UNIVERSITY OF CALIFORNIA,
IRVINE

Functional Characterization of Human LncRNA *JPX*

DISSERTATION

submitted in partial satisfaction of the
requirements for the degree of

DOCTOR OF PHILOSOPHY

In Biological Sciences

By

Heather Karner

Dissertation Committee:
Assistant Professor Sha Sun, Chair
Professor Ken Cho
Associate Professor Olga Razorenova
Associate Professor Rahul Warrior
Professor Kyoko Yokomori

2019

DEDICATION

To

My husband, Steve
Thank you for standing by me throughout this whole rigmarole. I love you!!!

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

I would like to thank my family, friends, and mentors who have provided support and guidance throughout my time here at UCI while obtaining my PhD.

Thank you to my PhD advisor, Dr. Sha Sun, who guided me and pushed me to become a better scientist, and supporting me through the rough times that tend to occur during the PhD process. I would also like to thank my thesis committee: Dr. Ken Cho, Dr. Olga Razorenova, Dr. Rahul Warrior, and Dr. Kyoko Yokomori. My committee has always challenged me to dig deeper in my research and provided helpful advice that guided the direction of my projects over the years.

Thank you to Dr. Justin Shaffer for being an amazing teaching mentor, and being the first person to educate me about effective teaching techniques and offering career advice that has helped me make decisions and take steps toward a career in academia. Thank you to Dr. Debra Mauzy-Melitz for being a wonderful GAANN faculty advisor, creating opportunities to grow in my teaching, and cultivating a group of amazing GAANN fellows who have helped me tremendously during my graduate studies. Thank you to Dr. De Gallow and Dr. Daniel Mann, my Pedagogical Fellow mentors, who have also helped mold me into the teacher that I am today by giving me the tools to create engaging lessons and the curiosity to continually assess and seek out effective methods for teaching. Additionally, thank you to my PFFs that I met during my time as a Pedagogical Fellow for helping me refine my teaching skills and all of the fascinating conversations we had during our time learning together.

Thank you to all of the Developmental and Cell Biology staff who have helped me during my time here as a graduate student run as smoothly as possible and always being willing to help with any problem, big or small.

I would like to give a very special thank you to my husband, Steve Karner, for all your help, love, and support during the good times and the tough times.

Last, but not least, thank you to everyone in the Sun Lab, who have directly contributed to my research. Thank you to my past undergraduate Bio199 students, especially Micaela Erhard and Najla Bernichi-Shilleh, for all of their hard work. Thank you to my past labmates, Ben Lin, Dr. Sarah Carmona, and Dr. Chiu-Ho Webb, for your advice and aid with my projects. I hope that all of the work we have done together has left a good foundation for the next Sun Lab members to make great discoveries!

Financial support was provided by the Graduate Assistance in Areas of National Need (GAANN) fellowship from the US Department of Education, grant number P200A120207. Portions of Chapter 2, and all of Chapter 3, have appeared in the following publication and are used in this thesis with the permission of Elsevier:

Karner, H.M., Webb, C.-H., Carmona, S., Liu, Y., Lin, B., Erhard, M., Chan, D., Baldi, P., Spitale, R.C., and Sun, S. (2019). Functional Conservation of LncRNA *JPX* Despite Sequence and Structural Divergence. *Journal of Molecular Biology*

CURRICULUM VITAE
Heather Karner

Education

Ph.D. Candidate in the Department of Developmental and Cell Biology **2013-2019**
University of California, Irvine (UCI)
Dissertation: Functional Characterization of Human LncRNA JPX

B.S. in Microbiology, Immunology, and Molecular Genetics **2005-2010**
University of California, Los Angeles (UCLA)

Research Experience

Ph.D. Candidate: Department of Developmental and Cell Biology at UCI **July 2014-**
Laboratory of Dr. Sha Sun **Nov 2019**

- Thesis lab
- Studying the structure and function of long noncoding RNA JPX within humans
- Training and mentoring Bio 199 research students

Staff Research Associate: Department of Neurosurgery at UCLA **May 2011-May 2013**
Laboratory of Dr. Isaac Yang

- Lab Manager
    - Prepared lab to begin wet lab and *in vivo* experiments
    - Maintained training records, lab supplies, and other basic lab duties
    - Trained and mentored medical and undergraduate students, and supervised student projects (15 students)
    - Assisted in the writing and submission of grants and publications
- Research Coordinator
    - Planned and supervised research projects
    - Arranged and conducted *in vivo* mouse experiments
    - Studied the efficacy of vault nanoparticles in activating an immune response against antigens for brain cancers such as glioblastoma
    - Studied CD133+ cancer stem-like cells and the feasibility of using a CD133+ pulsed dendritic cell vaccine to prevent recurrence of glioblastoma

Staff Research Associate: Department of Microbiology, **July 2010-May 2011**
Immunology and Molecular Genetics at UCLA
Laboratory of Dr. Imke Schroeder

- Studied the effects of mutations in select regulatory genes of the type III secretion system of bacterium *Burkholderia pseudomallei,* cause of the infectious disease Melioidosis
- Trained and mentored undergraduate students in the lab (2 students)

Publications

Note: Last name changed from Garcia to Karner in 2013

(In Press)

1. **Karner, H.M.**, Webb, C.-H., Carmona, S., Liu, Y., Lin, B., Erhard, M., Chan, D., Baldi, P., Spitale, R.C., and Sun, S. Functional Conservation of LncRNA *JPX* Despite Sequence and Structural Divergence. *Journal of Molecular Biology.* 2019. **Featured Article**.

2. **Karner H**, Sun S. lncRNA miR503HG is a new player in hepatocellular carcinoma metastasis. *Non-coding RNA Investigation*. 2018. **Editorial**.

3. Li C, Hong T, Webb C, **Karner H**, Sun S, Nie Q. A self-enhanced transport mechanism through long noncoding RNAs for X-chromosome inactivation. *Scientific Reports*. 2016.

4. Nagasawa DT, Spasic M, Choy W, Yew A, Trang A, **Garcia H**, Yang I. An analysis for the treatment and outcomes of intracranial epidermoid tumors with malignant transformation. *Clinical Neurology and Neurosurgery*. 2012.

5. Nagasawa DT, Spasic M, Choy W, Yew A, Trang A, Zarinkhou G, **Garcia H**, Yang I. Genetic expression profiles of adult and pediatric ependymomas: molecular pathways, prognostic indicators and therapeutic targets. *Clinical Neurology and Neurosurgery*. 2012.

6. Nagasawa DT, Fong C, Yew A, Spasic M, **Garcia HM**, Kruse CA, Yang I. Passive immunotherapeutic strategies for the treatment of malignant gliomas. *Neurosurg Clin N Am.* 2012 Jul; 23(3):481-95.

7. Yang J, Nagasawa DT, Spasic M, Amolis M, Choy W, **Garcia HM**, Prins RM, Liau LM, Yang I. Endogenous vaults and bioengineered vault nanoparticles for treatment of glioblastomas: implications for future targeted therapies. *Neurosurg Clin N Am*. 2012; 23(3):451-8.

8. Yew A, Trang A, Spasic M, Nagasawa DT, Choy W, **Garcia H**, Yang I. Chromosomal alterations, prognostic factors, and targeted molecular therapies for malignant meningiomas. *J Clin Neurosci*. 2012.

9. Thant AA, Wu Y, Lee J, Mishra DK, **Garcia H**, Koeffler HP, Vadgama JV. Role of caspases in 5-FU and selenium-induced growth inhibition of colorectal cancer cells. *Anticancer Res.* 2008.

Awards & Fellowships

UC CCRC (Cancer Research Coordinating Committee) **July 2019**
- Seed grant for lab resources and personnel
- Aided in the writing of this grant

UCI ACS (American Cancer Society) **June 2019**
- Seed grant for lab resources and personnel
- Aided in the writing of this grant

Campus Organization Excellence in Service Award **May 2018**
- Awarded to my service organization Reach Out Teach Out at the UCI Engage Celebration of Community Engagement of which I was Vice President at the time

GAANN Fellowship (Graduate Assistance in Areas of National Need) **Oct 2015, 2016, & 2017**
- Competitive fellowship from the Department of Developmental and Cell Biology at UCI given to graduate students who have demonstrated a desire to pursue a career in teaching
- Award Amount: Graduate student registration fees and stipend

Edward A. Steinhaus Teaching Award **June 2017**
- Competitive award from the Department of Developmental and Cell Biology at UCI that is presented to outstanding graduate students with promising futures as educators
- Award Amount: $750

Certificate of Teaching Excellence **May 2017**
- In recognition of advanced training in evidence-based pedagogy

UCI CIRTL Associate Level **April 2017**
- Successful completion of all UCI CIRTL associate level requirements

CCBS Opportunity Award on the topic of Mechanisms of Disease **June 2016**
- Project Title: Dynamics of IncRNAs in the X inactivation center in mouse ES and human breast cancer cells
- Collaboration project with the labs of Dr. Tom Schilling and Dr. Zeba Wunderlich in the Department of Developmental and Cell Biology at UCI
- Award Amount:
    - $10,000 for lab resources
    - $1000 for the students or postdocs to use

Pedagogical Fellow **Jan 2016**
- Competitive fellowship program through the Center for Engaged Instruction at UCI that instructs students in pedagogy and how to employ different teaching techniques in the classroom to engage students in learning
- Award Amount: $2000

CCBS Opportunity Award on the topic of Dynamics and Complexity **Aug 2015**

- Project Title: Dynamic Regulation and Noise Attenuation in lncRNA network for X-chromosome inactivation
- Collaboration project with the lab of Dr. Qing Nie in the Department of Mathematics at UCI
- Award Amount:
  - $10,000 for lab resources
  - $1000 for the students or postdocs to use

Conferences & Presentations

Symposium on Basic Cancer Research (UC Irvine) **May 2019**

- Co-presented poster
- Title: Downregulation of XIST Ovarian Cancer and Possible Mechanisms

SABER West (UC Irvine) **Jan 2018**

- Co-presented poster
- Title: Strategies for Undergraduate Scientific Success: Resolving the Disconnect

Center for Complex Biological Systems (CCBS) Retreat **Mar 2017**

- Gave a talk about the collaboration project done with the labs of Dr. Tom Schilling and Dr. Zeba Wunderlich in the Department of Developmental and Cell Biology at UCI as a result of receiving the CCBS Opportunity Award
- Title: Dynamics of X inactivation center lncRNAs in Human Cancer Cells

SABER West (UC Irvine) **Jan 2017**

- Attended workshops and presentations for research on teaching strategies

ABLE: Association for Biology Laboratory Education **July 2016**

- Selected to give a workshop
- Title: Awareness Makes You Friends: Addressing Diversity Concerns in the Classroom

Center for Complex Biological Systems (CCBS) Retreat **Mar 2016**

- Gave a talk about the collaboration project done with the lab of Dr. Qing Nie in the Department of Mathematics at UCI as a result of receiving the CCBS Opportunity Award
- Title: A self-enhanced transport mechanism through long noncoding RNAs for X-chromosome inactivation

SCalE: Southern California Evolutionary Genetics & Genomics meeting **Feb 2016**

- Selected to present a poster
- Title: Structure versus sequence conservation: Human long noncoding RNA JPX and its homology to murine Jpx

Professional Development

- ❖ Mentoring in Excellence Program (Certificate)
- ❖ Biological Sciences Graduate Teaching Fellow Program
- ❖ Graduate Public Speaking Skills (Drama 227)
- ❖ Pedagogical Fellows Program (University Studies 390A, 390B, & 390C)

Teaching Experience

Mentor for Bio Sci 199 Students at UCI **Fall 2014-Fall 2019**

- Currently mentoring four undergraduate students in our lab
- Students are taught how to conduct research as well as develop their own hypothesis-based research projects
- Students aid me on experiments for my thesis project
- Students are required to read scientific articles that either relate to our overall lab research or are specific to their own project, and then write reflections on what they have learned
- Students also help with basic maintenance of the lab, such as taking inventory of lab supplies, which teaches them the layout of the lab and what it takes to keep a lab functioning

Teaching assistant for Bio Sci 75 (Human Development: Conception to Birth) at UCI
Dr. Sha Sun **Spring Quarter 2019**

- 25% appointment
- Attended lecture and aided with in-class assignments
- Held weekly office hours
- Aided with developing exams and review materials
- Processed and input grades into gradebook
- **Guest lectured one class on mitosis and meiosis**

Teaching assistant for Bio Sci 100 (Scientific Writing) at UCI **Winter Quarter 2019**
Dr. Debra Mauzy-Melitz

- 25% appointment
- Administrative Teaching Assistant
- Graded writing assignments
- Processed and input grades into gradebook

Teaching assistant for Bio Sci 97 (Genetics) at UCI **Fall Quarter 2018**
Dr. Rahul Warrior, Dr. Zeba Wunderlich, Dr. Lee Bardwell, and Dr. Olivier Cinquin

- 50% appointment
- Administrative Teaching Assistant for two co-taught sections
- Processed and input grades into gradebook
- Proofread, proctored, and graded exams

Teaching assistant for Bio Sci 97 (Genetics) at UCI **Summer Session 2018**
Dr. Lee Bardwell

- 50% appointment
- Held two discussion sections a week in which I was responsible for creating material that aided student's understanding of the topics discussed in lecture
- Processed and input grades into gradebook
- Proofread, proctored, and graded exams
- Trained second TA in administrative and teaching duties

Teaching assistant for Bio Sci 93 (From DNA to Organisms) at UCI **Fall Quarter 2017**
Dr. Justin Shaffer

- 50% appointment
- Administrative Teaching Assistant
- Processed and input grades into gradebook
- Proofread, proctored, and graded exams
- Advised new TAs on administrative and class related issues
- Audited and gave feedback on fellow TA's discussion sections

Pedagogical Fellow for TAPDP Training at UCI **Sept 2017**

- Returned as a Pedagogical Fellow Alumni and conducted workshops to train new TAs from the School of Engineering during TAPDP training

Teaching assistant for Bio Sci 97 (Genetics) at UCI **Summer Session 2017**
Dr. Lee Bardwell

- 50% appointment
- Held two discussion sections a week in which I was responsible for creating material that aided student's understanding of the topics discussed in lecture
- Processed and input grades into gradebook
- Proofread, proctored, and graded exams

Teaching assistant for Bio Sci 93 (From DNA to Organisms) at UCI **Fall Quarter 2016**
Dr. Justin Shaffer

- 50% appointment
- Held three discussion sections a week in which I was responsible for helping students grasp the material taught in lecture through activities, worksheets, and quizzes I designed
- Graded writing assignments
- Attended lectures, aided students with in-class problems, and critiqued lectures
- Proofread and proctored exams
- Held office hours once a week
- Audited and gave feedback on fellow TA's discussion sections

Curriculum Design of University Studies 390C **Fall 2016**

- Designed the curriculum for University Studies 390C in conjunction with instructor Danny Mann and other pedagogical fellows
- Class was geared toward preparing pedagogical fellows for the job market
- Topics: cover letter, CV, research statement, diversity statement, networking, and social media presence

Pedagogical Fellow for TAPDP Training at UCI **Jan-Sept 2016**

- Took University Studies 390A & 390B as a part of the pedagogical fellows program and learned about research in pedagogy as well as how to create workshops for TAPDP (Teaching Assistant Professional Development Program)
- Developed and conducted workshops that were used to train new TAs from the School of Biological Sciences during TAPDP

Teaching assistant for Bio Sci D103 (Cell Biology) at UCI **Summer Session 2016**
Dr. Salme Taagepera

- 50% appointment
- Held two discussion sections a week in which I was responsible for going over the Problem Based Learning (PBL) worksheets associated with the lectures
- Attended lectures and aided students with in-class problems
- Proofread and proctored exams
- Graded exams, pre-class assignments, and in-class assignments
- Input grades into the gradebook

Guest lectured for Bio Sci 100 (Scientific Writing) at UCI **Winter Quarter 2016**
Dr. Debra Mauzy-Melitz

- Title: Clinical Research Design Flaws

Teaching assistant for Bio Sci 93 (From DNA to Organisms) at UCI **Fall Quarter 2015**
Dr. Justin Shaffer

- **Head TA** (help other TAs with creating teaching material for their discussion sections)
- 50% appointment
- Held three discussion sections a week in which I was responsible for helping students grasp the material taught in lecture through activities, worksheets, and quizzes I designed
- Attended lectures, aided students with in-class problems, and critiqued lectures
- Proofread and proctored exams
- Held office hours once a week
- Audited and gave feedback on fellow TA's discussion sections

Teaching assistant for Bio Sci 97 (Genetics) at UCI **Summer Session 2015**
Dr. David Camerini

- 50% appointment
- Held two discussion sections a week in which I was responsible for creating material that aided student's understanding of the topics discussed in lecture
- Processed and input grades into gradebook
- Proofread, proctored, and graded exams
- Held office hours once a week

Teaching assistant for Bio Sci 75 (Human Development: Conception to Birth) at UCI
Dr. Sha Sun **Spring Quarter 2015**

- 25% appointment
- Attended lecture and aided with in-class assignments
- Aided with grading exams
- Held office hours once a week
- **Guest lectured one class on mitosis and meiosis**

Teaching assistant for Bio Sci 93 (From DNA to Organisms) at UCI **Fall Quarter 2014**
Dr. Justin Shaffer

- **Biological Sciences Graduate Teaching Fellow Program**
- 50% appointment
- Held three discussion sections a week in which I was responsible for creating material that aided student's understanding of the topics discussed in lecture
- Attended lectures and aided students with in-class problems
- Critiqued lectures
- Proctored exams
- Held office hours once a week

Service

Reach Out Teach Out Club at UCI **Spring 2017-Spring 2018**

- Founded this club with another graduate student (Autumn Holms) at UCI
- Club is dedicated to introducing science and research to high school students in underrepresented areas around the community
- Students are given a mini-lecture on what one of our graduate student club members is working on for their degree followed by a lab tour and activity related to the graduate student's research
- Responsibilities:
    - Vice President
    - Attend regular meetings of officers to discuss financial and legal matters related to having high school students visiting on campus as well as meetings related to organizing event days

Breaker's Advanced Workshop **July 2017**

- Program for at-risk high school students from Laguna Beach High School
- Took students on a lab tour followed by a discussion on the ethics of animal research and a visual activity with ovarian cancer cell lines

Selection of 2017-2018 Pedagogical Fellows **Fall 2016**

- Participated in the selection process of the 2017-2018 cohort of pedagogical fellows
- Responsibilities:
    - Conducted pre- and post-consultations with applicants for their teaching video submission and provided a summary of their teaching strengths and weaknesses
    - Interviewed final round applicants and provided feedback on their performance and fit for the program

Strategies for Scientific Success: Amplify Your Portfolio **Oct 2016**

- Sponsored by GAANN
- Symposium for graduate students and postdocs interested in STEM field careers
- Topics:
    - Designing a class schedule, syllabus, and online supplement
    - Tips for writing a teaching philosophy and diversity statement
    - CV vs. Resume
    - Long-term goal setting using an Individual Development Plan
    - Career panel

Tech Trek **July 2015 & 2016**

- Program for underprivileged middle school girls that exposes them to science
- Our lab put on a hands-on demonstration that allowed the girls to make observations of different cancer cell lines, perform dilutions, and learn how DNA is run on agarose gels

2nd Annual Strategies for Scientific Success: Bio 199 & Beyond **April 2016**

- Sponsored by GAANN
- Symposium for undergraduate students interested in STEM fields
- Topics:
    - How to seek out and be successful in an undergraduate research position at UCI
    - Diversity in the classroom case studies
    - Current undergraduate researcher's panel
    - Career panel

The Baccalaureates **Fall 2014 & 2015**

- Outreach program that another UCI graduate student (Ankita Shukla) and I have developed
- Target group: high school seniors in underrepresented areas
- Offered to read and critique college admission essays for free
- **Created and presented college essay writing seminars at these high schools**

Ask-A-Scientist Night **Oct 2013, 2014, & 2015**

- Sponsored by the Irvine Unified School District
- K-12 students can ask scientists for advice on how to best execute and analyze their chosen science project

ESCAPE **Aug 2015**

- Program for 3rd-5th grade elementary school teachers in underserved areas that exposes them to scientists and their research
- Our lab put on a hands-on demonstration that allowed the teachers to extract DNA from strawberries; this experiment was made simple enough that they could use it in the classroom to explain DNA to their own students

Judge for the Intel International Science and Engineering Fair (Intel ISEF; Local District) **Mar 2015**

- Judged and selected projects to participate in the preliminary round of Intel ISEF
- Judged and selected winners who would be allowed to participate at the national level

Irvine Unified School District Science Fair Judge **Feb 2014**

- Interviewed K-12 students about their science fair projects and aided in the decision of the winners for 5th grade

ABSTRACT OF THE DISSERTATION

Functional Characterization of Human LncRNA *JPX*

By

Heather Karner

Doctor of Philosophy in Biological Sciences

University of California, Irvine, 2019

Assistant Professor Sha Sun, Chair

Long noncoding RNAs are present in all eukaryotes, but knowledge of their function and mechanisms are lacking. In this dissertation, I have worked toward characterizing a human lncRNA known as *JPX*. In mice, this lncRNA has been proposed to be the activator of the master regulator of X chromosome inactivation, lncRNA *Xist*. My research provides evidence that human lncRNA *JPX* is capable of this function as well, through comparative sequence, structural, and functional analyses. Human *JPX*, despite sequence and structural divergence from its mouse homolog, robustly binds CTCF—a protein that is known to sit on the *Xist* promoter and inhibit expression. Most interestingly, the human *JPX* complemented the deleterious effect of a heterozygous loss of *Jpx* in mouse embryonic stem cells and returned these cells to wild type viability and morphology, as well as increased *Xist* expression. The differences between the two lncRNAs begs the question of whether lncRNA *JPX* has gained new functions in humans. To study this, I investigated *JPX* and *XIST* in ovarian cancer. This cancer tends to be discovered at later, more aggressive stages. It was found that these later stages also tend to have decreased expression of *JPX* and *XIST* in patients. So far, my research indicates that these lncRNAs may be engaged in biological pathways involved in tumor suppression, and their loss could lead to the progression and metastasis of ovarian cancer.

Chapter 1: Introduction

1.1 Long Noncoding RNA History

As the central dogma would suggest, DNA leads to RNA and then to protein. However, with the discoveries from the human genome project, it was realized that only ~1.5% of the genome actually conscribes to this dogma. The majority of the genome does not code for proteins and instead produces noncoding transcripts of RNA. It was commonly assumed these transcripts that yield no protein were simply transcriptional noise and of no use to the cell (Kapranov et al., 2010; Ponjavic et al., 2007; Struhl, 2007). There are studies that have stated that only 8.5% of the genome is actually functional (Rands et al., 2014), which is in stark contrast to the predicted 80% functionality by the ENCODE project (Dunham et al., 2012). Due to the high frequency of noncoding RNAs discovered, in addition to several being involved in transcriptional regulation and development, it is more likely that much of the genome yields functional transcripts. Noncoding RNAs have also been found to be dysregulated in many diseases such as cancer, indicating their importance to maintaining the normal processes of the cell (Fatica and Bozzoni, 2014; Gutschner and Diederichs, 2012; Johnsson et al., 2014; Perry and Ulitsky, 2016; Rinn and Chang, 2012; Weakley et al., 2011).

Genes that produce a certain class of RNA transcripts known as long noncoding RNA (lncRNA) are thought to make up anywhere from 70%-90% of the genome (Kapranov et al., 2010). LncRNAs have been described as transcripts that are greater than 200 nucleotides, lack any significant protein-coding ability, are polyadenylated or non-polyadenylated, and are transcribed by RNA polymerase II. Furthermore, lncRNA

genes have epigenetic markers typical of protein-coding genes (Prensner and Chinnaiyan, 2011). Noncoding genes in general have a tendency to lack sequence conservation between species (Necsulea and Kaessmann, 2014; Ponjavic et al., 2007; Ponting and Lunter, 2006). This rapid evolution poses a difficult challenge for understanding how sequence changes relate to and affect noncoding RNA function.

In recent years, many lncRNAs have been located in the normal gene structure as well as in more novel locations, such as within the promoters of other genes (Wu et al., 2017). The focus of this project is on a class of lncRNAs known as long intergenic noncoding RNA (lincRNA). This class satisfies all of the criteria mentioned above for lncRNA; however, their location in the genome is what sets them apart from other lncRNAs. LincRNAs are transcribed from the regions between genes and have many characteristics in common with protein coding mRNA, just like other lncRNA classes (Cabili et al., 2011; Ulitsky et al., 2011). Their primary transcripts tend to be processed similarly and their loci are similar at the chromatin level when looking at common epigenetic patterns (Khalil et al., 2009; Ulitsky et al., 2011). These similar features aided in the discovery of many lincRNAs, making it one of the most studied classes of lncRNA (Wu et al., 2017). However, as with most lncRNAs, they do differ from mRNA in that their expression levels are much lower and they tend to be expressed in a tissue specific manner (Cabili et al., 2011; Ulitsky and Bartel, 2013). In addition, though lincRNAs have similar histone marks and transcription factors involved in their transcription, they appear at a lower rate on lincRNA promoters in comparison to mRNA protein coding promoters (or at a greater rate, in the case of H3K9me3) and the splicing of lincRNAs tends to be less efficient (Khalil et al., 2009; Ulitsky et al., 2011). Since

lncRNA nomenclature is still developing (Mattick and Rinn, 2015), and in order to avoid confusion, the long intergenic noncoding RNA discussed further in this dissertation will be referred to by the more general term of lncRNA.

1.2 LncRNA Evolution

The study of the evolution of lncRNA is still in its early stages. Many of the techniques that are used to study protein evolution do not work in the case of noncoding RNA (Necsulea and Kaessmann, 2014). As mentioned previously, lncRNAs tend to lack sequence conservation since there is rapid sequence evolution across species (Ulitsky et al., 2011). Though they share some similarities with protein-coding genes, they do not share the need to conserve their sequences in order to function (Cabili et al., 2011; Hezroni et al., 2015; Ulitsky et al., 2011; Wu et al., 2017). This lack of sequence conservation have made lncRNAs difficult to study through comparative genomics. Though some have detectable sequence similarity and some have conserved genomic locations, they are still difficult to find and annotate (Hezroni et al., 2015; Ulitsky et al., 2011). This has forced innovation in the field, and researchers have had to develop different sequencing and comparative sequence analysis techniques in order to find lncRNAs and uncover their homologs if present in other species (Hezroni et al., 2015; Wu et al., 2017). LncRNAs have exhibited many different conservation patterns, introducing further difficulty into their identification. Interesting conservation patterns that have been seen in lncRNAs are: 1) preservation of the intron-exon arrangement, 2) secondary structure conservation, 3) 5’-end biased conservation with 3’-end rapid evolution, 4) focal sequence conservation, and 5) transcriptional activity independent of

transcript (Hezroni et al., 2017; Ransohoff et al., 2018; Wu et al., 2017). This last conservation pattern is a new hypothesis that indicates the act of transcription is all that is required and the lncRNA transcript produced is actually dispensable. Each of these patterns implicate unique selective pressures that contribute to the different functional aspects we see for lncRNA.

An example of focal sequence conservation can be seen in a study by Hezroni *et al*., which took a genomics approach to identify lncRNAs that were derived from fragments of ancestral protein-coding genes that have lost their coding potential known as GLCPs (gene with lost coding potential). Their results indicated that the vast majority of lncRNA genes came from noncoding regions of the genome that acquired transcription ability, with only a small percentage arising from genes that were originally protein-coding. The main lncRNA subject of this dissertation, a lncRNA known as *JPX*, is among the GLCPs identified in this dataset. Hezroni *et al*. found an open reading frame (ORF) in the exon 1 of *JPX* containing a translation initiator sequence that represses translation of downstream ORFs. This sequence appears to be inherited from an ancestral protein-coding gene known as *UspL* in chickens (Hezroni et al., 2017).

1.3 LncRNA Function

Over the years, there have been many lncRNAs implicated in several biological processes. They have been shown to be involved in the regulation of chromatin structure and function as well as transcriptional regulation. Additionally, they can act as decoys and scaffolds for RNA and protein (Ransohoff et al., 2018). Due to how many biological processes lncRNAs are involved in, it makes sense that they have also been

implicated in all of the hallmarks of cancer (Gutschner and Diederichs, 2012; Schmitt and Chang, 2016), making the study of lncRNA function and mechanism crucial to the understanding of diseases such as cancer.

1.3.1 Function and Structure

Though there are various ways that lncRNA genes have gained transcriptional ability and a general lack of sequence conservation across species, there are still lncRNAs that retain function across species. However, the mechanism through which function is conferred is still being debated. One promising hypothesis is that secondary structure and secondary structural features could be what impart function rather than sequence in the case of lncRNAs. There have been several studies in recent years that support this hypothesis (Beniaminov et al., 2008; Fang et al., 2015; Ilik et al., 2013; Johnsson et al., 2014; Liu et al., 2017; Novikova et al., 2012; Smola et al., 2016; Uroda et al., 2019).

One such study by Beniaminov *et al*. investigated two lncRNAs known as *HAR1F* and *HAR1R* within humans and chimpanzees, and suggested that mutations in the human lineage of the two lncRNA's HAR1 (human accelerated region 1) segment stabilized their secondary structures (Beniaminov et al., 2008). In another recent study by Uroda *et al*., they used RNA probing and sequencing techniques to characterize the secondary and tertiary structure of a lncRNA known as *MEG3* (human maternally expressed gene 3). *MEG3* was found to contain two highly conserved regions with complementary sequences that interact in a pseudoknot kiss. When either of these regions were mutated, and the secondary and tertiary structure was disrupted, the

lncRNA lost its function *in vivo*. This indicates that these regions are what confer function for the lncRNA *MEG3* (Uroda et al., 2019).

However, in Chapter 3 of this dissertation you will find that it is possible for function to be retained despite lack of structural conservation. This is why there is a need for more comprehensive functional studies dedicated to uncovering how lncRNA function is retained across species and whether there is anything in the lncRNA structure—both secondary and tertiary—that is necessary for function.

1.3.2 Comprehensive Functional Studies and Cancer

In the past decade, though, there have been some excellent examples of comprehensive functional studies. Many have utilized global sequencing techniques to first identify noncoding RNAs, followed by functional studies of a few of the noncoding RNAs found. The studies described in this section have been done using cancer patient samples and cell lines allowing researchers to see marked differences in the global expression levels of lncRNAs, providing evidence for their functionality (Fish et al., 2018; Wang et al., 2018). Additionally, there have been studies that have uncovered new pathways connected to known and well-studied lncRNA such as *XIST* (X-inactive specific transcript), as discussed further below (Xing et al., 2018; Yildirim et al., 2013).

A study by Wang *et al*. compared hepatocellular carcinoma (HCC) patient samples to nearby tissues to determine whether there was dysregulation of any lncRNAs between the two samples. Initially, they found 713 differentially expressed lncRNAs and then narrowed this list to lncRNAs that were transcribed from microRNA host genes. This left 5 lncRNAs, with only one being significantly downregulated in the

HCC samples—lncRNA *miR503HG*. Through rigorous experimentation, this group was able to find the mechanism through which this lncRNA is involved in the suppression of HCC metastasis. In normal tissue, lncRNA *miR503HG* forms a complex with a heterogeneous nuclear protein (hnRNP) known as hnRNPA2B1. This protein is then marked for ubiquitination and can no longer stabilize p52 and p65 mRNAs, which code for two subunits of NF-κB, and in turn suppressing this metastasis-associated pathway. Wang *et al*. were able to find a lncRNA that suppresses metastasis in HCC and is a good HCC prognostic marker candidate that can potentially better inform treatment of patients (Karner and Sun, 2018; Wang et al., 2018).

Another comprehensive noncoding RNA study by Fish *et al*. discovered a new class of breast cancer specific small noncoding RNA they termed orphan noncoding RNA (oncRNA). They had performed small RNA-seq on 8 human breast cancer cell lines, compared them to human mammary epithelial cells (nontransformed), and discovered that this group of noncoding RNA were detectable only in the 8 breast cancer cell lines. One of these oncRNAs was expressed from the 3' end of *TERC* (the RNA component of telomerase), and was found to regulate gene expression and be prometastatic. They named this oncRNA *T3p*. Interestingly, Fish *et al*. were able to detect this oncRNA in the circulating and extracellular vesicle compartments shed from the cancer cells. *T3p* was present at a level that could be detected in liquid biopsies, making this a good diagnostic biomarker. Additionally, they took advantage of the expression profile difference of oncRNA in normal versus breast cancer patient samples, and through a machine learning approach, trained a gradient-boosted classifier (GBC) to identify these differences. The GBC was eventually able to identify

patient serum samples as either normal or as a patient with breast cancer the majority of the time (37 out of 40) just based on the oncRNA profiles detected (Fish et al., 2018). This study was a remarkable example of how noncoding RNA in general can be useful for diagnostic purposes, including possible early detection of cancer.

Yildirim *et al*. explored the role of the well-studied lncRNA *Xist* in the development of hematological cancer in mice. *Xist* is best known as the master regulator of X chromosome inactivation (XCI). *Xist* and the mechanism of XCI will be discussed in detail in section 1.4 of this chapter. This group challenged previous research that claimed *Xist* expression was not needed for the maintenance of XCI. These studies failed to account for the long-term effects of the loss of *Xist* expression. Yildirim *et al*. used conditional knockout methods in mice to show that loss of *Xist* even after the establishment of XCI led to the development of female-specific fully penetrant aggressive hematological cancer. This study demonstrated that *Xist* expression was indeed necessary for the long-term maintenance of XCI and that *Xist* acts as a potent tumor suppressor (Yildirim et al., 2013). This result sheds light on a recurrent problem seen in human female cancers where a loss of *XIST* expression is seen to be common in more severe stages of the cancer. Therefore, *Xist/XIST* could potentially be acting as a tumor suppressor in general.

A recent study by Xing *et al*. investigated this issue in breast cancer patients. The dysregulation of *XIST* expression was skewed in breast cancer patients with metastases to the brain. *XIST* was found to be lower in these patients in comparison to other breast cancer patient samples. Utilizing both *in vitro* and *in vivo* methods, they were able to determine that *XIST* contributed to the suppression of metastasis. In breast

cancer, the loss of *XIST* led to (1) increase in *MSN* expression and in turn, *c-MET*, (2) increase in *miR-503* expression, which reprograms microglia, and (3) stimulation of EMT (epithelial-mesenchymal transition). Altogether, loss of *XIST* expression in mammary tissues perturbs the tumor suppression response and leads to more aggressive metastasis, especially to the brain. Xing *et al*. also did a drug screen using an FDA-approved list of drugs to determine if any of them could selectively kill *XIST* low cells. Fludarabrine was shown to be significantly effective in the breast cancer brain metastasis mouse model used in this study (Xing et al., 2018).

These studies demonstrate the importance of studying noncoding RNAs and understanding their role in diseases such as cancer. Comprehensive functional studies are also necessary for finding new lncRNAs and the biological pathways to which they are linked. This could potentially lead to lncRNAs being exploited for diagnostic and treatment purposes.

1.4 X Chromosome Inactivation

1.4.1 Dosage Compensation: History, Mechanism, and Purpose

In 1961, Mary Lyon discovered in placental mammals a phenomenon known as X chromosome inactivation which occurs in order to enact dosage compensation of gene expression between XX females and XY males (Lyon, 1961). This type of dosage compensation leads to silencing of one X chromosome in females, since one X chromosome is condensed into what is historically known as the Barr-body but more well-known today as the inactive X chromosome (Xi) (Jégu et al., 2017). Dosage compensation has appeared in nature and changed over time in different species

through divergent evolution. So far, three mechanisms of X chromosome dosage compensation have been described in literature. These being the silencing of one X chromosome as seen in mammals (Jégu et al., 2017), the increase in the expression from a single hemizygous X as seen in *Drosophila melanogaster* (Conrad and Akhtar, 2012), and the decrease in expression from both X chromosomes as seen in *Caenorhabditis elegans* (Albritton and Ercan, 2018). Furthermore, the number of X chromosomes silenced appears to be determined by the ploidy of the autosomes present in the nucleus of an organism with only one active X (Xa) being allowed for every two autosomes. This is crucial in the cases of supernumerary Xs, as is found in Klinefelter syndrome and Triple X syndrome, even though silencing tends to be incomplete (Li et al., 2016; Otter et al., 2010; Viana et al., 2014).

XCI transcriptionally silences one of the two X chromosomes in placental mammalian females when the lncRNA known as *XIST* is expressed and coats the entire X chromosome in *cis* (Carmona et al., 2018). During mouse embryogenesis, silencing of one X chromosome can occur randomly (random XCI) in the inner cell mass, or only on the paternal X chromosome (imprinted XCI) which occurs solely in placental tissues. In the case of marsupials, imprinted XCI is the only mechanism of XCI (Duret et al., 2006; Jégu et al., 2017; Kung et al., 2013; Moreira de Mello et al., 2010; Sahakyan et al., 2018). Random XCI leads to females being a mosaic of two different cell populations due to either X chromosome being susceptible to silencing during embryonic development (Carrel and Willard, 2005; McMahon et al., 1983). During the initiation of XCI, *XIST* first targets gene rich islands that are in close proximity to one another. *XIST* then spreads across the entire chromosome, recruiting various silencing protein

complexes in order to heterochromatinize the future inactive X (Xi) (Costanzi and Pehrson, 1998; Engreitz et al., 2014; Heard et al., 2001; Jégu et al., 2017; Kung et al., 2013; Mira-Bontenbal and Gribnau, 2016; Plath et al., 2003; Simon et al., 2011, 2013; Zhao et al., 2008).

