

**The Function of the CLAMP Zinc Finger Protein in Targeting MSL Complex to
the X-chromosome during Dosage Compensation in *Drosophila Melanogaster***

Jessica Chery

B.S. and B.A., University of Rochester, 2009

A Dissertation Submitted in Partial Fulfillment of the Requirements for the Degree
of Doctor of Philosophy in the Department of Molecular Biology, Cell Biology, and
Biochemistry at Brown University

Providence, Rhode Island

May 2014

©Copyright 2014 by Jessica Chery

This dissertation by Jessica Chery is accepted in its present form by the Department of Molecular Biology, Cell Biology, and Biochemistry as satisfying the dissertation requirement for the degree of Doctor of Philosophy.

Date_____

Erica N. Larschan, Ph.D., Advisor

Recommended to the Graduate Council

Date_____

Judith Bender, Ph.D., Reader

Date_____

Alison Delong, Ph.D., Reader

Date_____

Richard Freiman, Ph.D., Reader

Date_____

Stephen Buratowski, Ph.D., Reader

Approved by the Graduate Council

Date_____

Peter Weber, Ph.D., Dean of the Graduate School

JESSICA CHERY

Molecular Biology, Cell Biology, and Biochemistry Department
Brown University Box G
Providence, Rhode Island 02912
(863) 651 7827 • Jessica_Chery@Brown.edu
Curriculum Vitae

EDUCATION

Brown University

Providence, Rhode Island

PhD Candidate

August 2009 - present

Molecular Biology, Cell Biology, and Biochemistry Department

Defense Date: April 18, 2014

University of Rochester

Rochester, New York

BS, Chemistry/BA, Biology*

May 2009

*Distinction in Research, Dean's List, Achievement in College Leadership

RESEARCH EXPERIENCE

Brown University

Providence, Rhode Island

- PhD Thesis Research: Dr. Erica Larschan, Ph.D. 2010 - present
"The Function of the CLAMP Zinc Finger Protein in Targeting MSL Complex to the X-chromosome during Dosage Compensation in *Drosophila melanogaster*"
- Rotation 2. Advisor: Dr. Richard Freiman, Ph.D. 2010
"The effect of BRD2 expression in neuronal and cancer cells" Established a new cell line to study the effects of BRD2 knockdown in neuronal and cancer cells
- Rotation 1. Advisor: Dr. Kimberly Mowry, Ph.D. 2009
"The role of Vg1 5' UTR in the regulation of translation" Designed constructs to perform translation assays using *in vitro* protein expression techniques

Johns Hopkins University

Baltimore, Maryland

- Bloomberg School of Public Health Summer Research Intern 2008
Conducted research in the Molecular Microbiology and Immunology Department studying the inflammatory role of IL-10 gene as an immunoregulatory cytokine using RNA extraction and cDNA production by reverse transcription in different cellular subsets (T-cells and macrophages)

University of Rochester

Rochester, New York

- McNair Research Summer Scholar 2007
Completed an individual research project “The Tryptophan Mutant in the Human Immunodeficiency Virus (HIV) Type 1 Reverse Transcriptase Active Site”. Wrote a research paper, completed poster and power point presentations to fellow undergraduate research scholars and professors from natural science departments.
- Senior Thesis: Independent Study with Dr. Baek Kim, PhD 2008-2009
“Antiviral effect of Single Chain Antibodies against Influenza A Virus Nucleoprotein”
- Research Assistant in laboratory of Dr. Baek Kim, PhD 2006-2008
Performed basic biochemistry lab techniques. Assisted in development and implementation of research techniques for study of HIV-1 Reverse Transcriptase in lab projects.
- Graduate Independent Study with Dr. Scott McIntosh, PhD 2008
Assisted in various community projects, such as, smoking cessation research, community health interventions, and environmental health project. Partnering with 4th year medical school student community health projects and assisting in grant writing and scientific literature research

Duke University

Durham, North Carolina

- School of Medicine Summer Intern 2006
Worked with peers on a research project analyzing potential biomedical solutions to the drop in testosterone levels in men with AIDS/HIV and (theoretically) devised a small, surgically implantable testosterone pump sensitive to drops in testosterone levels of individuals. Followed typical Duke medical student’s schedule with courses throughout the day related to academic concentration, lectures, workshops, and weekly clinical rotations lasting a couple hours (4-8 hours) or more.

INVITED TALKS

- Cold Spring Harbor New York, New York
Mechanisms of Eukaryotic Transcription August 27-31, 2013
“An evolutionarily conserved zinc finger protein promotes X-specific targeting of *Drosophila* dosage compensation”
- Brown University Providence, Rhode Island
Providence Area Aging Research Forum October 2, 2012
“Mechanisms of Coordinate Gene Regulation”

- Keystone Symposia
Chromatin Dynamics
"A novel zinc-finger protein regulates *Drosophila* Dosage Compensation"

Keystone, Colorado
January 17-22, 2012

PUBLICATIONS

- **Jessica Chery***, Jason Dobson*, Trevor Siggers, Lin Yang, Leila Reider, Sonia Boor, Liu Wei, Gerwald Jogl, Remo Rohs, Martha Bulyk, Erica Larschan. *Drosophila* CLAMP zinc-finger protein recognizes GA dinucleotide expansions that target dosage compensation to the male X-chromosome. 2014. *In preparation*
- **Jessica Chery**, Erica Larschan. X-marks the spots: X-chromosome identification during dosage compensation. 2013, *Biochim. Biophys. Acta* Volume 1839, Issue 3, 234-240
- Marcela M.L. Soruco*, **Jessica Chery***, Eric P. Bishop*, Trevor Siggers, Michael Y. Tolstorukov, Alexander R. Leydon, Arthur U. Sugden, Karen Goebel, Jessica Feng, Peng Xia, Anastasia Vedenko, Martha L. Bulyk, Peter J. Park, and Erica Larschan . The CLAMP protein links the MSL complex to the X-chromosome during *Drosophila* dosage compensation. 2013, *Genes and Development* 27, 1551-1556
- Erica Larschan, Marcela Soruco, Ok-Kyung Lee, Shouyong Peng, Eric Bishop, **Jessica Chery**, Karen Goebel, Peter J. Park, and Mitzi I. Kuroda. Identification of Chromatin-associated Regulators of MSL Complex Targeting in *Drosophila* Dosage Compensation. 2012, *PLoS Genetics* 8, e1002830
- ***Chery, Jessica**. Darwin Operario, Baek Kim. The Tryptophan 115 Mutant in the Human Immunodeficiency Virus (HIV) Type I Reverse Transcriptase Active Site. *Journal of Undergraduate Research (JUR)* Fall 2007 Volume 6 Issue 1

***First author(s) /These authors contributed equally to this work.**

HONORS and AWARDS

- XSEDE (Extreme Science and Engineering Discovery Environment) Scholar 2013
- Sigma Xi, the Scientific Honor Society (International Honor Society for Science and Engineering) 2012
- Keystone Symposia Future of Science Fund Scholarship (Chromatin Dynamics Jan. 17 – 22, 2012 Meeting) 2011

- Pre-Doctoral National Institutes of Health Ruth L Kirschstein National Research Service Award (NRSAP) F31 Predoctoral Fellowship – with score of 16: Awarded, Deferred
- University of Rochester Renaissance Scholar (Four-year full scholarship awarded to select number of students from top 1% of applicants based on academic promise) 2005-2009
- McNair Scholar (For students interested in pursuing advanced degrees of study) 2007- 2009
- David T. Kearns Scholar (Scholarship for students in math and science) 2006-2009
- National Scholars Honor Society (Honorary Recognition of academic collegiate achievement) 2009
- Iota Sigma Pi Scholar (National Honor Society for Women in Chemistry) 2008

TEACHING EXPERIENCE

Brown University Providence, Rhode Island

Teaching Fellow, Summer@Brown 2012-2013, 2013-2014

- Sex and Chromosomes: Genetics and Biochemistry of Development
Created this course to teach exceptional pre-college students

Teaching Fellow, Initiative to Maximize Student Development 2012-2013

- Scientific Writing: Key Principles, Fall 2013
- Designing and Delivering Scientific Presentations, Spring 2013
- Reading Scientific Publications, Fall 2012
Assisted faculty in developing and running focused workshops for PhD candidates to teach critical reading and writing and effective presentation skills in the life and physical sciences

Writing Associate 2012-2013

- Work with undergraduate and PhD students in one-on-one sessions helping develop effective writing skills such as clarity, logical flow, conciseness

Sheridan Center Teaching Certificate I 2012-2013

- Completed this certificate by meeting specific teaching requirements including Sheridan Teaching Seminars, workshops, developing a departmental microteaching session, and individual teaching consultation

Science Consultant of ARISE 2012-2013
• Advise Rhode Island high school science teachers on how to use various biology lab equipment and perform different labs with students through the ARISE (Advancing Rhode Island Science Education) program

Leadership Alliance Science Mentor 2013
• Train and advise undergraduate summer interns at Brown University through the Leadership Alliance program

Teaching Assistant 2010
Biology of Eukaryotic Cell
• Led weekly workshops with small groups of undergraduates and first-year PhD candidates from the class, tutored students one on-one, reviewed challenging concepts from lecture, held office hours, and graded exams

Mentor 2010 - 2014
• Trained 3 undergraduates, 1 rotating graduate student in widely used laboratory techniques, research methodology, and introductory biology concepts.

University of Rochester Rochester, New York

Teaching Assistant 2008 - 2009
• General Chemistry
• Biochemistry
Organized weekly workshops with small groups of students from class, tutored students one-on-one, reviewed homework and challenging concepts from lecture, held office hours, and graded exams

Student Center Assistant 2007-2008
• Physical Chemistry
Assisted students in the physical chemistry course to clarify lecture content and notes

UNIVERSITY SERVICE Providence, Rhode Island

BROWN University 2012- present
• Graduate Student Advisory Committee
• Graduate Student Council Representative

PROFESSIONAL ACTIVITIES

- Genetics Society of America
Public Policy Committee 2013 -present
General Member 2011 -present
- Annual Biomedical Research Conference for Minority Students (*ABRCMS*)
Judge 2012, 2013
Evaluated poster and oral presentations of undergraduates from colleges and universities all across the United States for awarding of prizes to best presenter
- American Association for Science Advancement 2010 – 2014
- Minority Association of Pre-medical Students (MAPS) 2007 – 2009
President, University of Rochester

POSTER PRESENTATIONS

- XSEDE (Extreme Science and Engineering Discovery Environment)
San Diego, CA 2013
- New England Science Symposium
Harvard Medical School, Boston, Massachusetts 2012, 2013
- National Institutes of Health National Human Genome Research Institute
(NHGRI) “Advances in Genomics Research”
Bethesda, Maryland 2012
- Keystone Symposia Chromatin Dynamics
Keystone, Colorado 2012
- Brown University Molecular Biology, Cell Biology, and Biochemistry
Department Retreat. Providence, Rhode Island 2010, 2011
- Chromatin and Epigenetic Regulation of Transcription 30th Summer
Symposium in Molecular Biology, Pennsylvania State University 2011
- *Drosophila* Research Conference
San Diego, California 2011
- Annual Biomedical Research Conference for Minority Students
(*ABRCMS*) 2010. Charlotte, North Carolina 2010

- Boston Regional Transcription Meeting
Tufts University, Boston, Massachusetts 2010

EMPLOYMENT

Rochester, New York

- **Residential Advisor** 2006-2009
University of Rochester
- **Meridian Society** 2005 - 2009
University of Rochester Office of Admissions

COMMUNITY OUTREACH

- ARISE Nature of Discovery Symposium Invited Speaker 2013
Brown University. Providence, Rhode Island
- Rhode Island Science Olympiad Volunteer 2013
Rhode Island College. Providence, Rhode Island
- Girls Interested in Real Science (GIRLS) Invited Speaker 2013
Ocean Explorium. New Bedford, Massachusetts
- Rhode Island Science and Engineering Fair Judge 2012, 2013
Community College of Rhode Island. Providence, Rhode Island
- “Sak Pase” Organizer 2011
Haiti Benefit Concert for January 10, 2010 Earthquake victims
Providence, Rhode Island
- Rhode Island Haiti Relief Foundation E-Board 2010
Providence, Rhode Island
- COMPEER Volunteer 2008
Rochester, New York
Mentored a child who had one parent incarcerated, serving as a role model through a friendship designed to inspire the child to believe in their capacity to overcome life’s obstacles to live and achieve their positive goals of success.

EXTRACURRICULAR ACTIVITIES

Rochester, New York

University of Rochester

- Student Organization for Caribbean Awareness Co-founder 2005-2007
- Afro-Expressions Dance Troupe Business Manager 2005-2006

SKILLS

- Software: MacVector, Adobe Illustrator, Microsoft Word & Excel, Keynote
- Language: French, Haitian Creole (Fluent)

DEDICATION

My Ph.D. dissertation is dedicated to my late mother, described as a warrior angel by those who knew her.

Thank you for living with honor, integrity, mercy, grace, generosity, forgiveness, faith, strength, and above all love. Thank you for being fierce! Thank you for always helping me believe I could accomplish anything I set my mind to. Thank you for sharing with me your wonder for and respect of life, curiosity about the world, service to others, and the drive to pursue my passions with excellence.

I thank her for teaching me all of value that I needed to know about life and living! Her excellent model of living well inspires me every day!

“Blessed is the man who remains steadfast under trial, for when he has stood the test he will receive the crown of life....” James 1:12

ACKNOWLEDGEMENTS

First, I thank GOD for giving me all I needed to complete graduate school.

I have many people I need to thank: people whose support has helped me achieve this dissertation.

I thank my PhD advisor, *Dr. Erica Larschan*, for her guidance in helping me produce this work. Thank you for a laboratory and space where I could develop my scientific abilities. Your enthusiasm for science is contagious. Your patience and generosity through my learning process have been good reminders of the attributes that inspire us all to brave learning something new. Through many experiments, I have learned to exercise my critical, creative, and independent thinking skills. Thank you to current and former *Larschan lab ladies* for a friendly working environment with laughter and baked goodies that sweeten the day!

I also want to thank my committee members: *Dr. Judith Bender*, *Dr. Alison Delong*, *Dr. Richard Freiman* for their scientific insight and advice throughout the years. Thank you to *Dr. Judith Bender* for teaching me the “logic tree”, which I will not forget. I am very thankful to my outsider reader *Dr. Stephen Buratowski* for taking the time to read my dissertation and taking an interest in my career development. Thank you, *Dr. Buratowski*, for the wonderful and unforgettable opportunity to present at the Cold Spring Harbor meeting!

I must say thank you to *Dr. Tricia Serio* for being an excellent advocate for the graduate students of MCB. Your dedication to student well-being was very much appreciated and sorely missed. I am also thankful to *Dr. Mike McKeown* for a great first interview and witty remarks later on. Our conversation helped me visualize coming to Brown. I also thank *Dr. Kimberly Mowry* for frank conversations: the very first of which helped me decide to come to Brown and later ones that helped me gain perspective on where I needed to grow as a scientist. Thank you very much to *Dr. Al Dahlberg* for reminding me that my basic research can have practical applications and letting me know that I could do great things!

Thank you to *Dr. Racquel Kim Sherwood* for very careful reading and editing of this complete dissertation! What a feat!

A great big thanks goes to *Kathy Cooper* and the *Graduate Medical and Christian Fellowship* (GMCF) community! Thank you so much for all your prayers, encouragement, and support! Your friendship and shared meals have added a spark of sunshine in many a bleak days!

Thank you to my graduate school friends *Dr. Noelle H. Kelso*, *Dr. Orly Clerge*, and *Heather L. Bennett* for all your insight, laughter, and encouragement! Orly, thank you for speaking positive words over seemingly negative situations. Noelle, your honesty and boldness are refreshing and inspiring.

Thank you ladies for sharing the world of half-marathon running with me! I have grown and learned so much from you two. *Heather*, you have seen me at my best and worst but still you see the good in me. I am so thankful for your friendship and hospitality through graduate school.

Thank you to my friend *Dr. Yen Lim* who appeared in my life like a shooting star and whose presence radiates warmth like the sun. Thank you for all the philosophical discussions, adventures, laughter, hugs, movie nights, soups for my sick days, and making sure I ate while writing my dissertation! Knowing you gives me hope for our world!

Thank you to all my fellow graduate students who shared their insights with me along this journey. I am very grateful to my undergraduate friends from the University of Rochester who have been there for me. Thank you for believing in me and being a text message and phone call away when I needed someone.

I especially want to thank advisors from my undergraduate years who introduced me to the world of research. Thank you to *Dr. Beth Olivares* and the *Kearns Scholars* group. Thank you to *Dr. Baek Kim* for showing me what a super cool and fun PI can be like and taking me on as a wide-eyed and inexperienced undergrad! Thank you to *Dr. Jay Bream* at Johns Hopkins University for encouraging me to pursue graduate school and letting me know that I could do it!

I am very thankful to *Dr. Noel Rose* at Johns Hopkins University for believing I can do far greater than I ever lay out, encouraging me in my pursuits, and being an eager and willing advocate.

Thank you to my high school English teacher *Mrs. Juanita Schnoor* for always saying “I expect to hear great things of you” and my Calculus teacher, *Mr. Chandler*, for first speaking to me about research.

Last but definitely not least, I thank my wonderful family whose love and support have been one of the best things about this journey. Thank you to my father for being an excellent listener and man of wisdom. Dad, your silent courage and respect for human life are a ray of sunshine. Thank you for instilling in me a love of learning, the importance of integrity and character, and a desire to do unto others as I would like done unto myself. Thank you for always pushing me to do all things with excellence and always praying for me!

Thank you to my eldest brother for always pushing the boundaries! Your constant search for novelty inspires me to think beyond my limits. Your relentless attitude encourages me to never give up! Your unwavering faith in my abilities is a rock! Thank you to my second brother for listening to and reasoning with me. Your belief in my ability to accomplish far more than I imagine encourages me. You challenge me to be magnanimous. Watching you grow into a splendid man has been a blessing. To my youngest brother, I cannot imagine our family without you.

No matter the situation you always make me laugh – without fail! You are so full of wisdom. Thank you for being there for me when I needed a shoulder to lean on.

Thank you to my extended *Chery*, non-*Chery* family, and the *Jeune* family! Your down-to-earth approach to life helps me stay focused on the things that truly matter! Thank you for your constant prayers, encouragement, and support through all the good and bad times. Your acts of love and care have been catalysts motivating me to be propelled by the good in our world!

Table of Contents	Page
Title Page.....	i
Copyright Page	ii
Signature Page	iii
Curriculum Vitae	iv
Acknowledgements.....	xii
Table of contents	xviii
List of Tables	xxiii
List of Figures	xxiii
Chapter 1: X-marks the spot: X-chromosome identification during dosage compensation.....	1
Abstract	3
Introduction	4
Sex Determination	6
Identification of the X-chromosome during dosage compensation	8
Long noncoding RNAs that function during dosage compensation	
in mammals and flies	9
DNA sequence elements that function during dosage compensation.....	13
The role of chromatin regulation during dosage compensation.....	15
Interactions between proteins and chromatin that are involved	
in dosage compensation	18
Conclusion.....	21
References.....	23

Figures and Figure Legends.....	30
Chapter 2: Identification of Chromatin-Associated Regulators of	
MSL Complex Targeting in <i>Drosophila</i> Dosage Compensation.....	32
Abstract	35
Author Summary	36
Introduction	37
Results	39
A genome-wide RNAi screen identifies genes that regulate an	
MSL-dependent reporter in male SL2 cells.....	39
Selection of NSL complex, PAF complex, and CG1832 as candidate direct	
regulators of MSL targeting or function	44
CG1832 and NSL complex modulate MSL complex recruitment to	
endogenous CES loci in vivo.....	46
Genome-wide occupancy of NSL components and CG1832 support	
a direct role in MSL targeting.....	48
Discussion	49
Methods	53
Acknowledgements	58
References	59
Figures and Figure Legends.....	62

Chapter 3: The CLAMP protein links the MSL complex to the

X chromosome during <i>Drosophila</i> dosage compensation.....	84
Abstract.....	87
Introduction	88
Results and Discussion.....	91
CLAMP directly recognizes MREs to tether MSL complex to the	
X-chromosome.....	91
CLAMP is required for MSL complex recruitment to the male	
X-chromosome <i>in vivo</i>	95
CLAMP is enriched on the X-chromosome independent of	
MSL complex.....	97
Acknowledgements.....	102
Methods.....	103
References.....	105
Figures and Figure Legends	110
Supplemental Methods	117
Supplemental Figures	132

Chapter 4: The *Drosophila* CLAMP zinc-finger protein recognizes

GA dinucleotide expansions that target dosage compensation to	
the male X-chromosome.....	141
Abstract.....	144
Introduction.....	145
Results and Discussion.....	148

The occupancy of the CLAMP protein is increased on the X-chromosome compared to autosomes due to increased number of motifs.....	148
CLAMP occupancy and motif distribution relative to genes differs 3' versus 5' on X-chromosome versus autosomes.....	149
Sequences flanking the core 8bp motif within the MRE are required for CLAMP recruitment.....	150
DNA Shape analysis reveals GA repeats as critical for CLAMP binding.....	153
Increased number of dinucleotide repeats increases CLAMP binding to DNA <i>in vitro</i> and <i>in vivo</i>	155
Conclusion.....	157
Materials and Methods	160
Figures and Figure Legends.....	172
References.....	203
Conclusions and Future Directions: Implications for the role of CLAMP in genomic targeting.....	206
Appendix 1: The glutamine domain of CLAMP is necessary for MSL Complex recruitment.....	227
Introduction.....	228
Results and Discussion.....	230
Conclusion.....	234
Future work.....	235

Materials and Methods.....	236
Figures.....	242
References.....	254
Appendix 2: CLAMP exhibits an <i>in vivo</i> interaction with the <i>roX2</i> RNA	
component of MSL complex.....	267
Introduction.....	268
Results	271
RNA Immunoprecipitation.....	271
Defining interactions between CLAMP and MSL complex	
components.....	272
CLAMP and MSL complex super-shift EMSAs.....	273
Discussion.....	274
CLAMP interacts with <i>roX2</i>	274
CLAMP co-immunoprecipitation with MSL complex components	275
The CLAMP zinc fingers may form a complex with both	
MSL complex and DNA.....	275
Conclusion/Future Directions.....	276
Materials and Methods.....	278
Figures and Figure Legends.....	288
References.....	296

List of Tables

Chapter 2

Table 1. Gene ontology categories that were enriched in the genome-wide RNAi screen.....	70
Table S1. Initial candidates identified from a genome-wide RNAi screen to identify novel regulators of dosage compensation.....	77
Table S2. Final candidates identified from a genome-wide RNAi screen.	83

List of Figures

Chapter 1

Figure 1. lncRNA transcription generates an environment favorable for coordinate regulation.....	30
Figure 2. MOF deposits H4K167acetylation, de-condensing chromatin to favor the X chromosome for active transcription.....	31

Chapter 2

Figure 1: An RNAi screening system in <i>Drosophila</i> SL2 cells to identify regulators of MSL complex recruitment and function.....	62
Figure 2: Candidate regulators of MSL complex identified from the genome-wide RNAi screen.....	64
Figure 3: NSL1 and CG1832 contribute to MSL complex recruitment.....	66

Figure 4: NSL complex and CG1832 enrichment at CES loci.....	68
Figure S1. mRNA levels of MSL complex components and candidates after CG1832,Nsl1, and Paf1 RNAi treatments.....	71
Figure S2. Western blots for CG1832 RNAi treatment and antibody validation.....	73
Figure S3. Validation of CG1832 antibody by ChIP at CES loci.....	74
Figure S4. Immunostaining of polytene chromosomes.....	75

Chapter 3

Figure 1. CLAMP localizes precisely at MREs at high levels when co-localized with MSL complex.....	110
Figure 2. CLAMP directly binds to MRE sequences.	112
Figure 3. CES have three classes of CLAMP occupancy patterns.....	113
Figure 4. Three subclasses of CES differ in their number of tandem MREs, two-dimensional clustering along the X-chromosome.....	115
Figure S1. Global comparisons of CLAMP and MSL complex occupancy at MREs.....	132
Figure S2. CLAMP and MSL complex are highly enriched when co-localized and CLAMP also localizes to the 5' ends of active genes genome-wide.....	134
Figure S3. CLAMP is required for roX gene transcription and MSL localization to the X.....	136
Figure S4. Comparison of CLAMP binding characteristics in the presence and absence of MSL complex.....	139

Chapter 4

Figure 1. CLAMP occupies sites across the genome	172
Figure 2. CLAMP binding sequences are enriched on the X chromosome relative to autosome.....	174
Figure 3A. CLAMP motifs are enriched in gene bodies and towards 3'ends on the X-chromosome.....	175
Figure 3B. On the X-chromosome, CLAMP occupancy is greater over gene bodies and towards 3' ends.....	176
Figure 4A. Probes used for customized Protein Binding Microarray.....	177
Figure 4B. Sequences flanking the core 8-bp region sequences increase binding.....	178
Figure 4C. Constant flanking DNA does not compete well with wild-type flank for binding.....	179
Figure 4D. Sequences flank the core 8 bp region increase binding <i>in vivo</i>	181
Figure 5. The strongest CLAMP binding sites have additional GA-repeats.....	183
Figure 5A. Motifs defined from custom PBMS reveal that the top motif with the best z-score (5%) has more GAs than modestly bound motif (50%).....	183
Figure 5B. Dinucleotides are the strongest predictors of CLAMP occupancy.....	184
Figure 6: Increased number of GA-repeats increases CLAMP occupancy.....	185
Figure 6A. Optimal GA repeat number for CLAMP binding <i>in vitro</i> is 26bp GA sequence.....	185
Figure 6B. X-chromosome has more GA repeats than autosomes.....	187

