenced genes, which suggests another Smchd1-independent pathway of DNA methylation(*25*). Therefore, further studies are required to identify the other components that are involved in the transition from the initiation state to the maintenance state of Xist-mediated silencing, which may provide us some insights into the mechanism of establishing XCI maintenance and the role Xist plays in this process.

In sum, by identifying Xist-interacting proteins, we show how a lncRNA can function as a scaffold which orchestrates various events through recruiting different proteins to specific

loci. The methods we used enable us reveal some unknown mechanism of Xist-mediated events that occur during XCI. In addition to Xist, we now have the opportunity to study the molecular mechanism of other lncRNAs with either known or unknown function. I believe that with Xist as an example, a door has been opened for us to study the functions of various lncRNAs and explore their molecular mechanisms!

References

1. Y. Hasegawa, N. Brockdorff, S. Kawano, K. Tsutui, S. Nakagawa, The matrix protein hnRNP U is required for chromosomal localization of Xist RNA. *Dev Cell* **19**, 469-476 (2010).
2. S. Mikami *et al.*, Structural insights into the recruitment of SMRT by the corepressor SHARP under phosphorylative regulation. *Structure* **22**, 35-46 (2014).
3. Y. Shi *et al.*, Sharp, an inducible cofactor that integrates nuclear receptor repression and activation. *Genes Dev* **15**, 1140-1151 (2001).
4. M. Ariyoshi, J. W. Schwabe, A conserved structural motif reveals the essential transcriptional repression function of Spen proteins and their role in developmental signaling. *Genes Dev* **17**, 1909-1920 (2003).
5. Z. Wang *et al.*, Genome-wide mapping of HATs and HDACs reveals distinct functions in active and inactive genes. *Cell* **138**, 1019-1031 (2009).
6. J. Li, Q. Lin, W. Wang, P. Wade, J. Wong, Specific targeting and constitutive association of histone deacetylase complexes during transcriptional repression. *Genes Dev* **16**, 687-692 (2002).
7. S. H. You *et al.*, Nuclear receptor co-repressors are required for the histone-deacetylase activity of HDAC3 in vivo. *Nature structural & molecular biology* **20**, 182-187 (2013).
8. M. G. Guenther, O. Barak, M. A. Lazar, The SMRT and N-CoR corepressors are activating cofactors for histone deacetylase 3. *Molecular and cellular biology* **21**, 6091-6101 (2001).
9. K. F. Toth *et al.*, Trichostatin A-induced histone acetylation causes decondensation of interphase chromatin. *Journal of cell science* **117**, 4277-4287 (2004).

10. S. M. Gorisch, M. Wachsmuth, K. F. Toth, P. Lichter, K. Rippe, Histone acetylation increases chromatin accessibility. *Journal of cell science* **118**, 5825-5834 (2005).

11. D. Lleres, J. James, S. Swift, D. G. Norman, A. I. Lamond, Quantitative analysis of chromatin compaction in living cells using FLIM-FRET. *J Cell Biol* **187**, 481-496 (2009).

12. K. Struhl, Histone acetylation and transcriptional regulatory mechanisms. *Genes Dev* **12**, 599-606 (1998).

13. M. H. Kuo, C. D. Allis, Roles of histone acetyltransferases and deacetylases in gene regulation. *BioEssays : news and reviews in molecular, cellular and developmental biology* **20**, 615-626 (1998).

14. S. B. Peeters, A. M. Cotton, C. J. Brown, Variable escape from X-chromosome inactivation: Identifying factors that tip the scales towards expression. *BioEssays* **36**, 746-756 (2014).

15. J. M. Engreitz *et al.*, The Xist lncRNA exploits three-dimensional genome architecture to spread across the X chromosome. *Science* **341**, 1237973 (2013).

16. L. Giorgetti *et al.*, Structural organization of the inactive X chromosome in the mouse. *Nature* **535**, 575 (2016).

17. Y. Zhang *et al.*, Genes that escape X-inactivation in humans have high intraspecific variability in expression, are associated with mental impairment but are not slow evolving. *Molecular biology and evolution* **30**, 2588-2601 (2013).

18. L. F. Lock, N. Takagi, G. R. Martin, Methylation of the Hprt gene on the inactive X occurs after chromosome inactivation. *Cell* **48**, 39-46 (1987).

19. D. P. Norris, N. Brockdorff, S. Rastan, Methylation status of CpG-rich islands on active and inactive mouse X chromosomes. *Mammalian Genome* **1**, 78-83 (1991).

20. M. Okano, D. W. Bell, D. A. Haber, E. Li, DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development. *Cell* **99**, 247-257 (1999).

21. C. A. McHugh *et al.*, The Xist lncRNA interacts directly with SHARP to silence transcription through HDAC3.

22. C. Chu *et al.*, Systematic discovery of Xist RNA binding proteins. *Cell* **161**, 404-416 (2015).

23. A. Minajigi *et al.*, Chromosomes. A comprehensive Xist interactome reveals cohesin repulsion and an RNA-directed chromosome conformation. *Science* **349**, (2015).

24. M. E. Blewitt *et al.*, SmcHD1, containing a structural-maintenance-of-chromosomes hinge domain, has a critical role in X inactivation. *Nat Genet* **40**, 663-669 (2008).

25. A.-V. Gendrel *et al.*, Smchd1-dependent and-independent pathways determine developmental dynamics of CpG island methylation on the inactive X chromosome. *Dev Cell* **23**, 265-279 (2012).