XIST is known as the master regulator of XCI and has been shown to be both necessary and sufficient for the initiation of X inactivation (Marahrens et al., 1997; Penny et al., 1996; Wutz and Jaenisch, 2000). The *XIST* gene produces a 17kb transcript and is located in the X inactivation center. The X inactivation center can be found in both mouse and human genomes, and contains various lncRNAs that are dedicated to the regulation of XCI (Figure 1.1) (Brockdorff et al., 1992; Brown et al., 1992; Jégu et al., 2017).

As mentioned previously, Yildirim *et al*. demonstrated in mice studied for 2 years that *Xist* is needed for the long-term maintenance of XCI (Yildirim et al., 2013), in contrast to the belief that it was dispensable after initiation of XCI (Brown and Willard, 1994; Csankovszki et al., 1999). Yildirim *et al*. developed mice that had either a heterozygous or a homozygous conditional knock out of the *Xist* gene in the emerging hematopoietic stem cells at E10.5. This time point is after XCI has already been established in order to avoid embryonic lethality. The loss of lncRNA *Xist* in both knockout conditions lead to the development of blood cancer related symptoms and eventual death in mice beginning at the 1.5 month mark, with only 10% of mice surviving to the 2-year endpoint (Yildirim et al., 2013). These results clearly showed the importance of *Xist* in the mechanism of XCI dosage compensation. This, along with various other studies of lncRNAs in the X inactivation center, are important examples of

the functional necessity of lncRNAs in the regulation of biological processes and pathways within the cell (Schmitt and Chang, 2016).

Figure 1.1

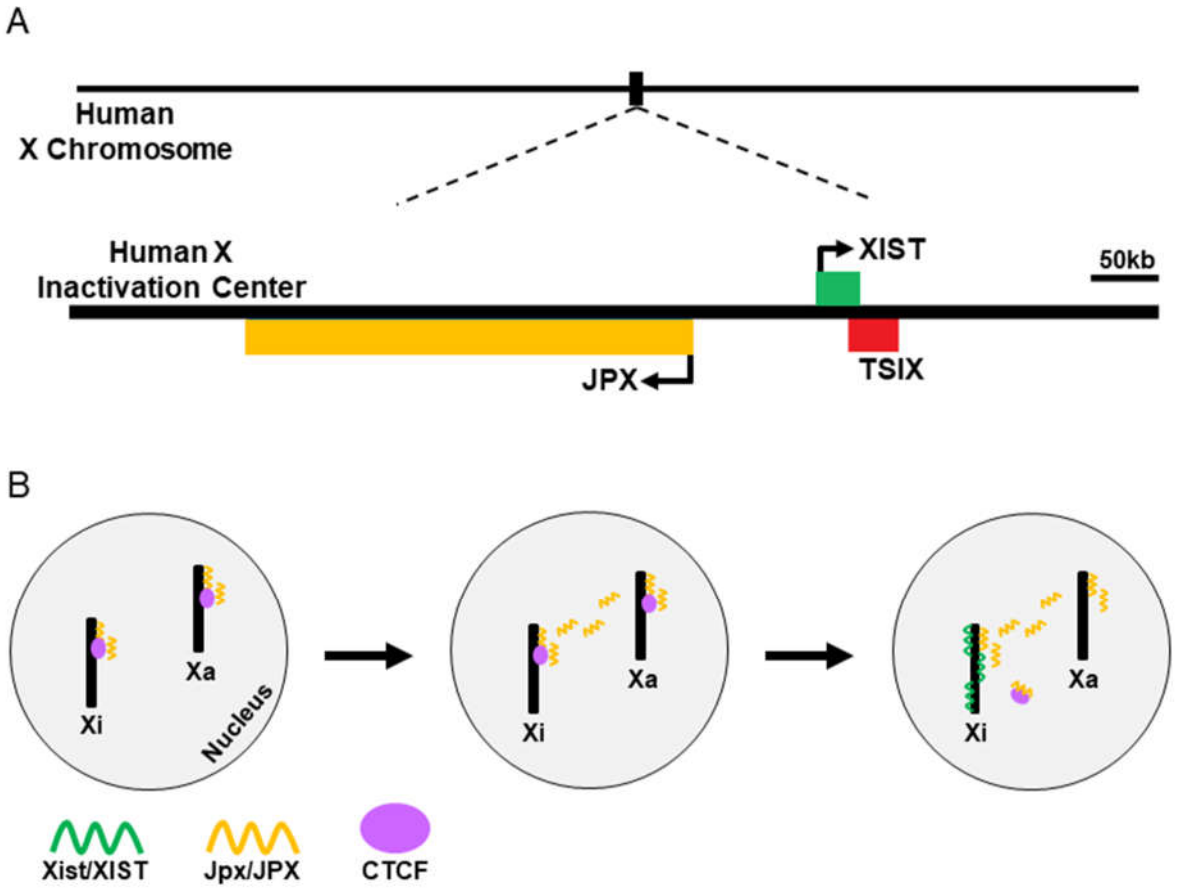


Figure 1.1: Proposed model of the initiation of X chromosome inactivation in mouse and human

(A) Simplified human X inactivation center. **(B)** Model of how lncRNA *Jpx/JPX* initiates *Xist/XIST* expression in mouse and human cells. LncRNA *Jpx/JPX* is expressed early on during embryogenesis and accumulates at the future inactive X. It then titrates away the insulator protein (CTCF) from the *Xist/XIST* promoter, allowing its expression and coating of the X chromosome to initiate XCI.

1.4.2 The Evolution of the X Inactivation Center and Dosage Compensation

As mentioned previously, dosage compensation in placental mammals is mediated through the XCI mechanism utilizing lncRNAs that originate from the X inactivation center to silence one X chromosome. The X inactivation center has been conserved in gene structure across species and is believed to have arisen from a gene cluster that was originally protein-coding (Duret et al., 2006; Elisaphenko et al., 2008; Hezroni et al., 2017; Horvath et al., 2011; Kung et al., 2013). Hezroni *et al*. describes the phenomenon of ancestral genes that lost coding potential (GLCPs). This is most likely what has occurred in the case of the X inactivation center. For example, the genes *XIST*, *FTX*, and *JPX* are the proposed lncRNA descendants of the genes *Lnx3*, *Wave 1*, and *UspL*, respectively (Hezroni et al., 2017). However, these ancestral genes do not seem to be involved in XCI and appear to perform different functions in other non-mammalian species. Despite this being the case, there have been gene discoveries that yield lncRNA transcripts which regulate X chromosome dosage compensation in other species. In *Drosophila*, *roX1* and *roX2* are the lncRNA equivalents of *Xist*. These two lncRNAs are a part of a complex known as the male-specific lethal (MSL) complex, which is responsible for dosage compensation in *Drosophila* and involves the increase in expression from the hemizygous X chromosome to match expression levels in females (Oh et al., 2003; Vensko and Stone, 2015). In opossum, a marsupial, it is interesting to note that though they are mammals, *Xist* is not present in this species as the master regulator of XCI. Instead, imprinted XCI is regulated by a lncRNA known as *Rsx* (RNA-on-the silent X). *Rsx* mediates imprinted XCI through a similar mechanism as *Xist/XIST* in mouse and human, whereby *Rsx* is solely expressed from the inactive X

chromosome and coats this same chromosome as well (Grant et al., 2012; Sado and Brockdorff, 2013). These lncRNAs appear to have developed through convergent evolution in order to regulate dosage compensation in different species without reliance on RNA sequence, demonstrating the importance of lncRNAs in a regulatory capacity.

Though dosage compensation among placental mammals in general occurs through random XCI, the mechanism of XCI over the course of embryogenesis can also be quite different between species. In mice, it has been shown that two mechanisms of XCI take place during embryogenesis. In the placental tissues, the paternal X chromosome is always silenced through imprinted XCI. The paternal X is reactivated in the inner cell mass and the two X chromosomes then undergo random XCI. In humans, only random XCI is the sole mechanism in all embryonic tissues, with both X chromosomes undergoing X chromosome dampening during preimplantation and then undergoing random XCI during post implantation (Moreira de Mello et al., 2010; Petropoulos et al., 2016; Sahakyan et al., 2018). So far, this dampening mechanism has only been reported in humans and *Caenorhabditis elegans* (Albritton and Ercan, 2018; Sahakyan et al., 2018). In comparing mouse and human X inactivation centers, one will see that the majority of the genes are in the same loci. When looking deeper though, the sequences of the genes are not conserved, which would normally mean a loss of function in the case of proteins. However, for lncRNA *Xist/XIST*, which has an overall homology of 49% between mouse and human (Nesterova et al., 2001), it is still capable of being the master regulator of XCI in both species. On the other hand, when studying lncRNA *Tsix*, a necessary negative regulator of XCI in mice (Lee et al., 1999), the gene in humans only produces a truncated transcript that seemingly plays no role in the

negative regulation of XCI (Chureau et al., 2002; Migeon et al., 2002). LncRNA regulatory networks appear to be more robust against mutations, supporting the hypothesis that lncRNAs may depend more on structure to retain function rather than sequence conservation (Beniaminov et al., 2008; Johnsson et al., 2014). These discrepancies between mouse and human bring up interesting questions about the evolution of lncRNAs and their role in regulation and disease.

1.4.3 Mouse vs. Human XCI Regulation

With the differences seen between the X inactivation centers and dosage compensation mechanisms of mouse and human, it is important to investigate how XCI is activated for the two species in order to fully understand its regulation. In mice, both Tian *et al*. and Sun *et al*. have worked out a potential mechanism of activation (Sun et al., 2013; Tian et al., 2010). It is believed that a zinc-finger protein known as CTCF has several binding sites within the promoter region of the *Xist* gene in mice (Bell et al., 1999; Essien et al., 2009; Navarro et al., 2006; Ohlsson et al., 2001; Pugacheva et al., 2005; Sheardown et al., 1997). There are two binding motifs for CTCF binding, one in the P2 promoter and one located 1kb downstream of the P1 promoter (Essien et al., 2009; Navarro et al., 2006; Sheardown et al., 1997). However, CTCF has also been found to bind the P1 promoter as well, though there is no CTCF binding motif present (Pugacheva et al., 2005; Sun et al., 2013). Sun *et al*. demonstrated that *in vitro* only the P2 and the downstream region of P1 were bound, while all three loci were bound *in vivo*. It was suggested that this binding may occur through some other methods like chromatin looping, bringing P2 and CTCF in close proximity to the P1 locus. P2 was

found to have allelic differences in CTCF occupation during XCI, with CTCF having preferential binding to P2 on the active X chromosome (Sun et al., 2013). This allelic preference indicated that CTCF plays a role in suppressing *Xist* expression. However, there was still the question of how is CTCF removed from the inactive X chromosome.

In 2010, Tian *et al*. published research indicating that an RNA-based switch controlled the expression of *Xist,* and in turn XCI. This RNA was a lncRNA known as *Jpx* (Just proximal to Xist). The *Jpx* gene and its transcript were investigated using a heterozygous knockout cell line and knockdown of the transcript was done with shRNA in another cell line, which demonstrated that downregulation of *Jpx* expression could cause dramatic phenotypes in mouse embryonic stem cells (mESCs). The phenotype presented as massive cell death, dysmorphic embryoid bodies, and downregulation of *Xist* expression—a surprising result since only one *Jpx* allele was removed and the second appeared to be unable to reach a threshold for activation of *Xist* transcription (Tian et al., 2010). Sun *et al*. utilized cell lines from this study to test the relationship between *Jpx* lncRNA mediated activation and CTCF occupation of the P2 *Xist* promoter. In a comparison of wild type cells to the *Jpx+/-* haploinsufficient cell line that had allele specific XCI, it was shown that CTCF had persistent binding to the P2 promotor in addition to being unable to upregulate *Xist* expression (Sun et al., 2013; Tian et al., 2010). When CTCF was overexpressed, this reduced *Xist* expression and inhibited embryoid body outgrowth. Transiently expressing *Jpx* rescued this phenotype (Sun et al., 2013). These findings indicated that *Jpx* and CTCF were working antagonistically to one another to regulate XCI activation. Further experiments revealed that *Jpx* was actually binding to CTCF and titrating it away from the *Xist* P2 promoter to allow for

transcription of the master regulator, and in turn XCI, to occur (Sun et al., 2013). These two studies together presented a model for the mechanism of activation of XCI in mice.

There have been some opposing views to this mechanism, and more specifically *Jpx* being necessary to XCI, as demonstrated previously (Sun et al., 2013; Tian et al., 2010). In a 2014 paper by Barakat *et al*., researchers knocked out *Rnf12* (a gene that is upstream of X inactivation center) and *Jpx*, along with several other genes located in the X inactivation center in mouse embryonic stem cells. These cells were still able to activate *Xist* expression and undergo XCI without any morphological defects (Barakat et al., 2014). However, with the loss of so many genes, including *Tsix*, from the X inactivation center in the heterozygous knockout cell line may have forced nonrandom XCI in the cell, bypassing the need for *Jpx* to activate XCI (Lee and Lu, 1999). Furthermore, when only *Jpx* and *Rnf12* were knocked out, cells suffered a decrease in *Xist* expression when compared to wild type cells and the *Rnf12* knockout cell line (Barakat et al., 2014). Later studies have indicated that Rnf12 protein is dispensable for random XCI and is only necessary for imprinted XCI (Shin et al., 2014; Wang et al., 2017). Therefore, *Jpx* still appears to be the best candidate for the activation of random XCI.

In humans, the mechanism of XCI is still being investigated. Though it is well known that *XIST* does still coat the X chromosome and is the master regulator of XCI, the regulation of this mechanism is poorly understood in humans (Figure 1.1). *Tsix* plays a crucial role in the negative regulation of *Xist* expression on the active X chromosome in mice (Lee and Lu, 1999). *TSIX* in humans on the other hand, does not play a role in the negative regulatory arm of *XIST* expression, with the gene only

producing a truncated transcript at very low levels (Chureau et al., 2002; Migeon et al., 2002). In the case of the positive regulatory arm of *XIST* expression, *JPX* gene structure is relatively conserved between human and mouse, and produces a 343nt transcript while the mouse *Jpx* gene produces a 383nt transcript (Chureau et al., 2002) (see also Chapter 3). Moreover, CTCF appears to have a binding site in promoter P1 of the human *XIST* gene, though its purpose in XCI is still not fully known. It has been hypothesized that it may have a positive regulatory function for *XIST* expression; however, this was once thought to be the case for mice and was determined to be incorrect (Pugacheva et al., 2005; Sun et al., 2013). Therefore, more research is needed to determine the activation arm of XCI in humans and whether it is conserved between mouse and humans. The results of such a study could influence how diseases are diagnosed and treated.

The research in this dissertation has been dedicated to increasing our understanding of the activation of XCI in humans. The focus of my research was to characterize the proposed activator of *XIST* expression, namely lncRNA *JPX*. Chapter 3 describes a study that we have published in the *Journal of Molecular Biology*, exploring human *JPX* function in comparison to its mouse homolog. Since it was very clear from alignments across the two species that nucleotide sequence was overall not conserved, we used selective 2'-hydroxyl acylation analyzed by primer extension (SHAPE) to probe the RNA structures of the two lncRNAs. In general, the two were highly divergent, but had highly structured secondary structures that consisted of mostly base-paired nucleotides and hairpin loops. Despite this divergence between the two lncRNAs,

human *JPX* was able to bind CTCF as well as complement the heterozygous loss of mouse *Jpx* in mouse embryonic stem cells.

The differences seen in the sequence and structure of human *JPX* is likely due to adaptive evolution in the human lineage and could indicate possible gain of function in humans. Chapter 4 describes the start of a study to explore this idea by studying *JPX* and *XIST* in human ovarian cancer. As mentioned above, there is a phenomenon seen in female cancers where patients with more severe grades of the cancer have lower expression of *XIST*. It has also been found in other cancers that downregulation of *JPX* expression is associated with poor prognosis. Therefore, I wanted to use this cancer model to uncover possible biological pathways that are being disrupted by the loss of these two lncRNAs. So far, through qPCR analysis of cancer-associated genes, I have found that in an immortalized ovarian cell line which has lost expression of *JPX* and *XIST*, that there is an increase in *MSN* and a decrease in *ATRX* expression. In contrast, *MSN* and *KDM6A* were downregulated when *XIST* alone was knocked down in an ovarian cancer cell line. An increase in proliferation and accelerated migration was still seen in these *XIST* knockdown cells, indicating a possible role as a tumor suppressor.

My research of human lncRNA *JPX* has led to a better understanding of its function and has laid the foundation for the exploration of its role in other biological pathways.

1.5 References

Albritton, S.E., and Ercan, S. (2018). Caenorhabditis elegans Dosage Compensation: Insights into Condensin-Mediated Gene Regulation. Trends Genet. *34*, 41–53.

Barakat, T.S., Loos, F., van Staveren, S., Myronova, E., Ghazvini, M., Grootegoed, J.A., and Gribnau, J. (2014). The Trans-Activator RNF12 and Cis-Acting Elements Effectuate X Chromosome Inactivation Independent of X-Pairing. Mol. Cell *53*, 965–978.

Bell, A.C., West, A.G., and Felsenfeld, G. (1999). The Protein CTCF Is Required for the Enhancer Blocking Activity of Vertebrate Insulators. Cell *98*, 387–396.

Beniaminov, A., Westhof, E., and Krol, A. (2008). Distinctive structures between chimpanzee and humanin a brain noncoding RNA. RNA *14*, 1270–1275.

Brockdorff, N., Ashworth, A., Kay, G.F., McCabe, V.M., Norris, D.P., Cooper, P.J., Swift, S., and Rastan, S. (1992). 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.

Brown, C.J., and Willard, H.F. (1994). The human X-inactivation centre is not required for maintenance of X-chromosome inactivation. Nature *368*, 154–156.

Brown, C.J., Hendrich, B.D., Rupert, J.L., Lafrenière, R.G., Xing, Y., Lawrence, J., and Willard, H.F. (1992). The human XIST gene: Analysis of a 17 kb inactive X-specific RNA that contains conserved repeats and is highly localized within the nucleus. Cell *71*, 527–542.

Cabili, M.N., Trapnell, C., Goff, L., Koziol, M., Tazon-Vega, B., Regev, A., and Rinn, J.L. (2011). Integrative annotation of human large intergenic noncoding RNAs reveals global properties and specific subclasses. Genes Dev. *25*, 1915–1927.

Carmona, S., Lin, B., Chou, T., Arroyo, K., and Sun, S. (2018). LncRNA Jpx induces Xist expression in mice using both trans and cis mechanisms. PLOS Genet. *14*, e1007378.

Carrel, L., and Willard, H.F. (2005). X-inactivation profile reveals extensive variability in X-linked gene expression in females. Nature *434*, 400–404.

Chureau, C., Prissette, M., Bourdet, A., Cattolico, L., Jones, L., Avner, P., and Duret, L. (2002). Comparative Sequence Analysis of the X-Inactivation Center Region in Mouse, Human, and Bovine. Genome Res. *12*, 894–908.

Conrad, T., and Akhtar, A. (2012). Dosage compensation in Drosophila melanogaster: epigenetic fine-tuning of chromosome-wide transcription. Nat. Rev. Genet. *13*, 123–134.

Costanzi, C., and Pehrson, J.R. (1998). Histone macroH2A1 is concentrated in the inactive X chromosome of female mammals. Nature *393*, 599–601.

Csankovszki, G., Panning, B., Bates, B., Pehrson, J.R., and Jaenisch, R. (1999). Conditional deletion of Xist disrupts histone macroH2A localization but not maintenance of X inactivation. Nat. Genet. *22*, 323–324.

Dunham, I., Kundaje, A., Aldred, S.F., Collins, P.J., Davis, C.A., Doyle, F., Epstein, C.B., Frietze, S., Harrow, J., Kaul, R., et al. (2012). An integrated encyclopedia of DNA elements in the human genome. Nature *489*, 57–74.

Duret, L., Chureau, C., Samain, S., Weissanbach, J., and Avner, P. (2006). The Xist RNA gene evolved in eutherians by pseudogenization of a protein-coding gene. Science *312*, 1653–1655.

Elisaphenko, E.A., Kolesnikov, N.N., Shevchenko, A.I., Rogozin, I.B., Nesterova, T.B., Brockdorff, N., and Zakian, S.M. (2008). A Dual Origin of the Xist Gene from a Protein-Coding Gene and a Set of Transposable Elements. PLOS One *3*, e2521.

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

Essien, K., Vigneau, S., Apreleva, S., Singh, L.N., Bartolomei, M.S., and Hannenhalli, S. (2009). CTCF binding site classes exhibit distinct evolutionary, genomic, epigenomic and transcriptomic features. Genome Biol. *10*, R131.

Fang, R., Moss, W.N., Rutenberg-Schoenberg, M., and Simon, M.D. (2015). Probing Xist RNA Structure in Cells Using Targeted Structure-Seq. PLOS Genet. *11*, e1005668.

Fatica, A., and Bozzoni, I. (2014). Long non-coding RNAs: new players in cell differentiation and development. Nat. Rev. Genet. *15*, 7–21.

Fish, L., Zhang, S., Yu, J.X., Culbertson, B., Zhou, A.Y., Goga, A., and Goodarzi, H. (2018). Cancer cells exploit an orphan RNA to drive metastatic progression. Nat. Med. *24*, 1743–1751.

Grant, J., Mahadevaiah, S.K., Khil, P., Sangrithi, M.N., Royo, H., Duckworth, J., McCarrey, J.R., VandeBerg, J.L., Renfree, M.B., Taylor, W., et al. (2012). Rsx is a metatherian RNA with Xist-like properties in X-chromosome inactivation. Nature *487*, 254–258.

Gutschner, T., and Diederichs, S. (2012). The hallmarks of cancer: a long non-coding RNA point of view. RNA Biol. *9*, 703–719.

Heard, E., Rougeulle, C., Arnaud, D., Avner, P., Allis, C.D., and Spector, D.L. (2001). Methylation of Histone H3 at Lys-9 Is an Early Mark on the X Chromosome during X Inactivation. Cell *107*, 727–738.

Hezroni, H., Koppstein, D., Schwartz, M.G., Avrutin, A., Bartel, D.P., and Ulitsky, I. (2015). Principles of Long Noncoding RNA Evolution Derived from Direct Comparison of Transcriptomes in 17 Species. Cell Rep. *11*, 1110–1122.

Hezroni, H., Ben-Tov Perry, R., Meir, Z., Housman, G., Lubelsky, Y., and Ulitsky, I. (2017). A subset of conserved mammalian long non-coding RNAs are fossils of ancestral protein-coding genes. Genome Biol. *18*, 162.

Horvath, J.E., Sheedy, C.B., Merrett, S.L., Diallo, A.B., Swofford, D.L., NISC Comparative Sequencing Program, Green, E.D., and Willard, H.F. (2011). Comparative analysis of the primate X-inactivation center region and reconstruction of the ancestral primate XIST locus. Genome Res. *21*, 850–862.

Ilik, I., Quinn, J., Georgiev, P., Tavares-Cadete, F., Maticzka, D., Toscano, S., Wan, Y., Spitale, R., Luscombe, N., Backofen, R., et al. (2013). Tandem Stem-Loops in roX RNAs Act Together to Mediate X Chromosome Dosage Compensation in Drosophila. Mol. Cell *51*, 156–173.

Jégu, T., Aeby, E., and Lee, J.T. (2017). The X chromosome in space. Nat. Rev. Genet. *18*, 377–389.

Johnsson, P., Lipovich, L., Grandér, D., and Morris, K. V (2014). Evolutionary conservation of long non-coding RNAs; sequence, structure, function. Biochim. Biophys. Acta *1840*, 1063–1071.

Kapranov, P., St Laurent, G., Raz, T., Ozsolak, F., Reynolds, C.P., Sorensen, P.H.B., Reaman, G., Milos, P., Arceci, R.J., Thompson, J.F., et al. (2010). The majority of total nuclear-encoded non-ribosomal RNA in a human cell is “dark matter” un-annotated RNA. BMC Biol. *8*, 149.

Karner, H.M., and Sun, S. (2018). LncRNA miR503HG is a new player in hepatocellular carcinoma metastasis. Non-Coding RNA Investig. *2*, 51–51.

Khalil, A.M., Guttman, M., Huarte, M., Garber, M., Raj, A., Rivea Morales, D., Thomas, K., Presser, A., Bernstein, B.E., van Oudenaarden, A., et al. (2009). Many human large intergenic noncoding RNAs associate with chromatin-modifying complexes and affect gene expression. Proc. Natl. Acad. Sci. *106*, 11667–11672.

Kung, J.T.Y., Colognori, D., and Lee, J.T. (2013). Long Noncoding RNAs: Past, Present, and Future. Genetics *193*, 651–669.

Lee, J.T., and Lu, N. (1999). Targeted Mutagenesis of Tsix Leads to Nonrandom X Inactivation. Cell *99*, 47–57.

Lee, J., Davidow, L.S., and Warshawsky, D. (1999). Tsix, a gene antisense to Xist at the X-inactivation centre. Nat. Genet. *21*, 400–404.

Li, C., Hong, T., Webb, C.-H., Karner, H., Sun, S., and Nie, Q. (2016). A self-enhanced transport mechanism through long noncoding RNAs for X chromosome inactivation. Sci. Rep. *6*, 31517.

Liu, F., Somarowthu, S., and Pyle, A.M. (2017). Visualizing the secondary and tertiary architectural domains of lncRNA RepA. Nat. Chem. Biol. *13*, 282–289.

Lyon, M.F. (1961). Gene Action in the X-chromosome of the Mouse (Mus musculus L.). Nature *190*, 372–373.

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

Mattick, J.S., and Rinn, J.L. (2015). Discovery and annotation of long noncoding RNAs. Nat. Struct. Mol. Biol. *22*, 5–7.

McMahon, a, Fosten, M., and Monk, M. (1983). X-chromosome inactivation mosaicism in the three germ layers and the germ line of the mouse embryo. J. Embryol. Exp. Morphol. *74*, 207–220.

Migeon, B.R., Lee, C.H., Chowdhury, A.K., and Carpenter, H. (2002). Species Differences in TSIX/Tsix Reveal the Roles of These Genes in X-Chromosome Inactivation. Am. J. Hum. Genet. *71*, 286–293.

Mira-Bontenbal, H., and Gribnau, J. (2016). New Xist -Interacting Proteins in X-Chromosome Inactivation. Curr. Biol. *26*, R338–R342.

Moreira de Mello, J.C., Araújo, É.S.S. de, Stabellini, R., Fraga, A.M., Souza, J.E.S. de, Sumita, D.R., Camargo, A.A., and Pereira, L. V. (2010). Random X Inactivation and Extensive Mosaicism in Human Placenta Revealed by Analysis of Allele-Specific Gene Expression along the X Chromosome. PLOS One *5*, e10947.

Navarro, P., Page, D.R., Avner, P., and Rougeulle, C. (2006). Tsix-mediated epigenetic switch of a CTCF-flanked region of the Xist promoter determines the Xist transcription program. Genes Dev. *20*, 2787–2792.

Necsulea, A., and Kaessmann, H. (2014). Evolutionary dynamics of coding and non-coding transcriptomes. Nat. Rev. Genet. *15*, 734–748.

Nesterova, T.B., Ya. Slobodyanyuk, S., Elisaphenko, E.A., Shevchenko, A.I., Johnston, C., Pavlova, M.E., Rogozin, I.B., Kolesnikov, N.N., Brockdorff, N., and Zakian, S.M. (2001). Characterization of the Genomic Xist Locus in Rodents Reveals Conservation of Overall Gene Structure and Tandem Repeats but Rapid Evolution of Unique Sequence. Genome Res. *11*, 833–849.

Novikova, I. V., Hennelly, S.P., and Sanbonmatsu, K.Y. (2012). Structural architecture of the human long non-coding RNA, steroid receptor RNA activator. Nucleic Acids Res. *40*, 5034–5051.

Oh, H., Park, Y., and Kuroda, M.I. (2003). Local spreading of MSL complexes from roX genes on the Drosophila X chromosome. Genes Dev. *17*, 1334–1339.

Ohlsson, R., Renkawitz, R., and Lobanenkov, V. (2001). CTCF is a uniquely versatile transcription regulator linked to epigenetics and disease. Trends Genet. *17*, 520–527.

Otter, M., Schrander-Stumpel, C.T., and Curfs, L.M. (2010). Triple X syndrome: a review of the literature. Eur. J. Hum. Genet. *18*, 265–271.

Penny, G.D., Kay, G.F., Sheardown, S.A., Rastan, S., and Brockdorff, N. (1996). Requirement for Xist in X chromosome inactivation. Nature *379*, 131–137.

Perry, R.B.-T., and Ulitsky, I. (2016). The functions of long noncoding RNAs in development and stem cells. Development *143*, 3882–3894.

Petropoulos, S., Edsgärd, D., Reinius, B., Deng, Q., Panula, S.P., Codeluppi, S., Plaza Reyes, A., Linnarsson, S., Sandberg, R., and Lanner, F. (2016). Single-Cell RNA-Seq Reveals Lineage and X Chromosome Dynamics in Human Preimplantation Embryos. Cell *165*, 1012–1026.

Plath, K., Fang, J., Mlynarczyk-Evans, S.K., Cao, R., Worringer, K.A., Wang, H., de la Cruz, C.C., Otte, A.P., Panning, B., and Zhang, Y. (2003). Role of histone H3 lysine 27 methylation in X inactivation. Science *300*, 131–135.

Ponjavic, J., Ponting, C.P., and Lunter, G. (2007). Functionality or transcriptional noise? Evidence for selection within long noncoding RNAs. Genome Res. *17*, 556–565.

Ponting, C.P., and Lunter, G. (2006). Signatures of adaptive evolution within human non-coding sequence. Hum. Mol. Genet. *15*, R170–R175.

Prensner, J.R., and Chinnaiyan, A.M. (2011). The Emergence of lncRNAs in Cancer Biology. Cancer Discov. *1*, 391–407.

Pugacheva, E.M., Tiwari, V.K., Abdullaev, Z., Vostrov, A.A., Flanagan, P.T., Quitschke, W.W., Loukinov, D.I., Ohlsson, R., and Lobanenkov, V. V. (2005). Familial cases of point mutations in the XIST promoter reveal a correlation between CTCF binding and pre-emptive choices of X chromosome inactivation. Hum. Mol. Genet. *14*, 953–965.

Rands, C.M., Meader, S., Ponting, C.P., and Lunter, G. (2014). 8.2% of the Human Genome Is Constrained: Variation in Rates of Turnover across Functional Element Classes in the Human Lineage. PLOS Genet. *10*, e1004525.

Ransohoff, J.D., Wei, Y., and Khavari, P.A. (2018). The functions and unique features of long intergenic non-coding RNA. Nat. Rev. Mol. Cell Biol. *19*, 143–157.

Rinn, J.L., and Chang, H.Y. (2012). Genome Regulation by Long Noncoding RNAs. Annu. Rev. Biochem. *81*, 145–166.

Sado, T., and Brockdorff, N. (2013). Advances in understanding chromosome silencing by the long non-coding RNA Xist. Philos. Trans. R. Soc. B *368*, 20110325.

Sahakyan, A., Yang, Y., and Plath, K. (2018). The Role of Xist in X-Chromosome Dosage Compensation. Trends Cell Biol. *28*, 999–1013.

Schmitt, A.M., and Chang, H.Y. (2016). Long Noncoding RNAs in Cancer Pathways. Cancer Cell *29*, 452–463.

Sheardown, S.A., Duthie, S.M., Johnston, C.M., Newall, A.E.T., Formstone, E.J., Arkell, R.M., Nesterova, T.B., Alghisi, G.-C., Rastan, S., and Brockdorff, N. (1997). Stabilization of Xist RNA Mediates Initiation of X Chromosome Inactivation. Cell *91*, 99–107.

Shin, J., Wallingford, M.C., Gallant, J., Marcho, C., Jiao, B., Byron, M., Bossenz, M., Lawrence, J.B., Jones, S.N., Mager, J., et al. (2014). RLIM is dispensable for X-chromosome inactivation in the mouse embryonic epiblast. Nature *511*, 86–89.

Simon, M.D., Wang, C.I., Kharchenko, P. V., West, J.A., Chapman, B.A., Alekseyenko, A.A., Borowsky, M.L., Kuroda, M.I., and Kingston, R.E. (2011). The genomic binding sites of a noncoding RNA. Proc. Natl. Acad. Sci. *108*, 20497–20502.

Simon, M.D., Pinter, S.F., Fang, R., Sarma, K., Rutenberg-Schoenberg, M., Bowman, S.K., Kesner, B. a, Maier, V.K., Kingston, R.E., and Lee, J.T. (2013). High-resolution Xist binding maps reveal two-step spreading during X-chromosome inactivation. Nature *504*, 465–469.

Smola, M.J., Christy, T.W., Inoue, K., Nicholson, C.O., Friedersdorf, M., Keene, J.D., Lee, D.M., Calabrese, J.M., and Weeks, K.M. (2016). SHAPE reveals transcript-wide interactions, complex structural domains, and protein interactions across the Xist lncRNA in living cells. Proc. Natl. Acad. Sci. *113*, 10322–10327.

Struhl, K. (2007). Transcriptional noise and the fidelity of initiation by RNA polymerase II. Nat. Struct. Mol. Biol. *14*, 103–105.

Sun, S., Del Rosario, B.C., Szanto, A., Ogawa, Y., Jeon, Y., and Lee, J.T. (2013). Jpx RNA Activates Xist by Evicting CTCF. Cell *153*, 1537–1551.

Tian, D., Sun, S., and Lee, J.T. (2010). The Long Noncoding RNA, Jpx, Is a Molecular Switch for X Chromosome Inactivation. Cell *143*, 390–403.

Ulitsky, I., and Bartel, D.P. (2013). lincRNAs: Genomics, Evolution, and Mechanisms. Cell *154*, 26–46.

Ulitsky, I., Shkumatava, A., Jan, C.H., Sive, H., and Bartel, D.P. (2011). Conserved Function of lincRNAs in Vertebrate Embryonic Development despite Rapid Sequence Evolution. Cell *147*, 1537–1550.

Uroda, T., Anastasakou, E., Rossi, A., Teulon, J.-M., Pellequer, J.-L., Annibale, P., Pessey, O., Inga, A., Chillón, I., and Marcia, M. (2019). Conserved Pseudoknots in lncRNA MEG3 Are Essential for Stimulation of the p53 Pathway. Mol. Cell *75*, 982-995.e9.

Vensko, S.P., and Stone, E.A. (2015). Recent progress and open questions in Drosophila dosage compensation. Fly *9*, 29–35.

Viana, J., Pidsley, R., Troakes, C., Spiers, H., Wong, C.C.Y., Al-Sarraj, S., Craig, I., Schalkwyk, L., and Mill, J. (2014). Epigenomic and transcriptomic signatures of a Klinefelter syndrome (47,XXY) karyotype in the brain. Epigenetics *9*, 587–599.

Wang, F., McCannell, K.N., Bošković, A., Zhu, X., Shin, J., Yu, J., Gallant, J., Byron, M., Lawrence, J.B., Zhu, L.J., et al. (2017). Rlim -Dependent and -Independent Pathways for X Chromosome Inactivation in Female ESCs. Cell Rep. *21*, 3691–3699.

Wang, H., Liang, L., Dong, Q., Huan, L., He, J., Li, B., Yang, C., Jin, H., Wei, L., Yu, C., et al. (2018). Long noncoding RNA miR503HG, a prognostic indicator, inhibits tumor metastasis by regulating the HNRNPA2B1/NF-κB pathway in hepatocellular carcinoma. Theranostics *8*, 2814–2829.

Weakley, S.M., Wang, H., Yao, Q., and Chen, C. (2011). Expression and Function of a Large Non-coding RNA Gene XIST in Human Cancer. World J. Surg. *35*, 1751–1756.

Wu, H., Yang, L., and Chen, L.-L. (2017). The Diversity of Long Noncoding RNAs and Their Generation. Trends Genet. *33*, 540–552.

Wutz, A., and Jaenisch, R. (2000). A Shift from Reversible to Irreversible X Inactivation Is Triggered during ES Cell Differentiation. Mol. Cell *5*, 695–705.

Xing, F., Liu, Y., Wu, S.-Y., Wu, K., Sharma, S., Mo, Y.-Y., Feng, J., Sanders, S., Jin, G., Singh, R., et al. (2018). Loss of XIST in Breast Cancer Activates MSN-c-Met and Reprograms Microglia via Exosomal miRNA to Promote Brain Metastasis. Cancer Res. *78*, 4316–4330.

Yildirim, E., Kirby, J.E., Brown, D.E., Mercier, F.E., Sadreyev, R.I., Scadden, D.T., and Lee, J.T. (2013). Xist RNA Is a Potent Suppressor of Hematologic Cancer in Mice. Cell *152*, 727–742.

Zhao, J., Sun, B.K., Erwin, J.A., Song, J.-J., and Lee, J.T. (2008). Polycomb Proteins Targeted by a Short Repeat RNA to the Mouse X Chromosome. Science *322*, 750–756.