Figure 7. Model: Increasing GA repeat number increases CLAMP binding	
elevating MSL complex recruitment	188
Supplemental 1. Sequences flanking the 8bp core of the CLAMP motif	
significantly contribute to CLAMP binding <i>in vitro</i> (PBM)	189
Supplemental Figure 2A. MBP-CLAMP binding is specific for CLAMP motif.....	190
Supplemental Figure 2B. MBP-CLAMP binding is specific for CLAMP motif.....	191
Supplemental Figure 3A. PBM probes from the X-chromosome have greater	
CLAMP binding than Autosomal probes.....	192
Supplemental Figure 3B. Histogram indicates mixed Gaussian in probe	
signal intensity.....	193
Supplemental Figure 3C. Classification of probes into weak (blue),	
medium (green), and strong (red) probes.....	194
Supplemental Figure 3D. Comparison of A vs X for probes of strong signal.....	195
Supplemental Figure 3E. Comparison of A vs X probes for weak signal intensity:	
Weak probes on X more strongly bound	196
Supplemental Figure 3F. Difference in shape features is more striking for weak	
probes instead of strong probes.....	197
Supplemental Figure 3G. Mann-Whitney test of sequence +shape versus sequence	
only model shows sequence only has greater effect on binding.....	202

Appendix 1

Figure 1. Western blot depicting Copper inducible expression of	
TAP-tagged CLAMP	242

Figure 2A. HA tagged CLAMP truncations were expressed from an actin promoter in male cells	243
Figure 2B. HA tagged CLAMP truncations were expressed under an actin promoter in male cells	244
Figure 3. CLAMP mRNA levels after transfection with dsRNA before transfecting in HA truncations.....	246
Figure 4. Luciferase activity after transfection with dsRNA before transfecting in HA truncations.....	247
Figure 5. Amino acids 161-562 and the zinc finger domain of CLAMP display the greatest loss in luminescence after dsRNA2 and dsRNA3 treatments.....	248
Figure 5A. Luciferase activity of amino acids 161-562 and 213-562 of CLAMP.....	248
Figure 5B. Luciferase activity of amino acids 1-287 and 1-260 of CLAMP	249
Figure 5C. Luciferase activity of zinc finger domains of CLAMP	250
Figure 6. Working Model.....	251
Figure 6a. Luciferase reporter used to assay MSL complex recruitment.....	251
Figure 6b. Expression of excess CLAMP zinc finger domain prohibits MSL complex recruitment.....	252
Figure 6c. Excess levels of 161-562 domain of CLAMP sequesters MSL complex.....	253

Appendix 2

Figure 1: CLAMP exhibits <i>in vivo</i> interaction with <i>roX2</i>	288
Figure 2: Co-Immunoprecipitation of CLAMP and MLE.....	289
Figure 2a. Co-immunoprecipitation of CLAMP and MLE when cells are crosslinked with formaldehyde.....	289
Figure 2b. Co-Immunoprecipitation with CLAMP antibody coupled beads.....	290
Figure 2c. Co-Immunoprecipitation without RNase treatment.....	291
Figure 2d. Co-Immunoprecipitation with RNase treatment.....	292
Figure 2e. CLAMP and MLE Co-Immunoprecipitation performed with Native conditions.....	293
Figure 3: Gel-shift with S2L (male) and KC (female) cell extracts.....	294

**CHAPTER 1: X-marks the spot: X-chromosome identification during dosage
compensation**

INTRODUCTION

X-marks the spot: X-chromosome identification during dosage compensation

The introduction portion of my thesis dissertation was published as a review article and appears below in its published journal format.

J. Chery, E. Larschan, "X-marks the spot: X-chromosome identification during dosage compensation", *Biochim. Biophys. Acta* (2014),

<http://dx.doi.org/10.1016/j.bbagr.2013.12.007>

Abstract

Dosage compensation is the essential process that equalizes the dosage of X-linked genes between the sexes in heterogametic species. Because all of the genes along the length of a single chromosome are co-regulated, dosage compensation serves as a model system for understanding how domains of coordinate gene regulation are established. Dosage compensation has been best studied in mammals, flies and worms. Although dosage compensation systems are seemingly diverse across species, there are key shared principles of nucleation and spreading that are critical for accurate targeting of the dosage compensation complex to the X-chromosome(s). We will highlight the mechanisms by which long non-coding RNAs function together with DNA sequence elements to tether dosage compensation complexes to the X-chromosome.

Introduction

Assuring that all genes are transcribed at the correct gene dosage across the genome is a key process because protein and RNA components of large macromolecular complexes are encoded on different chromosomes. Altering the stoichiometry of macromolecular complexes causes severe phenotypes (Disteche 2012). The most commonly known disease of gene dosage is Down's syndrome, which is caused by trisomy of chromosome 21 (Jacobs et al. 1959; Lejeune et al. 1959; Hattori et al. 2000). Furthermore, copy number variation throughout the genome has been linked to diverse diseases from autism to schizophrenia, and neurons are particularly sensitive to copy number changes (Girirajan et al. 2011).

There is a naturally occurring situation in which there is a seeming imbalance in gene dosage: the sex chromosomes in heterogametic species. Sex-chromosome dosage compensation is an essential process that evolved to regulate the levels of transcription of X-linked genes in heterogametic species (Beatriz Vicoso 2009; Mank 2009; Johnson and Lachance 2012). In the eutherian mammals (humans and mice) and *Drosophila*, the male is heterogametic (XY) and the female is homogametic (XX) (Disteche 2012). In contrast, *Caenorhabditis elegans* (*C. elegans*) hermaphrodites are XX and males are XY. In mammals, dosage compensation occurs by silencing one of the two female X-chromosomes (Lee and Bartolomei 2013). In *Drosophila*, the genes along the length of the single male X-chromosome are upregulated two-fold to equalize transcript levels between males and females (Belote and Lucchesi 1980a; Meller and Kuroda 2002). In *C. elegans*, the levels of

transcription from each of the two X-chromosomes in hermaphrodites are downregulated two-fold (Meyer 1997). Despite these diverse mechanisms, there are commonalities including nucleation of dosage compensation on the X-chromosome followed by spreading along its length (Kelley et al. 1997; Larschan et al. 2007; Alekseyenko et al. 2008)(*Drosophila*); (Pinter et al. 2012)(Mammals); (Csankovszki et al. 2004)(worm). Furthermore, recent work has revealed a key commonality among dosage compensation mechanisms in all organisms, which is the mechanism by which transcription levels of genes on the single active X-chromosome are equalized to those on autosomes: the acetylation of histone H4 at lysine 16 (Deng et al. 2011)(Deng X 2013).

In mammals, *C.elegans* and *Drosophila*, the process of dosage compensation is essential for proper development (Belote and Lucchesi 1980a; Hsu and Meyer 1994; Meyer 1997; Schulz and Heard 2013). In humans, imbalance in sex-chromosome number has strong phenotypic consequences including infertility (Heard and Turner 2011). Humans that lack one X-chromosome (XO) develop Turner syndrome and show symptoms of ovarian failure, infertility, short stature, and similar congenital malformations (Leppig and Distèche 2001). Males with an extra X-chromosome (XXY), termed Klinefelter syndrome, are sterile (Heard and Turner 2011). Similarly males with two Y chromosomes (YYX), Double Y Syndrome, are sterile if the extra Y is not lost early in sperm cell development. In addition, neurons are very sensitive to gene dosage and autism and schizophrenia have been linked to deletions of a single copy of several genetic loci that each contain many genes (Lee

and Bartolomei 2013). In *C. elegans* and *Drosophila*, dosage compensation is essential for viability.

Sex determination

The process of sex determination occurs upstream of dosage compensation and is directly linked to dosage compensation in *Drosophila* and *C. elegans*. In contrast, in mammals, it is not known how dosage compensation is linked to sex-determination. Master regulators such as *Sry* in mammals, *sex-lethal* (*sxl*) in *Drosophila*, *xol* in *C. elegans*, and *Dmrt1* conserved in flies, mammals, and worms regulate the process of sex determination. Because sex determination has been reviewed elsewhere (Sekido and Lovell-Badge 2009; Meyer 2010; Salz and Erickson 2010; Herpin and Schartl 2011), we will provide only a brief summary below.

In mammals the *Sry* gene on the Y-chromosome encodes the SRY protein that determines the male fate by acting on the *Sox9* gene. A positive feedback mechanism ensures continual expression of *Sox9* leading to testis differentiation. In the absence of SRY or temporal mis-expression of SRY, *Sox9* is silenced and follicle cell and ovary formation occurs.

In *Drosophila*, the ratio of the sex chromosomes to autosomes determines sex. The gene dosage of the *sisterless* genes that are encoded on the X-chromosome results in the differential regulation of the master regulator of sex determination, the *sex-lethal* gene. *sex-lethal* (*sxl*) determines sexual fate by using different promoters to produce sex-specific products at precise developmental stages, and the SXL protein to control alternative splicing of the female specific *tra* protein. The

Tra protein, in turn, regulates sexual dimorphism (Salz and Erickson 2010). Once initiated, the female fate is continually maintained through autoregulation of the *sxl* gene by the SXL protein. Furthermore, in females SXL translationally represses the *male-specific lethal-2* (MSL2) protein, a core component of the male specific lethal (MSL) complex that upregulates expression of the single X-chromosome in males (Kelley et al. 1997). In this way, sex determination is linked to dosage compensation.

Like *Drosophila*, in *Caenorhabditis elegans* (*C. elegans*), it is the ratio of sex chromosomes to autosomes that determines sex. The ratio of 2X:2A and 3X:4A defines the hermaphrodite, while the male is defined by 1X:2A and 2X:3A ratios (Meyer 2010). The key sex-determining factor is the XO-lethal gene (*xol*), which activates the cascade for the male fate. In hermaphrodites (2X:2A), *xol* is repressed by X-signal elements (XSEs) encoded on the X-chromosome. In hermaphrodites, repression of *xol* allows activation of the sex determination and dosage compensation complex (SDC-2), which is involved both in establishing hermaphrodite sexual development as well as dosage compensating the XX state (Yonker and Meyer 2003; Meyer 2010).

Dmrt1, a conserved transcription factor in worms, flies, and mammals, is critical for maintaining the male fate. In mice, *Dmrt1* inhibits activation of the female fate genetic network through suppression of key female fate factors such as the FOXL2 transcription factor and WNT4/ β -catenin (Herpin and Scharl 2011). In the absence of *Dmrt1* and upregulation of female fate genes, male specific cells such as the Sertoli cells, differentiate into female-specific granulosa cells producing estrogen

(Herpin and Scharf 2011). Subsequent to the initial steps of the sex determination process, dosage compensation mechanisms begin.

Identification of the X-chromosome during dosage compensation

Due to the haploinsufficiency of many loci throughout the genome, the cell is sensitive to dosage of a large number of genes including those that encode critical transcription and translation factors (Disteche 2012)(Girirajan et al. 2011). However, there is a widespread naturally occurring haploinsufficiency in all heterogametic species: genes that are encoded on the sex chromosomes. Dosage compensation mechanisms have evolved to specifically distinguish the X-chromosome from autosomes for subsequent transcriptional modulation. X-chromosome identification involves the following factors: 1) long non-coding RNAs that are likely to serve as nucleation centers (Batista and Chang 2013a) (Batista and Chang 2013b)(Rinn and Chang 2012)(Thomas Cremer 2010)(Derrien et al. 2012)(Dundr and Misteli 2010); 2) DNA sequence elements; 3) chromatin marks that alter packed chromatin; 4) proteins that read and write the chromatin marks (Thomas et al. 2011; Fisher et al. 2012; Guertin and Lis 2012; Lee 2012; Batista and Chang 2013b; Lusk and Eisen 2013). Below we discuss how each of these factors contributes to specifically target the dosage compensation machinery to the X-chromosome.

Long noncoding RNAs that function during dosage compensation in mammals and flies

Sex-chromosome dosage compensation in mammals, flies, and worms is achieved either by upregulation or downregulation of a single or two sex chromosomes. While there are differences between the dosage compensation mechanisms among these species, the establishment of nucleation centers is required to initially identify the whole X-chromosome for dosage compensation. In mammals and flies, a key structural scaffold of these nucleation centers is the use of long noncoding RNAs (lncRNAs) as targeting platforms (Engreitz et al. 2013),(Nora et al. 2012).

lncRNAs are defined as RNAs that are larger than 200 nucleotides and do not code for functional protein (Batista and Chang 2013b)(Orom et al. 2010; Rinn and Chang 2012). The ability for the secondary structure and location of lncRNAs to be preserved in the absence of sequence conservation (Derrien et al. 2012; Batista and Chang 2013b) make them attractive targeting platforms because this is likely to decrease the frequency of spatial temporal error in the establishment of a domain of coordinate gene regulation. lncRNAs also have the capacity to bring different factors in close proximity (Batista and Chang 2013b). Furthermore, lncRNAs can function in *cis* and *trans*, interact with multiple proteins, and shuttle between the nucleus and cytoplasm (Orom et al. 2010; Batista and Chang 2013b; Lai et al. 2013). For example, the *roX* lncRNAs (RNA on X) can induce transcription of nearby genes when ectopically inserted onto autosomes (Kelley et al. 1999; Park et al. 2010). Using the transcription of lncRNAs as a means of regulating expression of other

genes is also supported by the observation that promoters of lncRNAs are the regions of lncRNAs under the greatest selective pressure (Derrien et al. 2012; Batista and Chang 2013b). In summary, lncRNAs are an ideal system by which to nucleate chromatin domains in a spatially and temporally regulated manner (Fig. 1).

The dosage compensation system in mammals uses lncRNAs to control X-chromosome identification. In mammals, dosage compensation is achieved by silencing one of two X-chromosomes in female (XX), a process called X-chromosome Inactivation (XCI) (Lee 2012). XCI is controlled by the X inactivation center (Xic), which is a 100-500kb region found on the chromosome to be silenced, Xi, and the active chromosome, Xa (Lee and Bartolomei 2013). The Xic is a rich source of lncRNAs that are involved in X-chromosome choice, counting, pairing, and silencing (Lee and Bartolomei 2013). Xic regulation involves an intricate network of RNA and protein interactions. In the first steps of XCI, the X-specific-inactive-transcript (*Xist*) is transcribed only from the Xic on Xi into a 17-20kb noncoding RNA and coats Xi in *cis*. Silencing of Xi requires *Xist* expression as *Xist* utilizes its conserved repeat motif (Rep A) to bind directly to Polycomb repressive complex 2 (PRC2) to target PRC2 to Xi. PRC2 deposits the silencing mark trimethylation of histone H3 at lysine 27 (H3K27me3), facilitating chromatin compaction of DNA from active transcription (Lee 2012). *Xist* function however is regulated by its antisense transcript, *Tsix*, another lncRNA. *Xist* upregulation requires *Tsix* downregulation (Ko-Hsuan Hung 2013). *Tsix* inhibits *Xist* by silencing *Xist* through recruitment of a DNA methyltransferase (Dnmt3a) and inhibiting *Xist* binding to the PRC2 repressor protein (Zhao et al. 2008) (Lee 2012) (Lee and Bartolomei 2013). In addition to *Tsix*,

other lncRNAs encoded at the Xic repress *Xist* function. These include the X-inactivation intergenic transcription element (*Xite*) and the testis-specific X-linked gene (*Tsx*). A recently identified lncRNA called *Linx* (large intervening transcript in the Xic) may also have a role in *Tsix* regulation via modulating *Xite* (Nora et al. 2012). There are also a number of lncRNAs encoded at the Xic that activate *Xist*. These include the RepA RNA and the *Jpx* RNA that can function both in *cis* and *trans* to activate *Xist* (Ko-Hsuan Hung 2013). Overall, an intricate network of inter-dependent lncRNAs nucleates mammalian dosage compensation.

Drosophila sex-chromosome dosage compensation also utilizes lncRNAs to establish local neighborhoods of regulation. In contrast to mammals, *Drosophila* sex-chromosome dosage compensation is achieved by upregulating transcription of the single X-chromosome two-fold in males (Lucchesi 1998; Meller and Kuroda 2002). The Male Specific Lethal (MSL) complex, which releases paused RNA Polymerase into active elongation, mediates transcriptional upregulation by increasing histone H4 acetylation (Akhtar et al. 2000; Larschan et al. 2011b). As described earlier, the MSL2 core complex component is translationally repressed in females by SXL and is therefore only expressed in males (Bashaw and Baker 1997; Kelley et al. 1997). The MSL complex is a ribonucleoprotein complex composed of five proteins: MSL1, MSL2, MSL3 (Male-specific lethal 1,2, and 3), MOF (Males absent on the first), MLE (Maleless), and two lncRNAs: *roX1* and *roX2*, which are required for dosage compensation (Belote and Lucchesi 1980a; Meller 2003; Oh et al. 2003).

Although *roX1* and *roX2* differ in sequence and size (*roX1* is 3.7 kb and *roX2* is 0.5-1.2 kb), they are functionally redundant (Amrein and Axel 1997; Meller and

Rattner 2002b) (Smith et al. 2000). The presence of one *roX* RNA in the genome is required for MSL complex recruitment to the X-chromosome (Meller et al. 1997; Meller and Kuroda 2002; Meller and Rattner 2002b). However, unlike mammalian *Xist* RNA, *roX* RNAs can travel *in trans* to the X-chromosome from an autosomal location (Park et al. 2002). Furthermore, insertion of *roX* RNAs onto autosomes is sufficient to target MSL complex to autosomes leading to spreading of MSL complex and upregulation of autosomal genes within 200kb of the ectopically inserted *roX* gene (Kelley et al. 1999; Park et al. 2010). *roX1* and *roX2* RNAs are essential for MSL complex to spread from the initial sites of binding called Chromatin Entry Sites (CES) to regions of active transcription (Deng and Meller 2006). Although the precise mechanism by which *roX* RNAs distinguish the X-chromosome for MSL complex recruitment is unclear, it is thought MSL complex can be site-specifically targeted to the X-chromosome by co-transcriptionally incorporating the *roX* RNAs into the MSL complex (Park et al. 2002; Park et al. 2003; Li et al. 2008). Therefore, it is likely that the loci from which *roX* RNAs are transcribed are the initial platforms that target MSL complex to the X-chromosome. MSL complex spreads from these sites to coat the X-chromosome (see below)(Oh et al. 2003). This model for X-chromosome targeting is supported by the findings that *roX* RNAs localize to sites of MSL complex enrichment and the number of *roX* loci present and *roX* transcription levels alter both how much and how often MSL complex spreads (Park et al. 2002; Kelley et al. 2008; Gelbart and Kuroda 2009; Chu et al. 2011).

In summary, the studies of non-coding RNA function during mammalian and *Drosophila* dosage compensation have revealed the advantages of using lncRNAs as

regulatory factors to nucleate a chromatin domain on a single chromosome. lncRNAs have a short half-life and can site specifically target. The short half-life of some lncRNAs (e.g. 30 -60 min for *Tsix* (Lee 2012)) is likely to supply the maximum amount of RNA directly at the site of nucleation with minimal concern for diffusion, which could lead to mis-targeting. Since lncRNAs are restricted to specific sites in the genome, targeting of transcription complexes utilizing transcription of lncRNAs can occur locally and specifically at the site of lncRNA transcription (Derrien et al. 2012; Lee and Bartolomei 2013)(Batista and Chang 2013b; Batista and Chang 2013a).

DNA sequence elements that function during dosage compensation

Although lncRNAs play a central role in establishing nucleation centers where transcriptional regulation occurs, DNA elements are also critical for targeting to the appropriate sites. Dosage compensation in mammals, *Drosophila*, and worms utilize specific DNA sequences during the process of X-chromosome identification. In mammals, a tandem cluster of three of the binding motifs for the zinc finger protein YY1 is required for *Xist* RNA binding to Xi that nucleates the silencing process (Jeon and Lee 2011). Furthermore, long interspersed elements (LINES), enriched on the X-chromosome relative to autosomes facilitate the establishment of heterochromatin that forms the silent compartment of Xi (Chow et al. 2010). Transient expression of these LINES also appears to promote *Xist* mediated silencing of Xi (Chow et al. 2010).

In *Drosophila*, the MSL dosage compensation complex localizes most strongly to Chromatin Entry Sites (CES), also called High Affinity Sites (HAS), which are 1.5 kb regions that are sufficient to recruit MSL complex to autosomal regions (Dahlsveen et al. 2006; Alekseyenko et al. 2008; Straub et al. 2008). GA-rich motifs called MSL Recognition Elements (MREs) are present within almost all CES and are required for MSL complex recruitment across *Drosophila* species (Gallach et al. 2010). Recent work has indicated that MRE elements evolved from transposon insertions and by expansions of GA-rich sequences likely through replication errors (Ellison and Bachtrog 2013). MREs are enriched two-fold on the X-chromosome overall and approximately four-fold enriched at the 3' end of active genes (Alekseyenko et al. 2008). Although MREs are X-enriched, they are not X-specific sequences. Therefore, additional factors are required to increase X-specificity including a DNA binding protein that recognizes MREs (see protein-protein interactions section below).

Worms also utilize specific DNA motifs for identification of the X-chromosomes in hermaphrodites. Similar to *Drosophila*, which initially recruits MSL complex to CES that contain MREs followed by spreading in *cis*, *C.elegans* first recruit their dosage compensation complex (DCC), to recruitment elements on X (*rex* sites) that are found at only 200 sites on the X-chromosome (Jans et al. 2009; Meyer 2010). Clustered within the *rex* sites is a 12-bp consensus motif, MEX, which is enriched on the X-chromosome relative to autosomes and necessary for DCC recruitment to the X-chromosome (McDonel et al. 2006; Jans et al. 2009; Meyer 2010). Importantly, MEX motifs distinguish the *rex* sites as the regions to be

targeted (Meyer 2010). A second set of sites is important for recruiting the DCC; and these are the *dox* sites, (dependent on X)(McDonel et al. 2006; Jans et al. 2009; Meyer 2010). There are many more *dox* sites than *rex* sites and these sites do not have the MEX motif. However, DCC occupancy of *dox* sites requires binding to *rex* sites. While *rex* sites are mostly found in intergenic regions, *dox* sites are often located in actively transcribed promoters (Jans et al. 2009). *Rex* and *dox* sites are separated by 2-90kb distances suggesting that long-range interactions between the sites may facilitate dosage compensation across the whole chromosome (Meyer 2010). In summary, the use of specific DNA motifs in mammals, *Drosophila*, and *C. elegans* to recruit the dosage compensation machinery highlights the critical role of DNA elements in specifying domains of coordinate gene regulation (Xiao-Yong Li 2011).

The role of chromatin regulation during dosage compensation

While lncRNAs and DNA sequence elements are important for the initial identification of the X-chromosome, accessibility of these elements to dosage compensation complexes are regulated by chromatin (Guertin and Lis 2012). The genome-wide profiling of chromatin marks has revealed two types of chromatin marks: 1) Chromatin marks that are important for the identification of the X-chromosome; and 2) a chromatin mark that is associated with the dosage compensation process across species, H4K16ac (Fig 2).

The chromatin landscape of *Drosophila melanogaster* was analyzed on a genome-wide scale by comparing regions enriched for histone marks to domains with engaged RNA Polymerase II, lncRNAs, and DNA accessibility (Kharchenko et al. 2011; Guertin and Lis 2012). In this way, actively transcribed genes were found to have distinct chromatin features involving specific combinations of histone marks (Kharchenko et al. 2011). A computational model defined nine chromatin states distinguished by enrichment of histone marks associated either with activation such as H3K4me2/3 and H3K36me3 or repression such as H3K9me3 and H3K27me3. The MSL complex associates with the bodies of active genes and uses H3K36me3 as a means of spreading from CES to bodies of active genes (Larschan et al. 2007; Bell et al. 2008). The Polycomb repressive complex (PRC2), which facilitates Xi silencing in mammals, associates with H3K27me3 mark as a means of spreading from its initial sites of recruitment (Pinter et al. 2012)(Simon et al. 2013).

Although there are several chromatin marks that categorize active versus silent states, the H4K16ac mark is specifically associated with activation in the context of dosage compensation across species. This chromatin mark specifically disrupts inter-nucleosome interactions and therefore nucleosomes are more loosely packed in the presence of this mark (Dunlap et al. 2012)(Shogren-Knaak and Peterson 2006)(Shogren-Knaak et al. 2006).

In *Drosophila*, H4K16ac is deposited by the MOF protein that is part of the MSL dosage compensation complex (Smith et al. 2000; Gelbart et al. 2009; Conrad et al.