CHAPTER 2: Materials and Methods

2.1 Materials and Methods

Plasmid Preparation

The Tg(EF1α:JpxE1-E3) construct was generated by PCR-cloning the *Jpx* transcript out of the cDNA prepared from the total RNA of differentiated mES cells. The *Jpx E1-E3* isoform was cloned into the pEF1/V5-His mammalian expression vector (Invitrogen Cat# V92020), which contains an EF-1α promoter for mammalian expression and a T7 promoter for *in vitro* transcription. In parallel, the Tg(EF1α:JPXE1-E3) construct was generated by PCR-cloning the *JPX* transcript out of the cDNA prepared from the total RNA of human SKOV3iP1 ovarian cancer cells. The *JPX E1-E3* isoform was cloned into the pEF1/V5-His mammalian expression vector the same way as *Jpx E1-E3*.

RNA Electrophoretic Mobility Shift Assay (EMSA)

RNA EMSA was carried out as previously described (Cifuentes-Rojas et al., 2011; Hellman and Fried, 2007; Sun et al., 2013) using *in vitro* transcribed RNAs uniformly labeled with ATP[α-32p] and purified recombinant CTCF protein. Specifically, *Jpx E1-E3 1-383* and truncated RNAs, or *JPX E1-E3* full-length and truncated RNAs, were *in vitro* transcribed using T7 polymerase and DNA templates PCR-amplified from Tg(EF1α:JpxE1-E3) or Tg(EF1α:JPXE1-E3) cDNA plasmid. The control RNA, a 316nt *drz-Agam-2-1* ribozyme RNA, was prepared using genomic DNA of *Anopheles gambiae* (Webb et al., 2009). Primers used are as follows:

For *Jpx E1-E3 1-383* and truncated RNAs

JW_1F –

TTCCCGCGAAATTAATACGACTCACTATAGGGAGATGGCGGCGTCCACATGTAT

JW_2R – AGGTGGCAGGCAGCAGGCAT

JW_4R – ATAAGCAAGCTAGTACGCAC

JW_5F –

TTCCCGCGAAATTAATACGACTCACTATAGGGAGTGGCCAATTAATGAACAT

JW_21F –

TTCCCGCGAAATTAATACGACTCACTATAGGGAGCCACGGCACCACCAGGCTTC

JW_22R – GAGTTTATTTGGGCTTACAG

For *JPX E1-E3* full-length and truncated RNAs

hJPX-EMSA-T7+F2 –

TTCCCGCGAAATTAATACGACTCACTATAGGGAGGGAAGACTTAAGATGGCGGC

hJPX-EMSA-T7+F3 –

TTCCCGCGAAATTAATACGACTCACTATAGGGAGCTTACGGGGGTTGCAAG

hJPX-EMSA-R1 – CTGTAATCTCAGCTACTCGGGAG

hJPX-EMSA-R2 – GGTCATGCCATTGCATTCC

hJPX-EMSA-R3 – AGCCTGGGCAACAAGAG

hJPX-EMSA-R4 – TCGTCAGTAGAAGTTAGGCG

For *drz-Agam-2-1* ribozyme RNA (control)

JW_14F – TTCCCGCGAAATTAATACGACTCACTATA GCTCTGCAAATGGGGTAGGA

JW_24R – GTTTTTTCGTTTGCCGTTGAAGG

Recombinant CTCF protein was prepared and purified as previously described (Sun et al., 2013). Mouse CTCF cDNA corresponding to the full-length 736 amino acids was cloned with C-terminal 6xHis tag into pFLAG-2 (Sigma). FLAG-CTCF-6xHis protein was induced in Rosetta-Gami B-cells (EMD Millipore) with 0.2 M of IPTG at room temperature and was then purified with Ni-NTA resin (Qiagen) with 50 mM sodium phosphate (pH 8.0), 300 mM NaCl, and 250 mM imidazole. Eluates were dialyzed against 50 mM Tris-HCl (pH 7.5), 2.5 mM MgCl2, 150 mM NaCl, 0.1 mM ZnSO4, 1 mM DTT, 0.1% Tween-20, and 10% glycerol.

For gel shift (EMSA), RNAs were incubated with CTCF protein and the complexes were resolved in a 5% acrylamide gel. The gel was then exposed to a phosphorimage screen (Molecular Dynamics), scanned with Typhoon phosphorimager (GE Healthcare), and analyzed with ImageQuant software (GE Healthcare). The fraction of the RNA-protein complex was plotted against the concentration of CTCF and fit with a binding equation.

In Vitro and *In Vivo* RNA SHAPE (Selective 2'-Hydroxyl Acylation analyzed by Primer Extension)

RNA SHAPE was performed as previously described (Spitale et al., 2013; Wilkinson et al., 2006). Specifically, RNAs were *in vitro* transcribed from a plasmid expressing either *Jpx* or *JPX* (Tg(EF1α:JpxE1-E3) or Tg(EF1α:JPXE1-E3), respectively). Then 2pmol of purified RNA was denatured, folded, and modified with 1μL of 1M NAI. Immediately next was reverse transcription of the RNA and primer extension with γ-32P-ATP 5'-labeled reverse primers. Four reactions were set up for each primer

extension: 1) DMSO only negative control, 2) dideoxy-ATP (ddA) in DMSO, 3) dideoxy-CTP (ddC) in DMSO, and 4) NAI in DMSO. The γ-32P-ATP end-labelled primers and ~2pmol of RNA from the modification step were added and incubated at 95°C for the 2min annealing step, followed by a 2°C/sec step-down cooling to 4°C. Reverse transcription was performed using first-strand cDNA synthesis kit containing SuperScriptIII (2 units/μL; Invitrogen, Life Technologies).

In vivo SHAPE was performed as previously described with some modifications (Kwok et al., 2013; Lucks et al., 2011; Spitale et al., 2013). Briefly, HEK293T cells transiently transfected with Tg(EF1α:JPXE1-E3) for 24 hours were collected and incubated at 37°C with NAI at a final concentration of 100mM for every 4x10^6 cells. RNA was extracted using TRIzol (Life Technologies) according to manufacturer's instructions. Reverse transcription was performed with non-radiolabeled reverse primers. In order to enrich the presence of *JPX*, cDNA fragments generated during primer extension with cold primers were amplified using LMPCR (ligation mediated polymerase chain reaction) (Kwok et al., 2013; Lucks et al., 2011). A linker sequence with a 5' phosphate and 3' 3-carbon spacer group was added to the RNA using CircLigase ssDNA ligase (epicentre) according to manufacturer's instructions. PCR amplification was performed with a non-radiolabeled forward primer targeting the linker sequence and γ-32P-ATP end-labeled reverse primers.

Samples were resolved on a 10% denaturing poly acrylamide SHAPE Gel (0.4mm). The gel was dried and placed into a phosphor-imaging cassette for exposure overnight, and was scanned using a Typhoon phosphorimager. Band intensities and SHAPE reactivities were calculated using SAFA software (Das et al., 2005; Laederach

et al., 2008). After subtracting the DMSO background from the NAI band intensities, the average of the top 10% minus the top 2% reactivities was calculated and set to 1. This was then used to normalize the SHAPE reactivities (Ilik et al., 2013). When aligning SHAPE band positions to the transcript, the ladders (ddA and ddC) generated at the reverse transcription step are 1 nucleotide longer than the corresponding DMSO negative control and NAI samples (Wilkinson et al., 2006). RNA secondary structures are predicted with Vienna RNA Software (http://rna.tbi.univie.ac.at/cgi-bin/RNAWebSuite/RNAprobing.cgi) (Lorenz et al., 2011; Washietl et al., 2012). The extension primers used are as follows:

For *Jpx*

p2r – AGGTGGCAGGCAGCAGGCAT

p3r – CTTGAACTGATGGGTGCCAT

p4r – ATAAGCAAGCTAGTACGCAC

pex1r – GGGCATGTTCATTAATTGGCCAG

p25r – TGGCTAATCCCGGGAAGGAC

p26r – CTTCAAGTCCCTGCTTGAGTTTC

For *JPX*

pr R1 – CTGTAATCTCAGCTACTCGG

pr R1-2 – AGTGAGCCAAGGTCATGCCA

pr R2 – GAAGTTAGGCGATCAGCGAG

pr R2-3 – GAGACACAATACTATTAACTGGC

pr R3 – CATACTTCGGACGCCTTGCAAC

prR3-4 – CCCCGTAAGGACGCAGTGAT

LMPCR Linker Sequence –

AGATCGGAAGAGCGGTTCAGCAGGAATGCCGAGACCGATCTCGTATGCCGTCTTC

TGCTTG

Linker Primer F – GGAAGAGCGGTTCAGCAGGA

Transfection of *Jpx E1-E3* or *JPX E1-E3* constructs in mouse ES (mES) cells

For transient transfection, wild type 16.7 female mES cells and *Jpx-/+* mutant female mES cells (Tian et al., 2010) were cultured on feeder cells in media containing LIF (leukemia inhibitory factor). The cells were then collected on differentiation Day 0 and ~$1x10^6$ cells were seeded per well of a non-tissue culture treated 6-well plate for transfection. The mES cells were then transfected with 2μg of GFP plasmid (for visual confirmation of transfection) and 2.5μg of either empty vector, Tg(EF1α:JpxE1-E3), or Tg(EF1α:JPXE1-E3) plasmid per well of a 6-well plate according to Lipofectamine 2000 (Invitrogen) manufacturer's instructions. Cells were cultured in feeder-free and LIF-free conditions for differentiation. On Day 1 of mES differentiation, 2mL of the transfection media was gently refreshed with regular mES media. Cells were viewed on a fluorescent microscope to assess GFP fluorescence and confirmation of successful transfection. Once transfections were deemed successful using Lipofectamine 2000, the GFP plasmid was no longer included during the transient transfection. For analyses, cells were transferred to tissue culture treated plates on Day 4 for EB outgrowth and allowed to grow until Day 8, with the cells being collected on Day 4 and Day 8 to assess viability and gene expression.

To obtain stable transgenic mES cell lines with transfection, wild type 16.7 female mES cells and *Jpx-/+* mutant female mES cells (Tian et al., 2010) were seeded on feeder cells in 6-well plates at $3x10^5$ cells per well in media containing LIF. The next day, cells were transfected with 2.5µg of either empty vector, Tg(EF1α:JpxE1-E3), or Tg(EF1α:JPXE1-E3) plasmid per well of a 6-well plate according to Lipofectamine 2000 (Invitrogen) manufacturer's instructions while still in the stem cell state. After 24 hours, an entire well of cells were treated with 0.25% trypsin-EDTA and transferred to a 10cm dish of Neo-resistant feeder cells. Since the plasmids contain the Neomycin-resistance gene, media was switched to G418 (400ug/mL) + LIF selection media to select for cells that had stable integration of the transgenes. The cell's media was changed every day, and after 19 days, colonies were picked and treated with 0.25% trypsin-EDTA to break up the colonies and then transferred to a 24-well plate of Neo-resistant mEFs and grown out. When confluent, half of a 24-well was stocked and the other half was used for PCR screening of successful integration of plasmid. For differentiation, ~$5x10^5$ cells were seeded per well of a non-tissue culture treated 6-well plate without LIF and allowed to grow until Day 4 when they were transferred to tissue culture treated 6-well plates for EB outgrowth. Cells were collected on Day 4 and Day 8 to assess viability and qPCR analysis.

Cell Death Assay

On Day 4 of mES differentiation, supernatant and embryoid bodies (EBs) were collected, spun down, and broken up with 0.05% Trypsin-EDTA. Cells were then resuspended in 1mL of mES media and 20µL was taken from each sample for staining

with trypan blue. Cells were counted on a Countess II FL Automated Cell Counter and cell viability was recorded. For Day 8 of mES differentiation, supernatant and attached cells from EB outgrowth were collected, spun down, and treated with 0.05% Trypsin-EDTA before resuspended in 1mL of mES media for cell count.

Cell Culture of Cancer Cell Lines

Human epithelial ovarian adenocarcinoma cell lines, SKOV3iP1 and OVCAR3, were a generous gift from Dr. Olga Razorenova (University of California, Irvine). Human female embryonic kidney cell line, HEK293T, was a generous gift from Dr. Scott Atwood (University of California, Irvine). Human male osteosarcoma cell line, SJSA-1, was a generous gift from Dr. Claudia Benavente (University of California, Irvine). The HEK293T cell line was cultured in DMEM with 10% FBS, streptomycin (100 mg/mL), and penicillin (100 U/mL). SKOV3iP1, OVCAR3, and SJSA-1 cell lines were cultured in RPMI medium with 10% FBS, streptomycin (100 mg/mL), and penicillin (100 U/mL). All cultures were grown at 37^{o}C in a 5% CO2 atmosphere.

Quantitative Real-Time PCR

Cells from one well of a 6-well plate were spun down and resuspended in 1mL TRIzol (Life Technologies) for each respective sample. RNA was extracted and residual genomic DNA was removed with TURBO DNase (Ambion, Life Technologies) treatment according to manufacturer's instructions. Reverse transcription of RNA to cDNA was performed using SuperScriptIII (Invitrogen, Life Technologies) according to manufacturer's instructions. Real-time PCR for target genes was performed using FS

Universal SYBR Green Master (Rox) (Sigma-Aldrich) according to manufacturer's instructions under the following conditions: 95°C for 10mins, 95°C for 15secs, 58°C for 30secs, 72°C for 30secs, then repeat steps 2 through 4 for 39 cycles. Primers used for PCR were as follows:

Mouse

mJpx 76 (e1)-F – TTAGCCAGGCAGCTAGAGGA

mJpx 225 (ex2)-R – AGCCGTATTCCTCCATGGTT

XistNS33-F – CAGAGTAGCGAGGACTTGAAGAG

XistBP2F-R – CCCGCTGCTGAGTGTTTGATA

Gapdh-F – ATGAATACGGCTACAGCAACAGG

Gapdh-R – GAGATGCTCAGTGTTGGGGG

Human

hJPX E1-F – AATCACTGCGTCCTTACGGG

hJPX E3-R – GCAGGAGAACCACTTGAACT

hXIST-E1-F – TTGCCCTACTAGCTCCTCGGAC

hXIST-E3-R – TTCTCCAGATAGCTGGCAACC

ATRX-F – TCCTTGCACACTCATCAGAAGAATC

ATRX-R – CGTGACGATCCTGAAGACTTGG

KDM6A-F – TACAGGCTCAGTTGTGTAACCT

KDM6A-R – CTGCGGGAATTGGTAGGCTC

PIM2-F – GGCAGCCAGCATATGGG

PIM2-R – TAATCCGCCGGTGCCTGG

MSN-F – ACCGGGAAGCAGCTATTTGA

MSN-R – GAACTTGGCACGGAACTTAA

MIDI-F – CCTGTCAACATGTTGAAGTC

MIDI-R – GCAATCTGCTGAGCCAGTTT

MPP1-F – ATTGAATACTGTGACCGAGG

MPP1-R – TTCTAGGATCTCATCCCCCA

GAPDH-F – GAGTCAACGGATTTGGTCGT

GAPDH-R – GACAAGCTTCCCGTTCTCAG

Some human primer sequences were acquired from published data. Specifically, *ATRX* (Ritchie et al., 2008), *KDM6A* (Jiang et al., 2013), *PIM2* (Jiménez-García et al., 2016), *MSN*, *MIDI*, *MIPP1* (Xing et al., 2018), and *GAPDH* (Ritchie et al., 2008).

Gene Set Enrichment Analysis (GSEA)

GSEA was performed for low and high *XIST*, *JPX*, and *MALAT1* expression patients with previously described methods utilizing ovarian cancer patient data from the cancer genome atlas (TCGA) database (Reimand et al., 2019). We used the 189 oncogenic signatures (regulatory pathways that are dysregulated in cancer) that are available through GSEA's molecular signatures database (Liberzon et al., 2011, 2015; Subramanian et al., 2005).

RNA Fluorescent *in situ* Hybridization (FISH)

Human JPX Fluorescent Cyanine3 (Enzo Life Sciences) probes were made using a Nick Translation Kit (Roche) and column purified (GE Healthcare). Human XIST Fluorescein probes were designed and purchased from Integrated DNA Technologies

(IDT). RNA FISH was performed as described previously (Lee and Lu, 1999; Namekawa and Lee, 2011; Zhang et al., 2007). For each procedure, 50,000-100,000 cells were cytospun onto slides and fixed in 4% paraformaldehyde. Probes were incubated with cells on slides for 16 hours at 37°C and DAPI was used to stain the nuclei. Images were taken on a Zeiss LSM 700 or LSM 780 confocal microscope and analyzed with Volocity software (PerkinElmer).

Lenti-viral Knockdown of *XIST*

Knockdown of *XIST* was done using the XIST-set siRNA/shRNA/RNAi Lentivector (Human) and 2nd Generation Packaging System Mix purchased from Applied Biological Materials (abm) Inc. (Richmond, BC). The Scrambled siRNA GFP Lentivector control was also purchased from abm. Lentivectors were amplified according to manufacturer's instructions. HEK293T cells were plated in 10cm plates and transfected at ~70-80% confluency with vectors according to manufacturer's instructions using Lipofectamine 2000 for 24 hours and then transferred to normal media for another 24 hours. Media on the HEK293T cells were changed and old media was filtered with a 0.45μM filter and polybrene was added at 2μg/mL. This was then applied to OVCAR3 cells plated in 6-well plates which were spun down at 1000rpm for 1.5 hours and incubated at 37°C overnight. The next day, this media was switched out for new filtered viral vector media plus polybrene twice a day for 36 hours. GFP expression was seen after 24 hours. Cells were allowed to recover for 48 hours in normal media and then positive cells were selected for in media with puromycin at 2μg/mL. Control cells were dead after 48 hours and knockdown cells were then used in future experiments.

Crystal Violet Proliferation Assay

OVCAR3 cells were plated at 12,500 cells per well of a 24-well plate and grown for 96, 120, 144, and 168 hours. Cells were then fixed and stained with 0.05% crystal violet and allowed to dry overnight before images were taken. Methanol was used to destain the cells and absorbance was measured at 595nm on a microplate reader.

Wound Healing Scratch Assay

OVCAR3 cells were plated at 50,000 cells per well of a 6-well plate. When cells reached ~80% confluency, a P1000 pipet was used to scratch a line through the center of each well. Two brightfield images were taken at the 0hr, 24hr, and 48hr time points, and the area of the gap was then measured.

2.2 References

Cifuentes-Rojas, C., Kannan, K., Tseng, L., and Shippen, D.E. (2011). Two RNA subunits and POT1a are components of Arabidopsis telomerase. Proc. Natl. Acad. Sci. *108*, 73–78.

Das, R., Laederach, A., Pearlman, S.M., Herschlag, D., and Altman, R.B. (2005). SAFA: semi-automated footprinting analysis software for high-throughput quantification of nucleic acid footprinting experiments. RNA *11*, 344–354.

Hellman, L.M., and Fried, M.G. (2007). Electrophoretic mobility shift assay (EMSA) for detecting protein–nucleic acid interactions. Nat. Protoc. *2*, 1849–1861.

Ilik, I., Quinn, J., Georgiev, P., Tavares-Cadete, F., Maticzka, D., Toscano, S., Wan, Y., Spitale, R., Luscombe, N., Backofen, R., et al. (2013). Tandem Stem-Loops in roX RNAs Act Together to Mediate X Chromosome Dosage Compensation in Drosophila. Mol. Cell *51*, 156–173.

Jiang, W., Wang, J., and Zhang, Y. (2013). Histone H3K27me3 demethylases KDM6A and KDM6B modulate definitive endoderm differentiation from human ESCs by regulating WNT signaling pathway. Cell Res. *23*, 122–130.

Jiménez-García, M.P., Lucena-Cacace, A., Robles-Frías, M.J., Narlik-Grassow, M., Blanco-Aparicio, C., and Carnero, A. (2016). The role of PIM1/PIM2 kinases in tumors of the male reproductive system. Sci. Rep. *6*, 38079.

Kwok, C.K., Ding, Y., Tang, Y., Assmann, S.M., and Bevilacqua, P.C. (2013). Determination of in vivo RNA structure in low-abundance transcripts. Nat. Commun. *4*, 2971.

Laederach, A., Das, R., Vicens, Q., Pearlman, S.M., Herschlag, D., and Altman, R.B. (2008). Semi-automated and rapid quantification of nucleic acid footprinting and structure mapping experiments. Nat. Protoc. *3*, 1395–1401.

Lee, J.T., and Lu, N. (1999). Targeted Mutagenesis of Tsix Leads to Nonrandom X Inactivation. Cell *99*, 47–57.

Liberzon, A., Subramanian, A., Pinchback, R., Thorvaldsdottir, H., Tamayo, P., and Mesirov, J.P. (2011). Molecular signatures database (MSigDB) 3.0. Bioinformatics *27*, 1739–1740.

Liberzon, A., Birger, C., Thorvaldsdóttir, H., Ghandi, M., Mesirov, J.P., and Tamayo, P. (2015). The Molecular Signatures Database Hallmark Gene Set Collection. Cell Syst. *1*, 417–425.

Lorenz, R., Bernhart, S.H., Höner zu Siederdissen, C., Tafer, H., Flamm, C., Stadler, P.F., and Hofacker, I.L. (2011). ViennaRNA Package 2.0. Algorithms. Mol. Biol. *6*, 26.

Lucks, J.B., Mortimer, S.A., Trapnell, C., Luo, S., Aviran, S., Schroth, G.P., Pachter, L., Doudna, J.A., and Arkin, A.P. (2011). Multiplexed RNA structure characterization with selective 2'-hydroxyl acylation analyzed by primer extension sequencing (SHAPE-Seq). Proc. Natl. Acad. Sci. *108*, 11063–11068.

Namekawa, S.H., and Lee, J.T. (2011). Detection of nascent RNA, single-copy DNA and protein localization by immunoFISH in mouse germ cells and preimplantation embryos. Nat. Protoc. *6*, 270–284.

Reimand, J., Isserlin, R., Voisin, V., Kucera, M., Tannus-Lopes, C., Rostamianfar, A., Wadi, L., Meyer, M., Wong, J., Xu, C., et al. (2019). Pathway enrichment analysis and visualization of omics data using g:Profiler, GSEA, Cytoscape and EnrichmentMap. Nat. Protoc. *14*, 482–517.

Ritchie, K., Seah, C., Moulin, J., Isaac, C., Dick, F., and Bérubé, N.G. (2008). Loss of ATRX leads to chromosome cohesion and congression defects. J. Cell Biol. *180*, 315–324.

Spitale, R.C., Crisalli, P., Flynn, R.A., Torre, E.A., Kool, E.T., and Chang, H.Y. (2013). RNA SHAPE analysis in living cells. Nat. Chem. Biol. *9*, 18–20.

Subramanian, A., Tamayo, P., Mootha, V.K., Mukherjee, S., Ebert, B.L., Gillette, M.A., Paulovich, A., Pomeroy, S.L., Golub, T.R., Lander, E.S., et al. (2005). Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. Proc. Natl. Acad. Sci. *102*, 15545–15550.

Sun, S., Del Rosario, B.C., Szanto, A., Ogawa, Y., Jeon, Y., and Lee, J.T. (2013). Jpx RNA Activates Xist by Evicting CTCF. Cell *153*, 1537–1551.

Tian, D., Sun, S., and Lee, J.T. (2010). The Long Noncoding RNA, Jpx, Is a Molecular Switch for X Chromosome Inactivation. Cell *143*, 390–403.

Washietl, S., Hofacker, I.L., Stadler, P.F., and Kellis, M. (2012). RNA folding with soft constraints: reconciliation of probing data and thermodynamic secondary structure prediction. Nucleic Acids Res. *40*, 4261–4272.

Webb, C.-H.T., Riccitelli, N.J., Ruminski, D.J., and Luptak, A. (2009). Widespread Occurrence of Self-Cleaving Ribozymes. Science *326*, 953–953.

Wilkinson, K.A., Merino, E.J., and Weeks, K.M. (2006). Selective 2′-hydroxyl acylation analyzed by primer extension (SHAPE): quantitative RNA structure analysis at single nucleotide resolution. Nat. Protoc. *1*, 1610–1616.

Xing, F., Liu, Y., Wu, S.-Y., Wu, K., Sharma, S., Mo, Y.-Y., Feng, J., Sanders, S., Jin, G., Singh, R., et al. (2018). Loss of XIST in Breast Cancer Activates MSN-c-Met and Reprograms Microglia via Exosomal miRNA to Promote Brain Metastasis. Cancer Res. *78*, 4316–4330.

Zhang, L.-F., Huynh, K.D., and Lee, J.T. (2007). Perinucleolar Targeting of the Inactive X during S Phase: Evidence for a Role in the Maintenance of Silencing. Cell *129*, 693–706.

CHAPTER 3: Functional Conservation of LncRNA *JPX* Despite Sequence and Structural Divergence

3.1 Abstract

Long noncoding RNAs (lncRNAs) have been identified in all eukaryotes and are most abundant in the human genome. However, the functional importance and mechanisms of action for human lncRNAs are largely unknown. Using comparative sequence, structural, and functional analyses, we characterize the evolution and molecular function of human lncRNA *JPX*. We find that human *JPX* and its mouse homolog, lncRNA *Jpx*, have deep divergence in their nucleotide sequences and RNA secondary structures. Despite such differences, both lncRNAs demonstrate robust binding to CTCF, a protein that is central to *Jpx*'s role in X chromosome inactivation. In addition, our functional rescue experiment using *Jpx*-deletion mutant cells shows that human *JPX* can functionally complement the loss of *Jpx* in mouse embryonic stem cells. Our findings support a model for functional conservation of lncRNAs independent from sequence and structural divergence. This study provides mechanistic insight into the evolution of lncRNA function.

3.2 Introduction

Long noncoding RNAs (lncRNAs) are transcripts over 200 nucleotides in length that do not code for proteins. In contrast to protein-coding transcripts, which map to only about 1.5% of the human genome, DNA sequences for lncRNA transcripts are estimated to represent 70% to 90% of the genome (Kapranov et al., 2010). Due to their low levels of expression and the general lack of functional information, lncRNAs were

regarded as transcriptional noise. Only recently has the high frequency of their occurrence in the human genome and their direct relevance to various biological processes been recognized (Fatica and Bozzoni, 2014; Kopp and Mendell, 2018; Perry and Ulitsky, 2016; Rinn and Chang, 2012; Wu et al., 2017). Specifically, lncRNAs are known to be capable of scaffolding protein complexes and recruiting chromatin modifiers for transcriptional regulation (Kopp and Mendell, 2018; Mercer and Mattick, 2013; Ransohoff et al., 2018). Gene expression profiling has revealed highly tissue-specific transcription of lncRNAs and a large number of lncRNAs that are active during animal development in humans, mice, flies, and farm animals (Derrien et al., 2012; Kern et al., 2018; Sun et al., 2013a; Wen et al., 2016). Importantly, lncRNAs have been implicated in the evolution of new genes and associated with functions in sexual reproduction (Dai et al., 2008; Gao et al., 2014; Heinen et al., 2009; Wen et al., 2016). Furthermore, lncRNA functions have been shown to be conserved during embryonic development (Kapusta and Feschotte, 2014; Ulitsky et al., 2011).

LncRNAs are present in all eukaryotes; however, the function and evolution of the vast majority of lncRNA genes still remain elusive (Haerty and Ponting, 2014; Hezroni et al., 2017; Ling et al., 2015; Necsulea et al., 2014). Unlike protein-coding genes, in which functions are mostly defined by evolutionary conserved coding sequences and their flanking regulatory elements, lncRNAs are known to have poor sequence conservation (Cabili et al., 2011; Hezroni et al., 2015; Kirk et al., 2018). Hence, it has been a challenge to uncover conserved features of lncRNAs and determine underlying mechanisms for function. RNA secondary structure is one molecular feature that has recently been recognized to be important for lncRNA

functional conservation (Delli Ponti et al., 2018; Fang et al., 2015; Ilik et al., 2013; Liu et al., 2017; Novikova et al., 2012; Smola et al., 2016; Uroda et al., 2019). Yet, the direct connection between the structure and function of lncRNAs and their implications in molecular evolution remain unclear (Johnsson and Morris, 2014; Liu et al., 2017). Defining this connection has been notoriously difficult due to the lack of structural and biochemical analysis of evolutionarily related lncRNAs and the complex nature of their interactions with protein factors.

In this chapter, we focus on lncRNAs involved in a mechanism of dosage compensation known as X chromosome inactivation (XCI) to determine whether function is conserved and if molecular features such as RNA sequence and secondary structure influence conservation. XCI is the evolutionary solution to the 1X:2X dosage imbalance between XY male and XX female mammals. Outside the lineage of modern mammals, different mechanisms are used to balance the sex chromosome gene dosage. As previously mentioned, dosage compensation has evolved independently in divergent species and frequently uses lncRNAs as key regulators (Grant et al., 2012; Payer and Lee, 2008; Straub and Becker, 2007; Wutz et al., 2002). This suggests that RNA, unlike protein, carries functional advantages that aid, and are sometimes prerequisites, for biological processes such as gene dosage controls in developing embryos. As discussed in Chapter 1, XCI is controlled by a genomic region known as the X inactivation center that encodes a cluster of lncRNAs in both human and mouse genomes (Figure 3.1) (Brockdorff et al., 1992; Brown et al., 1992). This gene cluster has evolved from a group of protein-coding genes during the divergence of eutherians and marsupials to become the home of all lncRNAs involved in XCI (Casanova et al., 2016;

Duret et al., 2006; Elisaphenko et al., 2008; Horvath et al., 2011). It is worthwhile to note that marsupials (e.g., opossums; Figure 3.1B) do not have the master regulator lncRNA *Xist* for XCI. In the same chromosome locus, marsupials carry the protein-coding gene *Lnx3*, which does not possess dosage compensation functions or affect the sex chromosome (Duret et al., 2006; Elisaphenko et al., 2008). However, a marsupial lncRNA, *Rsx*, was discovered to play the role of silencing the X chromosome in opossums. Although *Rsx* has no obvious sequence homology with *Xist*, the function of the two genes appears to be equivalent (Grant et al., 2012; Lee and Bartolomei, 2013; Sado and Brockdorff, 2013). Thus, the use of lncRNAs for control of mammalian XCI represents convergent evolution of functions that may be independent of RNA nucleotide sequences.

We took advantage of a defined molecular mechanism in mouse XCI which involves the direct binding of a lncRNA known as *Jpx* with a specific chromatin insulator protein, CTCF, to initiate XCI (Figure 3.2A) (Sun et al., 2013b). This model allowed us to identify the molecular features underlying the function and evolution of *Jpx*. In mice, *Jpx* has been shown to activate *Xist* (Carmona et al., 2018; Li et al., 2016; Sun et al., 2013b; Tian et al., 2010). By contrast, the function of its human homolog, *JPX*, is unknown (de Hoon et al., 2017; Migeon, 2011). As a lncRNA in humans, *JPX* is expressed in early female human embryos (Figure 3.3A) (Petropoulos et al., 2016). This indicates that the gene has a role in early embryogenesis and likely functions similarly to *Jpx*. Here we will compare mouse lncRNA *Jpx* with human lncRNA *JPX* and determine their homology at the levels of nucleotide sequences, RNA secondary structures, and molecular functions.

Our results indicate that despite sequence and structural divergence, the two lncRNAs function through the same biochemical mechanism.

Figure 3.1

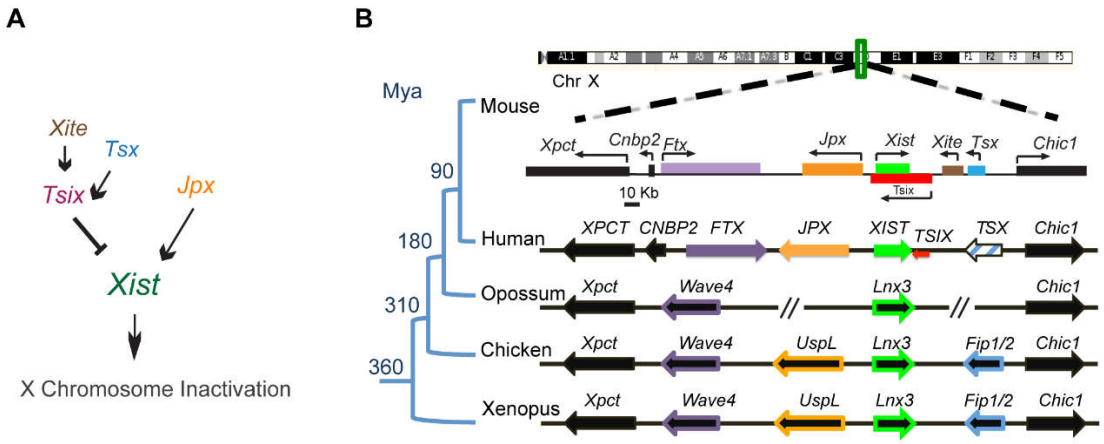


Figure 3.1: A cluster of lncRNAs at the X inactivation center locus control X chromosome inactivation

(A) *Xist* is regulated by positive and negative factors consisting of noncoding RNAs for XCI in the female mouse. **(B)** Genomic map of lncRNAs in the mouse X inactivation center locus and comparison with the orthologous region in human, opossum, chicken, and frog. LncRNA genes *Ftx*, *Jpx*, *Xist*, *Tsix*, and *Tsx* are shown in solid colors, with the *TSX* pseudogene in humans shown in hatched blue. Protein-coding genes *Xpct*, *Cnbp2*, *Chic1*, *Wave4*, *UspL*, *Lnx3*, and *Fip1/2* are in black with border-color matching the color of their homologous noncoding gene. Species divergence times are estimated in Mya (Million years ago), as indicated at the internal branches of the simplified phylogenetic tree. Orthologous genomes not drawn to scale; consistent with (Duret et al., 2006; Elisaphenko et al., 2008; Horvath et al., 2011).

Figure 3.2

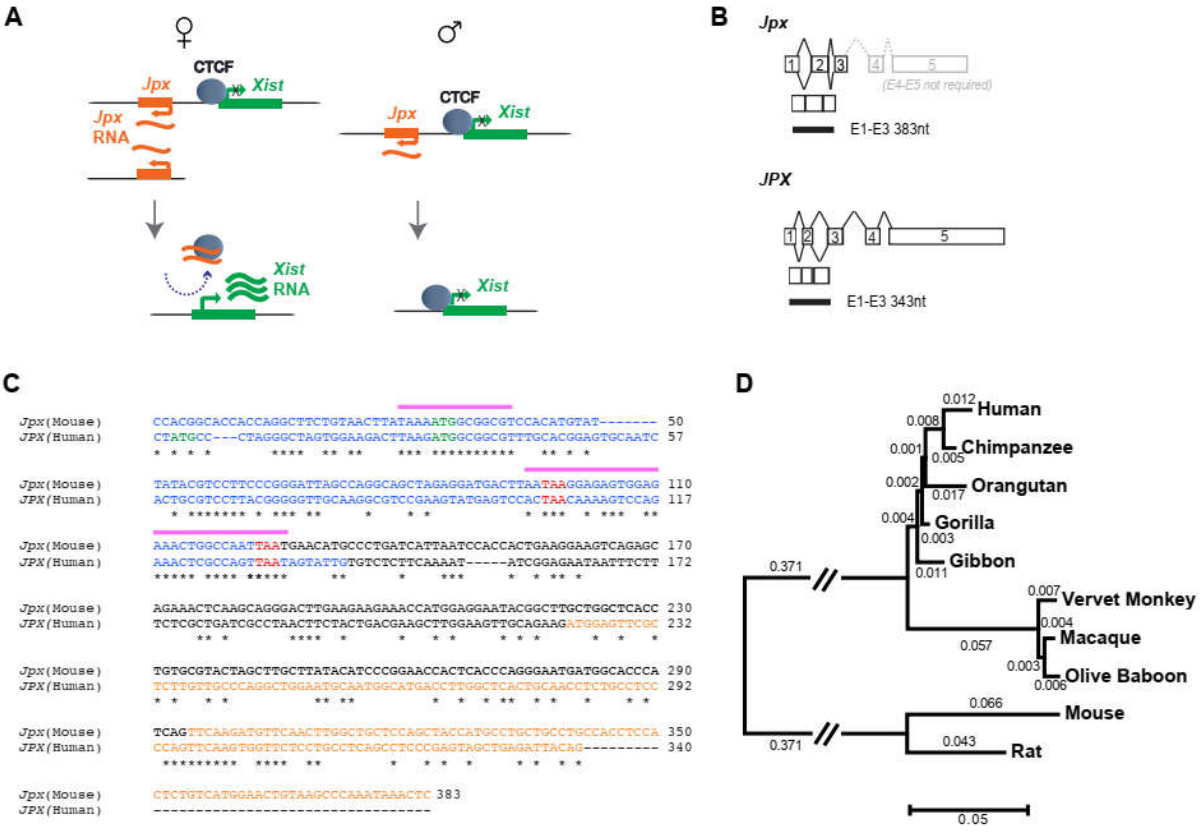
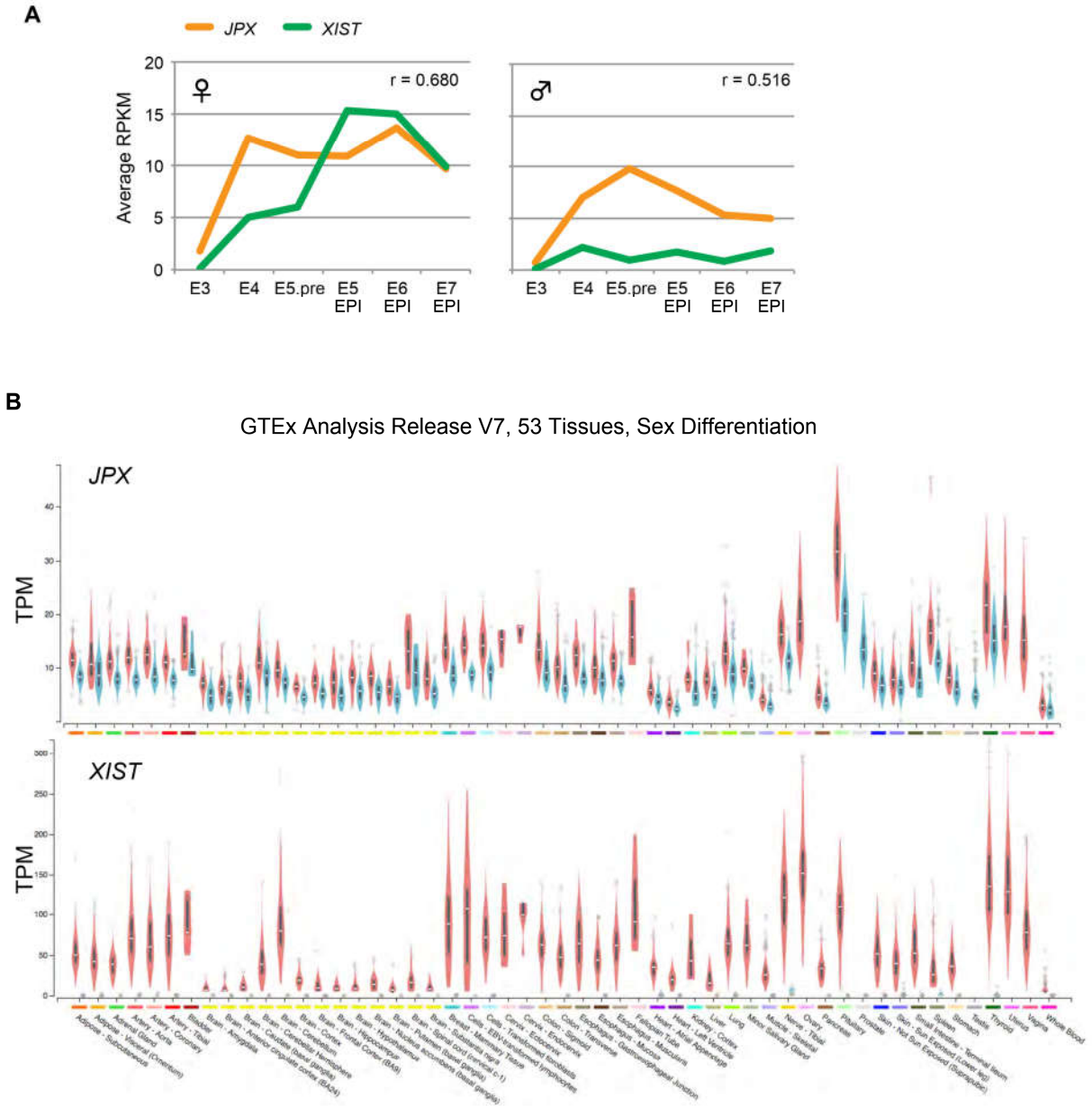

A
♀
♂
CTCF
Jpx
Xist
Jpx RNA
Xist RNA
B
Jpx
E1-E3 383nt
JPX
E1-E3 343nt
(E4-E5 not required)
C
Jpx(Mouse) CCACGGCACCACCAGGCTTCTGTAACTTATAAAATGGCGGCGTCCACATGTAT------- 50
JPX(Human) CTATGCC---CTAGGGCTAGTGGAAGACTTAAGATGGCGGCGTTTGCACGGAGTGCAATC 57
Jpx(Mouse) TATACGTCCTTCCCGGGATTAGCCAGGCAGCTAGAGGATGACTTAATAAGGAGAGTGGAG 110
JPX(Human) ACTGCGTCCTTACGGGGGTTGCAAGGCGTCCGAAGTATGAGTCCACTAACAAAAGTCCAG 117
Jpx(Mouse) AAACTGGCCAATTAATGAACATGCCCTGATCATTAATCCACCACTGAAGGAAGTCAGAGC 170
JPX(Human) AAACTCGCCAGTTAATAGTATTGTGTCTCTTCAAAAT-----ATCGGAGAATAATTTCTT 172
Jpx(Mouse) AGAAACTCAAGCAGGGACTTGAAGAAGAAACCATGGAGGAATACGGCTTGCTGGCTCACC 230
JPX(Human) TCTCGCTGATCGCCTAACTTCTACTGACGAAGCTTGGAAGTTGCAGAAGATGGAGTTCGC 232
Jpx(Mouse) TGTGCGTACTAGCTTGCTTATACATCCCGGAACCACTCACCCAGGGAATGATGGCACCCA 290
JPX(Human) TCTTGTTGCCCAGGCTGGAATGCAATGGCATGACCTTGGCTCACTGCAACCTCTGCCTCC 292
Jpx(Mouse) TCAGTTCAAGATGTTCAACTTGGCTGCTCCAGCTACCATGCCTGCTGCCTGCCACCTCCA 350
JPX(Human) CCAGTTCAAGTGGTTCTCCTGCCTCAGCCTCCCGAGTAGCTGAGATTACAG--------- 340
Jpx(Mouse) CTCTGTCATGGAACTGTAAGCCCAAATAAACTC 383
JPX(Human) ---------------------------------
D
Human
Chimpanzee
Orangutan
Gorilla
Gibbon
Vervet Monkey
Macaque
Olive Baboon
Mouse
Rat
0.012
0.008
0.001
0.005
0.002
0.017
0.004
0.003
0.011
0.371
0.007
0.004
0.057
0.003
0.006
0.066
0.371
0.043
0.05

Figure 3.2: Comparative sequence analysis of mouse lncRNA *Jpx* and human lncRNA *JPX*

(A) Model of mouse lncRNA *Jpx* mechanism. LncRNA *Jpx* is transcribed upstream of *Xist*, removes CTCF from the *Xist* promoter, and activates *Xist*. This activation is dose-dependent—*Jpx* titrates away CTCF only when present in 2-fold excess, such as in female cells, and is insufficient to activate *Xist* in male cells (Sun et al., 2013b). **(B)** Mouse *Jpx* (Top) and human *JPX* (Bottom) gene structures and transcript isoforms. The 383nt transcript of mouse *Jpx E1-E3* is required for function (Sun et al., 2013b), which corresponds to the 343nt transcript of human *JPX E1-E3*. **(C)** Sequence alignment of mouse *Jpx* (Top) and human *JPX* (Bottom) transcripts; exons 1 (Blue) – exon 2 (Black) – exon 3 (Orange) analyzed by Clustal 2.1 (Larkin et al., 2007) and the alignment was manually adjusted. Asterisks (*) mark the identical nucleotides. Pink bars label the highly conserved regions between *Jpx* and its ancestral protein homolog *UspL* (Hezroni et al., 2017). 'ATG' in Green and 'TAA' in Red mark potential start and stop codons, respectively. **(D)** Evolutionary relationships of taxa analyzed by MEGA7. Sequences were obtained from the UCSC whole genome assemblies and the evolutionary history was inferred using the Neighbor-Joining method (Saitou and Nei, 1987). The tree is drawn to scale, with branch lengths (next to the branches) in the same units as those of the evolutionary distances used to infer the phylogenetic tree. The evolutionary distances were computed using the Maximum Composite Likelihood method (Tamura et al., 2004) and are in the units of the number of base substitutions per site.

Figure 3.3



Correlations of expression *JPX* vs. *XIST*: r = 0.783, across 51 female tissues;
r = 0.784, within samples of Breast - Mammary tissues;
r = 0.690, within samples of Pituitary tissues.

Figure 3.3: *JPX* as a possible activator of *XIST* in human cells

(A) Single-cell transcript levels in human preimplantation embryos along embryonic days E3 to E7 for females (Left) and males (Right). Correlation of *JPX* and *XIST* expression shown as Pearson's *r*. Analysis performed using datasets available from Petropoulos et al. (Petropoulos et al., 2016). E5.pre = Embryonic day 5 preimplantation. E5 EPI = E5 Epiblast. **(B)** Expression of *JPX* and *XIST* across 53 human tissues as violin plots. TPM: Transcript Per Million. Female expression is in red; male expression is in blue. Figures obtained from the Genotype-Tissue Expression (GTEx) Project, GTEx Portal (latest release version, V7).

3.3 Results

Comparative sequence analysis suggests that human lncRNA *JPX* is functional

The molecular features and functional roles of human lncRNA *JPX* were not previously known. The GenBank annotation for human lncRNA *JPX* shows that the gene (Gene ID: 554203; RNA Sequence: NR_024582) contains five exons, similar to its mouse homolog, *Jpx* (Gene ID: 70252) (Figure 3.2B), which has also been shown to be composed of 5 exons, with the transcripts containing exons 4 and 5 being minor isoforms (Johnston et al., 2002; Kolesnikov and Elisaphenko, 2010; Tian et al., 2010; Tsuritani et al., 2007). PCR analysis of cDNA also showed that the primary isoform for human JPX was also generated from the first three exons as well (data not shown). Using total RNA from human ovarian cancer SKOV3iP1 cells, we were able to isolate a primary *JPX* transcript spanning exons 1-3, *JPX E1-E3*. In mice, it has been shown that the corresponding *Jpx E1-E3* is the primary isoform responsible for the function of *Jpx*

in XCI (Lee et al., 1999; Sun et al., 2013b). More precisely, nucleotides 1-383 of the *Jpx E1-E3* sequence are necessary and sufficient for mouse lncRNA *Jpx* binding to CTCF (Sun et al., 2013b). As lncRNA *JPX E1-E3* was detected in human cells, we suspected conservation of gene structure and nucleotide sequence between *JPX* and *Jpx*. To characterize their sequence homology, we performed a pairwise sequence alignment between the critical 383nt mouse *Jpx E1-E3* with the full-length (343nt) human *JPX E1-E3* (Figure 3.2C). Despite an overall similarity of gene structure, including five exons in both *Jpx* and *JPX* (Figure 3.2B), the exact sequence identity of exons 1-3 is approximately 40%, which is much less than the average nucleotide sequence identity (85%) for protein-coding sequences between humans and mice (one-tailed binomial test, $P < 10^{-6}$) (Gibbs et al., 2004; Makalowski et al., 1996). Interestingly, exon 1 of *Jpx* and *JPX* both contain remnants of protein-coding sequences similar to the chicken *UspL,* which have recently been reported as possible regulatory sequences for lncRNA function (Hezroni et al., 2017).

While a lack of sequence conservation is not surprising for noncoding genes (Necsulea and Kaessmann, 2014; Ponjavic et al., 2007; Ponting and Lunter, 2006), it is unknown how lncRNAs evolve and how nucleotide changes in noncoding genes affect the lncRNA function. Taking advantage of fully sequenced genomes of multiple vertebrates, we searched for other homologous sequences of *Jpx E1-E3* and *JPX E1-E3* in the UCSC genome assemblies. Appropriate exon sequences for alignment were obtained from primates and murine rodents, which allowed us to look into the phylogenetic history of *Jpx*. Using the Neighbor-Joining method (Saitou and Nei, 1987), we constructed the evolutionary tree for *Jpx E1-E3* and *JPX E1-E3* in ten species (2

rodents and 8 primates), with evolutionary distances calculated from Maximum Composite Likelihood method (Figure 3.2D) (Tamura et al., 2004). The rate of *Jpx* sequence evolution between mouse and rat is 0.109 substitution per site, which is ~44% lower than the estimated neutral evolution rate of 0.196 between these two rodents (one-tailed binomial test, $P = 0.000014$) (Cooper, 2004; Gibbs et al., 2004). This is consistent with our understanding that *Jpx* has an important functional role in mice, and thus the sequence changes have been under substantial constraint in the rodent lineages. By contrast, the evolutionary distance between rodent *Jpx* and human *JPX* is 0.796 substitution per site, which is ~74% higher than the neutral rate of 0.457 between humans and rodents (one-tailed binomial test, $P < 10^{-6}$) (Cooper, 2004). A more rapid nucleotide substitution between human and rodent suggests positive selection acting on the sequence of human *JPX*.

We next compared our *Jpx/JPX* gene tree to the recently published species tree of primates. The neutral substitution rates in the lineages leading to the hominoid (human, chimpanzee, and orangutan), as estimated from the common ancestor between hominoids and Old World Monkeys (vervet monkey, macaque, and olive baboon), are within the range of 0.026 – 0.027 (Moorjani et al., 2016). By contrast, the corresponding evolutionary rates of the *JPX* gene in the same lineages appear to be more variable, ranging from 0.009 to 0.027. Importantly in the human lineage, the nucleotide change rate is 0.012, which is two-fold higher than the neutral substitution rate of 0.0058 for humans (Moorjani et al., 2016) (one-tailed binomial test, $P = 0.13$). A larger than two-fold difference in the substitution rates is seen between the human and chimpanzee branches (bootstrap 80% over 500 replicates), which is notable given that

the rates of evolution on these two lineages are estimated to be very similar with only a 1.9% difference (Moorjani et al., 2016). Such observations suggest that adaptive nucleotide sequence changes have occurred in the hominoid lineages, which are supportive of a functional *JPX*, particularly in the human lineage.

Consistent with a possible role of *JPX* in regulating *XIST* within humans, re-analysis of available single-cell RNA-seq data revealed a positive correlation between *JPX* and *XIST* expression levels in human preimplantation embryos, especially in female cells of the epiblast lineage (Figure 3.3A; raw data obtained from (Petropoulos et al., 2016)). Such observations suggest that human lncRNA *JPX* likely functions as a positive regulator of *XIST*. Additionally, data from the Genotype-Tissue Expression (GTEx) Project showed sex-dependent expression and a positive correlation between *JPX* and *XIST* expression across 51 female samples (Pearson's correlation $r = 0.783$, $P < 10^{-6}$) (Figure 3.3B). Within specific tissue types, *JPX* and *XIST* activities in individual samples also showed positive correlations: $r = 0.784$ for the breast tissues (n = 290, $P < 10^{-6}$); $r = 0.690$ for the pituitary tissues (n = 183, $P < 10^{-6}$).

These results suggest that human lncRNA *JPX* is functionally important and that a detailed analysis of *Jpx/JPX* would provide a novel experimental model to understand conservation of function despite sequence diversity.

RNA structural probing reveals divergence of *Jpx/JPX* homologous lncRNAs

If human lncRNA *JPX* shares similar function with its mouse homolog, it is possible that there is conservation at the RNA structural level despite a nucleotide sequence divergence. Such conservation has been supported by previous RNA

structure-function studies on well-characterized noncoding RNAs, such as ribozymes and riboswitches, but thus far there has been limited analysis for lncRNAs (Ilik et al., 2013; Kirk et al., 2018; Webb et al., 2009). To explore how the sequence determines the secondary structure of RNA, we performed selective 2'-hydroxyl acylation analyzed by primer extension (SHAPE) RNA structural probing (Spitale et al., 2013; Wilkinson et al., 2006). For mouse *Jpx*, we focused on the functional lncRNA transcript *Jpx 34-347* and designed the reverse primers spanning the 314nt sequence (Figure 3.4A). Extension primers were also designed to probe the human lncRNA transcript *JPX 1-343* (Figure 3.4B). We used the SHAPE reagent, 2-methylnicotinic acid imidazolide (NAI), to modify structured *in vitro* transcribed RNA and map to residues that are accessible, such as unpaired or flexible bases (Figure 3.4C). As has been reported (Ilik et al., 2013; Wilkinson et al., 2006), sites of nucleotide modification can be identified as stops to primer extension by reverse transcriptase. Using radiolabeled reverse primers to generate cDNAs from NAI treated or DMSO treated RNAs, we were able to resolve the modified unpaired bases by running the reverse transcribed cDNAs through denaturing gel electrophoresis for sequencing. The intensities of the gel bands are positively correlated with the NAI modification strengths and thus reveal features of the RNA secondary structures at single-base resolution. In the SHAPE profiles for the 5' and 3' regions of *Jpx 34-347*, we observed segments of the base pairing in nucleotides G56-G66 and A71-C81 (Figure 3.4C, pex1r panel), and in nucleotides C309-C317 (Figure 3.4C, p2r panel), suggesting possible stem-loops at these sites. Similarly for the 5' and 3' regions of *JPX 1-343*, we observed the base pairing in nucleotides C38-C46 (Figure

3.4D, pR3-4), and in nucleotides C276-C285 and U306-C311 (Figure 3.4D, pR1), which suggest corresponding stem-loop features.

To systematically analyze the lncRNA SHAPE profiles, we calculated the SHAPE reactivity for each nucleotide after measuring band intensity, background (the DMSO lane) subtraction, and normalization (Ilik et al., 2013). By inputting all the SHAPE reactivity values into the ViennaRNA program for RNA secondary structure (Lorenz et al., 2011; Washietl et al., 2012), we were able to derive the most likely structure for *Jpx 34-347* and *JPX 1-323* based on a linear log model for pairing probabilities (Zarringhalam et al., 2012). As illustrated in Figure 3.5A, mouse *Jpx 34-347* RNA contains multiple stem-loops. Overall, about 50% of the nucleotides are base-paired, indicating that *Jpx 34-347* RNA is highly structured. Both the 5' and 3' nucleotides, 34-114 and 252-347 respectively, are involved in stem-loop formation, suggesting possible secondary structural configurations necessary for function. We asked what structural features human *JPX* may share with mouse *Jpx*. As shown in Figure 3.5B, *JPX 1-343* RNA also contains stem-loops with more than 50% of paired bases. However, the overall structure is obviously different from the mouse *Jpx* RNA structure (Figure 3.5).

To determine whether *in vitro* conditions may limit the RNA secondary structural probing, we performed SHAPE analysis *in vivo* with *JPX* lncRNA modified by NAI in human embryonic kidney (HEK293) cells. We note that *in vivo* RNA structural probing is particularly difficult on low abundance RNA for which results can be confounded due to the presence of more abundant RNA in the same sample (Kwok et al., 2013; Xue et al., 2008). To enrich *JPX* for *in vivo* SHAPE profiling, we adapted a cDNA amplification step using LMPCR (ligation mediated polymerase chain reaction) which has shown to be

instrumental for *in vivo* RNA structural probing (Kwok et al., 2013; Lucks et al., 2011). We chose the sequence domain *JPX 104-172*, which shares 50% nucleotide identity and corresponds to the mouse *Jpx* sequence essential for function (Sun et al., 2013b). A direct comparison between *in vitro* and *in vivo* SHAPE profiles of *JPX 104-172* showed overlapping segments representing single-stranded RNA domains (Figure 3.6A). There is an overall 70% exact matching between the *in vitro* and *in vivo* RNA structural predictions (Figure 3.6B), with 76% (22 out of 29) single-stranded nucleotides from the *in vivo* structure falling into the loop regions predicted by *in vitro* SHAPE reactivities (Figure 3.6C), supporting that *in vitro* SHAPE profiling is instructive to determine RNA secondary structural features. Based on the large differences revealed by the *in vitro* SHAPE reactivities for *JPX E1-E3* and *Jpx E1-E3* (Figure 3.6), we conclude that human lncRNA *JPX* has diverged from mouse lncRNA *Jpx* in their overall secondary RNA structures.

Figure 3.4

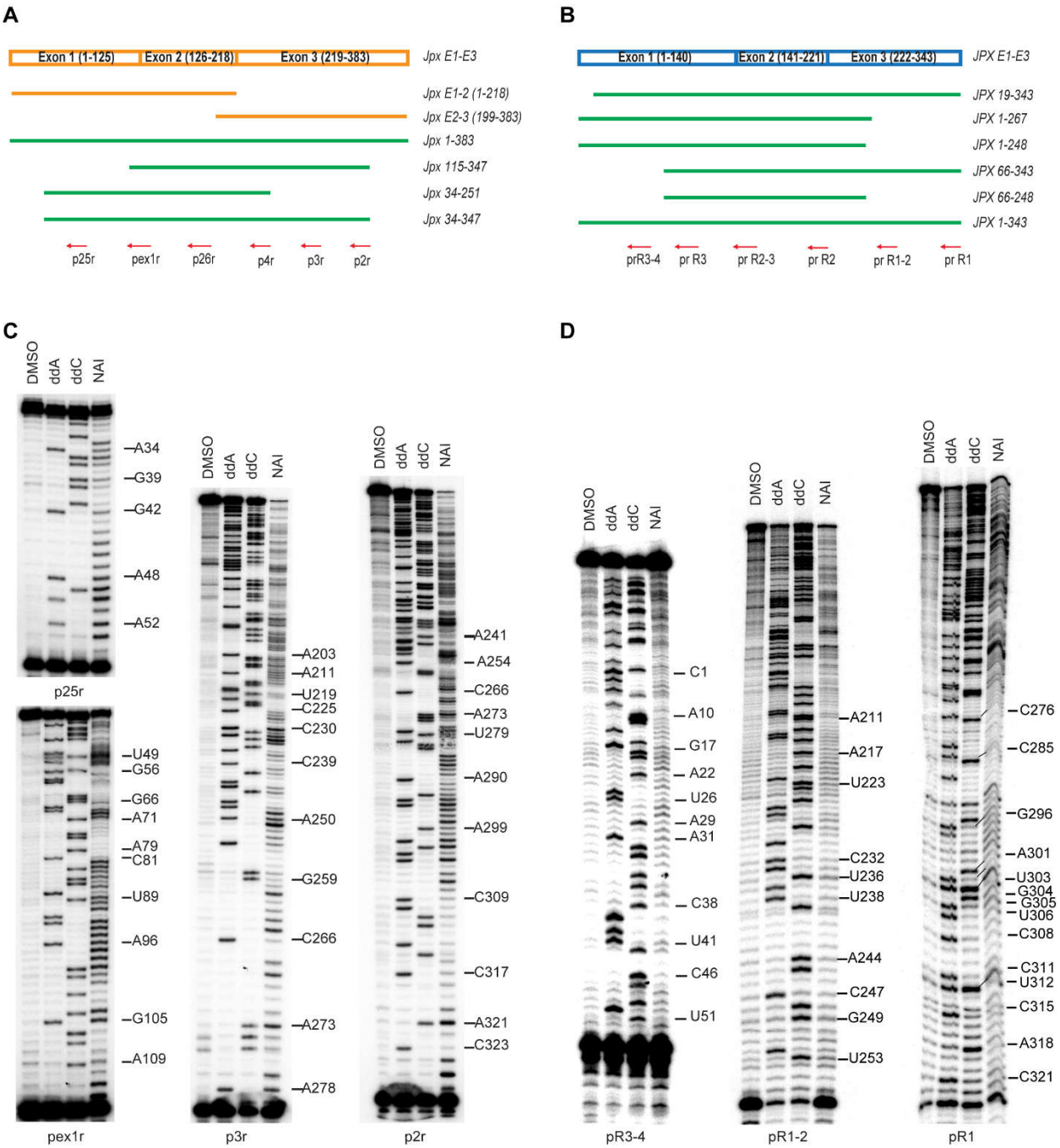

A
Exon 1 (1-125)
Exon 2 (126-218)
Exon 3 (219-383)
Jpx E1-E3
Jpx E1-2 (1-218)
Jpx E2-3 (199-383)
Jpx 1-383
Jpx 115-347
Jpx 34-251
Jpx 34-347
p25r
pex1r
p26r
p4r
p3r
p2r
B
Exon 1 (1-140)
Exon 2 (141-221)
Exon 3 (222-343)
JPX E1-E3
JPX 19-343
JPX 1-267
JPX 1-248
JPX 66-343
JPX 66-248
JPX 1-343
prR3-4
pr R3
pr R2-3
pr R2
pr R1-2
pr R1
C
DMSO
ddA
ddC
NAI
A34
G39
G42
A48
A52
p25r
U49
G56
G66
A71
A79
C81
U89
A96
G105
A109
pex1r
A203
A211
U219
C225
C230
C239
A250
G259
C266
A273
A278
p3r
A241
A254
C266
A273
U279
A290
A299
C309
C317
A321
C323
p2r
D
C1
A10
G17
A22
U26
A29
A31
C38
U41
C46
U51
pR3-4
A211
A217
U223
C232
U236
U238
A244
C247
G249
U253
pR1-2
C276
C285
G296
A301
U303
G304
G305
U306
C308
C311
U312
C315
A318
C321
pR1

Figure 3.4: Functional domain mapping and RNA structure probing

(A) Different length *in vitro* transcribed RNA corresponding to the coverage of *Jpx*. Mouse lncRNA 220nt E1-E2 or 183nt E2-E3 (Orange) is not sufficient for protein binding *in vitro* (Sun et al., 2013b). The mouse 383nt functional *Jpx* transcript and its truncated forms (Green) are assayed with RNA EMSA. Red arrows indicate positions of the reverse primers used in SHAPE. **(B)** Different length *in vitro* transcribed RNA corresponding to the coverage of *JPX*. A full-length 343nt *JPX* transcript and its truncated forms (Green) are assayed with RNA EMSA. Red arrows indicate positions of the reverse primers used in SHAPE. **(C)** Mouse *Jpx* RNA structure probing by SHAPE: polyacrylamide gel electrophoresis (PAGE) resolves RNA footprint after treatment of RNA by either DMSO (control) or SHAPE modification reagent NAI (Spitale et al., 2013), followed by RNA reverse transcription using primers indicated and RNA hydrolysis. **(D)** Human *JPX* RNA structure probing by SHAPE: as described in C. **(C-D)** At least two replicates were performed for each reaction and representative gel images are shown. Band intensity and corresponding nucleotide positions were integrated with SAFA software (Das et al., 2005; Laederach et al., 2008). SHAPE reactivities reflect single-stranded (highly reactive) and double-stranded (not reactive) states at individual nucleotides. Nucleotide labels correspond with NAI modified nucleotides.

Figure 3.5

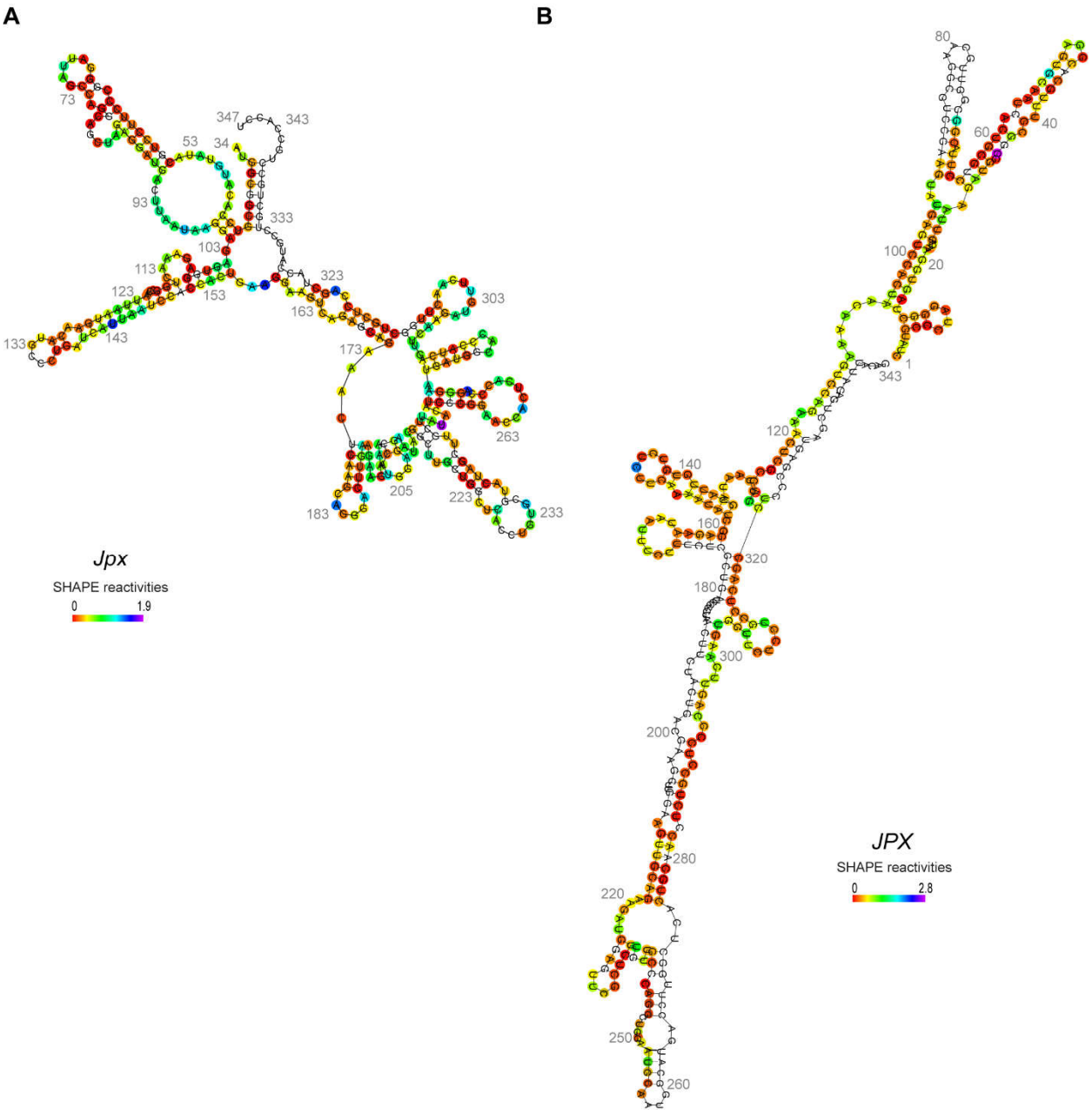

A
73
347
343
53
34
93
333
103
113
323
123
153
163
303
133
143
173
263
205
223
183
233
Jpx
SHAPE reactivities
0
1.9
B
80
60
40
100
20
1
343
120
140
160
320
180
300
200
280
220
250
260
JPX
SHAPE reactivities
0
2.8

Figure 3.5: RNA secondary structures of mouse *Jpx* and human *JPX* derived from SHAPE reactivity

(A) SHAPE reactivities at individual nucleotides of mouse *Jpx* were normalized to a scale of 0 to 1.9. **(B)** SHAPE reactivities of human *JPX* were normalized to a scale of 0 to 2.4. **(A-B)** Scales were denoted with color codes at individual nucleotides. The secondary structures were drawn with the RNA probing web server (http://rna.tbi.univie.ac.at) based on the fold algorithms in Lorenz et al. and Washietl et al. (Lorenz et al., 2011; Washietl et al., 2012)

Figure 3.6

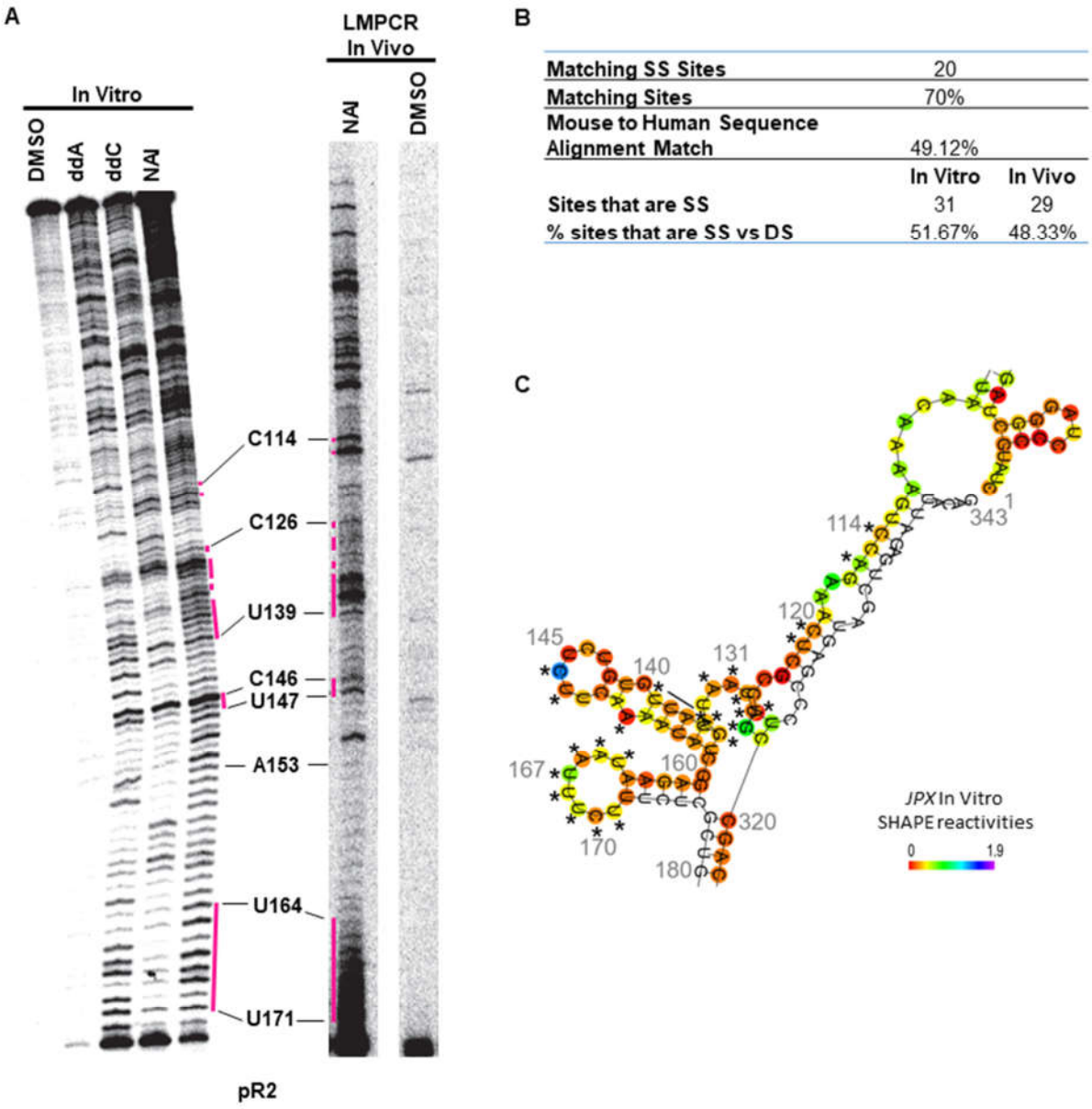

A
LMPCR
In Vivo
In Vitro
DMSO
ddA
ddC
NAI
NAI
DMSO
C114
C126
U139
C146
U147
A153
U164
U171
pR2
B
Matching SS Sites 20
Matching Sites 70%
Mouse to Human Sequence Alignment Match 49.12%
In Vitro
In Vivo
Sites that are SS 31 29
% sites that are SS vs DS 51.67% 48.33%
C
1
343
114
120
131
140
145
160
167
170
180
320
JPX In Vitro SHAPE reactivities
0
1.9

Figure 3.6: Human *JPX* RNA structure probing by SHAPE *in vitro* and *in vivo*

(A) Polyacrylamide gel electrophoresis (PAGE) of RNA footprint after treatment of RNA by either DMSO (control) or NAI *in vitro* and *in vivo* (Kwok et al., 2013; Lucks et al., 2011; Spitale et al., 2013), followed by reverse transcription using primer pR2. Dideoxy sequencing was done using *in vitro* transcribed RNA. Magenta lines indicate regions of similar NAI reactive patterns between *in vitro* and *in vivo* SHAPE. These represent the single-stranded regions with unpaired nucleotides of the lncRNA *JPX*. *In vivo* SHAPE was performed using LMPCR enrichment methods (Kwok et al., 2013; Lucks et al., 2011). **(B)** Table of percentages of matching sites between *in vitro* and *in vivo* NAI profiles and the overall sequence conservation of the probed lncRNA region. **(C)** RNA secondary structure of the probed *JPX* lncRNA region as derived from *in vitro* SHAPE. Asterisks (*) mark nucleotides reactive to NAI *in vivo* and indicated as single-stranded, as also shown in the *in vivo* SHAPE profile (A).

LncRNA-protein binding *in vitro* demonstrates sequence requirements for RNA function

Given the low conservation of nucleotide sequences and RNA secondary structures, we asked whether specific domains or motifs might be required for the molecular functions of *Jpx* and *JPX* lncRNAs. Utilizing the known molecular interaction between mouse *Jpx* and CTCF, we performed an *in vitro* RNA electrophoresis mobility shift assay (EMSA) and tested the binding capacity of *Jpx* RNA with regard to various truncation forms (Figure 3.4A). It has been reported that truncated *Jpx* RNAs, 220nt *Jpx E1-E2* and 183nt *Jpx E2-E3*, failed to bind CTCF (Sun et al., 2013b), suggesting that both halves of *Jpx E1-E3* are needed for it to function. We then removed segments of the 5' and 3' ends of *Jpx E1-E3* and characterized the binding kinetics of mutant *Jpx* RNAs in comparison to the 383nt *Jpx E1-E3*. For negative control, we used a 316nt *drz-Agam-2-1* ribozyme RNA from *Anopheles gambiae* (Webb et al., 2009). As shown in Figure 3.7A, with increasing CTCF concentration, the full-length *Jpx 1-383* (Red) and the truncated *Jpx 34-347* (Black) both exhibited robust binding. By contrast, the truncated *Jpx 115-347* (Blue) and *Jpx 34-251* (Pink) failed to bind CTCF. We conclude that both the 5' sequence of nucleotides 34-114 and the 3' sequence of nucleotides 252-347 are required for *Jpx* to bind CTCF, and that these regions may be responsible for its function.