2012). The opening of chromatin structure by H4K16ac is thought to be involved in the mechanism of transcriptional upregulation by promoting the entry of RNA Polymerase II into gene bodies (Larschan et al. 2011b). Similarly in mammals, H4K16ac, deposited by MOF upon the X-chromosome, is associated with increased transcription of the single active X-chromosome (Deng X 2013). *C. elegans* also modulate H4K16ac during dosage compensation. However, it is depletion rather than enhancement of H4K16ac that correlates with the downregulation of the hermaphrodite X-chromosomes (Wells et al. 2012). Like *Drosophila* and mammals, this acetylation is regulated in tandem with a methylation mark. In *C.elegans*, this decrease in H4K16ac is coupled with enrichment of monomethylation of lysine 20 at histone 4, H4K20me1 (Wells et al. 2012).

Ubiquitylation of histones and other proteins can change the chromatin landscape by altering the chromatin environment. Recently, ubiquitination has been linked to the dosage compensation process. The Rnfl12 ubiquitin ligase stabilizes the *Xist* RNA by degrading Rex1, a repressor of *Xist* transcription (Gontan et al. 2012). Also the RING finger of MSL2, which is required to initiate the upregulation of the X-chromosome in male *Drosophila*, was recently identified to be a histone ubiquitin E3 ligase with specificity for both histone H2B and ubiquitylation of MSL complex components (Lipeng Wu 2011)(Villa et al. 2012).

Recently, it was determined that higher order chromatin structure also contributes to establishing a chromatin domain during the dosage compensation process

(Engreitz et al. 2013). The X-chromosome in mammals is arranged into 200 kilobase-1Mb topologically associated domains (TADS) that align with chromatin marks and coordinately regulated genes (Nora et al. 2012). Interference with TAD structure leads to misregulation of gene expression through disruption of chromosomal contacts (Nora et al. 2012). Early sites of X-chromosome identification after induction of the *Xist* gene in male cells are primarily regions that are topologically associated with each other (Engreitz et al. 2013). In summary, individual chromatin marks and higher order chromatin structure contribute to both the identification of the X-chromosome and its subsequent regulation for dosage compensation.

Interactions between proteins and chromatin that are involved in dosage compensation

In addition to a permissive chromatin landscape, interactions between components of the dosage compensation complexes and DNA are required to establish dosage compensation. Chromatin interacting proteins are required to integrate all of these signals and mediate the process and accurate targeting of dosage compensation across species. Zinc finger proteins are a large class of transcriptional regulators that are attractive targeting factors given their ability to sequence specifically bind to DNA.

In mammals, both an initial targeting step involving the YY1 zinc finger protein and chromatin altering proteins are involved in X-inactivation. The YY1 protein tethers *Xist* RNA to the XIC to establish the nucleation centers (Jeon and Lee 2011). Furthermore, the *Xist* RNA recruits the PRC2 chromatin remodeling complex and this recruitment is stabilized by the YY1 protein. After tethering to the XIC by YY1, PRC2 spreads to an initial 150 “canonical” sites, which contain CpG islands and a high occupancy of EZH2, the PRC2 enzymatic subunit that deposits the H3K27me3 mark (Pinter et al. 2012). PRC2 further spreads by binding to 4,000 “noncanonical sites” that are devoid of H3K4me3 and CpG islands (Pinter et al. 2012). Therefore, the function of DNA, lncRNAs, chromatin remodelers and zinc finger proteins are integrated during the process of X-chromosome inactivation.

In *Drosophila*, the initial stage of MSL complex targeting involves the MSL1 and MSL2 proteins forming a core protein complex that is sufficient to identify the initial binding sites called CES (Li et al. 2005). Recent work has shown that MSL1 dimerizes and MSL2 binds this dimer to form the core complex that first targets the CES that recruit MSL complex (Hallacli et al. 2012). However, none of the well-studied components of the MSL complex directly binds to DNA with sequence specificity (Fauth et al. 2010). However, a recently identified zinc finger protein, called CLAMP (Chromatin Linked Adaptor for MSL Protein) is required to recruit MSL complex to the CES sites (Larschan et al. 2012; Soruco et al. 2013). CLAMP binds directly to the MRE sequences that are required for MSL complex recruitment (Soruco et al. 2013), and associates with the MSL complex (Soruco et al. 2013; Wang

et al. 2013a). Similar to YY1, this factor can bind sequence specifically to DNA, and tether a dosage compensation complex to its target sites of regulation.

The *Drosophila* MSL complex includes three proteins that mediate the spreading process and interactions with chromatin and lncRNAs: MSL3, MLE, and MOF. MSL3 is a chromo domain containing protein that binds to nucleosomes and H3K36me3, a mark of active transcription associated with active gene bodies, thereby targeting MSL complex to active gene bodies (Buscaino et al. 2006; Sural et al. 2008). MLE is an ATP-dependent RNA/DNA helicase that unwinds short double-stranded RNA or DNA/RNA hybrids (Morra et al. 2008; Dunlap et al. 2012). MLE specifically interacts with *roX* RNAs via known interaction domains and facilitates the interaction between *roX* RNAs and the MSL2 core MSL complex component (Morra et al. 2008)(Ilik et al. 2013). MOF is the histone acetyltransferase that deposits the H4K16ac mark at active genes to facilitate upregulation of the male X-chromosome (Smith et al. 2000; Kind et al. 2008; Gelbart et al. 2009; Prestel et al. 2010b).

In *C. elegans*, the Dosage Compensation Complex (DCC) shares protein subunits with complexes that are involved in condensing chromatin (Pferdehirt et al. 2011). This suggests that the DCC works similarly to condensin complexes to restructure chromosomes as a means of facilitating transcriptional regulation. For example, DPY-27, a protein involved in condensing chromatin, associates with the X-chromosome as part of the DCC (Chuang et al. 1994). DPY-30, an integral subunit of the DCC, is crucial for recruiting DCC to the X-chromosome (Hsu and Meyer 1994;

Pferdehirt et al. 2011). However, DPY-30 exhibits similar levels of binding to the X-chromosomes and autosomes and also functions in histone modification as part of the activator MLL/COMPASS complex (Pferdehirt et al. 2011). In contrast, DPY-27, which is specific to the DCC, exhibits greater binding on X than autosomes (Pferdehirt et al. 2011). While the DCC binds at low levels to promoters of expressed genes, optimum binding of DCC to the X-chromosomes requires the SDC-2 and SDC-3 subunits (Pferdehirt et al. 2011). Additionally, SDC-3 is a zinc finger protein involved in assembling the DCC complex (Davis and Meyer 1997). It is not known how the MEX DNA sequences that are important for dosage compensation are specifically recognized. In summary, mammals, *Drosophila* and *C. elegans* all use a combination of zinc finger proteins and chromatin interacting proteins to promote X-chromosome identification and spreading along the length of the X-chromosome.

Conclusion

The dosage compensation process is a paradigm for how non-coding RNAs, DNA elements, chromatin landscape and protein factors function together to regulate gene expression. There are several common principles that we have highlighted above: 1) long-noncoding RNAs that act as spatial and temporal platforms to nucleate dosage compensation in both mammals and *Drosophila*; 2) specific DNA sequences are involved in X-chromosome identification across species; 3) Chromatin landscape marks that facilitate spreading across the X-chromosome such as H3K27me3 in mammals and H3K36me3 in *Drosophila*; 4) Protein factors

that associate with chromatin, DNA, and RNAs. H4K16ac has been implicated in the transcriptional regulatory process of dosage compensation across species. In mammals, higher order chromatin structure is thought to contribute to X-chromosome identification. Protein-protein interactions including tethering of dosage compensation complexes to X-chromosomes by zinc finger proteins also contribute to X-chromosome identification. To conclude, dosage compensation serves as a paradigm for understanding how chromatin domains are established within highly complex eukaryotic genomes.

References

- 1 Disteche, C. M. Dosage compensation of the sex chromosomes. *Annu Rev Genet* **46**, 537-560, doi:10.1146/annurev-genet-110711-155454 (2012).
- 2 Hattori, M. *et al.* The DNA sequence of human chromosome 21. *Nature* **405**, 311-319, doi:10.1038/35012518 (2000).
- 3 Jacobs, P. A., Baikie, A. G., Court Brown, W. M. & Strong, J. A. The somatic chromosomes in mongolism. *Lancet* **1**, 710 (1959).
- 4 Lejeune, J., Gautier, M. & Turpin, R. [Study of somatic chromosomes from 9 mongoloid children]. *Comptes rendus hebdomadaires des seances de l'Academie des sciences* **248**, 1721-1722 (1959).
- 5 Girirajan, S., Campbell, C. D. & Eichler, E. E. Human copy number variation and complex genetic disease. *Annu Rev Genet* **45**, 203-226, doi:10.1146/annurev-genet-102209-163544 (2011).
- 6 Beatriz Vicoso, D. B. Progress and prospects toward our understanding of the evolution of dosage compensation. (2009).
<<http://link.springer.com/article/10.1007/s10577-009-9053-y/fulltext.html>>.
- 7 Johnson, N. A. & Lachance, J. The genetics of sex chromosomes: evolution and implications for hybrid incompatibility. *Ann N Y Acad Sci* **1256**, E1-22, doi:10.1111/j.1749-6632.2012.06748.x (2012).
- 8 Mank, J. E. The W, X, Y and Z of sex-chromosome dosage compensation. *Trends in genetics : TIG* **25**, 226-233, doi:10.1016/j.tig.2009.03.005 (2009).
- 9 Lee, J. T. & Bartolomei, M. S. X-Inactivation, Imprinting, and Long Noncoding RNAs in Health and Disease. *Cell* **152**, 1308-1323, doi:S0092-8674(13)00205-5 [pii] 10.1016/j.cell.2013.02.016 (2013).
- 10 Belote, J. M. & Lucchesi, J. C. Control of X-chromosome transcription by the maleless gene in *Drosophila*. *Nature* **285**, 573-575 (1980).
- 11 Meller, V. H. & Kuroda, M. I. Sex and the single chromosome. *Advances in genetics* **46**, 1-24 (2002).
- 12 Meyer, B. J. Sex Determination and X-chromosome Dosage Compensation. doi:NBK20081 [bookaccession] (1997).
- 13 Kelley, R. L., Wang, J., Bell, L. & Kuroda, M. I. Sex lethal controls dosage compensation in *Drosophila* by a non-splicing mechanism. *Nature* **387**, 195-199, doi:10.1038/387195a0 (1997).
- 14 Alekseyenko, A. A. *et al.* A sequence motif within chromatin entry sites directs MSL establishment on the *Drosophila* X-chromosome. *Cell* **134**, 599-609, doi:10.1016/j.cell.2008.06.033 (2008).
- 15 Larschan, E. *et al.* MSL complex is attracted to genes marked by H3K36 trimethylation using a sequence-independent mechanism. *Molecular cell* **28**, 121-133, doi:10.1016/j.molcel.2007.08.011 (2007).

- 16 Pinter, S. F. *et al.* Spreading of X-chromosome inactivation via a hierarchy of defined Polycomb stations. *Genome research* **22**, 1864-1876, doi:10.1101/gr.133751.111 (2012).
- 17 Csankovszki, G., McDonel, P. & Meyer, B. J. Recruitment and spreading of the *C. elegans* dosage compensation complex along X-chromosomes. *Science* **303**, 1182-1185, doi:10.1126/science.1092938 (2004).
- 18 Deng, X. *et al.* Evidence for compensatory upregulation of expressed X-linked genes in mammals, *Caenorhabditis elegans* and *Drosophila melanogaster*. *Nature genetics* **43**, 1179-1185, doi:ng.948 [pii] 10.1038/ng.948 (2011).
- 19 Deng X, B. J., Ma W, Nguyen DK, Hiatt JB, Noble WS, Shendure J, Disteche CM. Mammalian X Upregulation Is Associated with Enhanced Transcription Initiation, RNA Half-Life, and MOF-Mediated H4K16 Acetylation. *Developmental Cell* **25**, 1-14 (2013).
- 20 Schulz, E. G. & Heard, E. Role and control of X-chromosome dosage in mammalian development. *Current opinion in genetics & development*, doi:S0959-437X(13)00018-X [pii] 10.1016/j.gde.2013.01.008 (2013).
- 21 Hsu, D. R. & Meyer, B. J. The dpy-30 gene encodes an essential component of the *Caenorhabditis elegans* dosage compensation machinery. *Genetics* **137**, 999-1018 (1994).
- 22 Heard, E. & Turner, J. Function of the sex chromosomes in mammalian fertility. *Cold Spring Harbor perspectives in biology* **3**, a002675, doi:10.1101/cshperspect.a002675 (2011).
- 23 Leppig, K. A. & Disteche, C. M. Ring X and other structural X-chromosome abnormalities: X inactivation and phenotype. *Seminars in reproductive medicine* **19**, 147-157, doi:10.1055/s-2001-15395 (2001).
- 24 Herpin, A. & Schartl, M. Sex determination: switch and suppress. *Current biology : CB* **21**, R656-659, doi:S0960-9822(11)00825-6 [pii]10.1016/j.cub.2011.07.026 (2011).
- 25 Sekido, R. & Lovell-Badge, R. Sex determination and SRY: down to a wink and a nudge? *Trends in genetics : TIG* **25**, 19-29, doi:10.1016/j.tig.2008.10.008 (2009).
- 26 Salz, H. K. & Erickson, J. W. Sex determination in *Drosophila*: The view from the top. *Fly (Austin)* **4**, 60-70, doi:11277 [pii] (2010).
- 27 Meyer, B. J. Targeting X-chromosomes for repression. *Current opinion in genetics & development* **20**, 179-189, doi:S0959-437X(10)00042-0 [pii] 10.1016/j.gde.2010.03.008 (2010).
- 28 Yonker, S. A. & Meyer, B. J. Recruitment of *C. elegans* dosage compensation proteins for gene-specific versus chromosome-wide repression. *Development* **130**, 6519-6532, doi:10.1242/dev.00886 (2003).
- 29 Batista, P. J. & Chang, H. Y. Cytotopic localization by long noncoding RNAs. *Current opinion in cell biology* **25**, 195-199, doi:10.1016/j.ceb.2012.12.001 (2013).
- 30 Batista, P. J. & Chang, H. Y. Long Noncoding RNAs: Cellular Address Codes in Development and Disease. *Cell* **152**, 1298-1307, doi:10.1016/j.cell.2013.02.012 (2013).

- 31 Rinn, J. L. & Chang, H. Y. Genome regulation by long noncoding RNAs. *Annual review of biochemistry* **81**, 145-166, doi:10.1146/annurev-biochem-051410-092902 (2012).
- 32 Thomas Cremer, M. C. Chromosome Territories. *Cold Spring Harbor perspectives in biology* (2010).
- 33 Derrien, T. *et al.* The GENCODE v7 catalog of human long noncoding RNAs: analysis of their gene structure, evolution, and expression. *Genome research* **22**, 1775-1789, doi:10.1101/gr.132159.111 (2012).
- 34 Dundr, M. & Misteli, T. Biogenesis of nuclear bodies. *Cold Spring Harbor perspectives in biology* **2**, a000711, doi:10.1101/cshperspect.a000711 (2010).
- 35 Lee, J. T. Epigenetic regulation by long noncoding RNAs. *Science* **338**, 1435-1439, doi:338/6113/1435 [pii] 10.1126/science.1231776 (2012).
- 36 Lusk, R. W. & Eisen, M. B. Spatial promoter recognition signatures may enhance transcription factor specificity in yeast. *PloS one* **8**, e53778, doi:10.1371/journal.pone.0053778 (2013).
- 37 Fisher, W. W. *et al.* DNA regions bound at low occupancy by transcription factors do not drive patterned reporter gene expression in *Drosophila*. *Proceedings of the National Academy of Sciences of the United States of America* **109**, 21330-21335, doi:10.1073/pnas.1209589110 (2012).
- 38 Guertin, M. J. & Lis, J. T. Mechanisms by which transcription factors gain access to target sequence elements in chromatin. *Current opinion in genetics & development*, doi:10.1016/j.gde.2012.11.008 (2012).
- 39 Thomas, S. *et al.* Dynamic reprogramming of chromatin accessibility during *Drosophila* embryo development. *Genome biology* **12**, R43, doi:10.1186/gb-2011-12-5-r43 (2011).
- 40 Engreitz, J. M. *et al.* The Xist lncRNA exploits three-dimensional genome architecture to spread across the X-chromosome. *Science* **341**, 1237973, doi:10.1126/science.1237973 (2013).
- 41 Nora, E. P. *et al.* Spatial partitioning of the regulatory landscape of the X-inactivation centre. *Nature* **485**, 381-385, doi:10.1038/nature11049 (2012).
- 42 Orom, U. A. *et al.* Long noncoding RNAs with enhancer-like function in human cells. *Cell* **143**, 46-58, doi:10.1016/j.cell.2010.09.001 (2010).
- 43 Lai, F. *et al.* Activating RNAs associate with Mediator to enhance chromatin architecture and transcription. *Nature* **494**, 497-501, doi:10.1038/nature11884 (2013).
- 44 Kelley, R. L. *et al.* Epigenetic spreading of the *Drosophila* dosage compensation complex from roX RNA genes into flanking chromatin. *Cell* **98**, 513-522 (1999).
- 45 Park, S. W., Oh, H., Lin, Y. R. & Park, Y. MSL cis-spreading from roX gene up-regulates the neighboring genes. *Biochemical and biophysical research communications* **399**, 227-231, doi:10.1016/j.bbrc.2010.07.059 (2010).
- 46 Ko-Hsuan Hung, Y. W., Jin Crystal Zhao. Regulation of Mammalian Gene Dosage by Long Non-coding RNAs. *Biomolecules* **3(1)**, 124-142 (2013).

- 47 Zhao, J., Sun, B. K., Erwin, J. A., Song, J. J. & Lee, J. T. Polycomb proteins targeted by a short repeat RNA to the mouse X-chromosome. *Science* **322**, 750-756, doi:10.1126/science.1163045 (2008).
- 48 Lucchesi, J. C. Dosage compensation in flies and worms: the ups and downs of X-chromosome regulation. *Current opinion in genetics & development* **8**, 179-184 (1998).
- 49 Akhtar, A., Zink, D. & Becker, P. B. Chromodomains are protein-RNA interaction modules. *Nature* **407**, 405-409, doi:10.1038/35030169 (2000).
- 50 Larschan, E. *et al.* X-chromosome dosage compensation via enhanced transcriptional elongation in *Drosophila*. *Nature* **471**, 115-118, doi:10.1038/nature09757 (2011).
- 51 Bashaw, G. J. & Baker, B. S. The regulation of the *Drosophila* msl-2 gene reveals a function for Sex-lethal in translational control. *Cell* **89**, 789-798 (1997).
- 52 Meller, V. H. Initiation of dosage compensation in *Drosophila* embryos depends on expression of the roX RNAs. *Mechanisms of development* **120**, 759-767 (2003).
- 53 Oh, H., Park, Y. & Kuroda, M. I. Local spreading of MSL complexes from roX genes on the *Drosophila* X-chromosome. *Genes & development* **17**, 1334-1339, doi:10.1101/gad.1082003 (2003).
- 54 Amrein, H. & Axel, R. Genes expressed in neurons of adult male *Drosophila*. *Cell* **88**, 459-469 (1997).
- 55 Meller, V. H. & Rattner, B. P. The roX genes encode redundant male-specific lethal transcripts required for targeting of the MSL complex. *The EMBO journal* **21**, 1084-1091, doi:10.1093/emboj/21.5.1084 (2002).
- 56 Smith, E. R. *et al.* The *drosophila* MSL complex acetylates histone H4 at lysine 16, a chromatin modification linked to dosage compensation. *Molecular and cellular biology* **20**, 312-318 (2000).
- 57 Meller, V. H., Wu, K. H., Roman, G., Kuroda, M. I. & Davis, R. L. roX1 RNA paints the X-chromosome of male *Drosophila* and is regulated by the dosage compensation system. *Cell* **88**, 445-457 (1997).
- 58 Park, Y., Kelley, R. L., Oh, H., Kuroda, M. I. & Meller, V. H. Extent of chromatin spreading determined by roX RNA recruitment of MSL proteins. *Science* **298**, 1620-1623, doi:10.1126/science.1076686 (2002).
- 59 Deng, X. & Meller, V. H. roX RNAs are required for increased expression of X-linked genes in *Drosophila melanogaster* males. *Genetics* **174**, 1859-1866, doi:10.1534/genetics.106.064568 (2006).
- 60 Park, Y. *et al.* Sequence-specific targeting of *Drosophila* roX genes by the MSL dosage compensation complex. *Molecular cell* **11**, 977-986 (2003).
- 61 Li, F., Schiemann, A. H. & Scott, M. J. Incorporation of the noncoding roX RNAs alters the chromatin-binding specificity of the *Drosophila* MSL1/MSL2 complex. *Molecular and cellular biology* **28**, 1252-1264, doi:10.1128/MCB.00910-07 (2008).

- 62 Chu, C., Qu, K., Zhong, F. L., Artandi, S. E. & Chang, H. Y. Genomic maps of long noncoding RNA occupancy reveal principles of RNA-chromatin interactions. *Molecular cell* **44**, 667-678, doi:10.1016/j.molcel.2011.08.027 (2011).
- 63 Gelbart, M. E. & Kuroda, M. I. Drosophila dosage compensation: a complex voyage to the X-chromosome. *Development* **136**, 1399-1410, doi:10.1242/dev.029645 (2009).
- 64 Kelley, R. L., Lee, O. K. & Shim, Y. K. Transcription rate of noncoding roX1 RNA controls local spreading of the Drosophila MSL chromatin remodeling complex. *Mechanisms of development* **125**, 1009-1019, doi:10.1016/j.mod.2008.08.003 (2008).
- 65 Jeon, Y. & Lee, J. T. YY1 tethers Xist RNA to the inactive X nucleation center. *Cell* **146**, 119-133, doi:S0092-8674(11)00668-4 [pii] 10.1016/j.cell.2011.06.026 (2011).
- 66 Chow, J. C. *et al.* LINE-1 activity in facultative heterochromatin formation during X-chromosome inactivation. *Cell* **141**, 956-969, doi:10.1016/j.cell.2010.04.042 (2010).
- 67 Dahlsveen, I. K., Gilfillan, G. D., Shelest, V. I., Lamm, R. & Becker, P. B. Targeting determinants of dosage compensation in Drosophila. *PLoS genetics* **2**, e5, doi:10.1371/journal.pgen.0020005 (2006).
- 68 Straub, T., Grimaud, C., Gilfillan, G. D., Mitterweger, A. & Becker, P. B. The chromosomal high-affinity binding sites for the Drosophila dosage compensation complex. *PLoS genetics* **4**, e1000302, doi:10.1371/journal.pgen.1000302 (2008).
- 69 Gallach, M., Arnau, V., Aldecoa, R. & Marin, I. A sequence motif enriched in regions bound by the Drosophila dosage compensation complex. *BMC genomics* **11**, 169, doi:10.1186/1471-2164-11-169 (2010).
- 70 Ellison, C. E. & Bachtrog, D. Dosage compensation via transposable element mediated rewiring of a regulatory network. *Science* **342**, 846-850, doi:10.1126/science.1239552 (2013).
- 71 Jans, J. *et al.* A condensin-like dosage compensation complex acts at a distance to control expression throughout the genome. *Genes & development* **23**, 602-618, doi:10.1101/gad.1751109 (2009).
- 72 McDonel, P., Jans, J., Peterson, B. K. & Meyer, B. J. Clustered DNA motifs mark X-chromosomes for repression by a dosage compensation complex. *Nature* **444**, 614-618, doi:10.1038/nature05338 (2006).
- 73 Xiao-Yong Li, S. T., Peter J Sabo, Michael B Eisen, John A Stamatoyannopoulos and Mark D Biggin. The role of chromatin accessibility in directing the widespread, overlapping patterns of Drosophila transcription factor binding. *Genome biology* **12** (2011).
- 74 Kharchenko, P. V. *et al.* Comprehensive analysis of the chromatin landscape in Drosophila melanogaster. *Nature* **471**, 480-485, doi:10.1038/nature09725 (2011).