We next asked whether the molecular function of human *JPX* has diverged from mouse *Jpx* and performed EMSA on *JPX 1-343* RNA for its CTCF-binding capacity *in vitro* (Figures 3.7B-C and 3.8). We used purified recombinant CTCF with the full-length mouse CTCF protein of 736 amino acids, which is 98% identical with the human CTCF

protein. In vertebrates, CTCF is highly conserved and functioning as a global transcriptional regulator in all cell types (Ohlsson et al., 2001; Phillips and Corces, 2009). Therefore, whether human *JPX* binds CTCF the same as mouse *Jpx*, would inform functional importance and conservation at the molecular level. As shown in EMSA with increasing concentration of CTCF protein, *JPX 1-343* RNA was capable of binding the CTCF protein and was robustly shifted by CTCF (Figure 3.7C, left panel). By contrast, the 316nt control RNA of mosquito ribozyme showed weak interaction only at the highest concentration of CTCF (Figure 3.7C, right panel). We looked further into the binding kinetics of CTCF against various truncation forms of human lncRNA *JPX* (Figures 3.4B & 3.7B) with the goal of identifying RNA sequence domains critical for binding. As shown in Figure 3.7B, all *JPX* RNA truncations were able to bind CTCF in comparison to the control RNA (Green). Interestingly, the 5' truncation form, *JPX 19-343* (Black), and the 3' truncation form, *JPX 1-267* (Blue), showed stronger binding than the full-length *JPX 1-343* (Red). Our protein-binding assays therefore indicate that human *JPX* RNA is capable of binding CTCF, and that such a CTCF-*JPX* interaction can be robust against removal of *JPX* 5' or 3' RNA sequences. Overall, the EMSA results demonstrate that human *JPX* RNA has maintained, or may have even reinforced, its molecular binding capacity with CTCF protein.

Consistent with our observation of direct RNA-protein binding of *Jpx*-CTCF and *JPX*-CTCF *in vitro*, genome-wide studies have reported CTCF-RNA interactions in both mouse and human cells (Kung et al., 2015; Saldaña-Meyer et al., 2014). Notably, mouse *Jpx* RNA was identified as one of the locus-specific interacting RNAs of CTCF in mouse embryonic stem cells by CLIP-seq (cross-linking immunoprecipitation combined

with high-throughput sequencing) (Kung et al., 2015). In addition, we also found human *JPX* present as one of the RNA transcripts pulled-down with human CTCF by PAR-CLIP (photoactivatable ribonucleoside-enhanced crosslinking and immunoprecipitation) from human bone osteosarcoma U2OS cells (Supplementary data from (Saldaña-Meyer et al., 2014)). Moreover, RNA-binding regions (RBR) in CTCF have been recently reported and shown to be essential for the molecular interaction and function of CTCF in mouse and human cells (Hansen et al., 2018; Saldaña-Meyer et al., 2019), which confirms CTCF binding to endogenous RNAs and the functional importance of CTCF-RNA binding in gene regulation. Our evidence of direct binding of CTCF to human *JPX* RNA similar to CTCF-*Jpx* interaction in the mouse, thus supports the functional importance and conservation of human *JPX* to its mouse homolog.

Figure 3.7

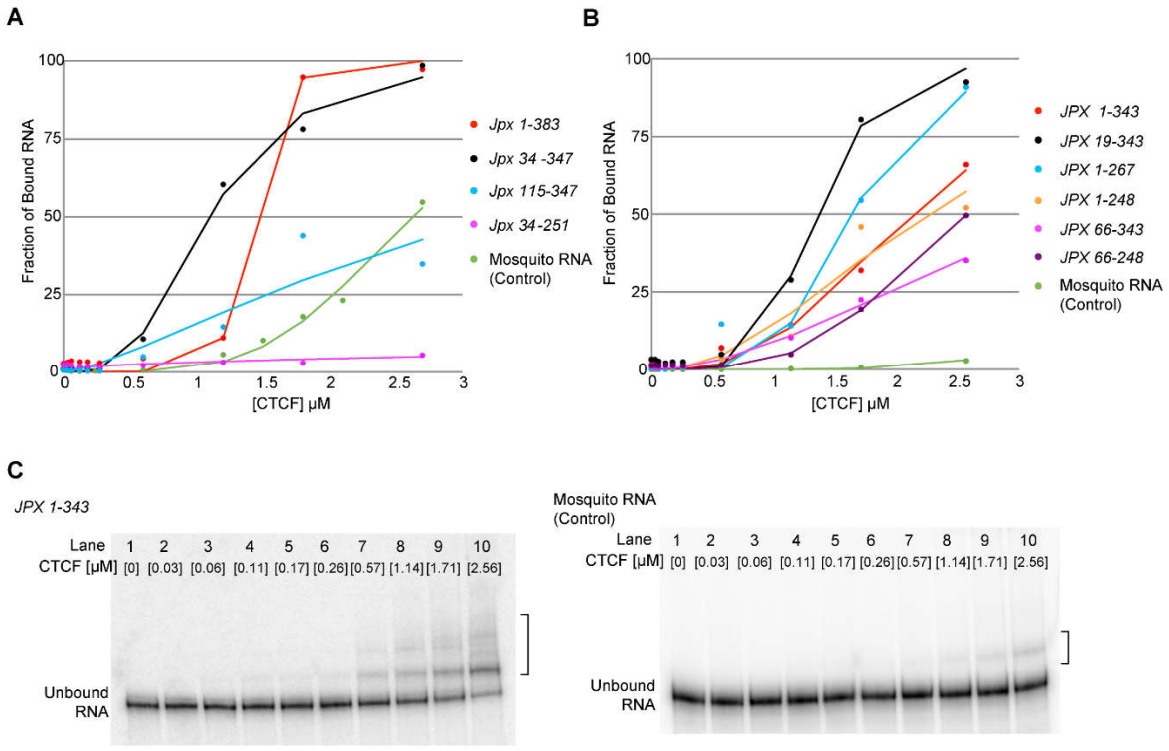

A
Fraction of Bound RNA
[CTCF] µM
Jpx 1-383
Jpx 34 -347
Jpx 115-347
Jpx 34-251
Mosquito RNA (Control)
B
Fraction of Bound RNA
[CTCF] µM
JPX 1-343
JPX 19-343
JPX 1-267
JPX 1-248
JPX 66-343
JPX 66-248
Mosquito RNA (Control)
C
JPX 1-343
Lane 1 2 3 4 5 6 7 8 9 10
CTCF [µM] [0] [0.03] [0.06] [0.11] [0.17] [0.26] [0.57] [1.14] [1.71] [2.56]
Unbound RNA
Mosquito RNA (Control)
Lane 1 2 3 4 5 6 7 8 9 10
CTCF [µM] [0] [0.03] [0.06] [0.11] [0.17] [0.26] [0.57] [1.14] [1.71] [2.56]
Unbound RNA

Figure 3.7: Mouse *Jpx* RNA and human *JPX* RNA are capable of binding to CTCF

(A) Binding isotherm for CTCF-*Jpx* from RNA EMSA. Binding curve was plotted as the percent bound against CTCF concentration and was fit by a nonlinear regression to a binding isotherm. The *Jpx 1-383* (Red) and the truncated *Jpx 34-347* (Black) both show robust binding, whereas 5'-truncated *Jpx 115-347* (Blue) and 3'-truncated *Jpx 34-251* (Pink) show weak or non-specific binding as compared to a 316nt control RNA (Green) from the malaria mosquito, *Anopheles gambiae* (Webb et al., 2009). **(B)** Binding isotherm for CTCF-*JPX* from RNA EMSA. All *JPX* RNA of different lengths demonstrated favorable binding as compared to the 316nt control RNA (Green). The 5'-truncated *JPX 66-343* (Pink) and *JPX 66-248* (Purple) showed weaker binding than the full-length *JPX 1-343* (Red). **(C)** Representative RNA EMSA gel image detecting direct binding of *JPX* RNA and CTCF protein *in vitro*. Left panel: binding of *JPX 1-343* RNA with CTCF protein at increasing concentrations. Right panel: binding of a 316nt control RNA (from the malaria mosquito, *Anopheles gambiae)* (Webb et al., 2009) with CTCF protein at increasing concentrations. RNA-protein shift is indicated by the bracket on the right side of the gel images.

Figure 3.8

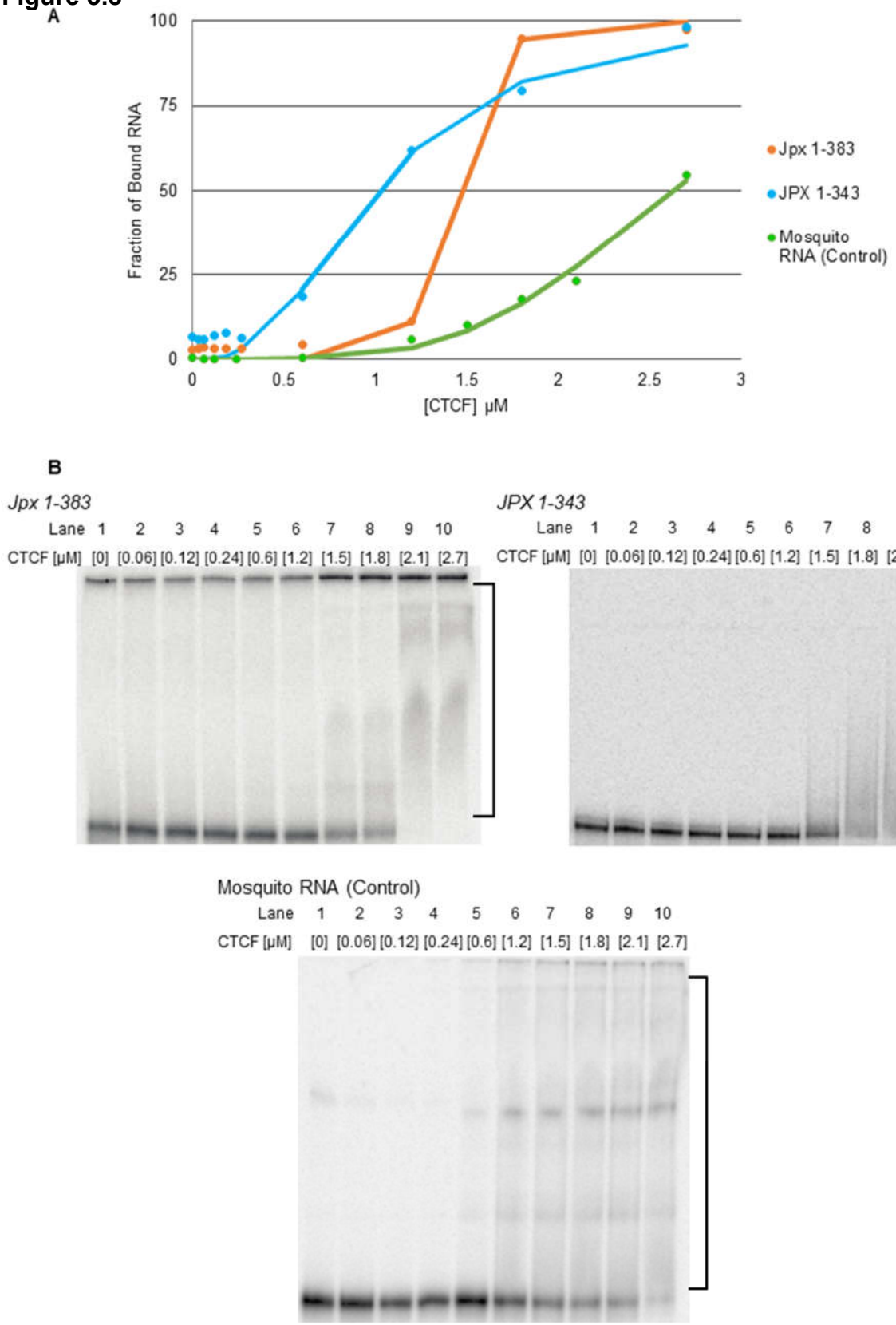

A
100
75
50
25
0
Fraction of Bound RNA
0
0.5
1
1.5
2
2.5
3
[CTCF] µM
Jpx 1-383
JPX 1-343
Mosquito RNA (Control)
B
Jpx 1-383
Lane 1 2 3 4 5 6 7 8 9 10
CTCF [µM] [0] [0.06] [0.12] [0.24] [0.6] [1.2] [1.5] [1.8] [2.1] [2.7]
JPX 1-343
Lane 1 2 3 4 5 6 7 8 9 10
CTCF [µM] [0] [0.06] [0.12] [0.24] [0.6] [1.2] [1.5] [1.8] [2.1] [2.7]
Mosquito RNA (Control)
Lane 1 2 3 4 5 6 7 8 9 10
CTCF [µM] [0] [0.06] [0.12] [0.24] [0.6] [1.2] [1.5] [1.8] [2.1] [2.7]

Figure 3.8: Mouse *Jpx 1-383* and full-length human *JPX* lncRNA binding to CTCF

(A) Binding isotherm for CTCF-*Jpx/JPX* RNA in comparison to the control with mosquito RNA. Binding curve was plotted as the percent bound against CTCF concentration and was fit by a nonlinear regression to a binding isotherm. *Jpx 1-383* (Orange) and *JPX 1-343* (Blue) both show robust binding as compared to 316nt control RNA (Green) from the malaria mosquito, *Anopheles gambiae* (Webb et al., 2009). **(B)** Representative RNA EMSA gel images detecting direct binding of *Jpx*, *JPX*, and control RNA to CTCF protein *in vitro*. Top left panel: binding of *Jpx 1-383* RNA with CTCF protein at increasing concentrations. Top right panel: binding of *JPX 1-343* RNA with CTCF protein at increasing concentrations. Bottom panel: binding of a 316nt control RNA with CTCF protein at increasing concentrations. RNA-protein shift is indicated by the bracket on the right side of the gel images.

Rescue of *Jpx-/+* mES cell viability and morphology by a human *JPX* transgene indicates functional complementation

Since mouse lncRNA *Jpx* activates XCI through its binding to CTCF (Figure 3.2A), we therefore asked whether human lncRNA *JPX* could function equivalently for XCI. Using a functional complementation test, we addressed whether the *in vivo* function of human lncRNA *JPX* is equivalent to mouse lncRNA *Jpx*. We used *Jpx-/+* female mouse embryonic stem (mES) cells for this purpose, an established *Jpx-/+* heterozygous knockout cell line. As previously reported, *Jpx-/+* female mES cells die during cell differentiation due to failed XCI, which is associated with morphology defects and a loss of *Xist* transcription (Tian et al., 2010). By introducing Tg(EF1α:JPXE1-E3) transgene expressing *JPX E1-E3* transcript into *Jpx-/+* mutant female mES cells, we asked whether the cell viability defects caused by loss of mouse lncRNA *Jpx* could be rescued by expression of human lncRNA *JPX*. As a reference in parallel, we also introduced Tg(EF1α:JpxE1-E3) transgene expressing *Jpx E1-E3* into *Jpx-/+* female mES cells. As shown in Figure 3.9A (second row), *Jpx-/+* female mES cells receiving the vector-only transgene were dying during mES differentiation, displaying irregular and disaggregated embryoid bodies (EBs) at Day 4 (white arrows). These phenotypes were rescued by transiently transfected transgenes overexpressing *Jpx E1-E3* (third row) or *JPX E1-E3* (bottom row), with which the mutant cells formed better EBs at Day 4 and showed less dissociated cells. At Day 8 of mES differentiation, the rescue effects were most obvious. While wild type female mES cells receiving the vector-only transgene showed EB attachment and cell outgrowth, the mutant *Jpx-/+* had no attached EBs or cell outgrowth, and instead, had mostly disintegrated EBs and floating

cells in the media. By contrast, *Jpx-/+* female mES cells receiving either *Jpx E1-E3* or *JPX E1-E3* showed clearly attached EBs and cell outgrowth comparable to the wild type control cells, suggesting complete reversal of cell lethality.

To validate the observed differences in EB morphology, we quantified the cell viability at Day 4. As shown in Figure 3.9B, mutant *Jpx-/+* mES cells exhibited viability defect with ~40% reduction as compared to wild type female control cells during ES cell differentiation (one-tailed Student *t*-test, $P < 0.01$). In comparison, mutant *Jpx-/+* mES cells receiving *Jpx E1-E3* or *JPX E1-E3* were rescued, and cell viabilities in both cases were significantly elevated (one-tailed Student *t*-test, $P < 0.05$), reaching 96% and 86% of the wild type level, respectively. To confirm that the phenotypic rescue of *Jpx-/+* mutant female cells was a response to the transgene expression, we assayed the *Jpx* and *JPX* RNA quantities in these transfected *Jpx-/+* mutant female mES cells (Figure 3.9C, Day 8 shown). Mutant *Jpx-/+* mES cells receiving *Jpx E1-E3* strongly expressed mouse *Jpx* but not human *JPX*, and cells that had received *JPX E1-E3* were the only ones strongly expressing human *JPX* RNA (Figure 3.9C). We did not observe any adverse defects when the same transgenes were expressed in wild type female mES cells (Figure 3.10). EB morphology and outgrowth were normal and cell viabilities were comparable to wild type control cells carrying the empty vector (Figure 3.10A-B). At the molecular level, the *Jpx E1-E3* transgene in wild type female mES cells induced higher *Xist* expression (Figure 3.10C, one-tailed Student *t*-test, $P < 0.01$), consistent with previous reports on the *trans* activation role of mouse *Jpx* on *Xist* (Carmona et al., 2018; Sun et al., 2013b). By contrast, the *JPX E1-E3* transgene did not enhance *Xist*

expression in wild type female mES cells, presumably due to the presence of intact endogenous mouse *Jpx* in these cells (Figure 3.10C).

A heterozygous *Jpx* deletion in the mouse female ES cell compromises overall *Jpx* expression, which leading to reduced *Xist* expression during ES differentiation (Tian et al., 2010). As shown in Figure 3.9D, mutant *Jpx-/+* mES cells at differentiation Day 8 had an overall lower level of *Xist* transcripts (45% of the wild type level). Expression of *Jpx E1-E3* transgene fully rescued *Xist* expression to 105% of the wild type level in *Jpx-/+* mES cells (one-tailed paired Student *t*-test, $P = 0.002$), consistent with the *trans* activation role of mouse *Jpx* on *Xist* (Carmona et al., 2018; Sun et al., 2013b). Expression of the human *JPX E1-E3* transgene in *Jpx-/+* mES cells rescued *Xist* expression to 93% of the wild type level (one-tailed paired Student *t*-test, $P = 0.108$). This is above and beyond the confidence interval of *Xist* expression in *Jpx-/+* female mES cells (13% – 76% of the wild type level), indicating that human *JPX E1-E3* RNA is capable of activating *Xist* to complement the loss of mouse *Jpx* RNA in the mES cells. Taken together, our results support that human lncRNA *JPX* is functionally homologous to mouse lncRNA *Jpx* in their molecular roles affecting XCI.

Figure 3.9

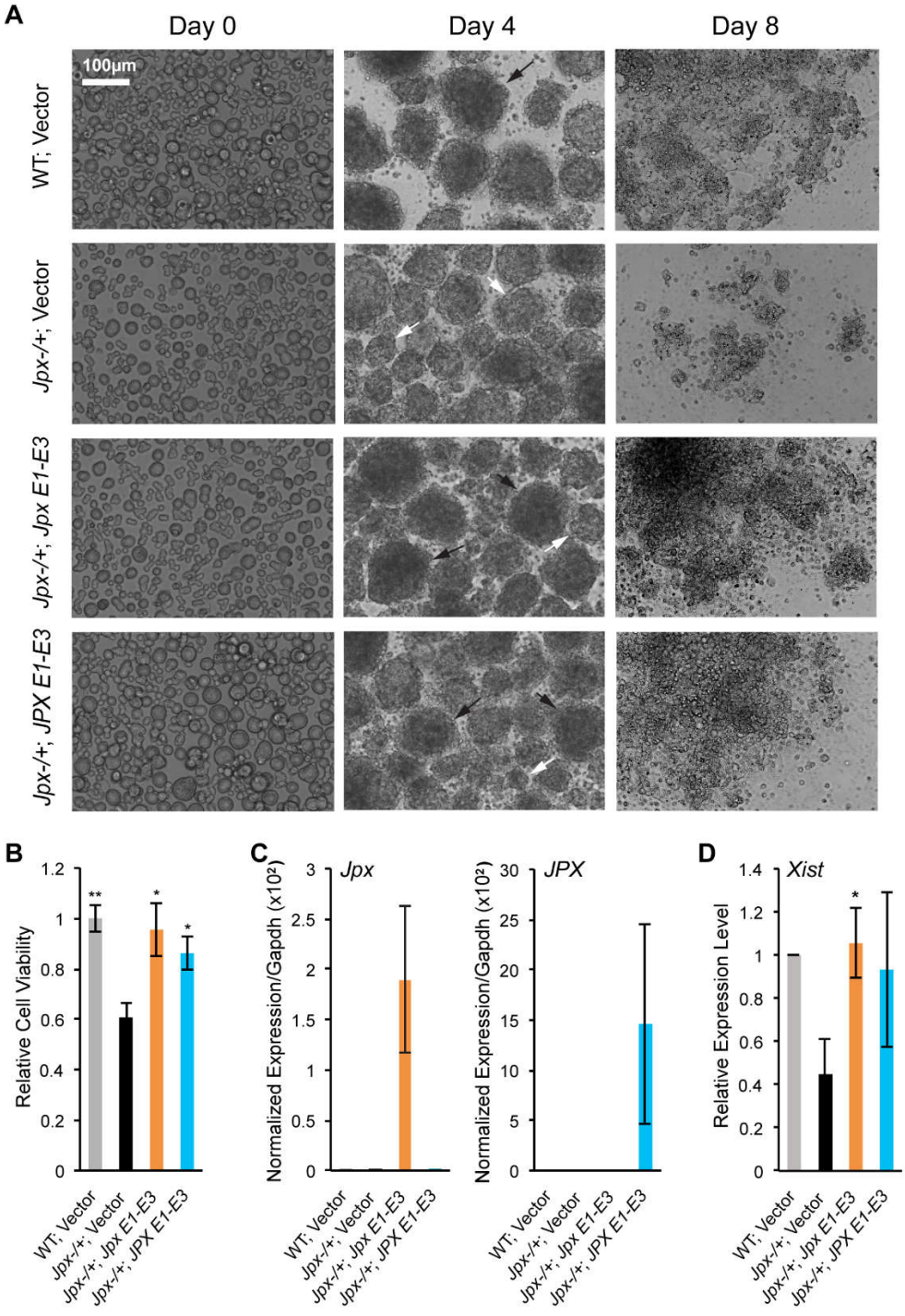

A
Day 0
Day 4
Day 8
100μm
WT; Vector
Jpx-/+; Vector
Jpx-/+; Jpx E1-E3
Jpx-/+; JPX E1-E3
B
Relative Cell Viability
C
Jpx
JPX
Normalized Expression/Gapdh (x10²)
D
Xist
Relative Expression Level
WT; Vector
Jpx-/+; Vector
Jpx-/+; Jpx E1-E3
Jpx-/+; JPX E1-E3

Figure 3.9: Rescue of *Jpx*-deletion in mouse ES cells by human lncRNA *JPX*

(A) Overexpression of *Jpx* and *JPX* RNA rescues outgrowth defect in *Jpx-/+* female mutant mES cells. Wild type (WT) control and *Jpx-/+* female mES cells transfected with vector only, Tg(EF1α:JpxE1-E3), or Tg(EF1α:JPXE1-E3). Representative brightfield images are shown of cultures on day 0, 4, and 8 of mES differentiation. Black arrows indicate normal EBs present in cultures. White arrows indicate disintegrating EBs in the cultures. **(B)** Rescue of the cell viability defect in the *Jpx-/+* mutant cells. At least three independent transfections were performed and average viability ± SEM is shown (*, $P < 0.05$ and **, $P < 0.01$ from one-tail paired Student *t*-tests in comparison to '*Jpx*-/+; Vector'). **(C)** Overexpression of *Jpx* and *JPX* achieved by Tg(EF1α:JpxE1-E3) and Tg(EF1α:JPXE1-E3), respectively, in the *Jpx-/+* mutant mES cells. At least three independent transfections were performed and qRT-PCR of *Jpx* expression (Left panel) and *JPX* expression (Right panel) were normalized to *Gapdh* mRNA. Average expression ± SEM is shown. **(D)** *Xist* RNA expression rescued by *Jpx* RNA overexpression. The qRT-PCR of *Xist* expression was normalized to *Gapdh* mRNA and is shown relative to the WT level (set to "1"). Average expression ± SEM is shown from at least three independent transfections (*, $P < 0.05$ and **, $P < 0.01$ from one-tail paired Student *t*-tests in comparison to '*Jpx-/+*; Vector')

Figure 3.10

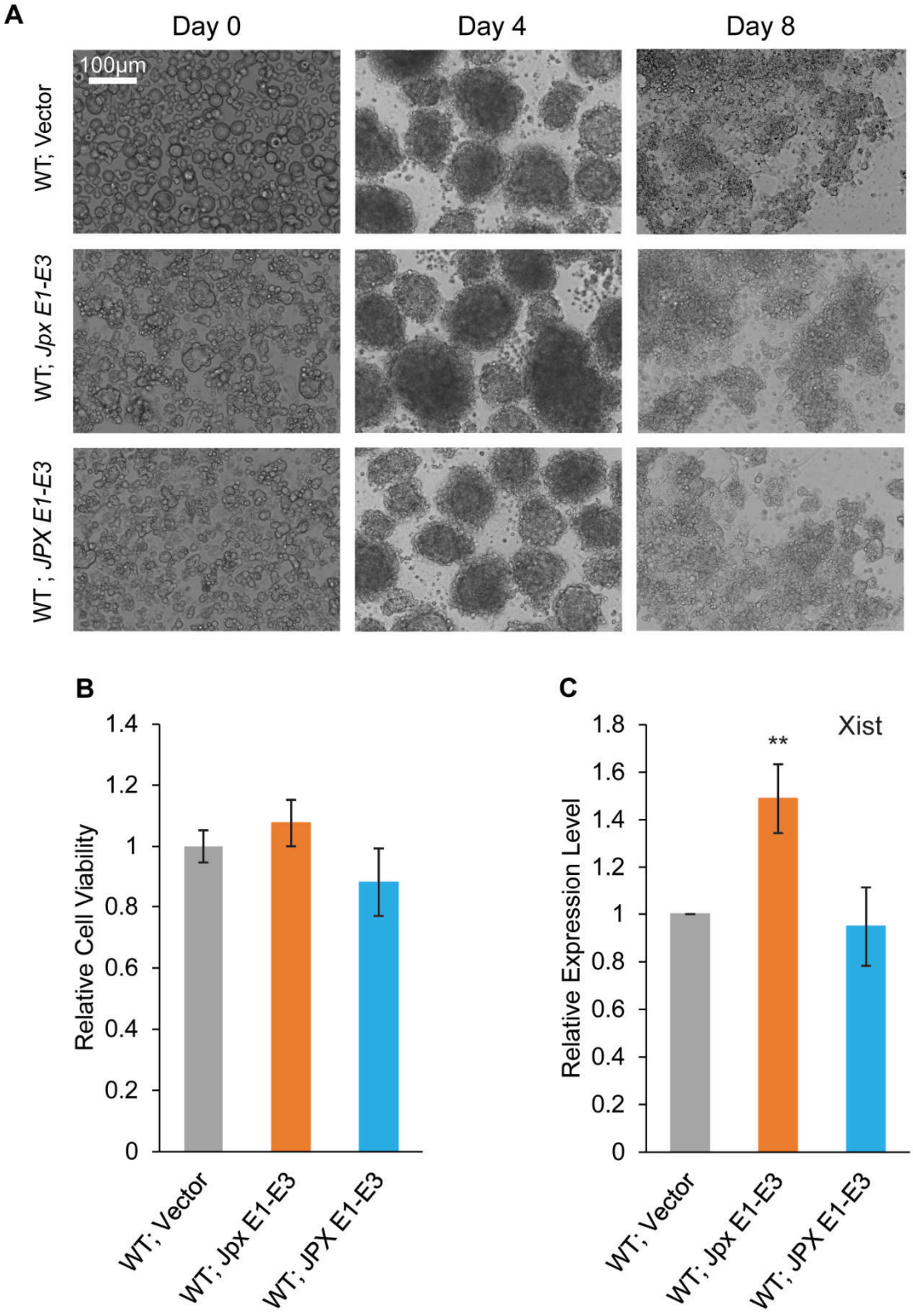

A
Day 0
Day 4
Day 8
100μm
WT; Vector
WT; Jpx E1-E3
WT ; JPX E1-E3
B
Relative Cell Viability
1.4
1.2
1
0.8
0.6
0.4
0.2
0
WT; Vector
WT; Jpx E1-E3
WT; JPX E1-E3
C
Xist
**
Relative Expression Level
1.8
1.6
1.4
1.2
1
0.8
0.6
0.4
0.2
0
WT; Vector
WT; Jpx E1-E3
WT; JPX E1-E3

Figure 3.10: No observable defect with overexpression of *Jpx/JPX* lncRNA in wild type mES cells

(A) No morphological defect detected in mES cells overexpressing *Jpx* or *JPX* RNA. Wild type (WT) control female mES cells transfected with vector only, Tg(EF1α:JpxE1-E3), or Tg(EF1α:JPXE1-E3). Images for WT with vector are the same ones shown in Figure 5, since the overexpression with WT mES cells was performed in parallel within the same experiments as the *Jpx-/+* mutant mES cells. **(B)** WT mES cells overexpressing *Jpx* or *JPX* RNA have comparable viability with that of WT mES cells transfected with vector only. At least three independent transfections were performed and average viability ± SEM is shown. **(C)** No reduction of *Xist* RNA expression caused by overexpression of *Jpx* or *JPX* RNA. The qRT-PCR of *Xist* expression was normalized to *Gapdh* mRNA and is shown relative to the WT level (set to "1"). Average expression ± SEM is shown from at least three independent transfections (**, $P < 0.01$ from one-tail paired Student *t*-tests in comparison to 'WT; Vector'). An increase of *Xist* expression by *Jpx* overexpression is consistent with previous reports (Carmona et al., 2018; Sun et al., 2013b).

3.4 Discussion

Our analyses comparing human lncRNA *JPX* with its mouse homolog lncRNA *Jpx* reveal large differences in their nucleotide sequences and RNA structures. Nevertheless, we see a conservation in their molecular functions—both human *JPX* and mouse *Jpx* lncRNAs bind CTCF, and human *JPX* can rescue *Jpx*-deletion defects in mouse embryonic stem cells. Therefore, evolutionary constraints on the sequence-structure-function linkage appear to be more relaxed and complex, consistent with current views on the genome-wide sequence evolution of noncoding RNA genes (Haerty and Ponting, 2014; Kirk et al., 2018; Necsulea and Kaessmann, 2014). A signature of positive selection acting on the human *JPX* sequence supports adaptive evolution of functional lncRNAs, which have been described in diverse organisms from *Drosophila* to mammalian species (Dai et al., 2008; Heinen et al., 2009; Kutter et al., 2012; Ponting and Lunter, 2006; Wen et al., 2016).

Early studies have reported that noncoding RNA structure can be retained across species and could possibly play a role in functional conservation, but similar analyses on lncRNAs have yet to be comprehensively performed (Ilik et al., 2013; Uroda et al., 2019; Webb et al., 2009). To address whether structural conservation contributes into the functional conservation of *Jpx/JPX*, *in vitro* SHAPE analysis of both mouse lncRNA *Jpx* and human lncRNA *JPX* was performed. Upon comparison, the two structures appear divergent, yet highly structured, with both having stem-loop formations and ~50% of their nucleotides base-paired. The divergence seen between these two lncRNAs may be the result of human *JPX* evolving due to positive selection (Johnsson et al., 2014). This is supported by our protein-binding assays, which show robust

interactions between CTCF and the human *JPX* RNA. A comparison of *in vivo* and *in vitro* SHAPE analyses on human *JPX 104-172* indicates a high percentage of RNA nucleotides reactive to NAI independently of the cellular microenvironment. This region corresponds to the E1-E2 junction that is relatively more conserved in sequence and shown to bind with CTCF, thus representing a possible lncRNA domain important for function. Consistent stem-loop features obtained from *in vivo* and *in vitro* predictions suggest that this RNA domain is structurally stable, and support the overall structural divergence observed *in vitro* for human *JPX E1-E3* and mouse *Jpx E1-E3*. RNA domains with conservation in both sequence and secondary structure have been reported for lncRNAs such as *roX*, *XIST*, *RepA*, and *MEG3* (Ilik et al., 2013; Liu et al., 2017; Lu et al., 2016; Uroda et al., 2019). We have identified regions of high sequence identity in *Jpx/JPX* sequence alignment (Figure 3.1C), which correspond to nucleotides with similar SHAPE reactivities indicative of similar secondary structures. However, a direct comparison of the RNA secondary structures for *Jpx* and *JPX* (Figure 3.5) shows no obvious correlation between nucleotide conservation and secondary structure match. It is possible that conserved nucleotides are important for a higher-order structure involving tertiary interactions between RNA domains. Further work combining mutagenesis of the conserved nucleotides in *Jpx/JPX* with structural and functional characterizations should help define the RNA motif conservation.

The lncRNA-binding capability of CTCF is supported by recent characterizations of CTCF as an RNA-binding protein with specific functions in mammalian cells (Hansen et al., 2018; Kung et al., 2015; Saldaña-Meyer et al., 2014, 2019). Our analyses of *Jpx/JPX* lncRNA-CTCF binding are also consistent with the earlier report that *Jpx* RNA

directly binds CTCF in activating *Xist* expression in mouse embryonic stem cells (Sun et al., 2013b). To determine whether specific RNA domains are responsible for binding to CTCF, we compare the binding affinities for the full-length *Jpx E1-E3* RNA and its mutant versions with 5' and/or 3' truncations, which show that both the 5' (bases 34-114) and 3' (bases 252-347) are necessary for CTCF-binding. Human *JPX E1-E3* RNA also binds to CTCF, and its binding capacity appears more robust against sequence deletions—the 5' truncation mutant *JPX 19-343* and the 3' truncation mutant *JPX 1-257* exhibit even stronger binding than the full length *JPX 1-343*. Together, our RNA EMSA results argue that CTCF-lncRNA binding with *Jpx/JPX* does not directly depend on sequence specificity.

Functional homology between human *JPX* and mouse *Jpx* is further supported by the complementary test in the mouse *Jpx-/+* mutant ES cells. Deleting a single copy of the *Jpx* gene in female mES cells disrupts *Xist* upregulation and leads to cell death during ES differentiation. Exogenous expression of either mouse *Jpx* or human *JPX* lncRNA in the mutant cells rescues *Xist* expression and cell viability. The mouse ES cell system is most suitable for the functional complementation experiment, because differentiation of mES cells faithfully recapitulates the upregulation of *Xist* and the XCI process; whereas human embryonic stem cells so far are not exhibiting the establishment of random XCI or the change of XCI state during human ES cell differentiation (Khan et al., 2017; Patel et al., 2017; Sahakyan et al., 2018). Transient transfection of mouse *Jpx E1-E3* or human *JPX E1-E3* was sufficient to increase *Xist* expression and rescue the phenotypic defects of *Jpx-/+* female mES cells, consistent with the *trans*-acting role of *Jpx/JPX* RNA on activating *Xist*. In contrast, transgenic

mES cell lines established from stable integrations of mouse *Jpx E1-E3* or human *JPX E1-E3* transgenes show poor EB morphologies during ES cell differentiation (Figure 3.11). Expression of the transgenic *Jpx* or *JPX* RNA appears less efficient and variable between independent clones, which leads to insufficient rescue of cell viability in *Jpx-/+* mutant female mES cells (Figure 3.12). The differences we observed between transiently transfected mES cells versus stable transgenic clones likely reflect the regulatory mechanism of *Jpx* that involves the *trans*-localization of lncRNA molecules and the quantitative threshold needed for activating *Xist* (Carmona et al., 2018; Li et al., 2016).

It is interesting to note that human *JPX E1-E3* is capable of rescuing *Jpx-/+* mES cell viability similarly to mouse *Jpx E1-E3* (Figure 3.9B; two-tailed paired student *t*-test, $P = 0.469$), but the efficiency of activating *Xist* is more variable with human *JPX E1-E3* than it is with mouse *Jpx E1-E3* (Figure 3.9D). This suggests that human *JPX* is complementary to mouse *Jpx* with regard to essential cellular functions. The genetic and structural divergence seen between the two may contribute to the differences in the specificity of molecular interactions, which affects the regulatory efficiency on the target gene (i.e., *Xist*). This is also consistent with results from the *in vitro* protein-binding assays using CTCF.

In conclusion, through comparative sequence and functional analyses involving the homologous human *JPX*, our results have demonstrated a convergent function of *Jpx/JPX* between mice and humans despite a rapid divergence in the nucleotide sequences and a change of the RNA secondary structures. Our findings suggest that lncRNAs are capable of maintaining essential roles in embryogenesis and such lncRNA

functions may be resistant to evolutionary constraints at both RNA sequence and structural levels.