- 75 Bell, O. *et al.* Transcription-coupled methylation of histone H3 at lysine 36 regulates dosage compensation by enhancing recruitment of the MSL complex in *Drosophila melanogaster*. *Molecular and cellular biology* **28**, 3401-3409, doi:10.1128/MCB.00006-08 (2008).
- 76 Simon, M. D. *et al.* High-resolution Xist binding maps reveal two-step spreading during X-chromosome inactivation. *Nature*, doi:10.1038/nature12719 (2013).
- 77 Dunlap, D. *et al.* Distinct contributions of MSL complex subunits to the transcriptional enhancement responsible for dosage compensation in *Drosophila*. *Nucleic acids research* **40**, 11281-11291, doi:10.1093/nar/gks890 (2012).
- 78 Shogren-Knaak, M. & Peterson, C. L. Switching on chromatin: mechanistic role of histone H4-K16 acetylation. *Cell Cycle* **5**, 1361-1365 (2006).
- 79 Shogren-Knaak, M. *et al.* Histone H4-K16 acetylation controls chromatin structure and protein interactions. *Science* **311**, 844-847, doi:10.1126/science.1124000 (2006).
- 80 Conrad, T. *et al.* The MOF chromobarrel domain controls genome-wide H4K16 acetylation and spreading of the MSL complex. *Dev Cell* **22**, 610-624, doi:10.1016/j.devcel.2011.12.016 (2012).
- 81 Gelbart, M. E., Larschan, E., Peng, S., Park, P. J. & Kuroda, M. I. *Drosophila* MSL complex globally acetylates H4K16 on the male X-chromosome for dosage compensation. *Nature structural & molecular biology* **16**, 825-832, doi:10.1038/nsmb.1644 (2009).
- 82 Wells, M. B., Snyder, M. J., Custer, L. M. & Csankovszki, G. *Caenorhabditis elegans* dosage compensation regulates histone H4 chromatin state on X-chromosomes. *Molecular and cellular biology* **32**, 1710-1719, doi:10.1128/MCB.06546-11 (2012).
- 83 Gontan, C. *et al.* RNF12 initiates X-chromosome inactivation by targeting REX1 for degradation. *Nature* **485**, 386-390, doi:10.1038/nature11070 (2012).
- 84 Lipeng Wu, B. M. Z., Yanming Wang, Benjamin A. Garcia, Yali Dou. The RING Finger Protein MSL2 in the MOF Complex Is an E3 Ubiquitin Ligase for H2B K34 and Is Involved in Crosstalk with H3 K4 and K79 Methylation. *Molecular cell* **43**, 132-144 (2011).
- 85 Villa, R. *et al.* MSL2 combines sensor and effector functions in homeostatic control of the *Drosophila* dosage compensation machinery. *Molecular cell* **48**, 647-654, doi:10.1016/j.molcel.2012.09.012 (2012).
- 86 Li, F., Parry, D. A. & Scott, M. J. The amino-terminal region of *Drosophila* MSL1 contains basic, glycine-rich, and leucine zipper-like motifs that promote X-chromosome binding, self-association, and MSL2 binding, respectively. *Molecular and cellular biology* **25**, 8913-8924, doi:10.1128/MCB.25.20.8913-8924.2005 (2005).
- 87 Hallacli, E. *et al.* Msl1-mediated dimerization of the dosage compensation complex is essential for male X-chromosome regulation in *Drosophila*. *Molecular cell* **48**, 587-600, doi:10.1016/j.molcel.2012.09.014 (2012).

- 88 Fauth, T., Muller-Planitz, F., Konig, C., Straub, T. & Becker, P. B. The DNA binding CXC domain of MSL2 is required for faithful targeting the Dosage Compensation Complex to the X-chromosome. *Nucleic acids research* **38**, 3209-3221, doi:10.1093/nar/gkq026 (2010).
- 89 Larschan, E. *et al.* Identification of chromatin-associated regulators of MSL complex targeting in Drosophila dosage compensation. *PLoS genetics* **8**, e1002830, doi:10.1371/journal.pgen.1002830 (2012).
- 90 Soruco, M. M. *et al.* The CLAMP protein links the MSL complex to the X-chromosome during Drosophila dosage compensation. *Genes & development* **27**, 1551-1556, doi:10.1101/gad.214585.113 (2013).
- 91 Wang, C. I. *et al.* Chromatin proteins captured by ChIP-mass spectrometry are linked to dosage compensation in Drosophila. *Nature structural & molecular biology* **20**, 202-209, doi:10.1038/nsmb.2477 (2013).
- 92 Sural, T. H. *et al.* The MSL3 chromodomain directs a key targeting step for dosage compensation of the Drosophila melanogaster X-chromosome. *Nature structural & molecular biology* **15**, 1318-1325, doi:10.1038/nsmb.1520 (2008).
- 93 Buscaino, A., Legube, G. & Akhtar, A. X-chromosome targeting and dosage compensation are mediated by distinct domains in MSL-3. *EMBO reports* **7**, 531-538, doi:10.1038/sj.embor.7400658 (2006).
- 94 Morra, R., Smith, E. R., Yokoyama, R. & Lucchesi, J. C. The MLE subunit of the Drosophila MSL complex uses its ATPase activity for dosage compensation and its helicase activity for targeting. *Molecular and cellular biology* **28**, 958-966, doi:10.1128/MCB.00995-07 (2008).
- 95 Ilik, I. A. *et al.* Tandem stem-loops in roX RNAs act together to mediate X-chromosome dosage compensation in Drosophila. *Molecular cell* **51**, 156-173, doi:10.1016/j.molcel.2013.07.001 (2013).
- 96 Prestel, M., Feller, C., Straub, T., Mitlohner, H. & Becker, P. B. The activation potential of MOF is constrained for dosage compensation. *Molecular cell* **38**, 815-826, doi:10.1016/j.molcel.2010.05.022 (2010).
- 97 Kind, J. *et al.* Genome-wide analysis reveals MOF as a key regulator of dosage compensation and gene expression in Drosophila. *Cell* **133**, 813-828, doi:10.1016/j.cell.2008.04.036 (2008).
- 98 Pferdehirt, R. R., Kruesi, W. S. & Meyer, B. J. An MLL/COMPASS subunit functions in the C. elegans dosage compensation complex to target X-chromosomes for transcriptional regulation of gene expression. *Genes & development* **25**, 499-515, doi:10.1101/gad.2016011 (2011).
- 99 Chuang, P. T., Albertson, D. G. & Meyer, B. J. DPY-27: a chromosome condensation protein homolog that regulates C. elegans dosage compensation through association with the X-chromosome. *Cell* **79**, 459-474 (1994).
- 100 Davis, T. L. & Meyer, B. J. SDC-3 coordinates the assembly of a dosage compensation complex on the nematode X-chromosome. *Development* **124**, 1019-1031 (1997).

Figure 1. lncRNA transcription generates an environment favorable for coordinate regulation

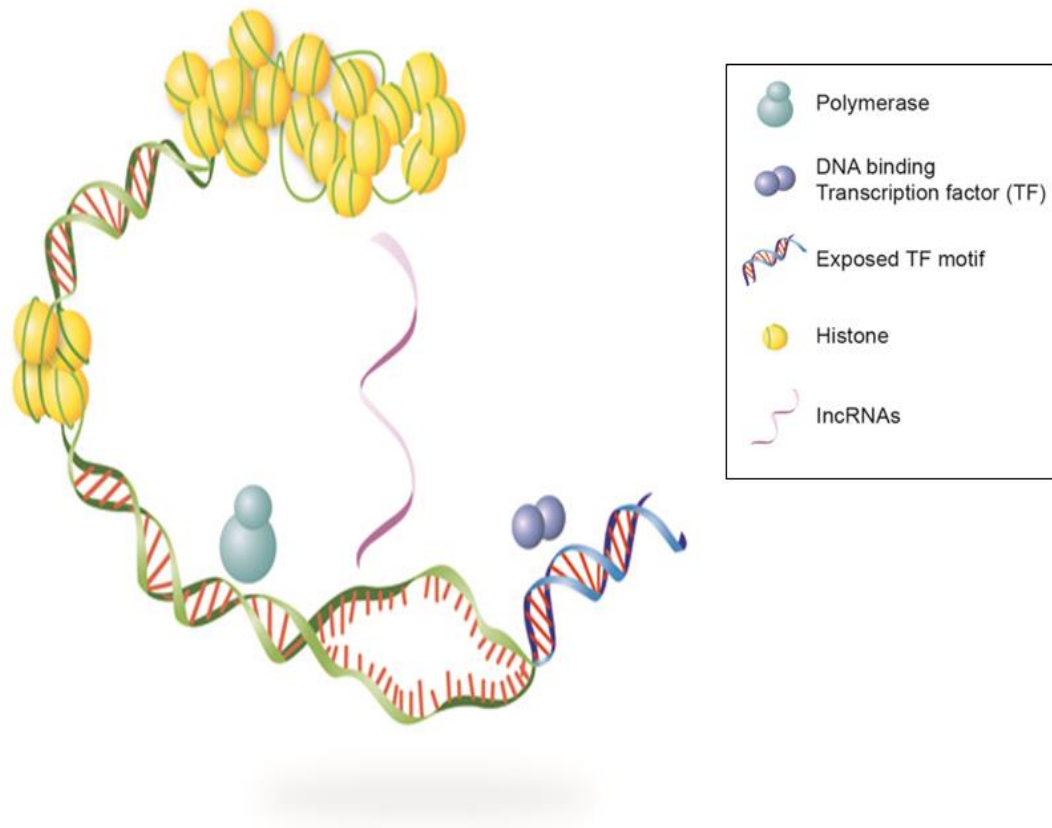


Figure 1. Recruitment of RNA polymerase for lncRNA transcription leads to a more open chromatin state. Nucleosomes are displaced and DNA is made more accessible, exposing transcription factor motifs for DNA binding. Once transcription factors bind to their motifs, this serves as a signal for regulation through recruitment of other co-activators and complex assembly. lncRNAs facilitate the process through their own transcription, sequestration of heterochromatin, creation of pockets of activity in the nucleus, and scaffolding different components into closer proximity to facilitate interactions favorable for regulation.

Figure 2. MOF deposits H4K16acetylation, de-condensing chromatin to favor the X-chromosome for active transcription

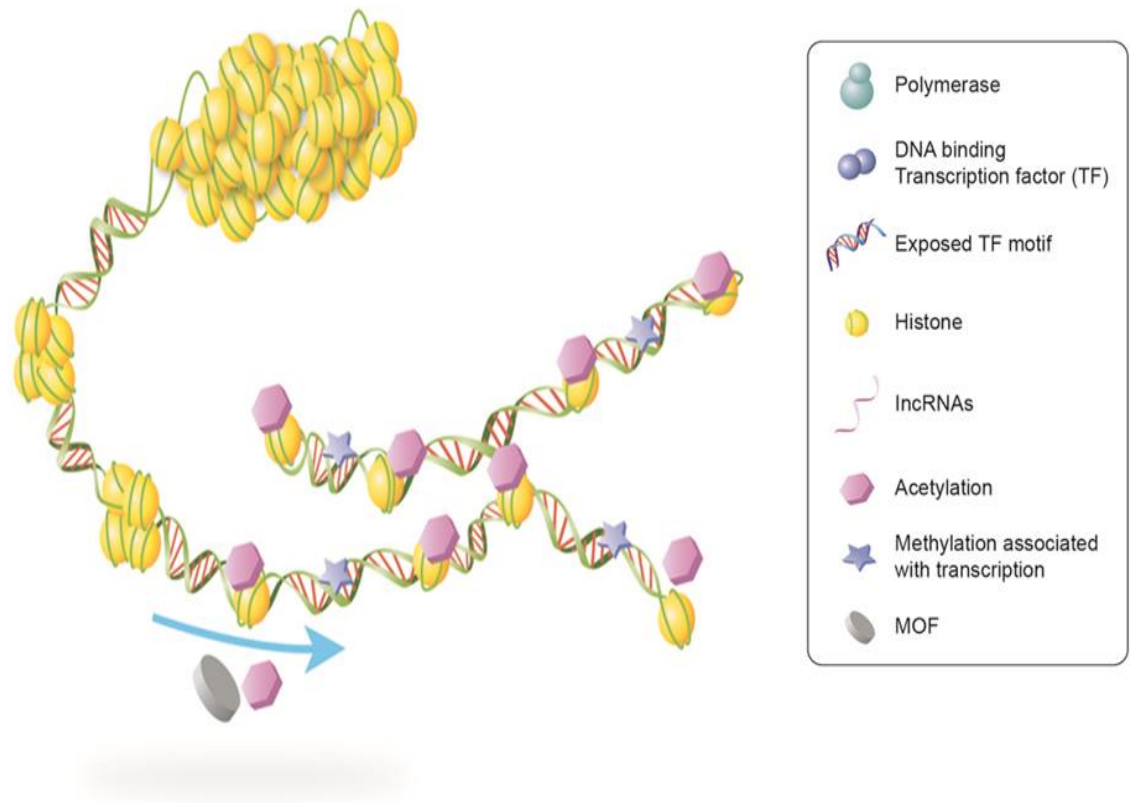


Figure 2. As nucleosomes are displaced and histones are acetylated, specifically H4K16acetylation in X-chromosome dosage compensation primes, the X-chromosome is primed for active transcription and upregulation.

Chapter 2: Identification of Chromatin-Associated Regulators of MSL Complex
Targeting in *Drosophila* Dosage Compensation

This chapter of my thesis dissertation as it appears in the following pages is a published manuscript for which I generated **Figure Supplemental 2**. My specific contributions to this paper included performing western blot analysis of the target protein, CLAMP, to confirm expression of the protein in both *Drosophila* male (SL2) and female (KC) cells. These Western blots also confirmed that CLAMP RNAi treatment was feasible and successful and validated the newly generated antibody for CLAMP. Our research group discovered CLAMP and I therefore validated our CLAMP antibody to be used for subsequent experiments, such as ChIP (Chromatin Immunoprecipitation) and immunostains.

After doing nuclear extraction of *Drosophila* male and female cells along with CLAMP RNAi treatment, I performed Western blots using CLAMP antibody. Westerns on whole cell and nuclear extracts with and without CLAMP RNAi demonstrated that CLAMP is a nuclear protein and CLAMP RNAi reduces CLAMP protein levels. Identifying CLAMP as a nuclear protein provided validation for identifying the protein as a transcription factor. Comparing Western blots of nuclear extracts from protein samples of *Drosophila* male and female cells indicated that CLAMP levels are similar in both males and females. These results allowed us to ascertain the effects that are due in particular to lack of CLAMP in MSL Chip and Chip-seq experiments.

**Identification of chromatin-associated regulators of MSL complex targeting in
Drosophila dosage compensation**

Erica Larschan, Marcela M.L. Soruco, Ok-Kyung Lee, Shouyong Peng, Eric Bishop,
Jessica Chery, Karen Goebel, Jessica Feng, Peter J. Park, and Mitzi I. Kuroda, 2012,
PLoS Genet 8(7): e10002830

Abstract

Sex chromosome dosage compensation in *Drosophila* provides a model for understanding how chromatin organization can modulate coordinate gene regulation. Male *Drosophila* increase the transcript levels of genes on the single male X approximately two-fold to equal the gene expression in females, which have two X-chromosomes. Dosage compensation is mediated by the Male-Specific Lethal (MSL) histone acetyltransferase complex. Five core components of the MSL complex were identified by genetic screens for genes that are specifically required for male viability, and dispensible for females. However, because dosage compensation must interface with the general transcriptional machinery, it is likely that identifying additional regulators that are not strictly male-specific will be key to understanding the process at a mechanistic level. Such regulators would not have been recovered from previous male-specific lethal screening strategies. Therefore, we have performed a cell culture-based, genome-wide RNAi screen to search for factors required for MSL targeting or function. Here we focus on the discovery of proteins that function to promote MSL complex recruitment to “chromatin entry sites”, which are proposed to be the initial sites of MSL targeting. We find that components of the NSL (Non-specific lethal) complex, and a previously unstudied zinc-finger protein, facilitate MSL targeting and display a striking enrichment at MSL entry sites. Identification of these factors provides new insight into how MSL complex establishes the specialized hyperactive chromatin required for dosage compensation in *Drosophila*.

Author Summary

Gene regulation is essential to all living things. For example, levels of gene expression in individual cells must be fine-tuned during development and in response to changing environmental conditions. Genes are regulated by DNA binding proteins and by factors that influence DNA packaging into chromatin. The MSL complex in *Drosophila melanogaster* is a chromatin-modifying complex that specifically regulates a large number of genes. The MSL complex targets active genes on the single male X chromosome to upregulate their output to match both female X chromosomes. How the MSL complex specifically targets the X and upregulates active genes is only partially understood. In order to increase our understanding of gene regulation at a mechanistic level, we performed a genome-wide genetic screen in male cells to identify factors that facilitate MSL targeting and function. Our results identify two chromatin-associated protein complexes and a new candidate DNA binding protein as key factors in MSL-based regulation. We also provide an extensive list of additional candidate genes to be examined in future studies.

Introduction

X-chromosome dosage compensation in male *Drosophila* provides a model for understanding how a large number of diversely regulated genes along the length of a single chromosome can be targeted for coordinate regulation. In *Drosophila*, the transcript levels of genes along the length of the single male X-chromosome are upregulated approximately two-fold, a process mediated by the MSL (Male-specific lethal) histone acetyltransferase complex (Gelbart and Kuroda 2009). A model, initially based on genetic observations (Kelley et al. 1999), posits that the MSL complex binds the X-chromosome in a two-step process. First, high levels of the MSL complex accumulate at approximately 150-300 “chromatin entry sites” (CES) containing GA-rich MRE (MSL Recognition Element) sequences that are ~two-fold enriched on the male X (Alekseyenko et al. 2008; Straub et al. 2008). Second, a sequence-independent spreading occurs in which the MSL complex associates with the bodies of active X-linked genes via general features associated with transcription, including H3K36me3 (Larschan et al. 2007). MSL complex increases transcript levels of its target genes by increasing the density of RNA Polymerase II (RNAP II) over transcription units, likely via MSL-dependent H4K16 acetylation (Larschan et al. 2011a).

While significant progress has been made towards understanding how MSL complex is targeted and functions, much remains to be understood. For example, since MREs are less than two-fold enriched on the X-chromosome (Alekseyenko et al.

2008) and the core MSL complex does not appear to have sequence-specific DNA binding activity (Fauth et al. 2010), it is likely that additional factors are involved in the key process of CES recognition.

Seminal genetic screens identified five core components of the MSL complex by isolating genes that were specifically required for male viability (Belote and Lucchesi 1980b; Hilfiker et al. 1997). However, these approaches would not recover a potentially key class of regulators that might be required for viability in both sexes, and act in dosage compensation by providing an interface with the core transcriptional machinery. In fact, based on experiments from budding yeast (Carrozza et al. 2005; Keogh et al. 2005) the essential SET2 H3K36 methyltransferase was identified as a potential regulator of MSL complex association with active genes (Larschan et al. 2007; Bell et al. 2008). SET2 contributes to targeting of the MSL complex to the bodies of active genes (Larschan et al. 2007; Bell et al. 2008), but it is likely that other factors also participate. Furthermore, there are no candidates for direct binding to the MRE sequence, which is proposed to be a key first step in MSL targeting.

Therefore, we designed a cell-based reporter system that specifically monitors MSL function at chromatin entry sites, and conducted a genome-wide RNAi screen to identify genes required to regulate this reporter. We identified the NSL (Mendjan et al. 2006; Prestel et al. 2010a) and PAF (Shi et al. 1996; Nagaike et al. 2011) general transcriptional regulators, as well as a previously unstudied zinc finger protein,

CG1832, that emerges as a strong candidate to function as the previously unknown link between MSL complex and MRE sequences within chromatin entry sites.

Results

A genome-wide RNAi screen identifies genes that regulate a MSL-dependent reporter in male SL2 cells

In order to identify genes that regulate MSL complex recruitment or function at CES loci, we performed a genome-wide RNAi screen using an MSL-dependent cell-based reporter system for which we previously reported functionality (Alekseyenko et al. 2008) (Fig. 1A). This reporter system includes the following three elements: 1) promoter of *roX2* gene (370 bp); 2) Firefly luciferase reporter gene; 3) MSL complex binding site for *roX2* (DNase I Hypersensitivity Site: 280 bp). The system is based on a similar one that functions as a male-specific MSL-dependent reporter *in vivo* (Bai et al. 2004). A genome-wide RNAi library generated by the *Drosophila* RNAi Screening Center (DRSC) that contains approximately 21,000 dsRNAs was screened in duplicate (www.flyrnai.org). To eliminate off-target effects, secondary screens were conducted using a set of validation RNAi constructs in addition to rescreening the original constructs.

In addition to direct regulators of MSL complex recruitment or function, several additional types of genes could be identified by our RNAi screen. Therefore, we

used the following steps to identify potential direct activators of MSL complex targeting: 1) To reduce the number of non-specific regulators of luciferase gene transcription and protein stability that were identified, we calculated the ratio of activity from our MSL-dependent Firefly luciferase reporter to a heterologous Renilla luciferase reporter. 2) We used multiple MSL-dependent reporters to further narrow the list of genes to those that function independently of the *roX2* promoter and specific CES used in the genome-wide screen; 3) We then used known functional information about the genes to narrow the list to those that have potential direct activating roles; 4) We determined whether new proteins affect MSL complex recruitment at endogenous sites without altering transcription of MSL complex components; 5) We performed genome-wide localization analysis to determine whether new proteins are enriched at CES loci.

As briefly described above, we first normalized the MSL-dependent Firefly luciferase reporter activity to an MSL-independent reporter including the Renilla luciferase gene that can be independently assayed using a different substrate. This dual Firefly/Renilla approach also controls for technical variation in transfection efficiencies among wells and differences in cell number caused by altered viability after RNAi treatments. We used a Renilla construct that was regulated by an Actin promoter as previously described (Nybakken and Perrimon 2002). Therefore, by calculating the ratio of Firefly/Renilla luciferase activity, RNAi conditions that specifically alter the MSL-dependent reporter but not the Actin-Renilla construct could be identified.

To conduct the RNAi screen, sixty-two screening plates containing approximately 21,000 dsRNAs arrayed in individual wells of a 384-well plate were assayed in duplicate. Each plate contained two positive control (*msl2* RNAi) and two negative control (GFP RNAi) constructs, that were previously validated. After confirming that at least a 5-fold dynamic range was observed in the expected activity of positive and negative controls on each plate, the Firefly/Renilla ratios of the two plates were averaged and screening hits were identified relative to the median Firefly/Renilla ratio for each plate (See Methods). We also conducted the same computational analysis for the Renilla values and removed hits in which the Renilla activity was dramatically altered by the RNAi treatment (See Methods).

Our screening approach identified 322 RNAi constructs that altered MSL-dependent reporter activity (Table S1). Importantly, we found all known components of the MSL complex in an unbiased way, validating our screen (Fig. 1B). We identified 254 RNAi constructs that reduced MSL-dependent reporter activity and 68 RNAi constructs that increased reporter activity. Gene Ontology analysis revealed several functional categories that were enriched including the following: protein synthesis, cell cycle control, and transcriptional regulation ($p < 0.05$) (Table 1). Similar analysis of publically available data sets from other RNAi screens performed using the same library indicated that the transcriptional regulation Gene Ontology category was not enriched in any other publically available screens. In contrast, the

protein synthesis category and cell cycle control categories were commonly enriched in the majority of screens conducted with this RNAi library.

Next, candidate genes were chosen for further analysis and validation with additional RNAi constructs. Three classes of genes were removed from the original list of candidates: 1) Ribosomal protein genes that are identified by many RNAi screens (41 genes); 2) Genes with a large number of predicted RNAi off targets using 17 bp as an overlap cutoff (Kulkarni et al. 2006) (19 genes) 3) RNAi amplicons that did not target a specific gene and/or target multiple predicted genes (21 amplicons). We rescreened the 241 remaining candidates using both original and validation RNAi constructs with the original screening system and two additional MSL-dependent reporters that we previously validated (Alekseyenko et al. 2008). These additional MSL-dependent reporters had a different promoter or CES compared with the original RNAi screening construct and were previously described (Fig. 1A). In this way, we identified 72 genes that were validated with an additional RNAi construct and regulated all three MSL-dependent reporters (Table S2). Notably, we did not identify the genes encoding HP1, SU(VAR)3-7, or ISWI, which each have a mutant phenotype that differentially affects the polytene X chromosome (Deuring et al. 2000; Spierer et al. 2005) , but which may not affect MSL function. Likewise, we did not identify JIL-1, SCF, NUP153, or Megator, all previously implicated in MSL interaction (Jin et al. 2000; Furuhashi et al. 2006; Mendjan et al. 2006).

To narrow our list to interesting candidate genes, we classified genes into categories based on their activity in our assay and previously determined functional information about the gene products as follows (Table S2): 1) **Class 1 (18 genes)**: General transcriptional regulators that *activate* our reporter; 2) **Class 2 (6 genes)**: Proteins that have potential sequence-specific DNA binding activity and *activate* our reporter; 3) **Class 3 (7 genes)**: General transcriptional regulators that *repress* our reporter; 4) **Class 4 (3 genes)**: Proteins that have potential sequence-specific DNA binding activity and *repress* our reporter; 5) **Class 5 (7 genes)**: Proteins that regulate RNA metabolism (e.g. splicing, non-sense mediated decay); 6) **Class 6 (26 genes)**: Proteins that are previously unstudied and/or have functions unrelated to transcription or RNA metabolism. The remaining genes are known MSL complex components. Members of each of these classes could directly modulate MSL recruitment or activity or indirectly affect MSL complex by altering complex levels.

Due to the large number of genes identified, we focused on potential direct regulators of MSL complex recruitment or function that had known roles in transcriptional regulation or potential DNA binding domains (Classes 1-4). This analysis identified 34 potential transcriptional regulators. We validated the activity of these candidate genes using our MSL-dependent reporter assay in a more accurate 96-well format (Fig. 2). Overall, we identified 24 activators (Classes 1 and 2) and 9 repressors (Classes 3 and 4) with known or predicted roles in transcriptional regulation.

In addition to identifying all core components of MSL complex as activators of our reporter system (Fig. 1B), we identified a known repressor of MSL function, NURF301, which we previously determined represses *roX* gene transcription *in vivo* (Bai et al. 2007) (Table S2). Also, NURF301 is known to associate with the *roX1* CES (Bai et al. 2007). Therefore, our RNAi screening strategy identified all MSL complex components and a known direct repressor of MSL-dependent transcription, validating our ability to identify activators and repressors with direct functions.