Figure 3.11

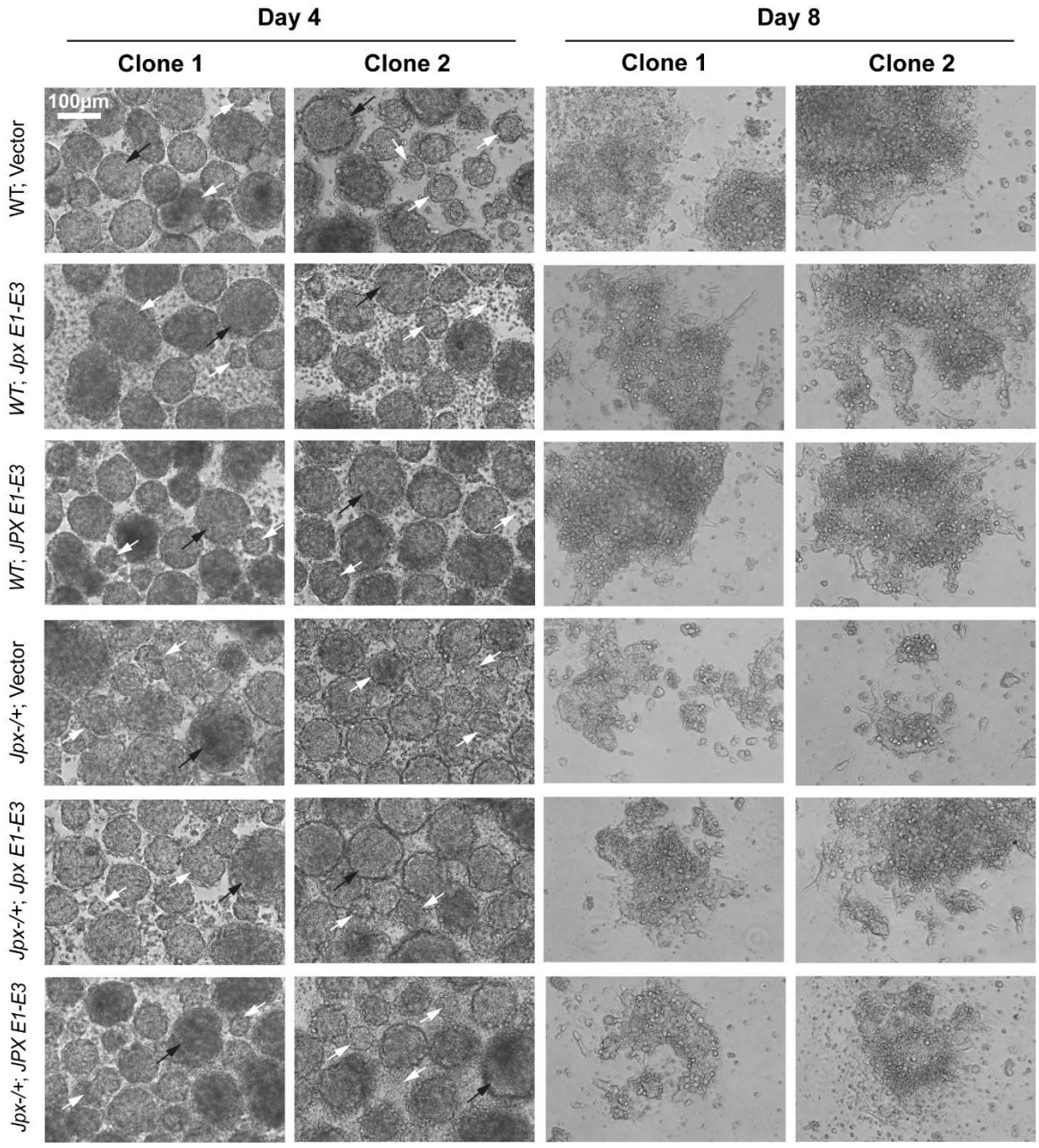


Figure 3.11: ES differentiation and EB outgrowth of stable transgenic mES cells with overexpression of *Jpx/JPX*

Two independent clones from each stable transfection of *Jpx* and *JPX* transgenes in wild type and *Jpx-/+* female mutant mES cells. Representative brightfield images for Day 4 (EB formation) and Day 8 (EB outgrowth) of differentiated mES cells are shown for wild type (WT) control and *Jpx-/+* female mES cells stably transfected with vector only, Tg(EF1α:JpxE1-E3), or Tg(EF1α:JPXE1-E3). Black arrows indicate normal EBs present in cultures. White arrows indicate disintegrating EBs in cultures.

Figure 3.12

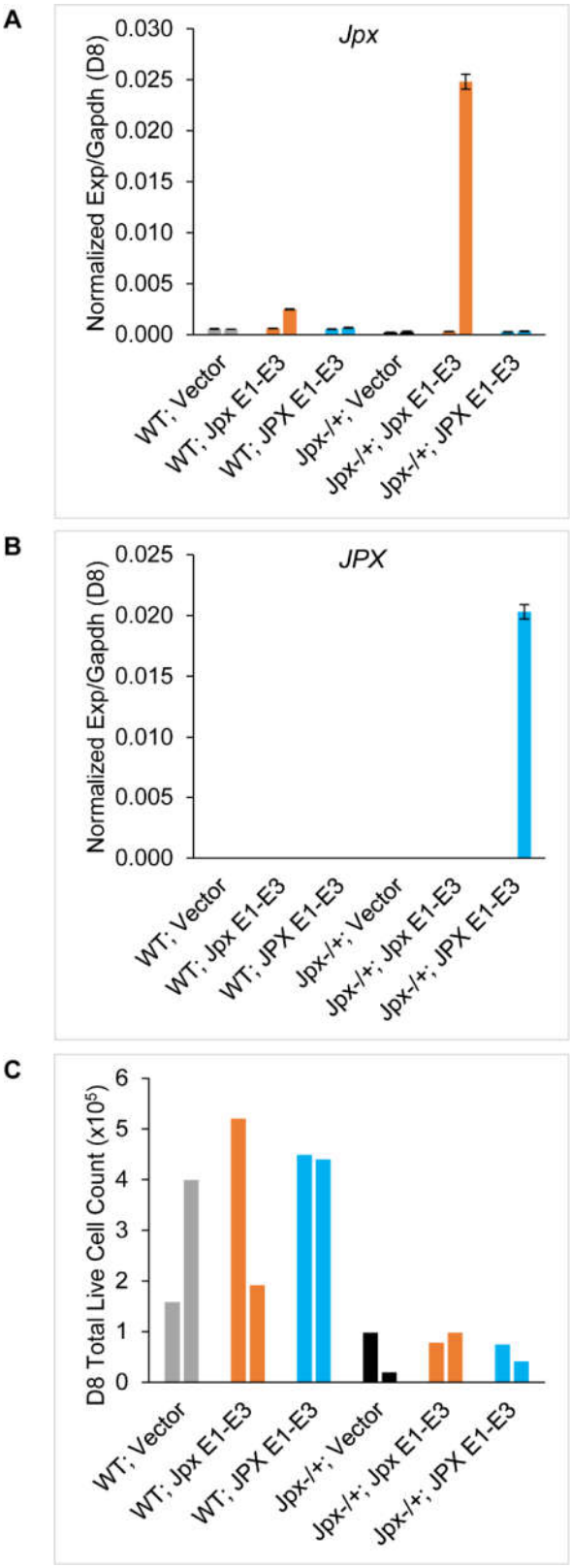

A
Jpx
Normalized Exp/Gapdh (D8)
0.030
0.025
0.020
0.015
0.010
0.005
0.000
WT; Vector
WT; Jpx E1-E3
WT; JPX E1-E3
Jpx-/+; Vector
Jpx-/+; Jpx E1-E3
Jpx-/+; JPX E1-E3
B
JPX
Normalized Exp/Gapdh (D8)
0.025
0.020
0.015
0.010
0.005
0.000
WT; Vector
WT; Jpx E1-E3
WT; JPX E1-E3
Jpx-/+; Vector
Jpx-/+; Jpx E1-E3
Jpx-/+; JPX E1-E3
C
D8 Total Live Cell Count (x10^5)
6
5
4
3
2
1
0
WT; Vector
WT; Jpx E1-E3
WT; JPX E1-E3
Jpx-/+; Vector
Jpx-/+; Jpx E1-E3
Jpx-/+; JPX E1-E3

Figure 3.12 Expression of *Jpx/JPX* and viability of cells in transgenic mES cells with stable transfection of *Jpx/JPX*

(A) Expression of *Jpx* in mES cells carrying stable transgenes. Two independent clones for each: wild type female mES cells (WT) with vector only; WT with Tg(EF1α:JpxE1-E3); WT with Tg(EF1α:JPXE1-E3), and mutant female mES cells (*Jpx-/+*) with vector only; *Jpx-/+* with Tg(EF1α:JpxE1-E3); *Jpx-/+* with Tg(EF1α:JPXE1-E3). Bars represent the average of at least two qRT-PCR plate replicates of *Jpx* expression normalized to *Gapdh* for each sample. Technical replicate average expression ± SEM is shown. **(B)** Expression of *JPX* in the same mES cells carrying stable transgenes as in (A). Bars represent the average of at least two qRT-PCR plate replicates of *JPX* expression normalized to *Gapdh* for each. Technical replicate average expression ± SEM is shown. **(C)** Total live cell count on ES differentiation Day 8 for the mES cells carrying stable transgenes. Two independent clones for each type were analyzed as in (A) & (B). All samples have started with the same number of ~$5x10^5$ undifferentiated mES cells.

3.5 References

Brockdorff, N., Ashworth, A., Kay, G.F., McCabe, V.M., Norris, D.P., Cooper, P.J., Swift, S., and Rastan, S. (1992). 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.

Brown, C.J., Hendrich, B.D., Rupert, J.L., Lafrenière, R.G., Xing, Y., Lawrence, J., and Willard, H.F. (1992). The human XIST gene: Analysis of a 17 kb inactive X-specific RNA that contains conserved repeats and is highly localized within the nucleus. Cell *71*, 527–542.

Cabili, M.N., Trapnell, C., Goff, L., Koziol, M., Tazon-Vega, B., Regev, A., and Rinn, J.L. (2011). Integrative annotation of human large intergenic noncoding RNAs reveals global properties and specific subclasses. Genes Dev. *25*, 1915–1927.

Carmona, S., Lin, B., Chou, T., Arroyo, K., and Sun, S. (2018). LncRNA Jpx induces Xist expression in mice using both trans and cis mechanisms. PLOS Genet. *14*, e1007378.

Casanova, M., Liyakat Ali, T.M., and Rougeulle, C. (2016). Enlightening the contribution of the dark matter to the X chromosome inactivation process in mammals. Semin. Cell Dev. Biol. *56*, 48–57.

Cooper, G.M. (2004). Characterization of Evolutionary Rates and Constraints in Three Mammalian Genomes. Genome Res. *14*, 539–548.

Dai, H., Chen, Y., Chen, S., Mao, Q., Kennedy, D., Landback, P., Eyre-Walker, A., Du, W., and Long, M. (2008). The evolution of courtship behaviours through the origination of a new gene in Drosophila. Proc. Natl. Acad. Sci. *105*, 7478–7483.

Das, R., Laederach, A., Pearlman, S.M., Herschlag, D., and Altman, R.B. (2005). SAFA: semi-automated footprinting analysis software for high-throughput quantification of nucleic acid footprinting experiments. RNA *11*, 344–354.

Delli Ponti, R., Armaos, A., Marti, S., and Tartaglia, G.G. (2018). A Method for RNA Structure Prediction Shows Evidence for Structure in lncRNAs. Front. Mol. Biosci. *5*, 111.

Derrien, T., Johnson, R., Bussotti, G., Tanzer, A., Djebali, S., Tilgner, H., Guernec, G., Martin, D., Merkel, A., Knowles, D.G., et al. (2012). The GENCODE v7 catalog of human long noncoding RNAs: Analysis of their gene structure, evolution, and expression. Genome Res. *22*, 1775–1789.

Duret, L., Chureau, C., Samain, S., Weissanbach, J., and Avner, P. (2006). The Xist RNA gene evolved in eutherians by pseudogenization of a protein-coding gene. Science *312*, 1653–1655.

Elisaphenko, E.A., Kolesnikov, N.N., Shevchenko, A.I., Rogozin, I.B., Nesterova, T.B., Brockdorff, N., and Zakian, S.M. (2008). A Dual Origin of the Xist Gene from a Protein-Coding Gene and a Set of Transposable Elements. PLOS One *3*, e2521.

Fang, R., Moss, W.N., Rutenberg-Schoenberg, M., and Simon, M.D. (2015). Probing Xist RNA Structure in Cells Using Targeted Structure-Seq. PLOS Genet. *11*, e1005668.

Fatica, A., and Bozzoni, I. (2014). Long non-coding RNAs: new players in cell differentiation and development. Nat. Rev. Genet. *15*, 7–21.

Gao, G., Vibranovski, M.D., Zhang, L., Li, Z., Liu, M., Zhang, Y.E., Li, X., Zhang, W., Fan, Q., VanKuren, N.W., et al. (2014). A long-term demasculinization of X-linked intergenic noncoding RNAs in Drosophila melanogaster. Genome Res. *24*, 629–638.

Gibbs, R.A., Weinstock, G.M., Metzker, M.L., Muzny, D.M., Sodergren, E.J., Scherer, S., Scott, G., Steffen, D., Worley, K.C., Burch, P.E., et al. (2004). Genome sequence of the Brown Norway rat yields insights into mammalian evolution. Nature *428*, 493–520.

Grant, J., Mahadevaiah, S.K., Khil, P., Sangrithi, M.N., Royo, H., Duckworth, J., McCarrey, J.R., VandeBerg, J.L., Renfree, M.B., Taylor, W., et al. (2012). Rsx is a metatherian RNA with Xist-like properties in X-chromosome inactivation. Nature *487*, 254–258.

Haerty, W., and Ponting, C.P. (2014). No Gene in the Genome Makes Sense Except in the Light of Evolution. Annu. Rev. Genomics Hum. Genet. *15*, 71–92.

Hansen, A.S., Hsieh, T.-H.S., Cattoglio, C., Pustova, I., Darzacq, X., and Tjian, R. (2018). An RNA-binding region regulates CTCF clustering and chromatin looping. BioRxiv.

Heinen, T.J.A.J., Staubach, F., Häming, D., and Tautz, D. (2009). Emergence of a New Gene from an Intergenic Region. Curr. Biol. *19*, 1527–1531.

Hezroni, H., Koppstein, D., Schwartz, M.G., Avrutin, A., Bartel, D.P., and Ulitsky, I. (2015). Principles of Long Noncoding RNA Evolution Derived from Direct Comparison of Transcriptomes in 17 Species. Cell Rep. *11*, 1110–1122.

Hezroni, H., Ben-Tov Perry, R., Meir, Z., Housman, G., Lubelsky, Y., and Ulitsky, I. (2017). A subset of conserved mammalian long non-coding RNAs are fossils of ancestral protein-coding genes. Genome Biol. *18*, 162.

de Hoon, B., Splinter, E., Eussen, B., Douben, J.C.W., Rentmeester, E., van de Heijning, M., Laven, J.S.E., de Klein, J.E.M.M., Liebelt, J., and Gribnau, J. (2017). X chromosome inactivation in a female carrier of a 1.28 mb deletion encompassing the human X inactivation centre. Philos. Trans. R. Soc. B Biol. Sci. *372*, 20160359

Horvath, J.E., Sheedy, C.B., Merrett, S.L., Diallo, A.B., Swofford, D.L., NISC Comparative Sequencing Program, Green, E.D., and Willard, H.F. (2011). Comparative analysis of the primate X-inactivation center region and reconstruction of the ancestral primate XIST locus. Genome Res. *21*, 850–862.

Ilik, I., Quinn, J., Georgiev, P., Tavares-Cadete, F., Maticzka, D., Toscano, S., Wan, Y., Spitale, R., Luscombe, N., Backofen, R., et al. (2013). Tandem Stem-Loops in roX RNAs Act Together to Mediate X Chromosome Dosage Compensation in Drosophila. Mol. Cell *51*, 156–173.

Johnsson, P., and Morris, K. V (2014). Expanding the functional role of long noncoding RNAs. Cell Res. *24*, 1284–1285.

Johnsson, P., Lipovich, L., Grandér, D., and Morris, K. V (2014). Evolutionary conservation of long non-coding RNAs; sequence, structure, function. Biochim. Biophys. Acta *1840*, 1063–1071.

Johnston, C.M., Newall, A.E.T., Brockdorff, N., and Nesterova, T.B. (2002). Enox, a Novel Gene That Maps 10 kb Upstream of Xist and Partially Escapes X Inactivation. Genomics *80*, 236–244.

Kapranov, P., St Laurent, G., Raz, T., Ozsolak, F., Reynolds, C.P., Sorensen, P.H.B., Reaman, G., Milos, P., Arceci, R.J., Thompson, J.F., et al. (2010). The majority of total nuclear-encoded non-ribosomal RNA in a human cell is “dark matter” un-annotated RNA. BMC Biol. *8*, 149.

Kapusta, A., and Feschotte, C. (2014). Volatile evolution of long noncoding RNA repertoires: mechanisms and biological implications. Trends Genet. *30*, 439–452.

Kern, C., Wang, Y., Chitwood, J., Korf, I., Delany, M., Cheng, H., Medrano, J.F., Van Eenennaam, A.L., Ernst, C., Ross, P., et al. (2018). Genome-wide identification of tissue-specific long non-coding RNA in three farm animal species. BMC Genomics *19*, 684.

Khan, S.A., Audergon, P.N.C.B., and Payer, B. (2017). X-chromosome activity in naive human pluripotent stem cells—are we there yet? Stem Cell Investig. *4*, 54–54.

Kirk, J.M., Kim, S.O., Inoue, K., Smola, M.J., Lee, D.M., Schertzer, M.D., Wooten, J.S., Baker, A.R., Sprague, D., Collins, D.W., et al. (2018). Functional classification of long non-coding RNAs by k-mer content. Nat. Genet. *50*, 1474–1482.

Kolesnikov, N.N., and Elisaphenko, E.A. (2010). Comparative organization and the origin of noncoding regulatory RNA genes from X-chromosome inactivation center of human and mouse. Russ. J. Genet. *46*, 1223–1228.

Kopp, F., and Mendell, J.T. (2018). Functional Classification and Experimental Dissection of Long Noncoding RNAs. Cell *172*, 393–407.

Kung, J.T., Kesner, B., An, J.Y., Ahn, J.Y., Cifuentes-Rojas, C., Colognori, D., Jeon, Y., Szanto, A., del Rosario, B.C., Pinter, S.F., et al. (2015). Locus-Specific Targeting to the X Chromosome Revealed by the RNA Interactome of CTCF. Mol. Cell *57*, 361–375.

Kutter, C., Watt, S., Stefflova, K., Wilson, M.D., Goncalves, A., Ponting, C.P., Odom, D.T., and Marques, A.C. (2012). Rapid Turnover of Long Noncoding RNAs and the Evolution of Gene Expression. PLOS Genet. *8*, e1002841.

Kwok, C.K., Ding, Y., Tang, Y., Assmann, S.M., and Bevilacqua, P.C. (2013). Determination of in vivo RNA structure in low-abundance transcripts. Nat. Commun. *4*, 2971.

Laederach, A., Das, R., Vicens, Q., Pearlman, S.M., Herschlag, D., and Altman, R.B. (2008). Semi-automated and rapid quantification of nucleic acid footprinting and structure mapping experiments. Nat. Protoc. *3*, 1395–1401.

Larkin, M.A., Blackshields, G., Brown, N.P., Chenna, R., McGettigan, P.A., McWilliam, H., Valentin, F., Wallace, I.M., Wilm, A., Lopez, R., et al. (2007). Clustal W and Clustal X version 2.0. Bioinformatics *23*, 2947–2948.

Lee, J.T., and Bartolomei, M.S. (2013). X-Inactivation, Imprinting, and Long Noncoding RNAs in Health and Disease. Cell *152*, 1308–1323.

Lee, J.T., Lu, N., and Han, Y. (1999). Genetic analysis of the mouse X inactivation center defines an 80-kb multifunction domain. Proc. Natl. Acad. Sci. *96*, 3836–3841.

Li, C., Hong, T., Webb, C.-H., Karner, H., Sun, S., and Nie, Q. (2016). A self-enhanced transport mechanism through long noncoding RNAs for X chromosome inactivation. Sci. Rep. *6*, 31517.

Ling, H., Vincent, K., Pichler, M., Fodde, R., Berindan-Neagoe, I., Slack, F.J., and Calin, G.A. (2015). Junk DNA and the long non-coding RNA twist in cancer genetics. Oncogene *34*, 5003–5011.

Liu, F., Somarowthu, S., and Pyle, A.M. (2017). Visualizing the secondary and tertiary architectural domains of lncRNA RepA. Nat. Chem. Biol. *13*, 282–289.

Lorenz, R., Bernhart, S.H., Höner zu Siederdissen, C., Tafer, H., Flamm, C., Stadler, P.F., and Hofacker, I.L. (2011). ViennaRNA Package 2.0. Algorithms Mol. Biol. *6*, 26.

Lu, Z., Zhang, Q.C., Lee, B., Flynn, R.A., Smith, M.A., Robinson, J.T., Davidovich, C., Gooding, A.R., Goodrich, K.J., Mattick, J.S., et al. (2016). RNA Duplex Map in Living Cells Reveals Higher-Order Transcriptome Structure. Cell *165*, 1267–1279.

Lucks, J.B., Mortimer, S.A., Trapnell, C., Luo, S., Aviran, S., Schroth, G.P., Pachter, L., Doudna, J.A., and Arkin, A.P. (2011). Multiplexed RNA structure characterization with selective 2'-hydroxyl acylation analyzed by primer extension sequencing (SHAPE-Seq). Proc. Natl. Acad. Sci. *108*, 11063–11068.

Makalowski, W., Zhang, J., and Boguski, M.S. (1996). Comparative analysis of 1196 orthologous mouse and human full-length mRNA and protein sequences. Genome Res. *6*, 846–857.

Mercer, T.R., and Mattick, J.S. (2013). Structure and function of long noncoding RNAs in epigenetic regulation. Nat. Struct. Mol. Biol. *20*, 300–307.

Migeon, B.R. (2011). The single active X in human cells: evolutionary tinkering personified. Hum. Genet. *130*, 281–293.

Moorjani, P., Amorim, C.E.G., Arndt, P.F., and Przeworski, M. (2016). Variation in the molecular clock of primates. Proc. Natl. Acad. Sci. *113*, 10607–10612.

Necsulea, A., and Kaessmann, H. (2014). Evolutionary dynamics of coding and non-coding transcriptomes. Nat. Rev. Genet. *15*, 734–748.

Necsulea, A., Soumillon, M., Warnefors, M., Liechti, A., Daish, T., Zeller, U., Baker, J.C., Grützner, F., and Kaessmann, H. (2014). The evolution of lncRNA repertoires and expression patterns in tetrapods. Nature *505*, 635–640.

Novikova, I. V., Hennelly, S.P., and Sanbonmatsu, K.Y. (2012). Structural architecture of the human long non-coding RNA, steroid receptor RNA activator. Nucleic Acids Res. *40*, 5034–5051.

Ohlsson, R., Renkawitz, R., and Lobanenkov, V. (2001). CTCF is a uniquely versatile transcription regulator linked to epigenetics and disease. Trends Genet. *17*, 520–527.

Patel, S., Bonora, G., Sahakyan, A., Kim, R., Chronis, C., Langerman, J., Fitz-Gibbon, S., Rubbi, L., Skelton, R.J.P., Ardehali, R., et al. (2017). Human Embryonic Stem Cells Do Not Change Their X Inactivation Status during Differentiation. Cell Rep. *18*, 54–67.

Payer, B., and Lee, J.T. (2008). X Chromosome Dosage Compensation: How Mammals Keep the Balance. Annu. Rev. Genet. *42*, 733–772.

Perry, R.B.-T., and Ulitsky, I. (2016). The functions of long noncoding RNAs in development and stem cells. Development *143*, 3882–3894.

Petropoulos, S., Edsgärd, D., Reinius, B., Deng, Q., Panula, S.P., Codeluppi, S., Plaza Reyes, A., Linnarsson, S., Sandberg, R., and Lanner, F. (2016). Single-Cell RNA-Seq Reveals Lineage and X Chromosome Dynamics in Human Preimplantation Embryos. Cell *165*, 1012–1026.

Phillips, J.E., and Corces, V.G. (2009). CTCF: Master Weaver of the Genome. Cell *137*, 1194–1211.

Ponjavic, J., Ponting, C.P., and Lunter, G. (2007). Functionality or transcriptional noise? Evidence for selection within long noncoding RNAs. Genome Res. *17*, 556–565.

Ponting, C.P., and Lunter, G. (2006). Signatures of adaptive evolution within human non-coding sequence. Hum. Mol. Genet. *15*, R170–R175.

Ransohoff, J.D., Wei, Y., and Khavari, P.A. (2018). The functions and unique features of long intergenic non-coding RNA. Nat. Rev. Mol. Cell Biol. *19*, 143–157.

Rinn, J.L., and Chang, H.Y. (2012). Genome Regulation by Long Noncoding RNAs. Annu. Rev. Biochem. *81*, 145–166.

Sado, T., and Brockdorff, N. (2013). Advances in understanding chromosome silencing by the long non-coding RNA Xist. Philos. Trans. R. Soc. B *368*, 20110325.

Sahakyan, A., Yang, Y., and Plath, K. (2018). The Role of Xist in X-Chromosome Dosage Compensation. Trends Cell Biol. *28*, 999–1013.

Saitou, N., and Nei, M. (1987). The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol. Biol. Evol. *4*, 406–425.

Saldaña-Meyer, R., Gonzalez-Buendia, E., Guerrero, G., Narendra, V., Bonasio, R., Recillas-Targa, F., and Reinberg, D. (2014). CTCF regulates the human p53 gene through direct interaction with its natural antisense transcript, Wrap53. Genes Dev. *28*, 723–734.

Saldaña-Meyer, R., Rodriguez-Hernaez, J., Nishana, M., Jacome-Lopez, K., Nora, E.P., Bruneau, B.G., Furlan-Magaril, M., Skok, J., and Reinberg, D. (2019). RNA interactions with CTCF are essential for its proper function. BioRxiv.

Smola, M.J., Christy, T.W., Inoue, K., Nicholson, C.O., Friedersdorf, M., Keene, J.D., Lee, D.M., Calabrese, J.M., and Weeks, K.M. (2016). SHAPE reveals transcript-wide interactions, complex structural domains, and protein interactions across the Xist lncRNA in living cells. Proc. Natl. Acad. Sci. *113*, 10322–10327.

Spitale, R.C., Crisalli, P., Flynn, R.A., Torre, E.A., Kool, E.T., and Chang, H.Y. (2013). RNA SHAPE analysis in living cells. Nat. Chem. Biol. *9*, 18–20.

Straub, T., and Becker, P.B. (2007). Dosage compensation: the beginning and end of generalization. Nat. Rev. Genet. *8*, 47–57.

Sun, J., Lin, Y., and Wu, J. (2013a). Long Non-Coding RNA Expression Profiling of Mouse Testis during Postnatal Development. PLOS One *8*, e75750.

Sun, S., Del Rosario, B.C., Szanto, A., Ogawa, Y., Jeon, Y., and Lee, J.T. (2013b). Jpx RNA Activates Xist by Evicting CTCF. Cell *153*, 1537–1551.

Tamura, K., Nei, M., and Kumar, S. (2004). Prospects for inferring very large phylogenies by using the neighbor-joining method. Proc. Natl. Acad. Sci. *101*, 11030–11035.

Tian, D., Sun, S., and Lee, J.T. (2010). The Long Noncoding RNA, Jpx, Is a Molecular Switch for X Chromosome Inactivation. Cell *143*, 390–403.

Tsuritani, K., Irie, T., Yamashita, R., Sakakibara, Y., Wakaguri, H., Kanai, A., Mizushima-Sugano, J., Sugano, S., Nakai, K., and Suzuki, Y. (2007). Distinct class of putative “non-conserved” promoters in humans: Comparative studies of alternative promoters of human and mouse genes. Genome Res. *17*, 1005–1014.

Ulitsky, I., Shkumatava, A., Jan, C.H., Sive, H., and Bartel, D.P. (2011). Conserved Function of lincRNAs in Vertebrate Embryonic Development despite Rapid Sequence Evolution. Cell *147*, 1537–1550.

Uroda, T., Anastasakou, E., Rossi, A., Teulon, J.-M., Pellequer, J.-L., Annibale, P., Pessey, O., Inga, A., Chillón, I., and Marcia, M. (2019). Conserved Pseudoknots in lncRNA MEG3 Are Essential for Stimulation of the p53 Pathway. Mol. Cell *75*, 982-995.e9.

Washietl, S., Hofacker, I.L., Stadler, P.F., and Kellis, M. (2012). RNA folding with soft constraints: reconciliation of probing data and thermodynamic secondary structure prediction. Nucleic Acids Res. *40*, 4261–4272.

Webb, C.-H.T., Riccitelli, N.J., Ruminski, D.J., and Luptak, A. (2009). Widespread Occurrence of Self-Cleaving Ribozymes. Science *326*, 953–953.

Wen, K., Yang, L., Xiong, T., Di, C., Ma, D., Wu, M., Xue, Z., Zhang, X., Long, L., Zhang, W., et al. (2016). Critical roles of long noncoding RNAs in Drosophila spermatogenesis. Genome Res. *26*, 1233–1244.

Wilkinson, K.A., Merino, E.J., and Weeks, K.M. (2006). Selective 2′-hydroxyl acylation analyzed by primer extension (SHAPE): quantitative RNA structure analysis at single nucleotide resolution. Nat. Protoc. *1*, 1610–1616.

Wu, H., Yang, L., and Chen, L.-L. (2017). The Diversity of Long Noncoding RNAs and Their Generation. Trends Genet. *33*, 540–552.

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

Xue, C., Li, F., and Li, F. (2008). Finding noncoding RNA transcripts from low abundance expressed sequence tags. Cell Res. *18*, 695–700.

Zarringhalam, K., Meyer, M.M., Dotu, I., Chuang, J.H., and Clote, P. (2012). Integrating Chemical Footprinting Data into RNA Secondary Structure Prediction. PLOS One *7*, e45160.

CHAPTER 4: LncRNA *JPX* and its Target, *XIST*, in Cancer

4.1 Abstract

The genetic complexity of cancer has influenced medical treatments to move toward more personalized diagnostics. There is an emerging class of gene regulators known as long noncoding RNA. The molecular functions and observed high-degree of expression specificity that have been uncovered for some lncRNAs have made them promising candidates to investigate as therapeutic targets in cancer. *XIST*, the master regulator of XCI, is one of the earliest studied cancer-associated lncRNAs and has been shown to be dysregulated in breast and ovarian cancers. Both *XIST* and its proposed activator, *JPX*, have been indicated as potential prognostic markers in hepatocellular carcinoma. The role that these lncRNAs play in cancer regulation and progression, though, are still mostly unknown. We have found that *JPX* and *XIST* expression is dysregulated in higher-grade ovarian cancer patients as well as in ovarian cancer cells lines. Additionally, the loss of *JPX* and *XIST* expression in an immortalized ovarian cell line implicates an inverse relationship between the expression of cancer metastasis associated gene, *MSN*, and the expression of *JPX* and *XIST*. However, knockdown of *XIST* alone resulted in downregulation of *MSN* and *KDM6A*, increase in proliferation, and accelerated migration. Our findings support the further investigation of these two lncRNAs in order to find how they are mechanistically involved in the progression of cancer and how to exploit this relationship in the treatment of cancer patients.

4.2 Introduction

Ovarian cancer is the 5th leading cause of cancer death in women in the United States, according to the NIH National Cancer Institute. The 5-year survival rate for ovarian cancer across all stages is 47%, with non-Hispanic Black women having the lowest survival rate of 35%. Detection of ovarian cancer in women is most often at later stages, with 34% being detected at stage III and 26% at stage IV. In contrast to breast cancer, which has many early screening options required by healthcare, there is no effective early screening method for ovarian cancer. Furthermore, ovarian cancer does not have any noticeable signs or symptoms at early stages. All of this contributes to ovarian cancer being found at later, more aggressive stages, when the cancer has a greater potential to metastasize (Torre et al., 2018). Therefore, it is imperative that not just a prognostic marker is needed to aid in diagnosis of these women, but also a biomarker that can inform a course of action at these later stages.

As more information about the complex genetic make-up of cancer is discovered, medicine has been moving toward more personalized diagnostics for treatment of cancer patients (Abramovitz et al., 2019; Kim et al., 2018). High-throughput sequencing has revealed that the majority of cancer related gene variations are in the noncoding regions of the human genome, with only a small amount found in protein-coding regions (Cheetham et al., 2013). The majority of the noncoding region of the human genome contains lncRNA genes (70-90%), an emerging class of gene regulators that consist of transcripts that are longer than 200 nucleotides (Kapranov et al., 2010). This class of transcripts have been recognized for their roles in various biological processes and

even in the hallmarks of cancer (Gutschner and Diederichs, 2012; Huarte, 2015; Schmitt and Chang, 2016).

With documented molecular functions and an observed high-degree of expression specificity, some lncRNAs have been proposed as promising biomarkers and therapeutic targets in cancer. In testicular cancer and breast cancer, *XIST* has been proposed as a prognostic marker and biomarker for treatment (Kawakami et al., 2004; Xing et al., 2018). Both *XIST*, and its potential activator in humans, *JPX*, have been studied as potential prognostic markers in hepatocellular carcinoma (Ma et al., 2017). Additionally, *PCA3* lncRNA has been shown to be more accurate than the conventional protein biomarkers for prostate cancer and is currently being used for prognostic screening in the clinic (Crawford et al., 2012; Lee et al., 2011). Despite this influx of information about lncRNAs, there is still much to be learned about the properties and functions of this class of transcripts in order to exploit them for use in cancer diagnosis and treatment. Understanding how potential markers influence hallmarks of cancer, such as metastasis and proliferation, is important to the selection of treatment. An ideal biomarker can indicate a pathway that can be exploited in a particular patient group as well as offer treatment options that selectively target these specific cells.

Xist/XIST is known as the master regulator of X chromosome inactivation, with *Jpx* being shown to be its activator in mice (Carmona et al., 2018; Sun et al., 2013; Tian et al., 2010). XCI is a developmentally regulated dosage compensation process that balances X chromosome expression between females and males. *XIST* is one of the earliest studied cancer-associated lncRNAs, and is dysregulated in breast and ovarian female cancers (Perez et al., 2008). A strong connection between *XIST* and cancer has

been further demonstrated by deleting *Xist* in the blood lineages of mice, which leads to female specific hematologic cancer (Yildirim et al., 2013). However, the way human *XIST* functions as a tumor suppressor is mostly unknown.

Our lab's research has been focused on the positive and negative regulators of *Xist* and, for the first time, reported detailed molecular mechanisms for the function of lncRNA *Jpx* in activating *Xist* and initiating XCI in mice (Fig. 4.1A-C) (Carmona et al., 2018; Sun et al., 2013). Still, it remains elusive how the molecular mechanism and regulatory elements change in cancer cells (Fig. 4.1D), and what role *XIST* plays in affecting cancer susceptibility, especially in females.

JPX and *XIST* have both been shown to be dysregulated in hepatocellular carcinoma (Ma et al., 2017), and low expression of *XIST* in breast cancer was indicative of aggressive cancer progression and decreased survival prognosis (Xing et al., 2018). Such dysregulation of XCI-associated lncRNAs may influence aggression and metastasis in a variety of cancers, making them potential biomarkers that can inform treatment plans.

In this study, we investigate whether *JPX* and *XIST* can function as tumor suppressors, influencing signaling pathways related to proliferation and metastasis. Our data indicates that *XIST* is significantly dysregulated in ovarian cancer patients, with a downregulation of expression correlated with neoplasm hematological grade. We have identified the tumor growth relevant pathways and cancer related genes that are associated with dysregulated expression of *JPX* and *XIST* in ovarian cancer patients and cell lines. We hope to utilize the results of this project to aid in the treatment of late

stage ovarian cancer patients, and further the understanding the roles *JPX* and *XIST* play in cancer.