Selection of NSL complex, PAF complex, and CG1832 as candidate direct regulators of MSL targeting or function

The transcription of MSL complex components could be modulated and thus indirectly affect its function. Therefore, we selected candidates and examined whether they function directly to recruit MSL complex to CES loci using several approaches: 1) Determine whether RNAi against candidate genes affects recruitment of MSL complex to endogenous CES; 2) Define whether RNAi knockdown of candidates alters transcription of MSL complex components; 3) Examine the recruitment of candidate proteins to CES loci. We hypothesized that direct regulators of MSL complex recruitment will be recruited to CES loci and alter MSL complex recruitment but not transcription of MSL complex components. Using these approaches, we examined the function of three factors: 1) the NSL complex; 2) the PAF complex; and 3) CG1832.

We selected NSL complex for further study because six of its components (NSL1, NSL3, CG1135, z4, Chro, and MBD-R2) were identified in our screen and NSL and MSL complexes share the MOF catalytic component. NSL1 is one of the structural components of NSL complex that directly contacts MOF using a similar protein-protein interaction surface as MSL1 (Kadlec et al. 2011). Although this biochemical relationship would predict that NSL complex would compete with MSL complex for the MOF subunit, NSL complex was identified as an activator of our MSL-dependent reporter. Therefore, we further characterized the relationship between the two complexes to determine whether NSL complex functions positively during MSL complex recruitment.

We also examined the PAF transcription complex that associates with RNAP II across gene bodies to promote transcription elongation (Adelman et al. 2006; Nagaike et al. 2011). Again, we identified multiple subunits of this complex in our screen, making the PAF complex a strong candidate. The PAF complex has been implicated in facilitating H2B ubiquitylation catalyzed by the *S. cerevisiae* SAGA complex component Ubp8 (Henry et al. 2003; Tanny et al. 2007). Intriguingly, recent work has identified a new activity for the MSL2 protein as an E3 ligase that targets H2BK34 for ubiquitylation (H2BK31 in *Drosophila*) (Wu et al. 2011). Therefore, it is possible that the PAF complex facilitates the H2B ubiquitin ligase activity of MSL2. Also, MSL complex associates with gene bodies of active genes like the PAF complex (Alekseyenko et al. 2006; Gilfillan et al. 2006).

The previously-unstudied CG1832 zinc finger protein was chosen for further study because it was identified three independent times using different double-stranded RNA constructs and exhibits a very strong reporter activation phenotype. Furthermore, unlike other potential DNA binding proteins, CG1832 RNAi does not decrease cell viability significantly and was not identified by other RNAi screens. Similarly, loss of MSL complex components does not significantly alter cell viability although it causes male-specific lethality in flies. The CG1832 protein has a glutamine rich N-terminus and a seven zinc finger domain at its C-terminus. Because each finger is likely to recognize three bases (Klug 2010), CG1832 is a candidate for recognition of a 21-mer sequence such as the MRE element. CG1832 is highly conserved among insect species and orthologs are present in other species such as mouse (znf80) and human (znf429). Furthermore, CG1832 is maternally loaded into the early embryo and ubiquitously expressed (FlyAtlas) and therefore it could potentially target MSL complex early in development. In addition, CG1832 is likely to be an essential gene because P-element insertions at its 5' end are lethal in both sexes (data not shown). Therefore, CG1832 would not have been identified by the seminal male-specific lethal screens *in vivo*.

CG1832 and NSL complex modulate MSL complex recruitment to endogenous CES loci *in vivo*

To determine whether candidate genes are needed for MSL complex recruitment to endogenous CES loci, we used chromatin immunoprecipitation (ChIP) of the MSL2

core component in male SL2 cells. We found that both NSL1 and CG1832 RNAi knockdowns significantly reduce MSL2 recruitment to several CES loci *in vivo* (Fig. 3A). Next, we conducted ChIP on the H4K16ac histone modification to examine the roles of NSL1 and CG1832 in depositing this key modulator of dosage compensation. In parallel with the effect on MSL2 recruitment, both NSL1 and CG1832 RNAi caused reduced H4K16ac at several CES loci. In contrast, PAF RNAi treatment did not alter MSL complex recruitment and therefore it is likely that PAF functions to alter MSL complex function at a step subsequent to recruitment (Fig. 3A).

To examine whether NSL1 and CG1832 RNAi treatments disrupt MSL complex recruitment due to regulation of MSL complex component mRNA levels, we assayed these by qRT-PCR (Fig. S1). *roX1* is not expressed at significant levels in SL2 cells (Hamada et al. 2005). We noted effects on *roX2* RNA levels as expected because *roX2* is activated by MSL complex, but little effect on the mRNA levels of other MSL complex components. Also, NSL1 and CG1832 RNAi treatments do not affect mRNA splicing of MLE as reported for the Znf72D zinc finger protein (data not shown) (Worringer and Panning 2007). MLE splicing was quantified using splice-site specific primers that were previously validated (Worringer and Panning 2007). Because improper targeting of MSL complex reduces complex stability (Copps et al. 1998), precise measurements of MSL protein levels do not address the point at which NSL1 or CG1832 acts to promote MSL complex activity. Instead, we tested the hypothesis that these proteins promote MSL recruitment by assessing their own localization to CES loci.

Genome-wide occupancy of NSL components and CG1832 support a direct role in MSL targeting

To define NSL complex and CG1832 localization on chromatin, we analyzed the genome-wide occupancy of both factors. Genome-wide localization data were available for the Chromator/Chriz and MBD-R2 components of the NSL complex (www.modENCODE.org). Therefore, we analyzed the enrichment of these proteins at CES loci compared with other sites that contain MRE sequences. Results indicate that there is a broad enrichment of NSL complex components within a 10 kb region surrounding CES compared to additional genomic loci on the X and autosomes that contain MREs (Fig. 4A), consistent with a local or regional function.

To assess the genome-wide occupancy of CG1832, we generated a polyclonal antibody that targets the N-terminus of the protein (see Methods). We validated that the CG1832 antibody recognizes a protein of the predicted size and its RNAi knockdown strongly reduces its protein and mRNA levels and association with chromatin (Figs. S1,2,3). CG1832 is expressed in both males and females and binds to polytene chromosomes in both sexes, consistent with it having a general, essential function in addition to its potential role in facilitation of MSL targeting (Fig. S2, S4). We performed ChIP-seq analysis on the CG1832 protein in male SL2 cells and mapped 2,695 CG1832 sites on the X and 10,009 on autosomes using inclusive statistical criteria for defining binding sites (see Methods). Overall, MREs on the X chromosome are approximately 1.5-fold more likely to be occupied by CG1832 than

MREs on autosomes (49.7% on X: 1832/3683; 32.5% on autosomes: 3787/11619). Although CG1832 occupancy is not X-specific, it is highly enriched at CES loci. In fact, 98.5% (135/137) of the genetically defined CES loci that contain MRE sequences (Alekseyenko et al. 2008) have a CG1832 binding site. Also, 92.5% (446/484) of all MSL complex binding sites that contain an MRE also have a CG1832 binding site. This expanded set would include likely CES not characterized previously. Most compelling, CG1832 occupancy at MREs is strongly enriched at CES loci compared with locations on the X and autosomes that contain MRE sequences but fail to recruit MSL complex, as visualized by heat maps for enrichment at various classes of MREs (Fig. 4B). The precise pattern of enrichment over MREs makes CG1832 a strong candidate to directly recognize these sequences. These results demonstrate that our genetic screen has identified at least two new regulators of MSL complex recruitment that are implicated in functioning directly at chromatin entry sites.

Discussion

Using a novel cell-based RNAi screening approach, we have identified new candidate regulators of MSL complex function. The screen took advantage of the ~five fold regulation of a *roX2*-luciferase reporter construct by MSL complex, rather than relying on a direct assay of dosage compensation (<two fold) or on more laborious assays for MSL occupancy. Many of the regulators that we have identified are likely to have additional functions in both male and female flies and therefore could not

have been recovered from classical male-specific lethal genetic screens. In addition to direct regulators of MSL complex recruitment, it is possible that transcriptional regulators identified from our screening approach could function indirectly to regulate transcription of MSL complex components themselves. Therefore, we used chromatin immunoprecipitation and existing modEncode data to demonstrate that two factors, the NSL complex and CG1832, are likely to have local roles in MSL complex recruitment to CES. Moreover, CG1832 is a strong candidate to recruit MSL complex directly to CES due to its strong enrichment at MRE sequences.

Interestingly, we identified a previously unstudied function of the highly-conserved NSL complex at CES loci. NSL and MSL complex share the MOF histone acetyltransferase as their catalytic component (Mendjan et al. 2006; Prestel et al. 2010b) and there are conserved MOF protein-protein interaction interfaces between MSL1 and NSL1 (Kadlec et al. 2011). Therefore, it is possible that MSL complex and NSL complex would compete the MOF subunit, suggesting that NSL complex would antagonize MSL complex. However, we instead found that NSL complex facilitates MSL complex recruitment to CES (Fig. 3). Therefore, it is possible that MOF may be recruited to MSL entry sites via NSL complex components such as MBD-R2 or RCD-5 that are known to be involved in MOF recruitment to 5' ends of active genes (Prestel et al. 2010b; Raja et al. 2010). Subsequently, MOF could be transferred to newly assembling MSL complexes at chromatin entry sites as previously suggested (Kadlec et al. 2011). Alternatively, since functional CES are

enriched in the vicinity of active genes (Alekseyenko AA 2012), this could indicate a role for NSL-dependent active gene expression in CES function.

Many additional activators were identified from our RNAi screening approach including all components of the PAF transcription complex that associates with RNAP II across gene bodies to promote transcription elongation and polyadenylation (Adelman et al. 2006; Nagaike et al. 2011). However, RNAi treatment targeting the PAF complex did not alter MSL complex recruitment to entry sites or MSL target genes (Fig. 3A). There are several possible mechanisms by which PAF complex could regulate our MSL-dependent reporter gene. The PAF complex has been implicated in facilitating H2B ubiquitylation catalyzed by the *S. cerevisiae* SAGA complex component Ubp8 (Henry et al. 2003; Tanny et al. 2007). H2B was also identified as an activator in our screen potentially implicating H2B ubiquitylation in facilitating MSL-dependent activity of the reporter (Table S1). Moreover, recent work has identified a new activity for the MSL2 protein as an E3 ligase that targets H2B for ubiquitylation (Wu et al. 2011). Therefore, it is possible that the PAF complex facilitates the H2B ubiquitin ligase activity of MSL2. Alternatively, PAF has been implicated in mediating H3K36me3 (Zhou et al. 2009) suggesting that it could also mediate MSL complex activity by modulating deposition of this known regulator of MSL complex recruitment to active genes.

In addition, to identifying activators of MSL-dependent reporter activity, we identified several repressors including NURF301, a known repressor of MSL

complex function at the *roX* CES loci (Bai et al. 2007). In this way, we validated that additional candidates have potential to directly function in repressing MSL complex recruitment or activity. Other repressors of known function include the *Sfmbt* protein that recruits the Polycomb repressor complex and the large TRRAP protein that mediates interactions between gene specific regulators and large transcription complexes (Nikiforov et al. 2002). It is possible that these proteins and other potential repressors directly antagonize MSL complex function or alter the chromatin environment around CES loci to reduce transcriptional activation.

Our most exciting new candidate is CG1832, a previously unstudied zinc-finger protein that is strongly and precisely enriched at MRE sequences within CES loci and is required for MSL complex recruitment to these sites. This protein has seven zinc-fingers at its C-terminus that could potentially bind DNA and a glutamine-rich N-terminus that may interact with MSL complex. Therefore, future analyses will focus on the exciting possibility that CG1832 functions as a previously unknown link between MSL complex and MRE sequences.

In summary, our novel screening approach identified two regulators that facilitate MSL complex recruitment at CES loci: NSL complex and CG1832. Furthermore, our genetic screen provides an extensive dataset of additional candidate genes that may facilitate MSL targeting and function. We expect that future genetic, genomic, and biochemical approaches will define new mechanisms by which many of these candidate genes modulate coordinate gene regulation.

Methods

Genome-wide RNAi screen

A genome-wide RNAi screen was performed in duplicate using the version 2 RNAi libraries generated by the Drosophila RNAi Screening Center (www.flyrnai.org) (Ramadan et al. 2007). Transient transfection of MSL-dependent and independent luciferase reporter plasmids was conducted as described previously (Alekseyenko et al. 2008). For screening, all volumes of plasmids and luciferase reagents were reduced by 4-fold to accommodate a 384-well format from the original 96-well format. As controls, GFP dsRNA was added to two wells and MSL2 dsRNA was added to two additional wells on every screening plate. GFP and MSL2 dsRNA was prepared as described previously (Gelbart et al. 2009). All RNAi treatments were conducted for 5 days.

Identification of candidate regulators of MSL complex targeting from RNAi screening data

First, the ratio of the Firefly luciferase to the Renilla luciferase activity was computed for each well in each of the 124 screening plates. Second, the Firefly/Renilla activity ratios for the GFP dsRNA wells and the MSL2 dsRNA wells were compared to assure that at least a 5-fold dynamic range was present for each plate. All plates fulfilled these criteria. Third, we calculated the median Firefly/Renilla ratio for each plate and its standard deviation. Our criterion for identifying a well as a positive screening hit was that its Firefly luciferase/Renilla luciferase ratio is more than two standard deviations above or below the plate

median for both replicate plates. Most 384-well plates had 2-5 screening hits. To eliminate screening hits that cause non-specific cell lethality or low transfection efficiency, we did not consider wells with Renilla luciferase values that were more than three-fold below the plate median. Secondary screens using validation dsRNAs and additional plasmids described in (Alekseyenko et al. 2008) were performed using the same protocol and analysis approach. All dsRNA amplicons are identified in Table S1 by their DRSC number.

mRNA analysis and Western blotting

mRNA analysis and qRT-PCR was performed after the following RNAi treatments: 1) DRSC03718 to target CG1832; 2) DRSC15625 to target NSL1 (CG4699); and 3) DRSC27502 to target PAF1 (*atms*). All RNAi treatments were conducted for 5 days in SL2 cells as previously described (Gelbart et al. 2009). An additional CG1832 RNAi construct (DRSC 29935) was also tested with similar results (data not shown).

Western blotting was conducted as previously described using standard protocols (Gelbart et al. 2009). We generated the rabbit polyclonal CG1832 antibody against part of the glutamine-rich domain of the protein (amino acids: #22-121) that was determined to be unique in the *Drosophila* genome. For Western blotting, the CG1832 antibody was used at a 1:1000 dilution, the tubulin antibody (Sigma) was used at a 1:10,000 dilution, and the histone H3 antibody (Abcam) was used at a 1:1000 dilution.

Chromatin Immunoprecipitation

Chromatin immunoprecipitation from SL2 cells was conducted as described previously (Gelbart et al. 2009) using antibodies targeting the MSL2 protein, H4K16ac (Millipore), H3 (Abcam), and CG1832 (SDI) proteins. RNAi treatments were performed using the following DRSC RNAi constructs (www.flyrnai.org): 1) DRSC03718 to target CG1832; 2) DRSC15625 to target NSL1 (CG4699); and 3) DRSC27502 to target PAF1 (*atms*). To assure reproducibility, large scale RNAi experiments were performed in which 225 ug of each dsRNA was added to a T225 flask containing 45 mls of SL2 cells. An additional CG1832 RNAi construct (DRSC 29935) was also tested with similar results (data not shown).

Three independent chromatin preparations were performed for each experiment and qPCR was performed on input and IP samples using primers that are specific for MSL target genes and CES loci. Primer sequences were previously described (Gelbart et al. 2009). Standard deviations among the three replicates were calculated after normalization to a sample that was treated with the GFP dsRNA construct as a control within each chromatin preparation. mRNA extraction followed by qRT-PCR was performed in parallel with each ChIP experiment to assure that the RNAi treatments were effective.

ChIP-seq

Preparation of ChIP-seq libraries

Duplicate ChIP-seq libraries from independent chromatin preparations were generated using a protocol adapted from the Illumina ChIP-seq sample preparation guide as follows (www.illumina.com). To obtain sufficient starting material, three IPs were pooled and concentrated using Qiagen MinElute columns such that 15 ng of starting material were obtained prior to library preparation. Libraries from matching input samples were also prepared for each chromatin preparation. Each enzyme reaction was followed by purification using AMPure XP Beads (Agilent). The fragmented DNA was 3' end repaired and 5' phosphorylated using End-it Kit (Epicentre). A 3' adenosine was added using Klenow Fragment (3'->5' exo) (NEB). The annealed adapters were ligated in a 1:50 dilution of the 10 mM stock using T4 DNA Ligase Quick Ligase Kit (Enzymatics). The libraries were PCR amplified using Phusion high fidelity polymerase (NEB) for 15 cycles. The libraries were then size selected on a 2% agarose gel using the E-Gel system (Invitrogen) selecting a range from 200 bp to 500 bp followed by concentration on a MinElute Column (Qiagen). Sequencing was performed on the Illumina GaIIx platform.

Analysis of ChIP-seq data

Reads were aligned to the dm3 assembly of the *D. melanogaster* genome using the Bowtie aligner (Langmead et al. 2009). A summary of the sequencing statistics is provided below:

Replicate #1 input: 29495696 raw reads, 17591247 aligned reads (59.6%)

Replicate #1 CG1832 IP: 27820142 raw reads, 19210177 aligned reads (69.1%)

Replicate #2 input: 22498098 raw reads, 13510882 aligned reads (60.1%)

Replicate #2 CG1832 IP: 28231400 raw reads, 15194724 aligned reads (53.8%)

Pearson R value for CG1832 ChIP-seq replicates, 1kb bins: 0.95

Reads aligned with more than one mismatch, and reads that did not align uniquely were discarded. Peaks were called in each ChIP-seq sample using the SPP software package with an FDR threshold of 0.05 (Kharchenko et al. 2008). For comparisons with MSL complex ChIP-seq data, MSL complex ChIP-seq data (Alekseyenko et al. 2008) were realigned with Bowtie and peaks were called with SPP in the same manner as the CG1832 ChIP-seq samples, for more direct comparison. The locations of CES loci and MRE sequences were previously reported (Alekseyenko et al. 2008). Data will be available at NCBI GEO Short Read Archive (SRA) upon publication.

Generation of heatmaps

The average enrichment profiles of proteins around the MREs (+/- 5kb) are shown in Fig. 4. The five columns in the heatmap are the following: 'MRE in CES' consists of experimentally obtained 140 MREs (located within 137 CES) described in (Alekseyenko et al. 2008); 'Best MRE on X' and 'Best MRE on 2L' consist of 150 MREs that have the best consensus motif match on the X or 2L, respectively; and 'Random MRE on X' and 'Random MRE on 2L' consist of 150 MREs randomly chosen

from the X or 2L, respectively. More details can be found in Alekseyenko et al, 2012 (submitted).

modENCODE ChIP-chip data processing

ChIP-chip data using Genomic DNA Tiling Arrays v2.0 (Affymetrix) are publicly accessible online through the modENCODE project (www.modencode.org). Data analysis was performed in R statistical programming environment (<http://www.r-project.org>). For the visualization of the heatmap (e.g., Figure 4), the +/-5kb region surrounding each MRE was separated into non-overlapping bins of 200bp. The smoothed probe value within each bin is averaged to obtain the enrichment value for that bin. Heatmaps were generated using the same classes of MRE containing loci as described above for the CG1832 ChIP-seq data.

Acknowledgements

We are grateful to Dr. Judith Bender for critical reading of the manuscript, to Dr. Joshua Ho for expert contributions to data visualization, and to the *Drosophila* RNAi Screening Center (DRSC) for advice on performing the genome-wide RNAi screen. Next Generation sequencing was performed at the Brown University Genomics core facility.

References

1. Gelbart ME, Kuroda MI (2009) *Drosophila* dosage compensation: a complex voyage to the X chromosome. *Development* 136: 1399-1410.
2. Kelley RL, Meller VH, Gordadze PR, Roman G, Davis RL, et al. (1999) Epigenetic spreading of the *Drosophila* dosage compensation complex from roX RNA genes into flanking chromatin. *Cell* 98: 513-522.
3. Alekseyenko AA, Peng S, Larschan E, Gorchakov AA, Lee OK, et al. (2008) A sequence motif within chromatin entry sites directs MSL establishment on the *Drosophila* X chromosome. *Cell* 134: 599-609.
4. Straub T, Grimaud C, Gilfillan GD, Mitterweger A, Becker PB (2008) The chromosomal high-affinity binding sites for the *Drosophila* dosage compensation complex. *PLoS Genet* 4: e1000302.
5. Larschan E, Alekseyenko AA, Gortchakov AA, Peng S, Li B, et al. (2007) MSL complex is attracted to genes marked by H3K36 trimethylation using a sequence-independent mechanism. *Mol Cell* 28: 121-133.
6. Larschan E, Bishop E, Kharchenko P, Park P, Kuroda M (2011) X chromosome dosage compensation via enhanced transcription elongation in *Drosophila* males. *Nature* In press.
7. Fauth T, Muller-Planitz F, Konig C, Straub T, Becker PB (2010) The DNA binding CXC domain of MSL2 is required for faithful targeting the Dosage Compensation Complex to the X chromosome. *Nucleic Acids Res* 38: 3209-3221.
8. Belote JM, Lucchesi JC (1980) Male-specific lethal mutations of *Drosophila melanogaster*. *Genetics* 96: 165-186.
9. Hilfiker A, Hilfiker-Kleiner D, Pannuti A, Lucchesi JC (1997) mof, a putative acetyl transferase gene related to the Tip60 and MOZ human genes and to the SAS genes of yeast, is required for dosage compensation in *Drosophila*. *Embo J* 16: 2054-2060.
10. Carrozza MJ, Li B, Florens L, Suganuma T, Swanson SK, et al. (2005) Histone H3 methylation by Set2 directs deacetylation of coding regions by Rpd3S to suppress spurious intragenic transcription. *Cell* 123: 581-592.
11. Keogh MC, Kurdistani SK, Morris SA, Ahn SH, Podolny V, et al. (2005) Cotranscriptional set2 methylation of histone H3 lysine 36 recruits a repressive Rpd3 complex. *Cell* 123: 593-605.
12. Bell O, Conrad T, Kind J, Wirbelauer C, Akhtar A, et al. (2008) Transcription-coupled methylation of histone H3 at lysine 36 regulates dosage compensation by enhancing recruitment of the MSL complex in *Drosophila melanogaster*. *Mol Cell Biol* 28: 3401-3409.
13. Mendjan S, Taipale M, Kind J, Holz H, Gebhardt P, et al. (2006) Nuclear pore components are involved in the transcriptional regulation of dosage compensation in *Drosophila*. *Mol Cell* 21: 811-823.
14. Prestel M, Feller C, Becker PB (2010) Dosage compensation and the global re-balancing of aneuploid genomes. *Genome Biol* 11: 216.

15. Shi X, Finkelstein A, Wolf AJ, Wade PA, Burton ZF, et al. (1996) Paf1p, an RNA polymerase II-associated factor in *Saccharomyces cerevisiae*, may have both positive and negative roles in transcription. *Mol Cell Biol* 16: 669-676.
16. Nagaike T, Logan C, Hotta I, Rozenblatt-Rosen O, Meyerson M, et al. (2011) Transcriptional activators enhance polyadenylation of mRNA precursors. *Mol Cell* 41: 409-418.
17. Bai X, Alekseyenko AA, Kuroda MI (2004) Sequence-specific targeting of MSL complex regulates transcription of the roX RNA genes. *Embo J* 23: 2853-2861.
18. Nybakken K, Perrimon N (2002) Hedgehog signal transduction: recent findings. *Curr Opin Genet Dev* 12: 503-511.
19. Kulkarni MM, Booker M, Silver SJ, Friedman A, Hong P, et al. (2006) Evidence of off-target effects associated with long dsRNAs in *Drosophila melanogaster* cell-based assays. *Nat Methods* 3: 833-838.
20. Spierer A, Seum C, Delattre M, Spierer P (2005) Loss of the modifiers of variegation Su(var)3-7 or HP1 impacts male X polytene chromosome morphology and dosage compensation. *J Cell Sci* 118: 5047-5057.
21. Deuring R, Fanti L, Armstrong JA, Sarte M, Papoulas O, et al. (2000) The ISWI chromatin-remodeling protein is required for gene expression and the maintenance of higher order chromatin structure in vivo. *Mol Cell* 5: 355-365.
22. Furuhashi H, Nakajima M, Hirose S (2006) DNA supercoiling factor contributes to dosage compensation in *Drosophila*. *Development* 133: 4475-4483.
23. Jin Y, Wang Y, Johansen J, Johansen KM (2000) JIL-1, a chromosomal kinase implicated in regulation of chromatin structure, associates with the male specific lethal (MSL) dosage compensation complex. *J Cell Biol* 149: 1005-1010.
24. Bai X, Larschan E, Kwon SY, Badenhorst P, Kuroda MI (2007) Regional control of chromatin organization by noncoding roX RNAs and the NURF remodeling complex in *Drosophila melanogaster*. *Genetics* 176: 1491-1499.
25. Kadlec J, Hallacli E, Lipp M, Holz H, Sanchez-Weatherby J, et al. (2011) Structural basis for MOF and MSL3 recruitment into the dosage compensation complex by MSL1. *Nat Struct Mol Biol* 18: 142-149.
26. Adelman K, Wei W, Ardehali MB, Werner J, Zhu B, et al. (2006) *Drosophila* Paf1 modulates chromatin structure at actively transcribed genes. *Mol Cell Biol* 26: 250-260.
27. Henry KW, Wyce A, Lo WS, Duggan LJ, Emre NC, et al. (2003) Transcriptional activation via sequential histone H2B ubiquitylation and deubiquitylation, mediated by SAGA-associated Ubp8. *Genes Dev* 17: 2648-2663.
28. Tanny JC, Erdjument-Bromage H, Tempst P, Allis CD (2007) Ubiquitylation of histone H2B controls RNA polymerase II transcription elongation independently of histone H3 methylation. *Genes Dev* 21: 835-847.
29. Wu L, Zee BM, Wang Y, Garcia BA, Dou Y (2011) The RING Finger Protein MSL2 in the MOF Complex Is an E3 Ubiquitin Ligase for H2B K34 and Is Involved in Crosstalk with H3 K4 and K79 Methylation. *Mol Cell* 43: 132-144.

30. Alekseyenko AA, Larschan E, Lai WR, Park PJ, Kuroda MI (2006) High-resolution ChIP-chip analysis reveals that the *Drosophila* MSL complex selectively identifies active genes on the male X chromosome. *Genes Dev* 20: 848-857.
31. Gilfillan GD, Straub T, de Wit E, Greil F, Lamm R, et al. (2006) Chromosome-wide gene-specific targeting of the *Drosophila* dosage compensation complex. *Genes Dev* 20: 858-870.
32. Klug A (2010) The discovery of zinc fingers and their development for practical applications in gene regulation and genome manipulation. *Q Rev Biophys* 43: 1-21.
33. Hamada FN, Park PJ, Gordadze PR, Kuroda MI (2005) Global regulation of X chromosomal genes by the MSL complex in *Drosophila melanogaster*. *Genes Dev* 19: 2289-2294.
34. Worringer KA, Panning B (2007) Zinc finger protein Zn72D promotes productive splicing of the maleless transcript. *Mol Cell Biol* 27: 8760-8769.
35. Copps K, Richman R, Lyman LM, Chang KA, Rampersad-Ammons J, et al. (1998) Complex formation by the *Drosophila* MSL proteins: role of the MSL2 RING finger in protein complex assembly. *Embo J* 17: 5409-5417.
36. Prestel M, Feller C, Straub T, Mitlohner H, Becker PB (2010) The activation potential of MOF is constrained for dosage compensation. *Mol Cell* 38: 815-826.
37. Raja SJ, Charapitsa I, Conrad T, Vaquerizas JM, Gebhardt P, et al. (2010) The nonspecific lethal complex is a transcriptional regulator in *Drosophila*. *Mol Cell* 38: 827-841.
38. Alekseyenko AA HJ, Peng S, Gelbart M, Tolstorukov M, Plachetka A, Kharchenko PV, Jung YL, Gorchakov AA, Larschan E, Gu T, Minoda A, Riddle NC, Schwartz YB, Elgin SCR, Karpen GH, Pirrotta V, Kuroda MI**, Park PJ**. (2012) Sequence-specific targeting of dosage compensation in *Drosophila* favors an active chromatin context. In press *PLoS Genetics*.
39. Zhou K, Kuo WH, Fillingham J, Greenblatt JF (2009) Control of transcriptional elongation and cotranscriptional histone modification by the yeast BUR kinase substrate Spt5. *Proc Natl Acad Sci U S A* 106: 6956-6961.
40. Nikiforov MA, Chandriani S, Park J, Kotenko I, Matheos D, et al. (2002) TRRAP-dependent and TRRAP-independent transcriptional activation by Myc family oncoproteins. *Mol Cell Biol* 22: 5054-5063.
41. Ramadan N, Flockhart I, Booker M, Perrimon N, Mathey-Prevot B (2007) Design and implementation of high-throughput RNAi screens in cultured *Drosophila* cells. *Nat Protoc* 2: 2245-2264.
42. Gelbart ME, Larschan E, Peng S, Park PJ, Kuroda MI (2009) *Drosophila* MSL complex globally acetylates H4K16 on the male X chromosome for dosage compensation. *Nat Struct Mol Biol* 16: 825-832.
43. Langmead B, Trapnell C, Pop M, Salzberg SL (2009) Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol* 10: R25.
44. Kharchenko PV, Tolstorukov MY, Park PJ (2008) Design and analysis of ChIP-seq experiments for DNA-binding proteins. *Nat Biotechnol* 26: 1351-1359.

Figure 1: An RNAi screening system in *Drosophila* SL2 cells to identify regulators of MSL complex recruitment and function

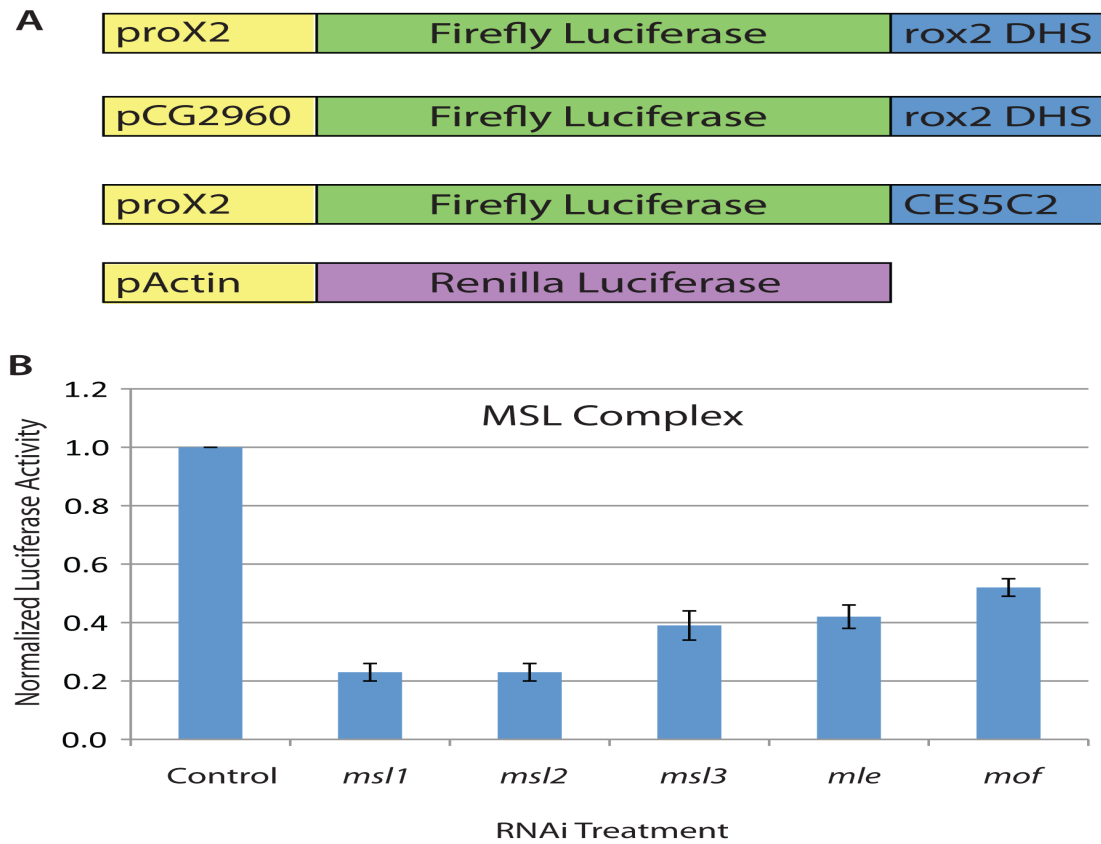


Figure 1: An RNAi screening system in *Drosophila* SL2 cells to identify regulators of MSL complex recruitment and function. **A)** Schematic of MSL-dependent reporter system for SL2 cell-based genome-wide RNAi screen. The screening system includes the following three elements: 1) promoter of *roX2* gene (370 bps); 2) Firefly luciferase reporter gene; 3) *roX2* DHS (DNase I Hypersensitivity Site: 280 bp). Two additional constructs were used for secondary screening in which the promoter or CES locus were altered. The alternative promoter is that of the CG2690 gene and the alternative CES locus is the minimal 150 bp fragment of CES5C2 that were previously described (Alekseyenko et al. 2008). **B)** Ratio of the Firefly luciferase activity to Renilla luciferase activity for MSL complex component RNAi treatments identified from the genome-wide screen. Data shown are the average and standard deviation of two replicates from the genome-wide screening plates. Control RNAi treatment targeted the GFP gene that is not present in SL2 cells.

Figure 2: Candidate regulators of MSL complex identified from the genome-wide RNAi screen.

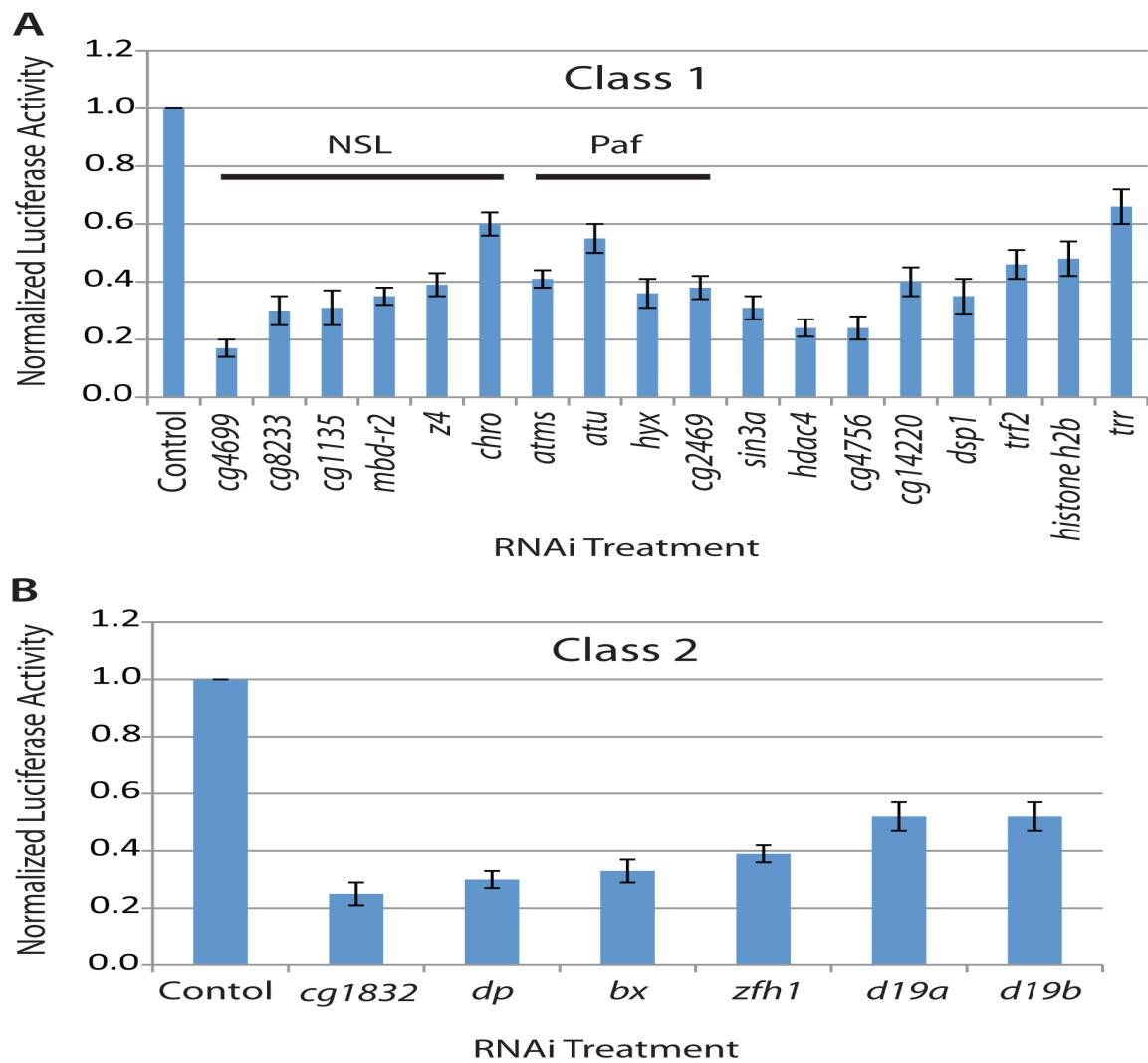


Figure 2: Candidate regulators of MSL complex identified from the genome-wide RNAi screen. The Firefly/Renilla luciferase activity ratio and its standard deviation was determined based on quadruplicate reactions in 96-well plate format after RNAi treatment targeting the following classes of candidate genes: **A)** Class 1: NSL complex, PAF complex and other general transcriptional regulators; **B)** Class 2: Activators with potential specific DNA binding activity like CG1832.

Figure 3: NSL1 and CG1832 contribute to MSL complex recruitment

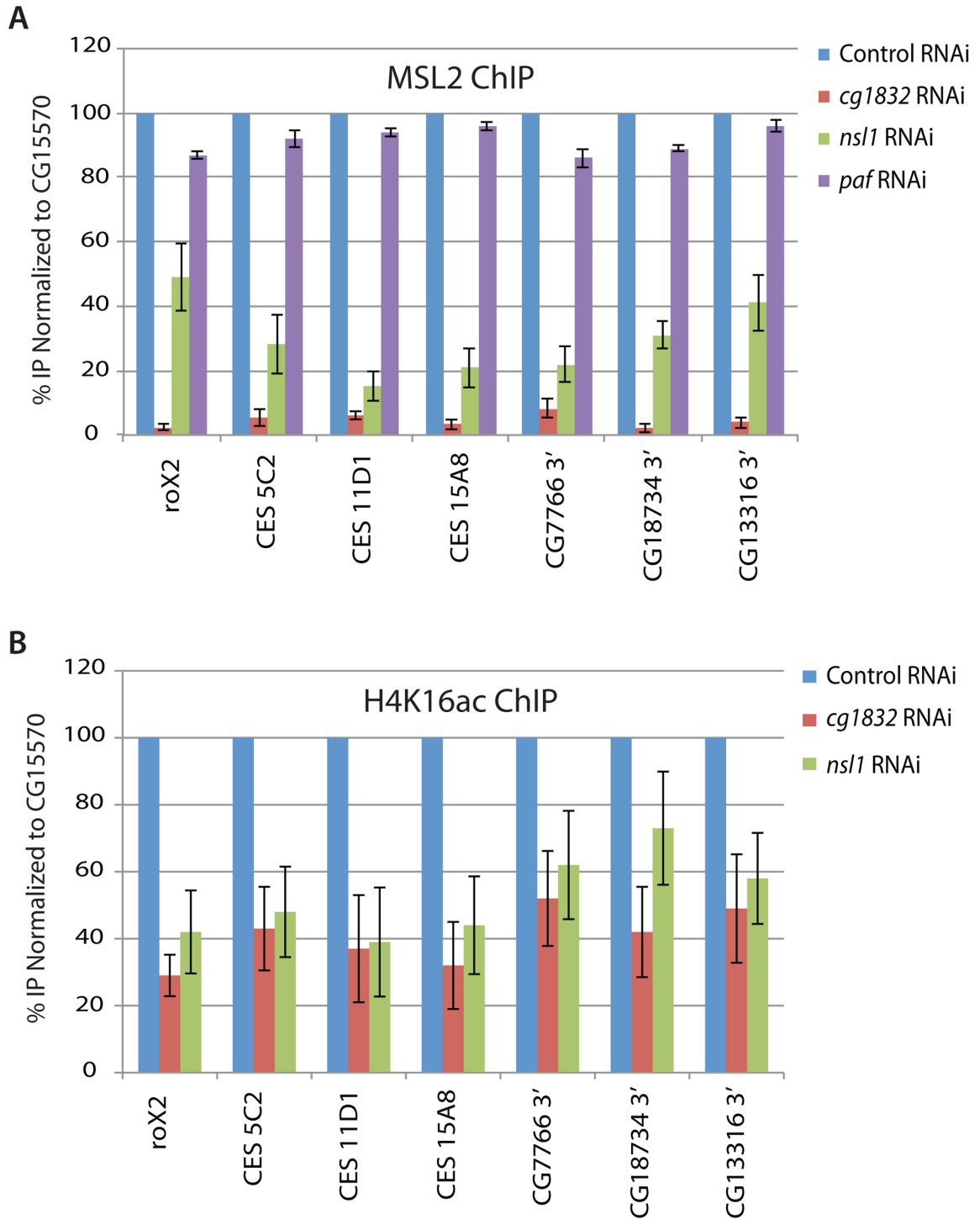


Figure 3: NSL1 and CG1832 contribute to MSL complex recruitment A)

Chromatin immunoprecipitation was conducted using an antibody specific for the MSL2 protein before and after RNAi treatment targeting the *paf1*, *ns1* and *CG1832* genes. Four CES loci and the 3' ends of three previously characterized MSL complex target genes were assayed (Larschan et al. 2007; Alekseyenko et al. 2008). Data shown are the average and standard deviation of at least two replicates. Normalization was conducted both to input DNA material and to the non-transcribed CG15570 gene as an internal control as previously described (Gelbart et al. 2009). **B)** Chromatin immunoprecipitation was conducted as described for (A) but an antibody targeting the H4K16ac histone modification (Millipore) was used instead of the MSL2 antibody. In addition to the normalization as described for (A), H4K16ac IPs were normalized to histone H3 occupancy as determined by H3 ChIP. Due to significant changes in histone H3 occupancy upon *paf1* RNAi treatment, these data were not included.

Figure 4: NSL complex and CG1832 enrichment at CES loci.

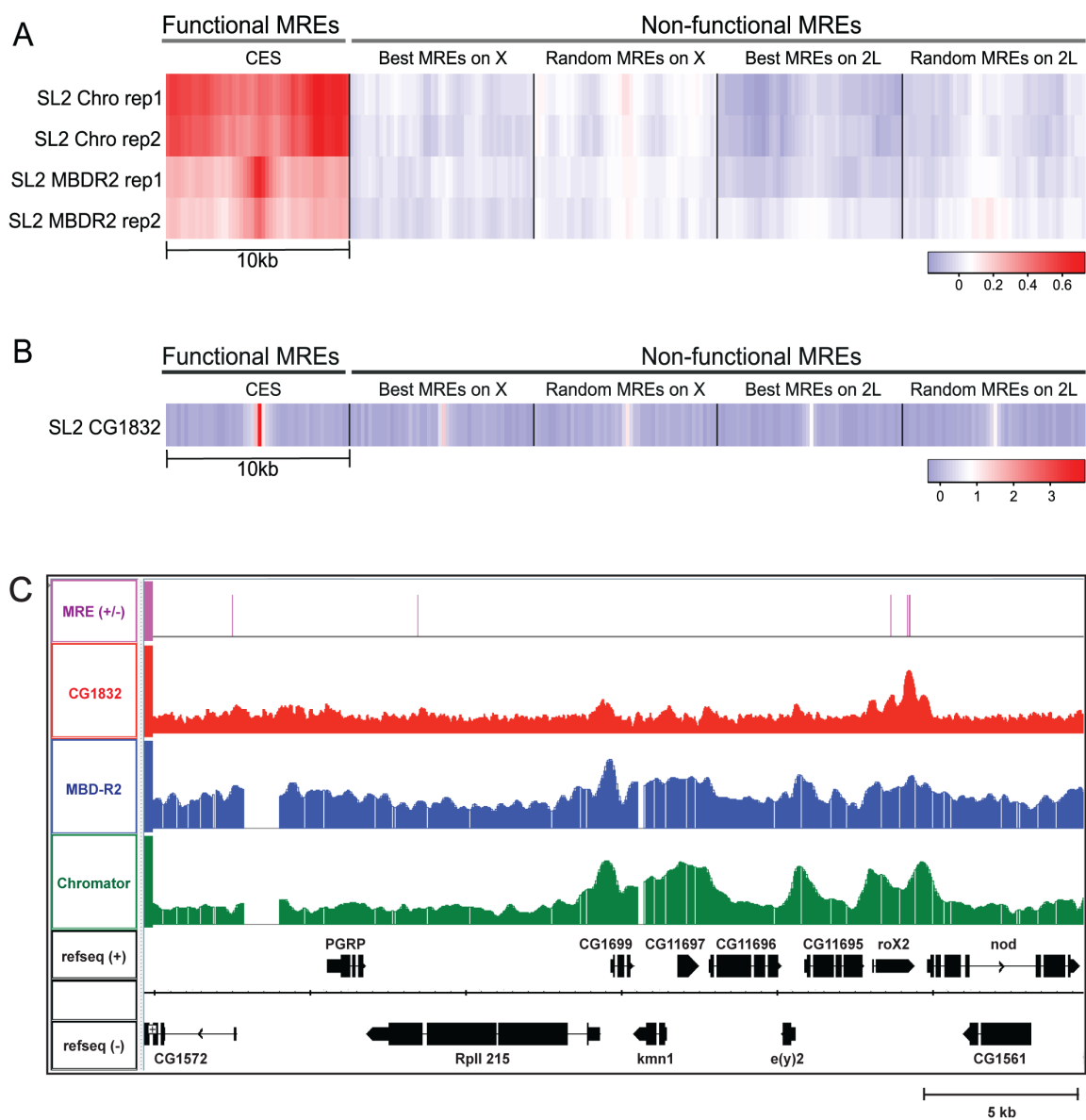


Figure 4: NSL complex and CG1832 enrichment at CES loci.

A) Heatmaps show the average enrichment of the Chromator/Chriz and MBD-R2 components of the NSL complex surrounding MRE sequences as assayed by ChIP-chip in SL2 cells (two replicates each; data available from <http://www.modENCODE.org>). The darker red color indicates greater occupancy, as measured in the $\log_2(\text{ChIP}/\text{input})$ scale. Heatmaps are shown for sets of 150 sites containing MREs that are defined as follows: MREs in chromatin entry sites (column 1); best MREs on the X chromosome or autosome 2L (MRE motifs closest to the consensus, but not utilized, columns 2 and 4); randomly chosen MREs on the X or autosome 2L (columns 3 and 5). B) Heatmaps show the enrichment of CG1832 surrounding MREs from ChIP-seq data in SL2 cells. Enrichment of CG1832 is shown surrounding the same MREs that were examined in 4A. The $\log_2(\text{ChIP}/\text{input})$ scale is used for binned read count; the enrichment difference is enhanced in the ChIP-seq data owing to its greater dynamic range.

Table 1. Gene ontology categories that were enriched in the genome-wide RNAi screen.

Functional Category	P-value
Protein biosynthesis	1.01E-22
Cell cycle	4.24E-07
Cell cycle control	1.18E-06
mRNA transcription	4.03E-06
Protein metabolism and modification	3.38E-05
Nucleoside, nucleotide and nucleic acid metabolism	1.03E-04
mRNA transcription regulation	2.99E-02
doi:10.1371/journal.pgen.1002830.t001	

Figure S1. mRNA levels of MSL complex components and candidates after CG1832, Nsl1, and Paf1 RNAi treatments.

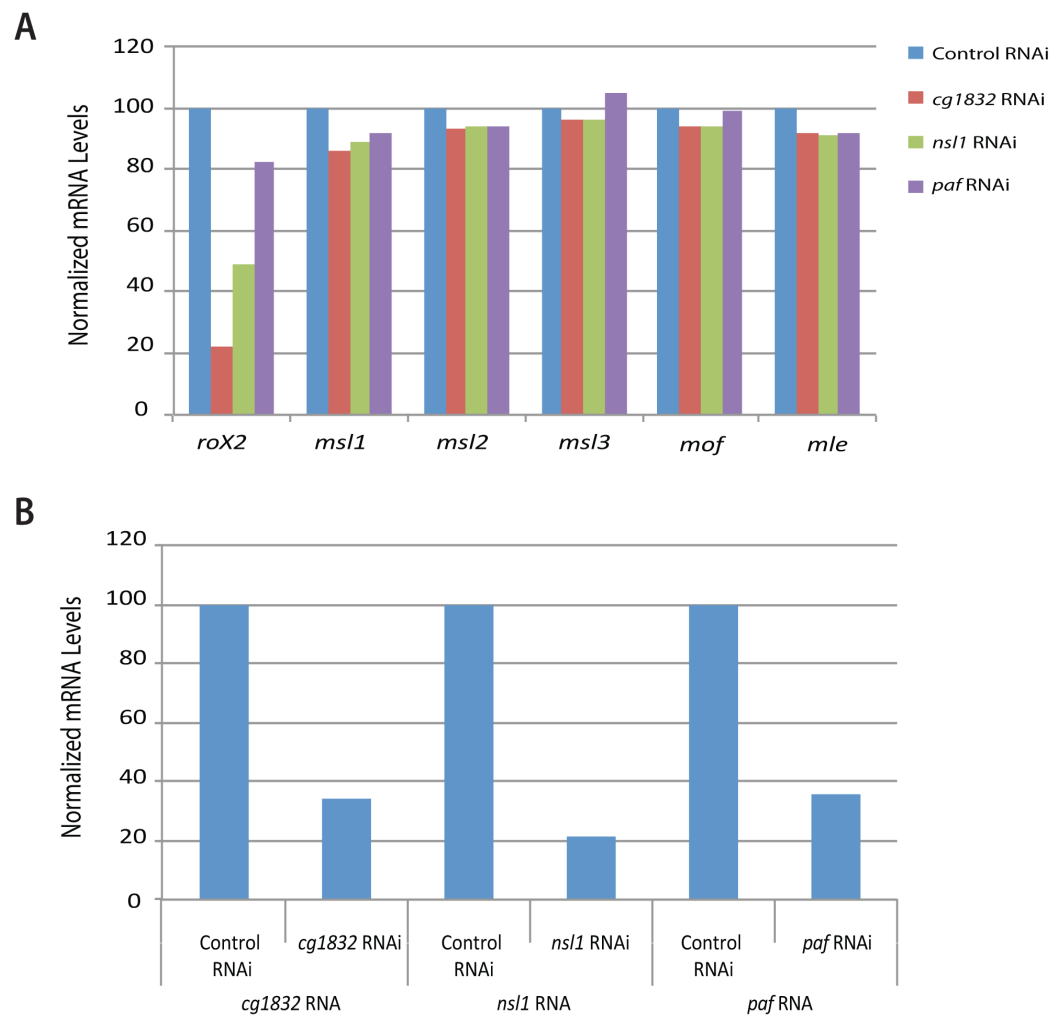


Figure S1. mRNA levels of MSL complex components and candidates after CG1832, Nsl1, and Paf1 RNAi treatments. A) qRT-PCR was used to assay the expression levels of all MSL complex components present in SL2 cells. roX2 levels are decreased likely because targeting of MSL complex to the roX2 CES locus is altered by the RNAi treatments as expected. Expression levels of other MSL complex components are unchanged. All data were normalized to the control RNAi treatment and are the average of two replicates. B) Effective RNAi treatments were validated by qRT-PCR and the results shown are the average of two replicates.