Figure 4.1

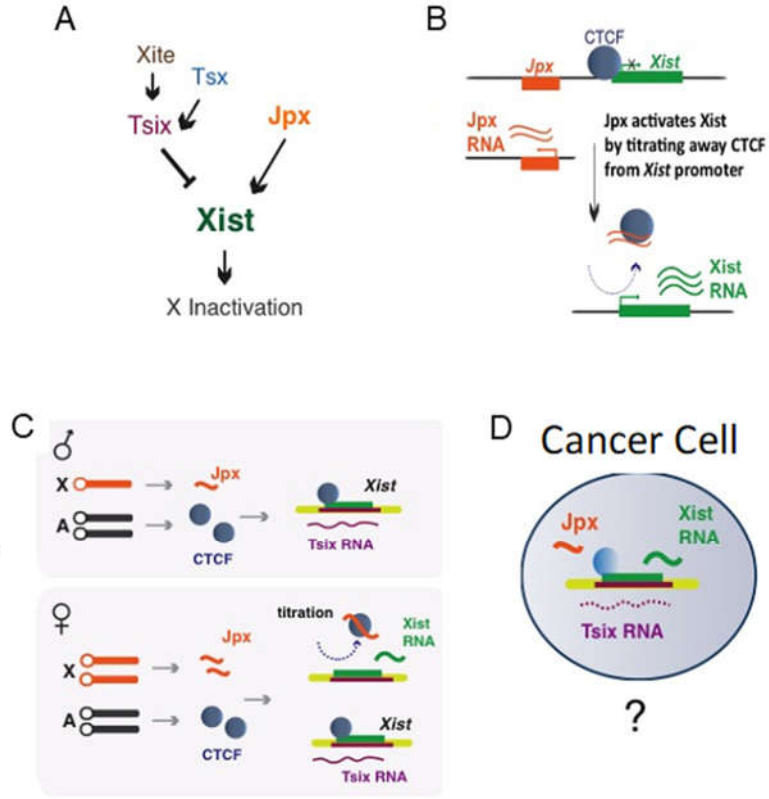


Figure 4.1: Regulatory control of *Xist*

(A) *Xist* is regulated by positive and negative factors, such as noncoding RNAs for X inactivation in the female. **(B)** *Jpx* RNA is transcribed upstream of *Xist*, removes CTCF from the *Xist* promoter, and activates *Xist*. **(C)** *Xist* activation is dose-dependent: *Jpx* titrates away CTCF only when present in 2-fold excess, such as in female somatic cells, and is insufficient to activate *Xist* in male somatic cells. **(D)** *Xist* and its regulatory RNAs are present in cancer cells, but neither their dysregulation nor function is understood.

4.3 Results

JPX and *XIST* expression is decreased in patients with higher grades of ovarian cancer

In a survey of 91 datasets from GenBank, 16 lncRNAs were found to have differential expression that was correlated with 12 different cancers in response to drug treatments or alteration in metastasis. Human *JPX* and *XIST* are among the listed lncRNAs (Table 4.1), which is consistent with previous reports that these lncRNAs are dysregulated in cancer and are potential prognostic markers (Kawakami et al., 2004; Ma et al., 2017; Xing et al., 2018).

Increased severity of cancer (higher neoplasm histological grades) are characterized as having higher rates of proliferation and metastasis. In order to see whether *JPX* and *XIST* were not just dysregulated in cancer patients, but differentially expressed in higher grades of ovarian cancer, we analyzed the Ovarian Serous Cystadenocarcinoma project (TCGA, PanCancer Atlas) dataset from cBioPortal (Cerami et al., 2012; Gao et al., 2013) and compared higher grades (G2 & G3) to borderline cancerous tissue (GB) to screen for expression differences (Figure 4.2A). Serous epithelial ovarian cancer is the most common, tends to be found at later neoplasm histological grades, and has a lower survival rate (Torre et al., 2018). We observed that higher grades of this cancer (G2 & G3) have significantly lower expression of *XIST* in comparison to GB (P = 0.01 for G2 vs. GB; P = 0.03 for G3 vs. GB). *JPX* expression in the higher grades also decreased in comparison to GB, though it was more variable and not statistically significant (P = 0.08 for G2 vs. GB; P = 0.05 for G3 vs. GB). *MALAT1* was included as a cancer associated lncRNA reference and showed no significant

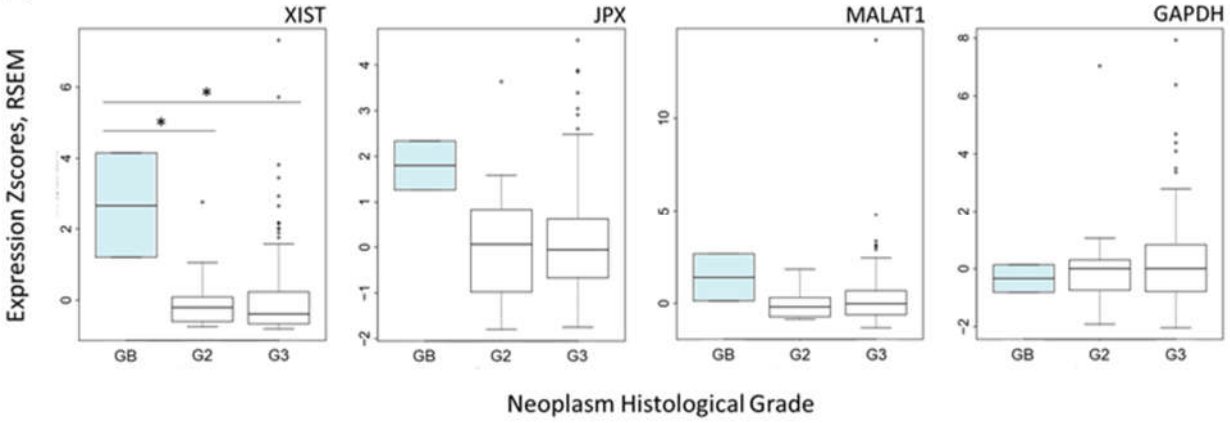
change of expression in higher grades (*P* = 0.17 for G2 vs. GB; *P* = 0.21 for G3 vs. GB), similar to our *GAPDH* reference gene (*P* = 0.81 for G2 vs. GB; *P* = 0.65 for G3 vs. GB). This data suggests downregulation of *JPX* and *XIST* in higher grades of epithelial serous ovarian cancer is XCI specific, providing support for these genes as prospective biomarkers for late stage treatment in ovarian cancer. Since *XIST* activation is necessary and essential for XCI in all female cells, its downregulation can influence expression of X-linked genes important to tumor metastasis suppression (Figure 4.2B).

Table 4.1: Cancer-associated lncRNAs from a survey of Gene Expression Omnibus (GEO) datasets

Cancer Type	Frequency	lncRNA with Modified Expression
Adenocarcinoma	4	XIST, JPX, MALAT1, MEG3
Bladder	3	PCGEM1
Breast	31	XIST, JPX, PCA3, HOTAIR, MALAT1, H19, GAS5, MEG3
Cervical	1	MEG3
Colorectal	9	JPX, PCA3, GAS5, MEG3
Glioma	6	JPX, MALAT1, GAS5, MEG3
Leukemia	4	PCA3, MALAT1, GAS5, MEG3
Lung	4	XIST, GAS5, MEG3
Ovarian	12	XIST, JPX, PCA3, HOTAIR, MALAT1, H19, GAS5, MEG3
Pancreatic	5	XIST, HOTAIR, GAS5, MEG3
Prostate	6	XIST, JPX, MALAT1, GAS5, MEG3, PRNCR1, PCGEM1
Sarcoidosis	1	PCA3

Figure 4.2

A



B

Wilcoxon-Mann-Whitney Rank Test of mRNA expression Zscores comparison

Cell Type Comparison	XIST Expression P-value	JPX Expression P-value	MALAT1 Expression P-value	GAPDH Expression P-value
GB vs G2	**0.01333**	0.08	0.1667	0.8067
GB vs G3	**0.03185**	0.05313	0.2053	0.6546
G2 vs G3	0.3939	0.9413	0.3555	0.4226

XIST vs JPX Pearson Correlation

XIST vs JPX Expression	r (Pearson Correlation)	P-value
G2 (n=23)	0.3978111	0.06012
G3 (n=160)	0.5006005	1.574e-11

Figure 4.2: Low expression of *JPX* and *XIST* correlates with higher grades of ovarian cancer

(A) Gene expression levels as Zscores from ovarian cancer patients in cBioPortal database (Ovarian Serous Cystadenocarcinoma (TCGA, PanCancer Atlas)). **(B)** Wilcoxon-Mann-Whitney Rank Tests and *XIST* vs. *JPX* Pearson correlations of patient data for indicated pairs. Analysis done in R.

Gene set enrichment analysis of ovarian cancer patients

Since later stages of cancer (higher grades) have metastatic properties, we incorporated pathway enrichment analysis (Reimand et al., 2019) for proliferation and metastasis gene pathways that are enriched in patients with low *JPX* or *XIST* expression. This helps uncover the possible pathways in which *JPX* and *XIST* may function in higher grades of ovarian cancer patients. For our preliminary analysis of pathways that are top enriched in *JPX* or *XIST* low expression patients, we used the 189 oncogenic signatures (regulatory pathways that are dysregulated in cancer) that are available through GSEA's (gene set enrichment analysis) molecular signatures database (MSigDB) (Liberzon et al., 2011, 2015; Subramanian et al., 2005). Additionally, cancer-related lncRNA *MALAT1* was also used as a reference. When comparing top enriched pathways in *XIST*, *JPX*, and *MALAT1* low expression ovarian cancer patients, we found that they all had different results (Figure 4.3).

Figure 4.3

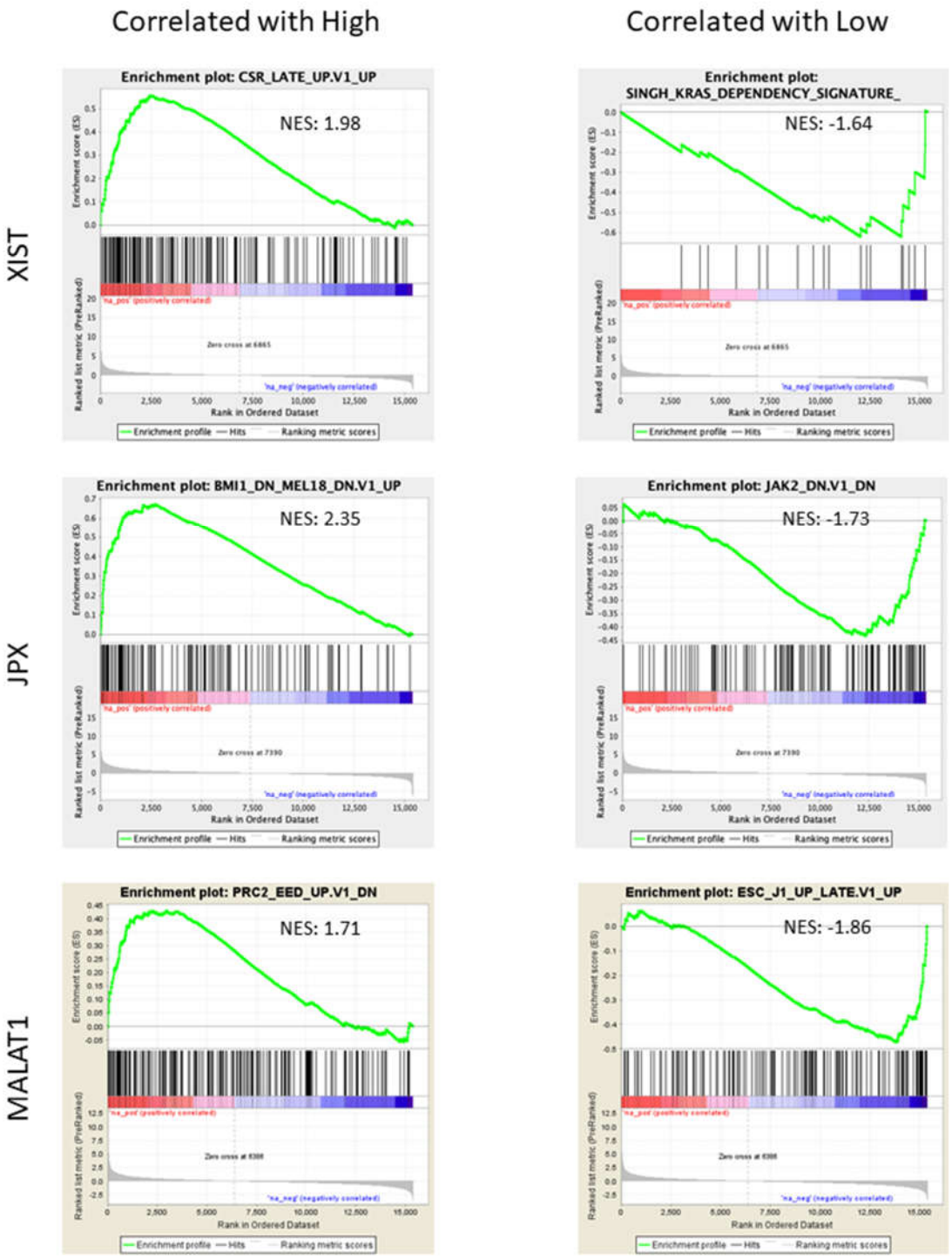

Correlated with High
Correlated with Low
XIST
JPX
MALAT1
Enrichment plot: CSR_LATE_UP.V1_UP
NES: 1.98
Enrichment plot: SINGH_KRAS_DEPENDENCY_SIGNATURE_
NES: -1.64
Enrichment plot: BMI1_DN_MEL18_DN.V1_UP
NES: 2.35
Enrichment plot: JAK2_DN.V1_DN
NES: -1.73
Enrichment plot: PRC2_EED_UP.V1_DN
NES: 1.71
Enrichment plot: ESC_J1_UP_LATE.V1_UP
NES: -1.86
Enrichment score (ES)
Ranked list metric (PreRanked)
Rank in Ordered Dataset
Enrichment profile
Hits
Ranking metric scores

Figure 4.3: Oncogenic pathways with highest NES score associated with high or low expression of *XIST*, *JPX*, or *MALAT1*

Examples of top enriched oncogenic pathways, as defined by the normalized enrichment score (NES), associated with respective gene expression from ovarian cancer patients. Pathways appear to be different for *JPX* and *XIST* high and low expression patients. *MALAT1* is a cancer lncRNA reference not associated with the X chromosome. Pathway enrichment analysis (Reimand et al., 2019) was done for proliferation and metastasis gene pathways that are enriched in patients with high or low *XIST*, *JPX*, or *MALAT1* expression.

Irregular expression pattern of *JPX* and diminished expression pattern of *XIST* in ovarian cancer cell lines

XIST is actively expressed in all female somatic cells. In order to test whether *XIST* can be visualized together with *JPX* and detect their expression changes in female cancer cells, we used ovarian cancer cell lines, SKOV3iP1 and OVCAR3. Figure 4.4 shows RNA fluorescence *in situ* hybridization (FISH) detecting *JPX* and *XIST* expression in several different human cell lines: female human epithelial kidney cell line, HEK293T (female cell line reference); male osteosarcoma cell line, SJSA-1 (male cell line control); and female epithelial ovarian adenocarcinoma cell lines, SKOV3iP1 and OVCAR3. *XIST* RNA (green) is typically visualized as a ‘cloud’ associated with the silent X chromosome in female cells and *JPX* RNA (red) is shown as a pinpoint, sometimes co-localizing with *XIST*. FISH shows that *JPX* RNA maintains a robust, yet inconsistent, expression in SKOV3iP1 and OVCAR3, with *XIST* RNA being obviously diminished in

comparison to HEK293T cells. This pattern can also be seen in the *JPX* and *XIST* expression levels measured with qRT-PCR (Figure 4.5).

Decrease of *JPX* and *XIST* expression linked to metastasis-associated gene dysregulation

A list of X-linked cancer associated genes was compiled (Table 4.2) (Dunford et al., 2017; Liu et al., 2012; Spatz et al., 2004; Xing et al., 2018) in order to cross-reference with the list of cancer pathways enriched in the GSEA of *JPX* and *XIST* low expression patients. Three genes were identified (*ATRX*, *KDM6A*, and *PIM2*) and investigated through qRT-PCR for their expression in ovarian cancer cell lines. Genes *MID1*, *MPP1*, and *MSN* from Xing *et al*. (2018) were also included since they were associated with *XIST* low breast cancer patients. *MSN* and *ATRX* were the only genes to show interesting expression patterns (Figure 4.5 and 4.6). RNA from an immortalized ovarian cell line was added to the qRT-PCR analysis as a control. However, it was discovered that the cell line from which the RNA came from had lost *JPX* and *XIST* expression. The expression levels for these two lncRNAs were below or comparable to that of the male cancer cell line control, which, as expected from a male cell line, did not express *XIST* and has *JPX* expression levels close to that of the ovarian cancer cell lines. Figure 3.3 also shows that ovaries have one of the highest levels of both *JPX* and *XIST* in comparison to other tissues, supporting that the immortalized cell line did indeed lose expression of these two lncRNAs. Therefore, this has inadvertently confirmed the inverse relationship between decrease in *JPX* and *XIST* expression and the expression of *MSN*, as was already seen when comparing HEK293T and the

ovarian cancer cell lines (Figure 4.5). Another potential pattern was seen in *ATRX* expression. *ATRX* is considered an escaping gene, in that it is expressed from the X chromosome and in turn escapes from XCI. Interestingly though, there appears to be a decrease in *ATRX* expression in the ovarian cancer cell lines, and a near loss of expression in the immortalized cell line. However, *KDM6A*, another escaping gene, did not show any distinct pattern.

Figure 4.4

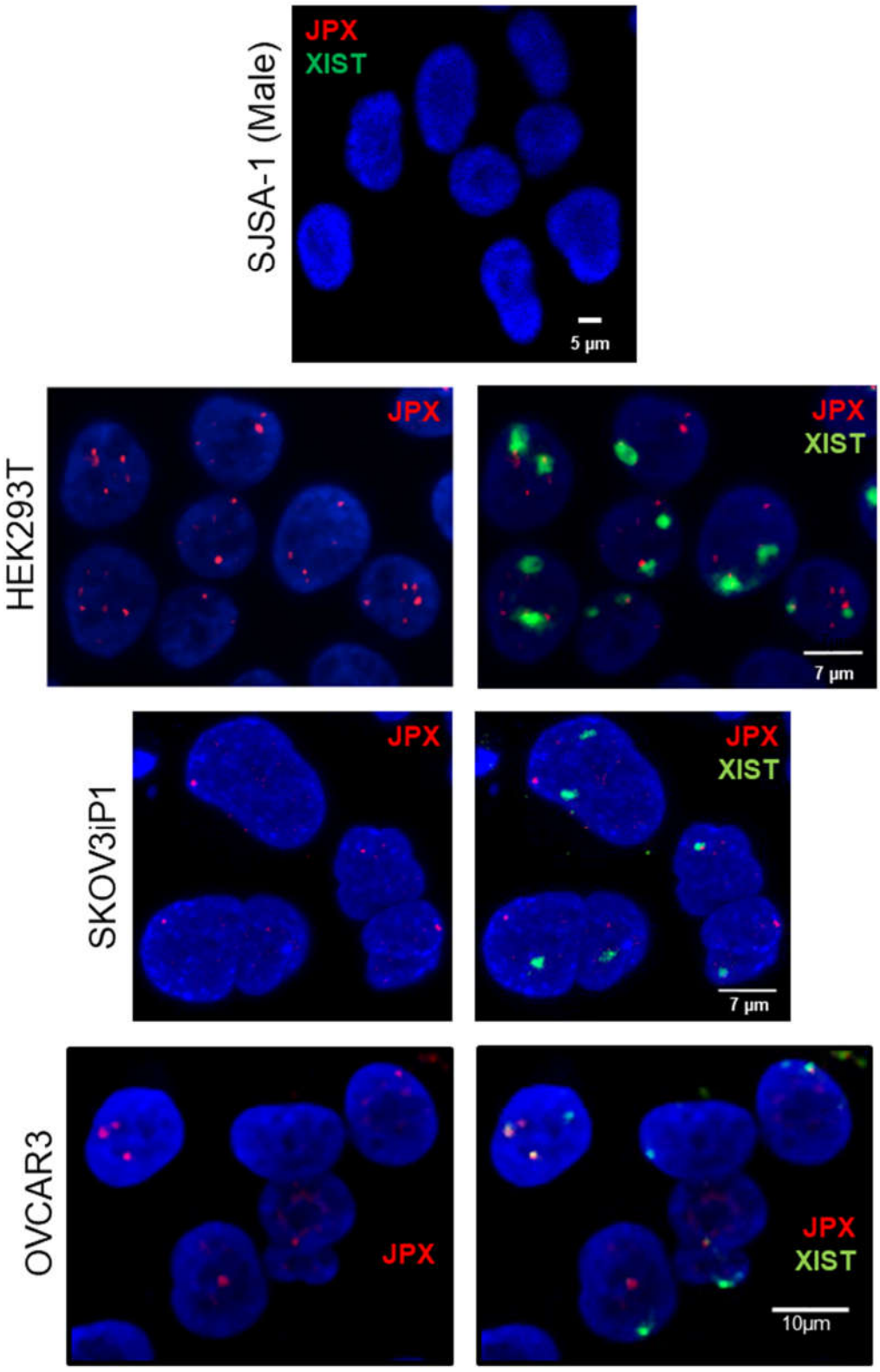

SJSA-1 (Male)
JPX
XIST
5 µm
HEK293T
JPX
JPX
XIST
7 µm
SKOV3iP1
JPX
JPX
XIST
7 µm
OVCAR3
JPX
JPX
XIST
10µm

Figure 4.4: Abnormal *JPX* and *XIST* RNA FISH signal patterns

XIST RNA is visualized as a green cloud and *JPX* RNA as a red pinpoint. SJSA-1 is a human male osteosarcoma control, representing cells that do not undergo XCI. HEK293T is a female human embryonic kidney cell line that represents the normal size of *XIST* clouds and *JPX* pinpoints. SKOV3iP1 and OVCAR3 are two ovarian cancer cell lines that have smaller *XIST* clouds and abnormal *JPX* pinpoint patterns.

Figure 4.5

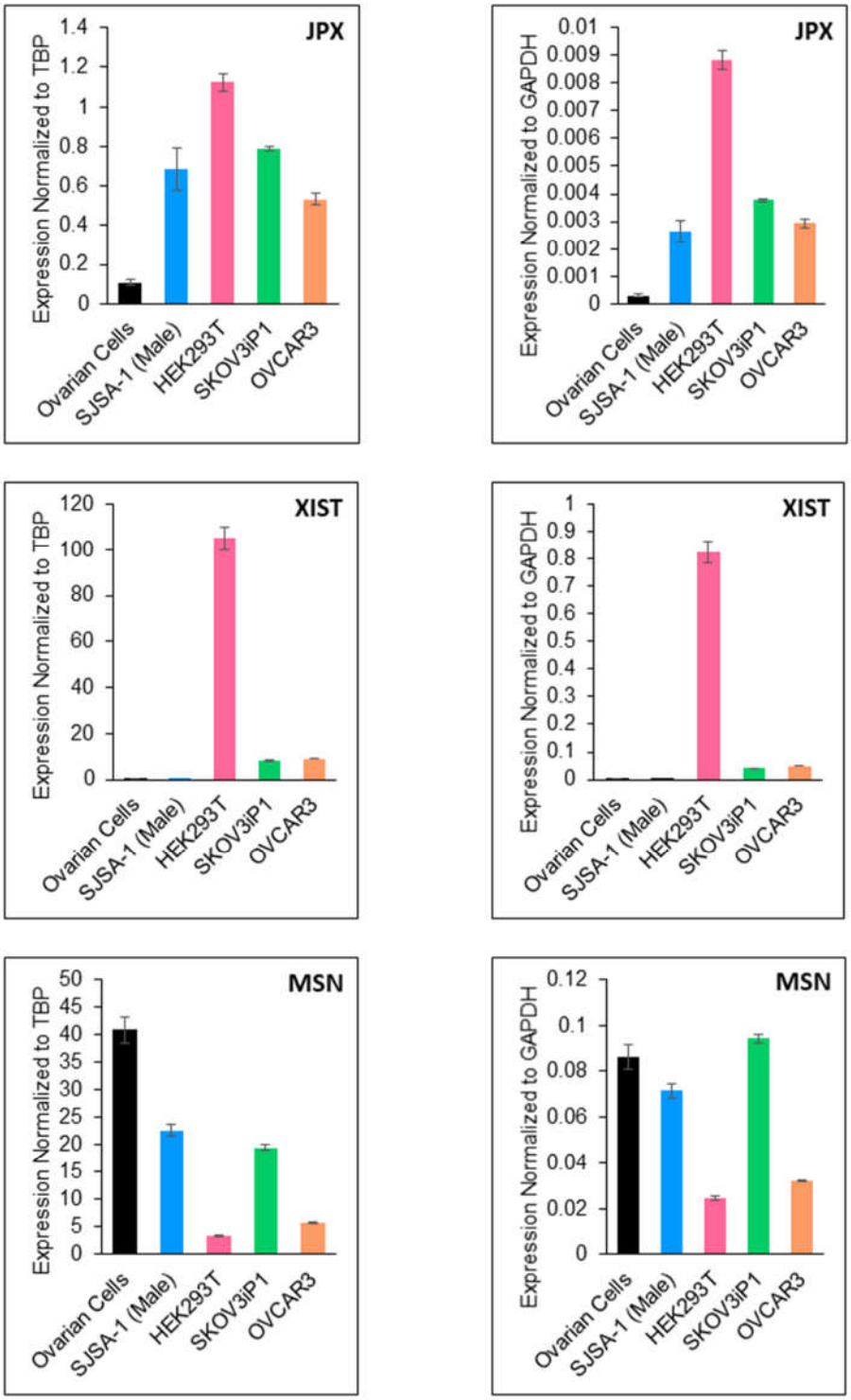

JPX
Expression Normalized to TBP
1.4
1.2
1
0.8
0.6
0.4
0.2
0
Ovarian Cells
SJSA-1 (Male)
HEK293T
SKOV3iP1
OVCAR3
JPX
Expression Normalized to GAPDH
0.01
0.009
0.008
0.007
0.006
0.005
0.004
0.003
0.002
0.001
0
Ovarian Cells
SJSA-1 (Male)
HEK293T
SKOV3iP1
OVCAR3
XIST
Expression Normalized to TBP
120
100
80
60
40
20
0
Ovarian Cells
SJSA-1 (Male)
HEK293T
SKOV3iP1
OVCAR3
XIST
Expression Normalized to GAPDH
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
Ovarian Cells
SJSA-1 (Male)
HEK293T
SKOV3iP1
OVCAR3
MSN
Expression Normalized to TBP
50
45
40
35
30
25
20
15
10
5
0
Ovarian Cells
SJSA-1 (Male)
HEK293T
SKOV3iP1
OVCAR3
MSN
Expression Normalized to GAPDH
0.12
0.1
0.08
0.06
0.04
0.02
0
Ovarian Cells
SJSA-1 (Male)
HEK293T
SKOV3iP1
OVCAR3

Figure 4.5: *MSN* is inversely expressed in comparison to *JPX* and *XIST* expression

qRT-PCR of *JPX*, *XIST*, and *MSN* expression normalized to either *TBP* (TATA binding protein) or *GAPDH* RNA. Ovarian cells represent RNA (purchased from abm) that came from an immortalized cell line which appears to have lost *JPX* and *XIST* expression. Remaining cell lines are the same as from Figure 4.4. Average expression ± SEM is shown.

Table 4.2: Cancer-associated X-linked genes

AR (Spatz 2004)	ELK1 (Spatz 2004)	**MID1 (Xing 2018)**
ARAF1 (Spatz 2004)	FAM123B (Liu 2012)	**MPP1 (Xing 2018)**
ATRX (Liu 2012, Dunford 2017)	FGF13 (Spatz 2004)	**MSN (Spatz 2004, Xing 2018)**
BMP15 (Spatz 2004)	FOXP3 (Liu 2012)	MST4 (Spatz 2004)
BMX (Spatz 2004)	GPC3 (Spatz 2004)	PHF6 (Liu 2012)
BZX (Spatz 2004)	GRPR (Spatz (2004)	**PIM2 (Spatz 2004)**
CD99 (Spatz 2004, Liu 2012)	HDGF (Spatz 2004)	PIR (Spatz 2004)
CNKSR2 (Dunford 2017)	HPCX (Spatz 2004)	RBBP7 (Liu 2012, Spatz 2004)
DDX3X (Dunford 2017)	KDM5C (Dunford 2017)	RPL10 (Spatz 2004, Liu 2012)
DKC1 (Liu 2012)	**KDM6A (Dunford 2017)**	RPS6KA6 (Liu 2012)
EDA2R (Liu 2012)	LDOC1 (Liu 2012)	TGCT1 (Spatz 2004)
EFNB1 (Spatz 2004)	MAGEC3 (Dunford 2017)	TIMP1 (Spatz 2004, Liu 2012)
ELF4 (Liu 2012)	MECP2 (Spatz 2004)	UBE1 (Spatz 2004)

Figure 4.6

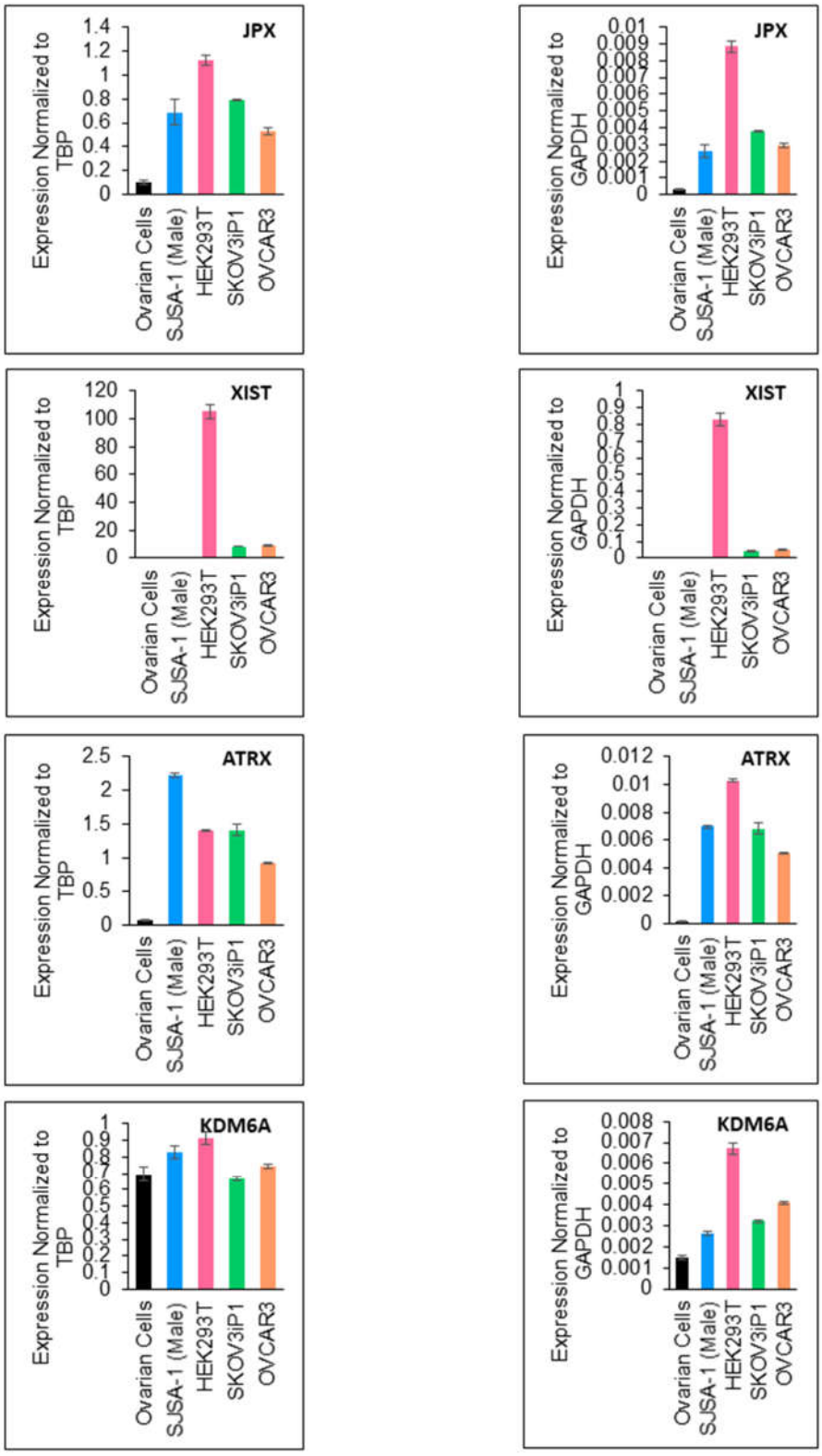

JPX
Expression Normalized to TBP
1.4
1.2
1
0.8
0.6
0.4
0.2
0
Ovarian Cells
SJSA-1 (Male)
HEK293T
SKOV3iP1
OVCAR3
JPX
Expression Normalized to GAPDH
0.01
0.009
0.008
0.007
0.006
0.005
0.004
0.003
0.002
0.001
0
Ovarian Cells
SJSA-1 (Male)
HEK293T
SKOV3iP1
OVCAR3
XIST
Expression Normalized to TBP
120
100
80
60
40
20
0
Ovarian Cells
SJSA-1 (Male)
HEK293T
SKOV3iP1
OVCAR3
XIST
Expression Normalized to GAPDH
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
Ovarian Cells
SJSA-1 (Male)
HEK293T
SKOV3iP1
OVCAR3
ATRX
Expression Normalized to TBP
2.5
2
1.5
1
0.5
0
Ovarian Cells
SJSA-1 (Male)
HEK293T
SKOV3iP1
OVCAR3
ATRX
Expression Normalized to GAPDH
0.012
0.01
0.008
0.006
0.004
0.002
0
Ovarian Cells
SJSA-1 (Male)
HEK293T
SKOV3iP1
OVCAR3
KDM6A
Expression Normalized to TBP
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
Ovarian Cells
SJSA-1 (Male)
HEK293T
SKOV3iP1
OVCAR3
KDM6A
Expression Normalized to GAPDH
0.008
0.007
0.006
0.005
0.004
0.003
0.002
0.001
0
Ovarian Cells
SJSA-1 (Male)
HEK293T
SKOV3iP1
OVCAR3

Figure 4.6: *ATRX* may be affected by decrease in *JPX* and *XIST* expression

qRT-PCR of *JPX*, *XIST*, *ATRX*, and *KDM6A* expression normalized to either *TBP* (TATA binding protein) or *GAPDH* RNA. Cell lines are the same as those in Figure 4.5. *ATRX* and *KDM6A* are cancer-associated genes that under normal conditions escape XCI. *ATRX* appears to have a decreased expression in relation to decreased *JPX* and *XIST* expression. *KDM6A* does not appear to be affected by the decrease in *JPX* and *XIST* expression. Average expression ± SEM is shown.

XIST knockdown increases proliferation and accelerates migration

Interestingly, the RNASeq data for 68 ovarian cancer cell lines in the Cancer Cell Line Encyclopedia (CCLE) database indicates that human epithelial ovarian adenocarcinoma cell lines, SKOV3iP1 and OVCAR3, are amongst the highest *JPX* and *XIST* expressing cell lines. SKOV3iP1 and OVCAR3 were both derived from epithelial tissues, with SKOV3iP1 being characterized as non-serous carcinoma and OVCAR3 as a serous carcinoma (Hallas-Potts et al., 2019). As mentioned previously, epithelial cancers are one of the most common forms of ovarian cancer, with serous accounting for over half of cases reported. Epithelial ovarian cancers also tend to be found at later and more aggressive stages (Torre et al., 2018).

Since *JPX* and *XIST* expression is downregulated in patients with higher grades of ovarian cancer, I wanted to test whether they are behaving as tumor suppressors. So far, we successfully knocked down *XIST* alone and were able to get two OVCAR3 knockdown clones (X1-OV and X2-OV). Expression of *ATRX*, *KDM6A,* and *MSN* were checked with qRT-PCR. Only *KDM6A* was downregulated in both clones and *MSN* was

downregulated in only one clone. *ATRX* did not show any significant difference in either clone (Figure 4.7A).

Knockdown of *XIST* caused an increase in proliferation and migration. At all time points, proliferation was shown to be significantly higher than that of the GFP-scramble control (Figure 4.7B). After 24 hours, there was a significant increase in migration of knockdown cells as shown by the wound healing scratch assay. However, the 48-hour time point results were too variable, and earlier time points may need to be taken in the future to account for these rapidly growing clones (Figure 4.7C-D).