Figure S2. Western blots for CG1832 RNAi treatment and antibody validation.

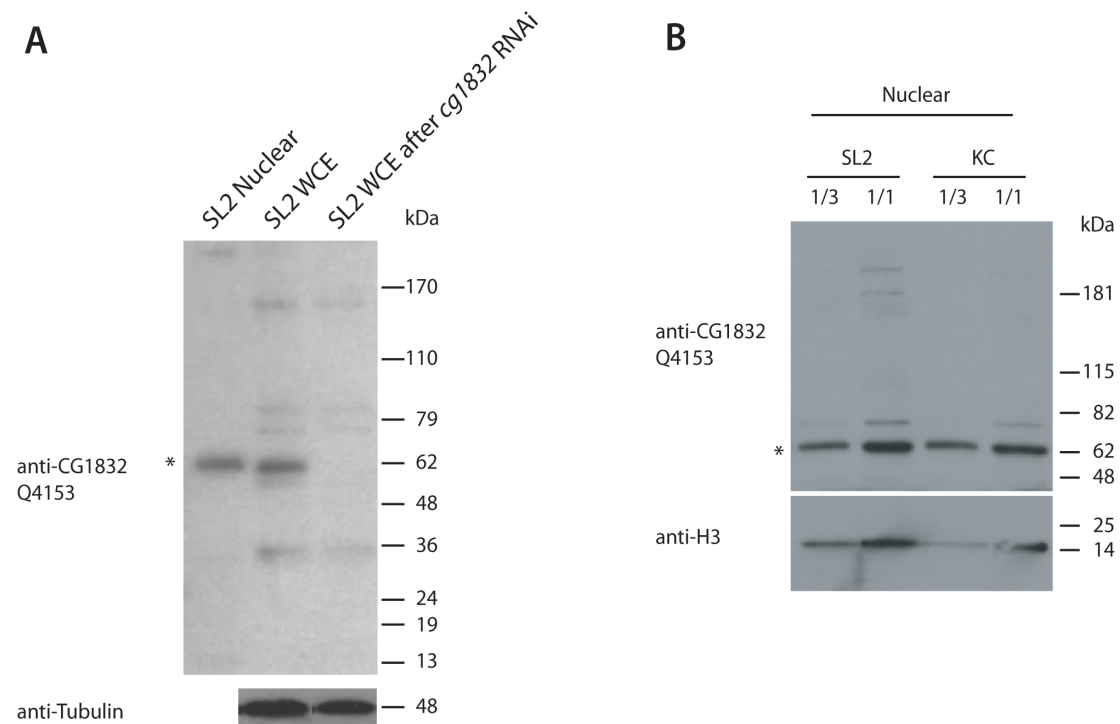


Figure S2. Western blots for CG1832 RNAi treatment and antibody validation.

A) Western blots on whole cell and nuclear extracts indicate that CG1832 is a nuclear protein and the CG1832 RNAi treatment reduces CG1832 protein levels. The asterik indicates the location of the CG1832 protein (60 kDa) and tubulin was used as a loading control. B) Nuclear extraction was performed on protein samples from SL2 (male) and Kc (female) cells followed by Western blotting. The histone H3 antibody was used as a loading control. Levels of CG1832 are similar in males and females.

Figure S3. Validation of CG1832 antibody by ChIP at CES loci.

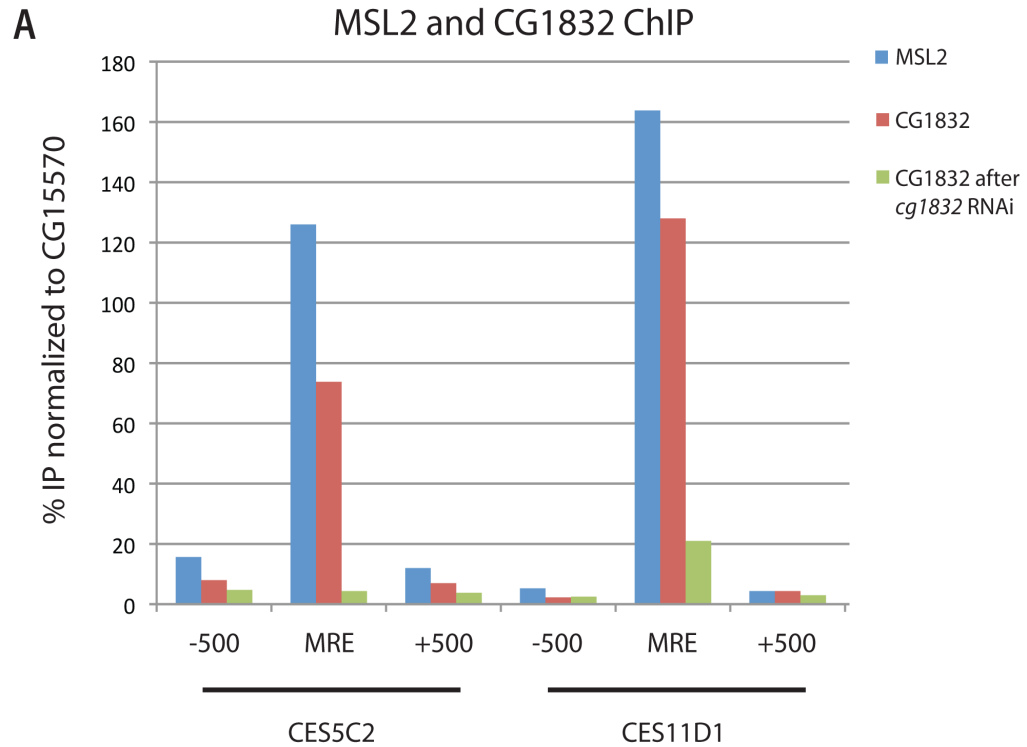
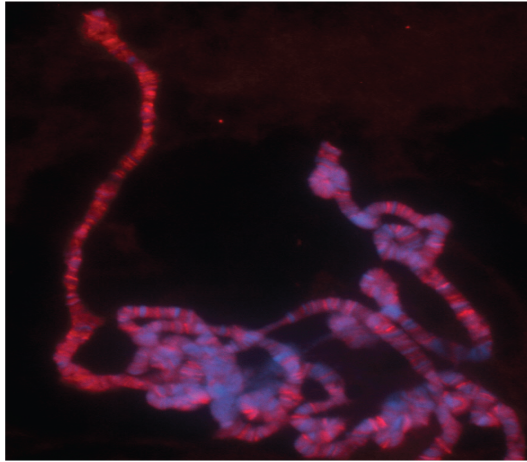


Figure S3. Validation of CG1832 antibody by ChIP at CES loci. qPCR was performed on CG1832 ChIP samples before and after CG1832 RNAi treatment. Three independent CES loci were assayed and CG1832 RNAi strongly reduced the levels of CG1832 protein on chromatin at all three loci. An average of two independent experiments is shown.

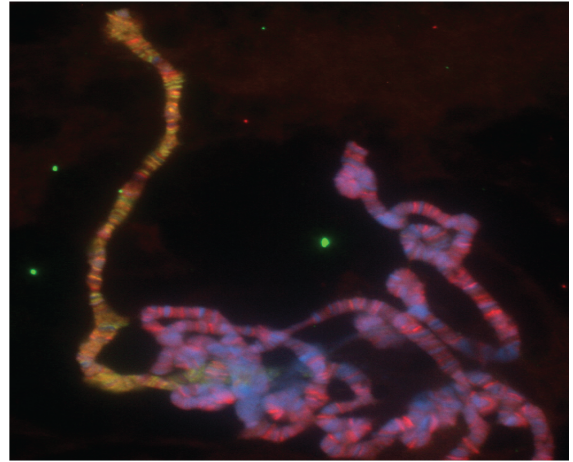
Figure S4. Immunostaining of polytene chromosomes.

A

Male
Hoechst (blue)
anti-CG1832 (red)



Male
Hoechst (blue)
anti-CG1832 (red)
anti-MSL3 (green)



B

Female
Hoechst (blue)
anti-CG1832 (red)

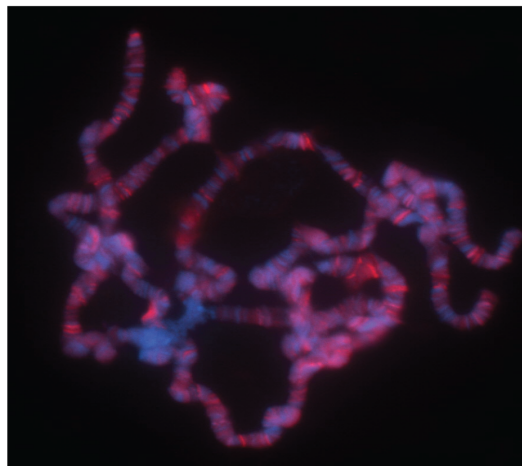


Figure S4. Immunostaining of polytene chromosomes. Polytene staining using the CG1832 antibody (red) in male (A) and female (B) larvae. Co-staining was performed with an anti-MSL3 antibody (green). Co-localization of the two proteins is shown in yellow.

Table S1. Initial candidates identified from a genome-wide RNAi screen to identify novel regulators of dosage compensation.

Key

Passes all secondary screens
Do not pass all secondary screens
Not rescreened

DRSC Amplicon	Ratio Firefly/Renilla normalized to plate median	Chrom	Gene	Human Homologene	17 bp Matches	19 bp Matches	proX2-DHS secondary screen	proX2-CES13	pCG2960-roX2-DHS	Passes all secondary screens
DRSC27502	0.41	3R	atms	PAF1	1	0	YES	YES	YES	YES
DRSC12123	0.55	3R	Atu	LEO1	3	0	YES	YES	YES	YES
DRSC23048	0.33	X	Bx	LMO1	20	1	YES	YES	YES	YES
DRSC29957	1.99	X	caz	FUS	1	0	YES	YES	YES	YES
DRSC18450	0.33	X	Cbp80	NCBP1	2	0	YES	YES	YES	YES
DRSC03504	0.48	2L	cdc2	CDC2	1	0	YES	YES	YES	YES
DRSC09735	0.52	3L	CG10486		2	0	YES	YES	YES	YES
DRSC27788	0.31	3L	CG1135	MCRS1	4	0	YES	YES	YES	YES
DRSC11663	0.44	3L	CG11451		2	0	YES	YES	YES	YES
DRSC26345	0.4	X	CG14220	SUD53	7	0	YES	YES	YES	YES
DRSC24312	1.73	2L	CG16975	L3MBTL2	3	0	YES	YES	YES	YES
DRSC29935	0.25	2L	CG1832		5	0	YES	YES	YES	YES
DRSC27308	0.38	3L	CG2469	CTR9	5	0	YES	YES	YES	YES
DRSC15344	0.65	3R	CG31211	C6orf111	62	3	YES	YES	YES	YES
DRSC29707	0.52	3R	CG31258		2	0	YES	YES	YES	YES
DRSC10042	0.52	3L	CG32369		2	0	YES	YES	YES	YES
DRSC04360	2.86	2R	CG3363		1	0	YES	YES	YES	YES
DRSC15625	0.17	3R	CG4699		7	0	YES	YES	YES	YES
DRSC19916	0.24	X	CG4756	SAP30L	9	0	YES	YES	YES	YES
DRSC10484	0.66	3L	CG5146		18	2	YES	YES	YES	YES
DRSC24273	0.41	3R	CG5514		5	0	YES	YES	YES	YES
DRSC15948	0.27	3R	CG6015	CDC40	1	0	YES	YES	YES	YES
DRSC20094	1.85	X	CG8211	INTS2	6	0	YES	YES	YES	YES
DRSC07116	0.3	2R	CG8233	FLJ10081	12	2	YES	YES	YES	YES
DRSC29282	0.52	X	CG8636		7	0	YES	YES	YES	YES
DRSC05953	0.36	2R	CG8878		5	0	YES	YES	YES	YES
DRSC11630	0.6	3L	Chro		2	0	YES	YES	YES	YES
DRSC27185	0.48	2R	cul-4	CUL4B	5	0	YES	YES	YES	YES
DRSC03295	0.26	2L	CycE	CCNE1	2	0	YES	YES	NO	YES
DRSC11133	0.52	3L	D19A		16	4	YES	YES	YES	YES
DRSC11134	0.52	3L	D19B		12	1	YES	YES	YES	YES
DRSC16932	0.57	3R	dmt		1	0	YES	YES	YES	YES
DRSC04559	1.73	2R	dom		10	1	YES	YES	YES	YES
DRSC23285	0.39	3L	dos		6	0	YES	YES	YES	YES
DRSC07402	0.3	2R	Dp	TFDP2	3	0	YES	YES	YES	YES
DRSC23078	0.35	X	Dsp1	HMGB2	71	6	YES	YES	YES	YES
DRSC07088	0.36	2R	dup	CDT1	10	0	YES	YES	YES	YES
DRSC08583	1.51	3L	E(bx)	BPTF	12	0	YES	YES	YES	YES
DRSC12339	0.64	3R	eIF3-S10	EIF3S10	4	1	YES	YES	YES	YES
DRSC06900	0.54	2R	eIF3-S9	EIF3S9	5	0	YES	YES	YES	YES
DRSC29596	0.36	3L	fol	SLC39A10	4	0	YES	YES	YES	YES
DRSC04984	0.65	2R	geminin		13	2	YES	YES	YES	YES
DRSC21874	0.11	3R	gdfz		0	0	YES	YES	YES	YES
DRSC25237	0.24	X	HDAC4	HDAC4	3	0	YES	YES	YES	YES
DRSC28348	0.43	2L	Hel25E	BAT1	6	0	YES	YES	YES	YES

Table S1 (continued)

DRSC Amplicon	Ratio Firefly/Renilla normalized to plate median	Chrom	Gene	Human Homologene	17 bp Matches	19 bp Matches	proX2-DHS secondary screen	proX2-CES13	pCG2960-roX2-DHS	Passes all secondary screens
DRSC21265	0.48	2L	His2B: CG33870, His2B: CG33868, His2B: CG33882, His2B: CG33894, His2B: CG33910, His2B: CG33892, His2B: CG33876, His2B: CG33898, His2B: CG33872, His2B: CG33908, His2B: CG33888, His2B: CG33902, His2B: CG33890, His2B: CG33886, His2B: CG33874, His2B: CG33884, His2B: CG33878,	His2B: CG33870: HIST2H2BF; His2B: CG33868: HIST2H2BF; His2B: CG33882: HIST2H2BF; His2B: CG33894: HIST2H2BF; His2B: CG33910: HIST2H2BF; His2B: CG33892: HIST2H2BF; His2B: CG33876: HIST2H2BF; His2B: CG33898: HIST2H2BF; His2B: CG33872: HIST2H2BF; His2B: CG33908: HIST2H	8	1	YES	YES	YES	YES
DRSC03546	0.25	2L	hoip	NHP2L1	8	0	YES	YES	YES	YES
DRSC14462	0.36	3R	hvx	CDC73	4	0	YES	YES	YES	YES
DRSC16977	2.31	3R	kay		24	0	YES	YES	YES	YES
DRSC07159	0.44	2R	l(2)05070	PSMB6	8	0	YES	YES	NO	YES
DRSC00708	2.03	2L	lilli		6	0	YES	YES	YES	YES
DRSC14180	0.35	3R	MBD-R2		4	0	YES	YES	YES	YES
DRSC04921	0.42	2R	mle	DHX9	4	0	YES	YES	YES	YES
DRSC18795	0.52	X	mof	MYST1	28	2	YES	YES	YES	YES
DRSC03573	0.23	2L	msl-1		9	0	YES	YES	YES	YES
DRSC28208	0.23	2L	msl-2		12	0	YES	YES	YES	YES
DRSC11371	0.39	3L	msl-3	MSL3L1	3	0	YES	YES	YES	YES
DRSC26774	1.33	X	nej	EP300	2	0	YES	YES	YES	YES
DRSC07738	1.85	2R	Nipped-A	TRRAP	0	0	YES	YES	YES	YES
DRSC23350	0.29	2R	Pcf11		5	0	YES	YES	YES	YES
DRSC26308	0.43	3R	puc	DUSP10	4	0	YES	YES	YES	YES
DRSC03599	0.43	2L	raw		3	0	YES	YES	YES	YES
DRSC07530	1.62	2R	Rho1	RHOC	1	0	YES	YES	YES	YES
DRSC07552	0.31	2R	Sin3A	SIN3A	4	0	YES	YES	YES	YES
DRSC18366	0.47	X	Smg1	SMG1	3	0	YES	YES	YES	YES
DRSC29336	1.95	3R	srp		0	0	YES	NO	YES	YES
DRSC29152	0.46	X	Trf2	TBPL1	9	0	YES	YES	YES	YES
DRSC28776	0.66	X	trr		3	0	YES	YES	YES	YES
DRSC19667	0.34	X	Upf1	UPF1	5	0	YES	YES	YES	YES
DRSC27785	0.34	X	Upf2	UPF2	4	0	YES	YES	YES	YES
DRSC26931	0.39	3R	zfh1	ZEB1	5	0	YES	YES	YES	YES
DRSC11848	0.39	3L	Z4		1	0	YES	YES	YES	YES
DRSC25024	0.32	X	Act5C	ACTG1	4	3	YES	YES	NO	NO
DRSC00396	0.3	2L	dpr3		9	2	YES	NO	YES	NO
DRSC16370	0.63	3R	14-3-3epsilon	YWHAE	6	0	YES	NO	NO	NO
DRSC22850	0.45	3L	AcCoAS	ACSS2	4	1	NO	YES	NO	NO
DRSC04835	0.26	2R	Act42A	ACTG2	10	5	NO	NO	NO	NO
DRSC04042	0.35	2R	Act57B	ACTB	36	10	NO	NO	NO	NO
DRSC11604	0.27	3L	Act79B	ACTA2	20	7	NO	NO	NO	NO
DRSC14105	0.31	3R	Act88F	LOC728378	20	6	NO	NO	NO	NO
DRSC20308	0.18	X	ari-1	ARIH1	9	1	YES	NO	NO	NO
DRSC26196	0.48	3L	bhr	C6orf166	2	0	NO	NO	NO	NO
DRSC11330	0.66	3L	brm	SMARCA4	4	1	NO	YES	YES	NO
DRSC20017	0.47	X	CCKLR-17D3	CCKAR	13	3	NO	YES	NO	NO
DRSC16921	0.33	3R	cdc2c	CDK2	1	0	NO	NO	NO	NO
DRSC23301	0.39	2L	Cdk5alpha		6	0	NO	NO	NO	NO
DRSC12517	2.78	3R	CG10277	RNF13	19	1	NO	NO	NO	NO
DRSC09744	0.47	3L	CG10575	COASY	5	0	NO	YES	NO	NO
DRSC09769	0.68	3L	CG10660		9	2	NO	NO	YES	NO
DRSC04085	1.8	2R	CG10955	RTF1	2	0	NO	NO	NO	NO
DRSC09807	0.56	3L	CG10960	SLC2A8	4	0	YES	NO	YES	NO
DRSC28079	0.56	3L	CG11006	SAP130	5	0	NO	NO	YES	NO
DRSC06042	0.63	2R	CG11125		2	0	NO	NO	NO	NO
DRSC19385	0.56	X	CG11138	C14orf4	43	5	NO	NO	YES	NO
DRSC14329	0.39	3R	CG11575		2	0	NO	YES	YES	NO
DRSC23384	0.27	X	CG11700		3	3	NO	NO	NO	NO
DRSC14379	0.51	3R	CG11771		5	1	NO	NO	YES	NO
DRSC20678	0.55	X	CG12655		2	0	NO	YES	NO	NO
DRSC17876	0.33	X	CG12685		8	1	NO	NO	NO	NO
DRSC06222	0.43	2R	CG12907		62	12	YES	NO	YES	NO
DRSC21857	0.49	2R	CG12928		3	0	NO	YES	NO	NO
DRSC06342	1.85	2R	CG13220		3	0	NO	YES	NO	NO
DRSC09992	0.63	3L	CG13445		73	14	NO	NO	NO	NO
DRSC10080	0.61	3L	CG14075		4	0	NO	NO	NO	NO
DRSC10131	0.67	3L	CG14127	MGC20983	18	4	NO	YES	NO	NO
DRSC14789	1.53	3R	CG14310		4	0	NO	NO	NO	NO
DRSC11710	0.64	3L	CG14453		27	6	NO	YES	YES	NO
DRSC20506	1.68	X	CG14584		0	0	NO	NO	NO	NO

Table S1 (continued)