Figure 4.7

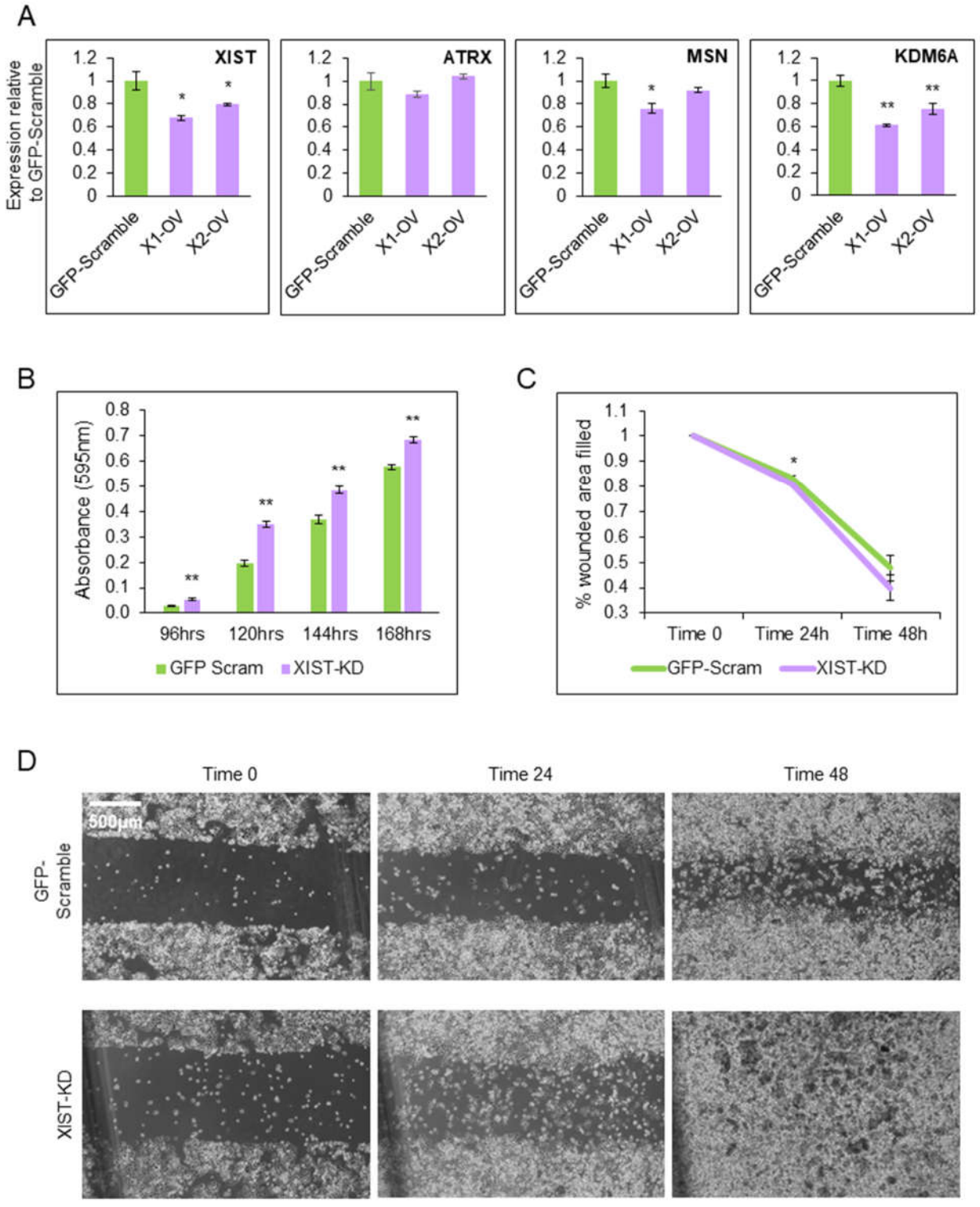

A
Expression relative to GFP-Scramble
XIST
ATRX
MSN
KDM6A
GFP-Scramble
X1-OV
X2-OV
B
Absorbance (595nm)
96hrs
120hrs
144hrs
168hrs
GFP Scram
XIST-KD
C
% wounded area filled
Time 0
Time 24h
Time 48h
GFP-Scram
XIST-KD
D
Time 0
Time 24
Time 48
500µm
GFP-Scramble
XIST-KD

Figure 4.7: *XIST* knockdown in high-expressing ovarian cancer cell line increases proliferation and accelerates migration

(A) qRT-PCR of *XIST*, *MSN*, and *KDM6A* expression normalized to *GAPDH* RNA. *MSN* and *KDM6A* appear to have a decreased expression in relation to *XIST* knockdown. X1-OV and X2-OV are OVCAR3 *XIST* knockdown clones. Average expression relative to GFP-scramble control ± SEM is shown. One-tailed Student *t*-test, *P<0.05 and ** P<0.01. **(B)** Increase in proliferation for knockdown clones measured by crystal violet staining and measuring absorbance. XIST-KD is the average of the *XIST* knockdown clones. GFP-scram is the average of the GFP-scramble controls. Cells were plated at equal density and allowed to grow to a higher density before staining. Measurements were taken for the 96hr, 120hr, 144hr, and 168hr time points. Average proliferation ± SEM is shown. One-tailed Student *t*-test, ** P<0.01. **(C)** Accelerated migration for knockdown clones was shown through a wound healing scratch assay. XIST-KD is the average of the *XIST* knockdown clones. GFP-scram is the average of the GFP-scramble controls. Cells were plated at equal density and scratched with a p1000 pipet tip when they reached ~80% confluency. Measurements were taken at 0hr, 24hr, and 48hr time points after scratching. Average of two clones (3 replicates each) ± SEM is shown. **(D)** Representative brightfield scratch test images. XIST-KD is the average of the *XIST* knockdown clones.

4.4 Discussion

JPX and *XIST* have previously been proposed as prognostic markers for hepatocellular carcinoma (Ma et al., 2017), and are dysregulated in ovarian and breast

cancer (Table 4.1) (Prensner and Chinnaiyan, 2011; Xing et al., 2018). Our study has confirmed a significant downregulation of *XIST* in patients with higher grades of ovarian cancer, which is consistent with it being a potential prognostic marker, *JPX* was reduced in these patients as well. This pattern of downregulation could also be seen in ovarian cancer cell lines, SKOV3iP1 and OVCAR3, and this downregulation appeared to have an inverse relationship with *MSN* expression—a gene that is associated with breast cancer metastasis (Xing et al., 2018). This data suggests that *JPX* and *XIST* could be behaving as tumor suppressors in the cell, and their loss can cause an increase in the cancer's severity.

Cancer is known to affect numerous biological pathways in the cell, with a single gene and its product (e.g., *P53*) having the potential to be involved with several pathways (Beckerman and Prives, 2010; Lin et al., 2019). We therefore performed a GSEA analysis of *JPX* and *XIST* low expression patients using the oncogenic signatures (c6) in the MSigDB in order to see what pathways could be associated with these patients. This information also provides a list of genes that may contribute to the increase in ovarian cancer aggressiveness correlated with these patients. Comparing the gene lists for the pathways from the GSEA to known X-linked cancer associated genes revealed two genes (*ATRX* and *KDM6A*) from the *XIST* low analysis and one gene (*PIM2*) from the *JPX* low analysis. These three genes, along with the three that have been previously associated with *XIST* low cancer patients (*MID1*, *MPP1*, and *MSN*) (Xing et al., 2018), were evaluated in ovarian cancer cell lines, SKOV3iP1 and OVCAR3. The two genes that had the most notable differences in these cell lines, and

in the immortalized ovarian cell line that had lost *JPX* and *XIST* expression, were *MSN* and *ATRX*.

MSN, also known as moesin, is an ERM (ezrin, radixin, and moesin) family protein that is a key player in the c-MET pathway. Normally, this pathway controls biological processes related to motility, proliferation, survival, and migration. However, in cancer cells these processes can go awry and lead to increased tumor growth and metastasis in patients (Organ and Tsao, 2011; Orian-Rousseau et al., 2002; Trusolino et al., 2010). Our ovarian cancer cell lines had low *JPX* and *XIST* expression levels, which coincided with the increase of *MSN* expression in these cell lines. There was also a drastic increase of *MSN* expression in the immortalized ovarian cell line, confirming this inverse relationship. This gene has been associated with increased metastasis to the brain in *XIST* low breast cancer patients (Xing et al., 2018). Therefore, it is not surprising to see that it may be involved in *XIST* low ovarian cancer severity. In contrast, when we knocked down *XIST* alone in OVCAR3 cells, there was a significant decrease in *MSN* expression in one of the clones. If *XIST* is directly related to *MSN* in ovarian cancer, then one possible reason for these results could be that the loss of both *JPX* and *XIST* is required for the increase in *MSN* expression. Another possibility is that *JPX* alone plays a role in the suppression of *MSN*, and knockdown of *JPX* is what led to the decrease in the immortalized cell line. A knockdown of *JPX* alone, as well as both *JPX* and *XIST*, in conjunction with rescue experiments, will be required to determine whether there is a direct connection between *JPX* and/or *XIST* and *MSN*.

ATRX is a gene that escapes from XCI and is a known tumor suppressor. This gene, however, decreased in expression within the ovarian cancer cell lines and the

immortalized cell line while another tumor suppressor escaping gene, *KDM6A*, maintained expression. ATRX binds *XIST* lncRNA to recruit PRC2 (polycomb repressive complex 2) to the X chromosome to facilitate gene silencing (Oppel et al., 2019; Sarma et al., 2014). The decrease of *ATRX* expression could potentially be related to the abnormally small *XIST* clouds seen in the ovarian cancer cell lines, suggesting a possible feedback loop that helps to maintain XCI. In contrast, *ATRX* maintained its expression and *KDM6A* decreased in expression when *XIST* alone was knocked down in OVCAR3 cells. Once again, more research will be needed to determine whether this is due to only *XIST* being knocked down, what a more severe knockdown would cause, and whether there is a direct connection between these lncRNAs and the cancer-associated XCI escaping genes.

Knockdown of *XIST* did cause a significant increase in proliferation and showed accelerated migration, despite the low severity of the knockdown. This is indicative of *XIST* playing some role in the progression of ovarian cancer. The pathways that it affects will still need to be investigated further through the use of sequencing techniques in the knockdown cell lines to better determine what other X-linked genes are being affected by the loss of *XIST*.

Overall, our data suggest that the lncRNAs *JPX* and *XIST* may be involved in different oncogenic pathways, with some possible overlap. Both are dysregulated in ovarian cancer patients and are possibly acting as tumor suppressors. The loss of *JPX* and *XIST* in ovarian cancer patients is potentially leading to an increase in the severity of cancer in these patients. Therefore, it is imperative that we continue to study these lncRNAs in the context of cancer and their use as prognostic and treatment markers.

4.5 References

Abramovitz, M., Williams, C., De, P.K., Dey, N., Willis, S., Young, B., Andreopoulou, E., Symmans, W.F., Sicklick, J.K., Kurzrock, R., et al. (2019). Personalized Cancer Treatment and Patient Stratification Using Massive Parallel Sequencing (MPS) and Other OMICs Data. In Predictive Biomarkers in Oncology, S. Badve, and G.L. Kumar, eds. (Cham: Springer International Publishing), pp. 131–147.

Beckerman, R., and Prives, C. (2010). Transcriptional Regulation by P53. Cold Spring Harb. Perspect. Biol. *2*, a000935.

Carmona, S., Lin, B., Chou, T., Arroyo, K., and Sun, S. (2018). LncRNA Jpx induces Xist expression in mice using both trans and cis mechanisms. PLOS Genet. *14*, e1007378.

Cerami, E., Gao, J., Dogrusoz, U., Gross, B.E., Sumer, S.O., Aksoy, B.A., Jacobsen, A., Byrne, C.J., Heuer, M.L., Larsson, E., et al. (2012). The cBio Cancer Genomics Portal: An Open Platform for Exploring Multidimensional Cancer Genomics Data. Cancer Discov. *2*, 401–404.

Cheetham, S.W., Gruhl, F., Mattick, J.S., and Dinger, M.E. (2013). Long noncoding RNAs and the genetics of cancer. Br. J. Cancer *108*, 2419–2425.

Crawford, E.D., Rove, K.O., Trabulsi, E.J., Qian, J., Drewnowska, K.P., Kaminetsky, J.C., Huisman, T.K., Bilowus, M.L., Freedman, S.J., Glover, W.L., et al. (2012). Diagnostic Performance of PCA3 to Detect Prostate Cancer in Men with Increased Prostate Specific Antigen: A Prospective Study of 1,962 Cases. J. Urol. *188*, 1726–1731.

Dunford, A., Weinstock, D.M., Savova, V., Schumacher, S.E., Cleary, J.P., Yoda, A., Sullivan, T.J., Hess, J.M., Gimelbrant, A.A., Beroukhim, R., et al. (2017). Tumor-suppressor genes that escape from X-inactivation contribute to cancer sex bias. Nat. Genet. *49*, 10–16.

Gao, J., Aksoy, B.A., Dogrusoz, U., Dresdner, G., Gross, B., Sumer, S.O., Sun, Y., Jacobsen, A., Sinha, R., Larsson, E., et al. (2013). Integrative Analysis of Complex Cancer Genomics and Clinical Profiles Using the cBioPortal. Sci. Signal. *6*, pl1.

Gutschner, T., and Diederichs, S. (2012). The hallmarks of cancer: a long non-coding RNA point of view. RNA Biol. *9*, 703–719.

Hallas-Potts, A., Dawson, J.C., and Herrington, C.S. (2019). Ovarian cancer cell lines derived from non-serous carcinomas migrate and invade more aggressively than those derived from high-grade serous carcinomas. Sci. Rep. *9*, 5515.

Huarte, M. (2015). The emerging role of lncRNAs in cancer. Nat. Med. *21*, 1253–1261.

Kapranov, P., St Laurent, G., Raz, T., Ozsolak, F., Reynolds, C.P., Sorensen, P.H.B., Reaman, G., Milos, P., Arceci, R.J., Thompson, J.F., et al. (2010). The majority of total nuclear-encoded non-ribosomal RNA in a human cell is “dark matter” un-annotated RNA. BMC Biol. *8*, 149.

Kawakami, T., Okamoto, K., Ogawa, O., and Okada, Y. (2004). XIST unmethylated DNA fragments in male-derived plasma as a tumour marker for testicular cancer. Lancet *363*, 40–42.

Kim, S., Han, Y., Kim, S.I., Kim, H.-S., Kim, S.J., and Song, Y.S. (2018). Tumor evolution and chemoresistance in ovarian cancer. Npj Precis. Oncol. *2*, 20.

Lee, G.L., Dobi, A., and Srivastava, S. (2011). Diagnostic performance of the PCA3 urine test. Nat. Rev. Urol. *8*, 123–124.

Liberzon, A., Subramanian, A., Pinchback, R., Thorvaldsdottir, H., Tamayo, P., and Mesirov, J.P. (2011). Molecular signatures database (MSigDB) 3.0. Bioinformatics *27*, 1739–1740.

Liberzon, A., Birger, C., Thorvaldsdóttir, H., Ghandi, M., Mesirov, J.P., and Tamayo, P. (2015). The Molecular Signatures Database Hallmark Gene Set Collection. Cell Syst. *1*, 417–425.

Lin, T., Hou, P.-F., Meng, S., Chen, F., Jiang, T., Li, M.-L., Shi, M.-L., Liu, J.-J., Zheng, J.-N., and Bai, J. (2019). Emerging Roles of p53 Related lncRNAs in Cancer Progression: A Systematic Review. Int. J. Biol. Sci. *15*, 1287–1298.

Liu, R., Kain, M., and Wang, L. (2012). Inactivation of X-linked tumor suppressor genes in human cancer. Futur. Oncol. *8*, 463–481.

Ma, W., Wang, H., Jing, W., Zhou, F., Chang, L., Hong, Z., Liu, H., Liu, Z., and Yuan, Y. (2017). Downregulation of long non-coding RNAs JPX and XIST is associated with the prognosis of hepatocellular carcinoma. Clin. Res. Hepatol. Gastroenterol. *41*, 163–170.

Oppel, F., Tao, T., Shi, H., Ross, K.N., Zimmerman, M.W., He, S., Tong, G., Aster, J.C., and Look, A.T. (2019). Loss of atrx cooperates with p53-deficiency to promote the development of sarcomas and other malignancies. PLOS Genet. *15*, e1008039.

Organ, S.L., and Tsao, M.-S. (2011). An overview of the c-MET signaling pathway. Ther. Adv. Med. Oncol. *3*, S7–S19.

Orian-Rousseau, V., Chen, L., Sleeman, J.P., Herrlich, P., and Ponta, H. (2002). CD44 is required for two consecutive steps in HGF/c-Met signaling. Genes Dev. *16*, 3074–3086.

Perez, D.S., Hoage, T.R., Pritchett, J.R., Ducharme-Smith, A.L., Halling, M.L., Ganapathiraju, S.C., Streng, P.S., and Smith, D.I. (2008). Long, abundantly expressed non-coding transcripts are altered in cancer. Hum. Mol. Genet. *17*, 642–655.

Prensner, J.R., and Chinnaiyan, A.M. (2011). The Emergence of lncRNAs in Cancer Biology. Cancer Discov. *1*, 391–407.

Reimand, J., Isserlin, R., Voisin, V., Kucera, M., Tannus-Lopes, C., Rostamianfar, A., Wadi, L., Meyer, M., Wong, J., Xu, C., et al. (2019). Pathway enrichment analysis and visualization of omics data using g:Profiler, GSEA, Cytoscape and EnrichmentMap. Nat. Protoc. *14*, 482–517.

Sarma, K., Cifuentes-Rojas, C., Ergun, A., del Rosario, A., Jeon, Y., White, F., Sadreyev, R., and Lee, J.T. (2014). ATRX Directs Binding of PRC2 to Xist RNA and Polycomb Targets. Cell *159*, 869–883.

Schmitt, A.M., and Chang, H.Y. (2016). Long Noncoding RNAs in Cancer Pathways. Cancer Cell *29*, 452–463.

Spatz, A., Borg, C., and Feunteun, J. (2004). X-Chromosome Genetics and Human Cancer. Nat. Rev. Cancer *4*, 617–629.

Subramanian, A., Tamayo, P., Mootha, V.K., Mukherjee, S., Ebert, B.L., Gillette, M.A., Paulovich, A., Pomeroy, S.L., Golub, T.R., Lander, E.S., et al. (2005). Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. Proc. Natl. Acad. Sci. *102*, 15545–15550.

Sun, S., Del Rosario, B.C., Szanto, A., Ogawa, Y., Jeon, Y., and Lee, J.T. (2013). Jpx RNA Activates Xist by Evicting CTCF. Cell *153*, 1537–1551.

Tian, D., Sun, S., and Lee, J.T. (2010). The Long Noncoding RNA, Jpx, Is a Molecular Switch for X Chromosome Inactivation. Cell *143*, 390–403.

Torre, L.A., Trabert, B., DeSantis, C.E., Miller, K.D., Samimi, G., Runowicz, C.D., Gaudet, M.M., Jemal, A., and Siegel, R.L. (2018). Ovarian cancer statistics, 2018. CA. Cancer J. Clin. *68*, 284–296.

Trusolino, L., Bertotti, A., and Comoglio, P.M. (2010). MET signalling: principles and functions in development, organ regeneration and cancer. Nat. Rev. Mol. Cell Biol. *11*, 834–848.

Xing, F., Liu, Y., Wu, S.-Y., Wu, K., Sharma, S., Mo, Y.-Y., Feng, J., Sanders, S., Jin, G., Singh, R., et al. (2018). Loss of XIST in Breast Cancer Activates MSN-c-Met and Reprograms Microglia via Exosomal miRNA to Promote Brain Metastasis. Cancer Res. *78*, 4316–4330.

Yildirim, E., Kirby, J.E., Brown, D.E., Mercier, F.E., Sadreyev, R.I., Scadden, D.T., and Lee, J.T. (2013). Xist RNA Is a Potent Suppressor of Hematologic Cancer in Mice. Cell *152*, 727–742.

CHAPTER 5: Future Directions and Conclusions

5.1 Future Directions

Follow-up to functional conservation of lncRNA *JPX* despite sequence and structural divergence

Though my research on lncRNA *JPX* function and homology to mouse lncRNA *Jpx* resulted in a publication, there are still questions left to be answered. Below, I describe several studies that can be done to further investigate human lncRNA *JPX* structure *in vivo*, and expand the knowledge of its function and connection to other proteins in the cell.

SHAPE-seq and resolving Jpx/JPX *structure* in vivo

A major roadblock to studying lncRNA *JPX* is that it is considered a low abundance transcript, making its *in vivo* structure difficult to resolve without amplification techniques. Even though there are examples of lncRNA structure resolution *in vivo*, the lncRNAs studied tend to be naturally abundant within cells (Kwok et al., 2013; Lucks et al., 2011; Spitale et al., 2013). As mentioned in Chapter 3, I was able to manually resolve a portion of the most stable region of human *JPX* lncRNA structure *in vivo*, but was unable to resolve more of the structure of either the mouse or human *Jpx/JPX* lncRNAs. *In vivo* structural analysis of lncRNA is notoriously difficult due to low abundance (Kubota et al., 2015; Kwok et al., 2013; Xue et al., 2008), and techniques to improve the resolution of *in vivo* SHAPE (Kwok et al., 2013; Lucks et al., 2011) were still unable to aid in fully resolving the *Jpx/JPX* lncRNA structures.

However, there have been improvements in the use of next generation sequencing (NGS) techniques in conjunction with SHAPE RNA probing that aid in the resolution of secondary and tertiary structures of RNA (Uroda et al., 2019; Watters et al., 2016). Watters *et al*. describes in detail how to perform *in vivo* SHAPE-seq, and incorporates methods that allow for the amplification of target RNA and improve sequencing library quality (Watters et al., 2016). More recently though, Uroda *et al*. utilized SHAPE-seq techniques to uncover the functional secondary and tertiary structures of human lncRNA maternally expressed gene 3 (*MEG3*) in a cell line that had negligible expression of *MEG3* (Uroda et al., 2019).

Therefore, it would be an interesting next step to explore whether the latest technologies for SHAPE-seq RNA probing can be used to resolve the *in vivo* secondary structure of lncRNA *JPX*. It is my hope that secondary and tertiary structural probing technologies advance even further, and methods will be developed that are truly effective for low expression lncRNAs such as *JPX*.

Mutated CTCF and binding to lncRNA JPX

Since human *JPX* demonstrated robust binding capabilities in spite of truncating its transcript, it would be interesting to see whether mutation of CTCF instead could perturb binding. Saldaña-Meyer *et al*. demonstrated that CTCF contains not only a DNA binding region, but an RNA binding region as well. In their study of CTCF and its interactions with RNA, they were able to demonstrate that mutating the RNA binding region disrupts binding of CTCF to RNA and destabilizes CTCF binding of chromatin (Saldaña-Meyer et al., 2014). In a more recent study by the same group, CTCF was

implicated as a possible transcription factor. However, they also admit that more research and technological breakthroughs will be needed to fully understand CTCF and RNA interactions, and their importance in humans (Saldaña-Meyer et al., 2019). Therefore, a collaboration to use these CTCF mutants and mutant cell lines would allow us to test if they are still capable of binding lncRNA *JPX*, and whether XCI is perturbed in anyway.

Uncovering other proteins capable of binding lncRNA JPX

A recent bioRxiv publication by Rosspopoff *et al*. studied lncRNA *JPX* in a human primed stem cell that had been reset to a naïve pluripotent stem cell. This system, however, can still suffer from XCI degradation. Their data suggests that *JPX* lncRNA in humans is nonfunctional, and only the act of transcription from the gene is necessary (Rosspopoff et al., 2019). Though we agree that the *JPX* gene is functional in humans, it is too soon to say that the *JPX* transcript is nonfunctional. If the act of transcription alone was what is solely necessary for activation of *XIST* expression, then this still leaves the question of why does *JPX* expression in male cells not trigger *XIST* expression as well? Furthermore, if *JPX* is dispensable in the context of human XCI, a redundant pathway could potentially step in to rescue activation of *XIST* expression. Importantly, the *JPX* transcript could also be connected to other pathways in different tissues. In Chapter 3, I showed that human *JPX* is capable of complementing the loss of mouse *Jpx*, demonstrating that it has retained its function. However, human *JPX* has rapidly evolved in comparison to its mouse homolog and the lncRNA function has likely undergone some change.

In order to understand and find what other pathways human *JPX* may be connected to, further study on whether there are proteins other than CTCF which are capable of binding lncRNA *JPX in vivo*. Methods such as those used by Chu *et al*. and McHugh *et al*. allow for RNA-directed pull down of proteins that interact with target RNA (Chu et al., 2015; McHugh et al., 2015). Both of these groups took advantage of proven crosslinking techniques in conjunction with target RNA specific biotinylated probes to pull down RNA and protein complexes. This is then followed up with mass spectrometry to identify the interacting proteins. Interestingly, RAP-MS (RNA Antisense Purification Mass Spectrometry) used by McHugh *et al*. incorporated a technique known as SILAC (stable isotope labelling by amino acids in culture) to facilitate quantitative comparisons of protein interactions, measuring which proteins interact with the target RNA more frequently (McHugh et al., 2015). These techniques have been tested on noncoding RNA that is more abundant in the nucleus and it would be intriguing to see if they also work for less abundant lncRNA.

Continuation of experiments for the study of lncRNA *JPX* and its target, *XIST*, in cancer

There is still much work to be done in order to explore the roles of *JPX* and *XIST* in ovarian cancer. I describe below several experiments that will contribute to a publication of this study that addresses whether these two lncRNAs act as tumor suppressors within the cell and if their loss leads to the increase in the severity of cancer in patients. If this is the case, then a final experiment has been proposed that aids in finding a drug that can be effective against cancers of this nature. Continuation

of this study will uncover pathways in which *JPX* and *XIST* are involved as well as help to inform the treatment of ovarian cancer patients.

Global expression profile of low and high JPX *and* XIST *expressing cell lines*

Since there is no normal tissue samples and a general lack of samples for the different grades of ovarian cancer in the TCGA database, it would be beneficial to find collaborators through the UCI Medical Center and other hospitals to obtain ovarian cancer patient samples with matching adjacent ovarian tissues. This would allow for a global sequencing analysis and comparison between normal tissue and the different ovarian cancer grades. A more comprehensive analysis of *JPX* and *XIST* low and high expression patients can then be done to see what other genes are affected disproportionately in these patients, and if there is any compounding effects when both *JPX* and *XIST* are lost.

Additionally, a global sequencing comparison of our *JPX* and *XIST* high expression ovarian cancer cell lines, SKOV3iP1 and OVCAR3, to a low expression cell line like TOV112D, could also reveal important differences in expression profiles. These profiles will also aid in future downstream analyses to uncover the possible roles lncRNAs *JPX* and *XIST* play in cancer.

Cancer-associated genes and the loss of JPX *and* XIST *expression*

As a follow-up to the initial qRT-PCR analysis of ovarian cancer cell lines, SKOV3iP1 and OVCAR3, further knockdown experiments of *JPX* and *XIST* will need to

be done to better assess how loss of these lncRNAs affect global expression as well as proliferation and metastasis.

It is expected that the loss of *JPX* and *XIST* together in these high expression cell lines will increase *MSN* expression and decrease *ATRX* expression, since this is what has occurred in the immortalized cell line. Expression results from *XIST*-only knockdown experiments in OVCAR3 suggest that the two lncRNAs may have different affects when knocked out separately. Therefore, it would be good to perform an in-depth RNA-seq analysis in order to uncover what genes are differentially expressed when *JPX* and *XIST* are knocked down together and separately in SKOV3iP1 and OVCAR3. Following knockdown experiments, overexpression rescue experiments should be performed to test whether supplying *JPX* and/or *XIST* back into these knockdown cell lines can rescue dysregulated expression of genes that were implicated as potential targets of *JPX* and/or *XIST*.

***Proliferation and metastasis in* JPX *and* XIST *knockdown cancer cell lines* in vitro**

Further analysis of proliferation and metastasis needs to be done on the SKOV3iP1 and OVCAR3 knockdown cell lines described above. Using crystal violet staining, proliferation will be assessed over a 72-hour period to measure whether loss of *JPX* and/or *XIST* can increase proliferation of these cell lines. In order to assess metastatic ability, *in vitro* migration and invasion assays should be done since they tend to correlate with *in vivo* metastasis. Wound healing scratch tests on the knockdown cell lines will show how loss of either *JPX* and/or *XIST* affect cell migration. Additionally, the

Boden Chamber Assay can be used to monitor the effect of *JPX* and *XIST* on cell invasion.

As shown in Chapter 4, OVCAR3 *XIST* knockdown cell lines already experienced an increase in proliferation and accelerated migration. Therefore, it is expected that this will be the case when an even greater knockdown of *XIST* is achieved in the OVCAR3 cell line as well as in the SKOV3iP1 cell line. Furthermore, supplying *JPX* and/or *XIST* back into these knockdown cell lines, or a *JPX* and *XIST* low expression cell line such as TOV112D, should attenuate these effects.

***Metastasis and proliferation of ovarian cancer knockdown cell lines* in vivo**

Metastasis of ovarian cancer occurs more easily than classic metastasis seen in other cancers. Ovarian cancer does not undergo rounds intravasion and extravasion in order to relocate, instead taking advantage of passive transport via the intraperitoneal fluid to the omentum and peritoneum. These are the most common secondary sites of ovarian cancer metastasis, aside from the fallopian tube and contralateral ovary (Lengyel, 2010).

As stated in Chapter 4, higher grades of ovarian carcinoma are correlated with low *XIST* expression, and low expression of *JPX* and *XIST* has been associated with more aggressive stages in other cancers (Ma et al., 2017; Xing et al., 2018). In order to study whether these lncRNAs behave as tumor suppressors and whether loss of expression is a contributing factor to the increase in metastasis *in vivo*, analysis in nude mice should be done. Luciferase-labeled *JPX* and *XIST* low and high expression cell lines, as well as high expression cell lines that have *JPX* and/or *XIST* knocked down,

can be used to monitor tumor progression and metastasis through bioluminescence tracking.

Drug screen for selective inhibition of JPX *and* XIST *low expression cells*

If loss of *JPX* and/or *XIST* has a clear phenotype indicative of contribution to cancer severity, it would be informative for treatment of these patients to identify an FDA-approved drug that selectively targets these cells. FDA-approved drug libraries are available through SelleckChem and can be customized to include drugs approved for cancer treatment. If promising, the drug that is effective on *JPX* and/or *XIST* knockdown cell lines can be assessed in mice to see if they significantly prolong survival.

Broader Future Directions

Aneuploidy is a characteristic found in many cancers and can lead to an increase in the severity of that cancer (Ben-David and Amon, 2019). For example, basal-like breast cancer, a class with one of the worst prognosis, has a propensity for X chromosome aneuploidy. Some patients can lose the inactive X chromosome and gain a second active X chromosome. Others can have three X chromosomes, with only one X being silenced when normally two would need to be silenced for proper dosage compensation (Richardson et al., 2006).

In the case of pre-existing X-linked aneuploidy, there is information on the cancer incidence in females with a single X chromosome (Turner Syndrome) and males with an extra X chromosome (Klinefelter Syndrome), but a severe lack of information for females with trisomy X (Triple X Syndrome). A national cohort study done in Great

Britain of cancer incidence in females with Turner Syndrome (45, X0) found that this group has an increased frequency of ovarian cancer but a lower risk of breast cancer (Schoemaker et al., 2008). In the case of males with Klinefelter Syndrome (47, XXY), there is an increased risk of patients developing breast cancer (Richardson et al., 2006). This could also be the case for females with Triple X Syndrome (48, XXX) and warrants further investigation to determine the cancer incidence in these patients. Additionally, the XCI status in Klinefelter Syndrome and Triple X Syndrome cancer patients should be assessed in order to determine if erosion of XCI could be a contributing factor to the progression of cancer in these patients. Lastly, it would be interesting to expand upon the study of Turner Syndrome patients done in Great Britain to determine if the findings of this study are seen in other countries, or if there is any skewing based on region, race, or ethnicity.

5.2: Conclusions

In summary, I have presented in this dissertation that the human long noncoding RNA known as *JPX* is functional and may play a role in ovarian cancer. LncRNAs are known to lack sequence conservation, and human lncRNA *JPX* is no exception, with 40% conservation between human and mouse. SHAPE analysis has revealed that the secondary structures of these two homologs are also divergent. However, despite the vast difference between sequence and structure, EMSA and complement assays revealed that human lncRNA *JPX* is able to robustly bind CTCF and complement mouse lncRNA *Jpx* function in *Jpx-/+* mutant female mES cells. These findings show

that not only is human *JPX* functional, but may also play roles in pathways other than XCI, as supported by my studies of *JPX* and *XIST* in ovarian cancer.

Patient data from cBioPortal shows that there is a significant loss of *XIST* and a decrease in *JPX* expression in patients with higher grades of ovarian cancer. Expression of these two lncRNAs tend to be high in tissues of the ovary. Preliminary analysis of ovarian cancer cell lines, SKOV3iP1 and OVCAR3, show that even in these high *JPX* and *XIST* expressing cell lines there is still a disruption of XCI, as shown by RNA FISH. Moreover, knockdown of *XIST* alone in the OVCAR3 cell line resulted in downregulation of two X-linked and cancer-associated genes—*MSN* and *KDM6A*. Knockdown clones also displayed increased proliferation and accelerated migration, supporting the possible role of *XIST* as a tumor suppressor.

This work has provided several major contributions to the field of lncRNA and cancer genetics. First, the structure of both mouse and human lncRNA *Jpx/JPX* has been resolved and determined to be divergent between the two species. Second, I have demonstrated that human *JPX* is functional within humans. Third, human *JPX* can complement mouse *Jpx* lncRNA function in *Jpx-/+* mutant mES cells, implicating homologous function in humans. Fourth, human *JPX* may have other roles in the cell as indicated by its binding kinetics with CTCF, divergent structure, variable rescue of *XIST* expression in mice, and possible connection to different oncogenic pathways from *XIST* in ovarian cancer patients. Lastly, I have shown that *JPX* and *XIST* are dysregulated in ovarian cancer patients and they are implicated in pathways related to suppression of cancer metastasis. The contributions in this dissertation allow for better understanding of XCI in humans and the roles that *JPX* and *XIST* play in female cancers.

5.3 References

Ben-David, U., and Amon, A. (2019). Context is everything: aneuploidy in cancer. Nat. Rev. Genet.

Chu, C., Zhang, Q.C., da Rocha, S.T., Flynn, R.A., Bharadwaj, M., Calabrese, J.M., Magnuson, T., Heard, E., and Chang, H.Y. (2015). Systematic Discovery of Xist RNA Binding Proteins. Cell *161*, 404–416.

Kubota, M., Tran, C., and Spitale, R.C. (2015). Progress and challenges for chemical probing of RNA structure inside living cells. Nat. Chem. Biol. *11*, 933–941.

Kwok, C.K., Ding, Y., Tang, Y., Assmann, S.M., and Bevilacqua, P.C. (2013). Determination of in vivo RNA structure in low-abundance transcripts. Nat. Commun. *4*, 2971.

Lengyel, E. (2010). Ovarian Cancer Development and Metastasis. Am. J. Pathol. *177*, 1053–1064.

Lucks, J.B., Mortimer, S.A., Trapnell, C., Luo, S., Aviran, S., Schroth, G.P., Pachter, L., Doudna, J.A., and Arkin, A.P. (2011). Multiplexed RNA structure characterization with selective 2'-hydroxyl acylation analyzed by primer extension sequencing (SHAPE-Seq). Proc. Natl. Acad. Sci. *108*, 11063–11068.

Ma, W., Wang, H., Jing, W., Zhou, F., Chang, L., Hong, Z., Liu, H., Liu, Z., and Yuan, Y. (2017). Downregulation of long non-coding RNAs JPX and XIST is associated with the prognosis of hepatocellular carcinoma. Clin. Res. Hepatol. Gastroenterol. *41*, 163–170.

McHugh, C.A., Chen, C.-K., Chow, A., Surka, C.F., Tran, C., McDonel, P., Pandya-Jones, A., Blanco, M., Burghard, C., Moradian, A., et al. (2015). The Xist lncRNA interacts directly with SHARP to silence transcription through HDAC3. Nature *521*, 232–236.

Richardson, A.L., Wang, Z.C., De Nicolo, A., Lu, X., Brown, M., Miron, A., Liao, X., Iglehart, J.D., Livingston, D.M., and Ganesan, S. (2006). X chromosomal abnormalities in basal-like human breast cancer. Cancer Cell *9*, 121–132.

Rosspopoff, O., Huret, C., Collier, A.J., Casanova, M., Rugg-Gunn, P.J., Ouimette, J.-F., and Rougeulle, C. (2019). Mechanistic diversification of XIST regulatory network in mammals. BioRxiv.

Saldaña-Meyer, R., Gonzalez-Buendia, E., Guerrero, G., Narendra, V., Bonasio, R., Recillas-Targa, F., and Reinberg, D. (2014). CTCF regulates the human p53 gene through direct interaction with its natural antisense transcript, Wrap53. Genes Dev. *28*, 723–734.

Saldaña-Meyer, R., Rodriguez-Hernaez, J., Nishana, M., Jacome-Lopez, K., Nora, E.P., Bruneau, B.G., Furlan-Magaril, M., Skok, J., and Reinberg, D. (2019). RNA interactions with CTCF are essential for its proper function. BioRxiv.

Schoemaker, M.J., Swerdlow, A.J., Higgins, C.D., Wright, A.F., and Jacobs, P.A. (2008). Mortality in women with Turner syndrome in Great Britain: A national cohort study. J. Clin. Endocrinol. Metab. *93*, 4735–4742.

Spitale, R.C., Crisalli, P., Flynn, R.A., Torre, E.A., Kool, E.T., and Chang, H.Y. (2013). RNA SHAPE analysis in living cells. Nat. Chem. Biol. *9*, 18–20.

Uroda, T., Anastasakou, E., Rossi, A., Teulon, J.-M., Pellequer, J.-L., Annibale, P., Pessey, O., Inga, A., Chillón, I., and Marcia, M. (2019). Conserved Pseudoknots in lncRNA MEG3 Are Essential for Stimulation of the p53 Pathway. Mol. Cell *75*, 982-995.e9.

Watters, K.E., Yu, A.M., Strobel, E.J., Settle, A.H., and Lucks, J.B. (2016). Characterizing RNA structures in vitro and in vivo with selective 2′-hydroxyl acylation analyzed by primer extension sequencing (SHAPE-Seq). Methods *103*, 34–48.

Xing, F., Liu, Y., Wu, S.-Y., Wu, K., Sharma, S., Mo, Y.-Y., Feng, J., Sanders, S., Jin, G., Singh, R., et al. (2018). Loss of XIST in Breast Cancer Activates MSN-c-Met and Reprograms Microglia via Exosomal miRNA to Promote Brain Metastasis. Cancer Res. *78*, 4316–4330.

Xue, C., Li, F., and Li, F. (2008). Finding noncoding RNA transcripts from low abundance expressed sequence tags. Cell Res. *18*, 695–700.