DRSC Amplicon	Ratio Firefly/Renilla normalized to plate median	Chrom	Gene	Human Homologene	17 bp Matches	19 bp Matches	proX2-DHS secondary screen	proX2-CES13	pCG2960-roX2-DHS	Passes all secondary screens
DRSC12227	0.6	3R	CG14641	RBM22	5	1	YES	NO	YES	NO
DRSC12261	0.65	3R	CG14675		1	0	YES	YES	NO	NO
DRSC17957	2.08	X	CG14799		24	5	NO	YES	YES	NO
DRSC19620	0.14	X	CG15044		7	1	NO	YES	NO	NO
DRSC25087	1.54	X	CG15321		5	0	NO	YES	YES	NO
DRSC02551	0.42	2L	CG16974		6	2	YES	NO	YES	NO
DRSC21723	0.15	2L	CG17293		1	0	NO	NO	NO	NO
DRSC18134	0.23	X	CG17263		11	1	NO	NO	NO	NO
DRSC29783	0.65	3R	CG17816		15	0	NO	NO	YES	NO
DRSC22371	0.38	X	CG1826	BTBD9	2	0	NO	NO	NO	NO
DRSC28176	1.42	X	CG1908		3	0	NO	YES	YES	NO
DRSC24707	1.54	3R	CG1979		2	0	NO	YES	YES	NO
DRSC02705	1.76	2L	CG2042		2	0	NO	NO	NO	NO
DRSC25407	1.62	4	CG2052		4	0	NO	YES	NO	NO
DRSC29255	1.56	3R	CG2097	SYMPK	3	0	NO	NO	YES	NO
DRSC14280	1.94	3R	CG31119	HDAC11	7	1	NO	NO	NO	NO
DRSC15361	1.74	3R	CG31121		4	0	NO	NO	NO	NO
DRSC16143	1.75	3R	CG31159	GFM2	1	0	NO	YES	NO	NO
DRSC25666	0.44	3R	CG31177		0	0	NO	NO	NO	NO
DRSC15115	1.75	3R	CG31223		3	0	NO	YES	NO	NO
DRSC01361	3.62	2L	CG31769		1	0	NO	NO	NO	NO
DRSC22067	0.39	2L	CG31832		3	0	YES	NO	NO	NO
DRSC09265	1.8	3L	CG32121		6	0	NO	NO	YES	NO
DRSC00576	0.43	2L	CG3213		17	2	NO	NO	NO	NO
DRSC15468	0.68	3R	CG3223		20	4	NO	YES	NO	NO
DRSC08413	0.42	3L	CG32486		78	21	YES	NO	YES	NO
DRSC19690	1.96	X	CG32647		2	0	NO	NO	NO	NO
DRSC17990	1.9	X	CG32683		69	14	NO	YES	NO	NO
DRSC10585	0.65	3L	CG33269		59	16	NO	NO	YES	NO
DRSC10584	0.44	3L	CG33270		51	13	YES	NO	YES	NO
DRSC28742	1.78	2L	CG33302		7	0	NO	NO	NO	NO
DRSC12574	1.48	3R	CG34127		18	6	NO	NO	NO	NO
DRSC22012	0.42	2R	CG3493	GOLGA4	14	2	NO	NO	NO	NO
DRSC15241	1.69	3R	CG3678, CG17556	CG3678; TTC35; CG17556; TTC35	2	0	NO	NO	NO	NO
DRSC26013	0.5	2R	CG3825		2	0	YES	NO	YES	NO
DRSC10386	0.52	3L	CG3893		9	0	NO	YES	NO	NO
DRSC22579	0.49	2R	CG40130		0	0	NO	YES	YES	NO
DRSC04459	0.64	2R	CG4286		3	1	NO	YES	NO	NO
DRSC19896	1.72	X	CG4400	BRMS1L	6	0	YES	NO	NO	NO
DRSC24444	2.09	3L	CG4641		9	0	NO	YES	YES	NO
DRSC02784	0.59	2L	CG4839		9	0	NO	NO	NO	NO
DRSC02820	1.84	2L	CG5155	ARMC4	5	0	NO	NO	NO	NO
DRSC11778	0.67	3L	CG5585	RBBP5	3	0	NO	YES	NO	NO
DRSC02876	2.09	2L	CG5681		8	2	NO	YES	NO	NO
DRSC25145	0.36	2R	CG5726		5	0	YES	NO	NO	NO
DRSC02888	3.05	2L	CG5755		1	0	NO	YES	NO	NO
DRSC02889	0.48	2L	CG5758		63	13	YES	NO	YES	NO
DRSC19970	1.64	X	CG6023		9	0	NO	YES	NO	NO
DRSC23372	0.41	3L	CG7177		4	0	YES	NO	NO	NO
DRSC20029	0.54	X	CG7274	ARID4B	3	0	NO	NO	NO	NO
DRSC16237	0.52	3R	CG7379	ING1	25	2	NO	NO	YES	NO
DRSC10880	0.48	3L	CG7564	LUC7L2	5	3	YES	YES	NO	NO
DRSC28378	1.88	3R	CG7568	WDR69	2	0	NO	NO	YES	NO
DRSC29623	0.1	3L	CG7597	CDC2L5	3	0	NO	NO	NO	NO
DRSC09681	0.55	3L	CG7600	FAM91A1	6	3	NO	NO	NO	NO
DRSC04893	0.125	2R	CG7843	ARS2	1	0	NO	NO	NO	NO
DRSC27212	1.64	3L	CG7853	SLC35B3	5	0	NO	NO	NO	NO
DRSC10997	0.34	3L	CG8564		1	0	NO	NO	YES	NO
DRSC11023	0.58	3L	CG8629		6	3	YES	YES	NO	NO
DRSC26566	1.92	2R	CG8726	PXK	1	0	YES	NO	NO	NO
DRSC20179	0.36	X	CG9281	ABCF2	10	2	NO	NO	NO	NO
DRSC16485	0.31	3R	CG9386	C4orf23	0	0	NO	YES	YES	NO
DRSC20690	0.28	X	CG9572		5	2	NO	YES	NO	NO
DRSC19345	0.32	X	CG9634		7	2	NO	NO	NO	NO
DRSC22667	0.3	3R	CG9775		11	0	NO	NO	YES	NO
DRSC29073	0.49	3R	CG9797		6	0	NO	YES	YES	NO
DRSC22961	1.52	3L	CR32314		6	1	NO	YES	NO	NO
DRSC22664	0.49	2L	Csi4	EXOSC1	4	0	NO	YES	YES	NO
DRSC16616	1.78	3R	CstF-64	CSTF2T	27	2	NO	YES	YES	NO
DRSC07599	1.61	2R	csul	PRMT5	5	0	NO	YES	NO	NO

Table S1 (continued)

DRSC Amplicon	Ratio Firefly/Renilla normalized to plate median	Chrom	Gene	Human Homologene	17 bp Matches	19 bp Matches	proX2-DHS secondary screen	proX2-CES13	pCG2960-roX2-DHS	Passes all secondary screens
DRSC09132	0.42	3L	CycA	CCNA2	1	0	YES	NO	YES	NO
DRSC03714	0.2	2L	CycK		4	0	NO	NO	NO	NO
DRSC03307	1.68	2L	Cyp6a16Psi		9	0	NO	NO	NO	NO
DRSC11140	0.41	3L	Dab	DAB2	7	0	NO	NO	YES	NO
DRSC14188	0.56	3R	dpr15		16	2	NO	NO	YES	NO
DRSC25267	1.66	3R	dpr9		14	0	YES	NO	YES	NO
DRSC14986	0.45	3R	dpr9		32	9	YES	NO	YES	NO
DRSC16655	0.42	3R	E2f		11	1	YES	NO	YES	NO
DRSC23583	0.46	3L	eIF-4E	EIF4E	1	0	NO	NO	YES	NO
DRSC07608	1.5	2R	elk	KCNH8	1	0	NO	NO	NO	NO
DRSC22553	0.48	2R	Gr59d		2	0	NO	YES	YES	NO
DRSC26546	1.58	3R	Gr98c		2	0	NO	YES	NO	NO
DRSC20167	0.48	X	Grip128	TUBGCP5	4	0	NO	YES	NO	NO
DRSC16689	0.64	3R	GstD8		11	1	NO	NO	NO	NO
DRSC12352	0.56	3R	Hdac3	HDAC3	7	0	NO	NO	YES	NO
DRSC03760	0.31	2L	His-Psi:CR31616, His-Psi:CR31615		77	67	NO	NO	NO	NO
DRSC03547	1.99	2L	homer	HOMER2	13	1	NO	NO	NO	NO
DRSC16971	0.5	3R	hyd	EDD1	2	1	YES	NO	NO	NO
DRSC00652	0.44	2L	ia2	PTPRN	78	9	NO	NO	YES	NO
DRSC27606	1.28	2R	Jra	JUND	2	0	YES	YES	NO	NO
DRSC10983	1.77	3L	kto	MED12L	0	0	NO	YES	NO	NO
DRSC26021	0.5	2R	l(2)dtl		1	0	YES	NO	YES	NO
DRSC26919	1.53	2L	Liprin-alpha	PPFIA1	2	0	NO	NO	NO	NO
DRSC27862	0.43	2R	lola		11	0	YES	NO	YES	NO
DRSC29636	0.52	X	m		38	0	YES	YES	NO	NO
DRSC04702	0.59	2R	mAcR-60C	CHRM3	84	24	NO	YES	NO	NO
DRSC06309	0.46	2R	MCPH1		4	2	YES	YES	NO	NO
DRSC06960	1.97	2R	mip120	LIN54	10	3	NO	NO	YES	NO
DRSC18644	2	X	mus81	MUS81	2	1	NO	NO	NO	NO
DRSC01946	5.57	2L	nAcRalpha-34E	CHRNA7	2	0	NO	NO	NO	NO
DRSC16184	1.55	3R	Nrx-1		5	0	NO	YES	NO	NO
DRSC28865	0.52	2R	Nxt1	NXT1	0	0	YES	YES	YES	NO
DRSC28374	2	3R	Osi2		2	0	NO	NO	YES	NO
DRSC27176	0.45	3R	Osi22		4	0	NO	YES	NO	NO
DRSC17219	2.76	4	pho		4	0	NO	NO	NO	NO
DRSC12323	0.62	3R	Pl4KIalpha	PI4K2B	18	4	NO	NO	YES	NO
DRSC03591	0.32	2L	pim		4	0	NO	NO	NO	NO
DRSC23503	0.24	2L	PNUTS	PPP1R10	5	0	NO	NO	NO	NO
DRSC23312	0.4	2R	PpD5		3	0	NO	NO	NO	NO
DRSC07519	0.36	2R	Psi	FUBP1	18	2	NO	NO	NO	NO
DRSC20364	0.56	X	ran	RAN	7	1	NO	NO	YES	NO
DRSC23541	0.46	2R	RnrS	RRM2	2	0	YES	NO	NO	NO
DRSC25010	0.36	X	RpS6	RPS6	2	0	NO	YES	YES	NO
DRSC29803	0.61	X	sbr	NXF1	1	0	YES	NO	YES	NO
DRSC16845	0.43	3R	SF2	SFRS1	3	0	NO	YES	NO	NO
DRSC19564	0.41	X	Shawn, Tyler	Shawn: SLC25A40	2	1	NO	NO	NO	NO
DRSC11898	1.88	3L	skd	THRAP2	6	0	NO	YES	YES	NO
DRSC27336	0.5	2R	skf	MPP7	4	0	NO	NO	NO	NO
DRSC27834	0.22	2L	Smg5	SMG5	1	0	NO	NO	NO	NO
DRSC27703	0.25	3R	Smg6	SMG6	2	0	NO	YES	YES	NO
DRSC25781	0.38	X	Smr		8	0	YES	NO	YES	NO
DRSC06458	1.67	2R	snoRNA:U31:54Ec, snoRNA:snR38:54Eb, snoRNA:U27:54Eb, snoRNA:snR38:54Ec, snoRNA:U27:54Ec, snoRNA:snR38:54Ea, snoRNA:U31:54Ed, Uhg1, snoRNA:U31:54Eb, snoRNA:U29:54Ed, snoRNA:U27:54Ea		13	4	NO	NO	NO	NO
DRSC28782	0.53	3R	ssh		1	0	NO	YES	NO	NO
DRSC25879	0.42	X	sta	RPSA	3	0	NO	NO	NO	NO
DRSC15332	0.44	3R	stumps		17	2	NO	NO	YES	NO
DRSC18839	0.17	X	su(s)		6	0	NO	NO	NO	NO
DRSC09741	0.58	3L	Tektin- C	TEKT1	14	1	YES	NO	YES	NO
DRSC11303	0.28	3L	TfIIIEalpha	GTF2E1	9	1	NO	NO	NO	NO
DRSC17083	1.6	3R	tin		7	0	NO	NO	NO	NO
DRSC26920	1.43	3L	Unr	CSDE1	2	0	NO	YES	NO	NO
DRSC19968	0.23	X	upd3		10	0	YES	NO	NO	NO
DRSC27316	0.38	2R	Upf3	UPF3B	2	0	NO	NO	NO	NO
DRSC17097	0.47	3R	yemalpha	FLJ25778	11	1	NO	NO	NO	NO

Table S1 (continued)

DRSC Amplicon	Ratio Firefly/Renilla normalized to plate median	Chrom	Gene	Human Homologene	17 bp Matches	19 bp Matches	proX2-DHS secondary screen	proX2-CES13	pCG2960-roX2-DHS	Passes all secondary screens
DRSC00267	0.46	2L	MED15	PCQAP	117	45	N/A	N/A	N/A	N/A
DRSC09795	0.69	3L	CG10743	PPF1BP1	93	13	N/A	N/A	N/A	N/A
DRSC11697	0.61	3L	CG32428		132	20	N/A	N/A	N/A	N/A
DRSC13053	0.26	3R	CG3563		248	117	N/A	N/A	N/A	N/A
DRSC09450	0.43	3L	Lmpt	FHL2	258	28	N/A	N/A	N/A	N/A
DRSC05106	0.33	2R	mXr	GRM8	101	11	N/A	N/A	N/A	N/A
DRSC03502	0.49	2L	cad	CDX2	279	146	N/A	N/A	N/A	N/A
DRSC10133	0.65	3L	CG14129		195	30	N/A	N/A	N/A	N/A
DRSC18492	0.48	X	CG3056		223	52	N/A	N/A	N/A	N/A
DRSC11048	0.6	3L	CG32133		437	49	N/A	N/A	N/A	N/A
DRSC22786	0.16	3L	Deaf1	DEAF1	58	12	N/A	N/A	N/A	N/A
DRSC14703	0.59	3R	epsin-like	CLINT1	117	15	N/A	N/A	N/A	N/A
DRSC08235	0.48	3L	MED14	CRSP2	100	6	N/A	N/A	N/A	N/A
DRSC13670	0.47	3R	nAcRalpha-96Aa	CHRNA3	234	105	N/A	N/A	N/A	N/A
DRSC04050	0.64	2R	otp	OTP	191	19	N/A	N/A	N/A	N/A
DRSC09263	0.54	3L	sens		810	477	N/A	N/A	N/A	N/A
DRSC16884	1.53	3R	TFIIFbeta	GTF2F2	82	6	N/A	N/A	N/A	N/A
DRSC00256	0.47	2L	CG15628		448	152	N/A	N/A	N/A	N/A
DRSC02519	0.44	2L	CG16812		107	15	N/A	N/A	N/A	N/A
DRSC00033	0.4	2L	HDC00125 ('+' in Stolc et al) ('+' in Hild et al)		34	9	N/A	N/A	N/A	N/A
DRSC05237	0.42	2R	HDC05653 ('+' in Stolc et al) ('+' in Hild et al)		236	116	N/A	N/A	N/A	N/A
DRSC21860	0.43	2R	HDC05694 ('+' in Hild et al)		5	0	N/A	N/A	N/A	N/A
DRSC05425	0.48	2R	HDC06207 ('+' in Stolc et al) ('+' in Hild et al)		892	645	N/A	N/A	N/A	N/A
DRSC05715	0.33	2R	HDC07108 ('+' in Stolc et al) ('+' in Hild et al)		244	93	N/A	N/A	N/A	N/A
DRSC09294	0.39	3L	HDC10050 ('+' in Stolc et al) ('+' in Hild et al)		30	9	N/A	N/A	N/A	N/A
DRSC12785	0.3	3R	HDC12852 ('+' in Stolc et al)		36	4	N/A	N/A	N/A	N/A
DRSC13340	0.39	3R	HDC14424 ('+' in Stolc et al) ('+' in Hild et al)		806	488	N/A	N/A	N/A	N/A
DRSC13727	0.48	3R	HDC15639 ('+' in Stolc et al)		35	9	N/A	N/A	N/A	N/A
DRSC26753	1.64	3R	HDC16185 ('+' in Stolc et al)		8	0	N/A	N/A	N/A	N/A
DRSC17554	0.44	X	HDC17958 ('+' in Stolc et al) ('+' in Hild et al)		697	433	N/A	N/A	N/A	N/A
DRSC19217	0.12	X	HDC19473 ('+' in Stolc et al) ('+' in Hild et al)		13	2	N/A	N/A	N/A	N/A
DRSC22703	0.45	X	HDC19721 ('+' in Hild et al)		5	2	N/A	N/A	N/A	N/A
DRSC20849	0.53	2Rh	HDC20253		0	0	N/A	N/A	N/A	N/A
DRSC21202	0.25	U	HDC20586 ('+' in Hild et al)		2	0	N/A	N/A	N/A	N/A
DRSC21223	0.23	2Rh	HDC20606		1	0	N/A	N/A	N/A	N/A
DRSC04624	1.77	2R	Mov34	PSMD7	11	1	N/A	N/A	N/A	N/A
DRSC11947	0.67	3L	Qm		0	0	N/A	N/A	N/A	N/A
DRSC10798	0.57	3L	Rpl10Ab	RPL10A	12	3	N/A	N/A	N/A	N/A
DRSC26217	0.43	2R	Rpl11	RPL11	1	0	N/A	N/A	N/A	N/A
DRSC03416	0.53	2L	Rpl13	RPL13	1	0	N/A	N/A	N/A	N/A
DRSC11269	0.63	3L	Rpl14	RPL14	5	0	N/A	N/A	N/A	N/A
DRSC20963	0.44	3Lh	Rpl15		3	0	N/A	N/A	N/A	N/A
DRSC29971	0.6	X	Rpl17	RPL17	3	0	N/A	N/A	N/A	N/A
DRSC22323	0.38	3L	Rpl18	RPL18	6	0	N/A	N/A	N/A	N/A
DRSC07538	0.38	2R	Rpl18A	LOC285053	7	0	N/A	N/A	N/A	N/A
DRSC04649	0.49	2R	Rpl19	RPL19	3	1	N/A	N/A	N/A	N/A
DRSC03704	0.63	2L	Rpl21	RPL21, LOC729402	2	0	N/A	N/A	N/A	N/A
DRSC25025	0.36	X	Rpl22	RPL22	3	0	N/A	N/A	N/A	N/A
DRSC04648	0.63	2R	Rpl23	RPL23	2	0	N/A	N/A	N/A	N/A
DRSC25828	0.41	3L	Rpl23A	hCG_16001	5	0	N/A	N/A	N/A	N/A
DRSC10726	0.61	3L	Rpl26	RPL26	9	0	N/A	N/A	N/A	N/A
DRSC25890	0.4	3R	Rpl27	RPL27	2	0	N/A	N/A	N/A	N/A
DRSC08293	0.53	3L	Rpl28	RPL28	3	1	N/A	N/A	N/A	N/A
DRSC16835	0.53	3R	Rpl32	RPL32	6	1	N/A	N/A	N/A	N/A
DRSC12302	0.51	3R	Rpl35A	RPL35A	1	0	N/A	N/A	N/A	N/A
DRSC21646	0.34	X	Rpl37a	RPL37	4	0	N/A	N/A	N/A	N/A
DRSC25032	0.48	3R	Rpl4	RPL4	2	0	N/A	N/A	N/A	N/A
DRSC29038	0.63	2L	Rpl5		1	0	N/A	N/A	N/A	N/A
DRSC25982	0.4	3R	Rpl6	RPL6	4	0	N/A	N/A	N/A	N/A
DRSC29340	0.51	2L	Rpl7	RPL7, HCG_31916	6	0	N/A	N/A	N/A	N/A
DRSC28833	0.38	X	Rpl7A	RPL7A	1	0	N/A	N/A	N/A	N/A
DRSC29905	0.65	3L	Rpl8	RPL8	4	0	N/A	N/A	N/A	N/A
DRSC11878	0.61	3L	RplP0	RPLP0	2	0	N/A	N/A	N/A	N/A
DRSC23411	0.39	2L	RplP1	RPLP1	2	0	N/A	N/A	N/A	N/A
DRSC07539	0.57	2R	RplP2	RPLP2	27	3	N/A	N/A	N/A	N/A
DRSC22700	0.29	2R	RpS11	RPS11	0	0	N/A	N/A	N/A	N/A
DRSC03419	0.54	2L	RpS13	RPS13	4	1	N/A	N/A	N/A	N/A
DRSC23462	0.36	2R	RpS15	RPS15	5	0	N/A	N/A	N/A	N/A
DRSC19831	0.57	X, 2R	RpS15Ab, RpS15Aa		6	0	N/A	N/A	N/A	N/A
DRSC04442	0.65	2R	RpS16	RPS16	2	1	N/A	N/A	N/A	N/A
DRSC11271	0.63	3L	RpS17	RPS17	4	0	N/A	N/A	N/A	N/A
DRSC07540	0.43	2R	RpS18	RPS18	1	0	N/A	N/A	N/A	N/A
DRSC28412	0.43	2R	RpS24	RPS24	5	0	N/A	N/A	N/A	N/A
DRSC29946	0.6	3R	RpS27	RPS27	7	0	N/A	N/A	N/A	N/A
DRSC16433	0.64	3R	RpS29, snoRNA:783	RpS29: RPS29	3	0	N/A	N/A	N/A	N/A
DRSC11272	0.57	3L	RpS4	RPS4X	1	0	N/A	N/A	N/A	N/A
DRSC00445	0.26	2L			13	3	N/A	N/A	N/A	N/A
DRSC14104	0.29	NA			11	5	N/A	N/A	N/A	N/A
DRSC21210	0.2	NA			19	7	N/A	N/A	N/A	N/A
DRSC36508	0.39	2L			3	0	N/A	N/A	N/A	N/A
DRSC03788	0.35	2L			0	0	N/A	N/A	N/A	N/A

Table S1

Initial candidates identified from a genome-wide RNAi screen to identify novel regulators of dosage compensation.

Key	
Blue	Passes all secondary screens
Green	Do not pass all secondary screens
Orange	Not rescreened

Table S2. Final candidates identified from a genome-wide RNAi screen.

Gene	Human homologue	Yeast Homologue	Firefly/Renilla	Description	Class
CG4699			0.17	NSL1	1
CG4756	SAP30L		0.24	Sin3-associated protein	1
HDAC4	HDAC4		0.24	histone deacetylase	1
CG8233	FLJ10081		0.3	NSL3	1
CG1135	MCRS1		0.31	NSL component	1
Sin3A	SIN3A	SIN3	0.31	Sin3 corepressor	1
Dsp1	HMG2		0.35	HMG box protein	1
MBO-R2			0.35	DNA binding TF	1
hvx	CDC73		0.36	paf component	1
CG2469	CTR9	CTR9	0.38	paf component	1
Z4			0.39	z4	1
CG14220	SUD53		0.4	Sin3A complex member	1
atms	PAF1		0.41	Paf1	1
Trf2	TBPL1		0.46	TBP-like protein	1
His2B:CG3387	His2B:CG33870: HIST2	His2B:CG33870: HT	0.48	H2B	1
Atu	LEO1	LEO1	0.55	Leo1 Paf complex	1
Chro			0.6	chromatin architecture	1
trr			0.66	K4 methyltransferase	1
CG1832			0.23	transcription factor	2
Dp	TFDP2		0.3	Transcription factor related to E2F	2
Bx	LMO1		0.33	transcription factor	2
zfh1	ZEB1		0.39	zinc-finger transcription factor	2
D19A			0.52	transcription factor	2
D19B			0.52	transcription factor	2
nei	EP300		1.33	CBP/p300	3
E(bx)	BPTF		1.51	Nurf301	3
CG16975	L3MBTL2		1.73	transcription factor repressor	3
dom		SWR1	1.73	SWR1 ATP-dependent helicase	3
CG8211	INTS2		1.85	Integrator complex RNA Pol II associated 3' end formation	3
Nipped-A	TRRAP	TRA1	1.85	TRA1	3
caz	FUS		1.99	TFIID complex	3
srp			1.99	DNA-binding TF	4
lilli			2.03	DNA binding TF involved in neural development similar to AFF4	4
kay			2.31	DNA binding TF involved in neural development dpp pathway	4
hoip	NHP2L1	SNU13	0.25	mRNA splicing/export	5
Pcf11			0.29	mRNA cleavage factor	5
Cbp80	NCBP1		0.33	RNA cap-binding protein	5
Upf1	UPF1	NAM7	0.34	helicase involved in nonsense-mediated decay	5
Upf2	UPF2	NMD2	0.34	nonsense-mediated decay	5
Hel25E	BAT1	SUB2	0.43	Sub2 helicase involved in TREX complex	5
Smg1	SMG1		0.47	nonsense-mediated decay and microRNA function	5
glzf			0.11	glutathione transferase activity	6
CycE	CCNE1		0.26	cyclin E	6
CG6015	CDC40		0.27	Splicing factor	6
CG8878			0.36	kinase activity	6
dup	CDT1		0.36	DNA replication factor	6
foi	SLC39A10		0.36	zinc ion transporter	6
dos			0.39	SH2/SH3 protein receptor	6
CG5514			0.41	putative cell wall component	6
puc	DUSP10		0.43	phosphatase involved in dorsal closure	6
raw			0.43	autophagic cell death	6
CG11451			0.44	cell communication	6
l(2)05070	PSMB6	PRE3	0.44	proteasome/endopeptidase	6
cdc2	CDC2		0.48	cdc2	6
cul-4	CUL4B		0.48	ubiquitin-ligase	6
CG10486			0.52	membrane cation transport	6
CG31258			0.52	No known function	6
CG32369			0.52	proteolysis	6
CG8636			0.52	mitotic spindle formation	6
eIF3-S9	EIF3S9	PRT1	0.54	translation initiation factor	6
dmt			0.57	nuclear nervous system development	6
eIF3-S10	EIF3S10	RPG1	0.64	translation initiation factor	6
CG31211	C6orf111		0.65	No known function	6
geminin			0.65	repressor of DNA replication	6
CG5146			0.66	No known function	6
Rho1	RHOC		1.62	GTPase activity	6
CG3363			2.86	No known function	6
msl-1			0.23	MSL1	MSL
msl-2			0.23	MSL2	MSL
msl-3	MSL3L1		0.39	MSL3	MSL
mle	DHX9		0.42	mle	MSL
mof	MYST1		0.52	mof	MSL

Class 1: General transcriptional regulators that activate
Class 2: Proteins that have potential specific DNA binding activity and activate our reporter
Class 3: General transcriptional regulators that repress our reporter
Class 4: Proteins that have potential specific DNA binding activity and repress our reporter
Class 5: Proteins that regulate RNA metabolism
Class 6: Proteins that are previously unstudied and/or have functions unrelated to transcription or RNA metabolism

Table S2

Final candidates identified from a genome-wide RNAi screen. Candidates are color-coded and divided into functional categories described at the bottom of the table

**Chapter 3: The CLAMP protein links the MSL complex to the
X-chromosome during *Drosophila* dosage compensation**

This chapter is a co-first authored article in the journal *Genes and Development* and remains in its published format. My specific contributions to this article were developing a system to successfully express the CLAMP protein and performing PBM (Protein Binding Microarray) experiments with CLAMP to determine the precise DNA sequence that CLAMP binds to *in vitro*. After expressing both the 4 and the 6 C-terminal C2H2 zinc fingers of CLAMP in separate constructs, I tested their binding specificity on the PBMs using a fluorescently labeled antibody to the GST tag that was fused to the N-terminus of these proteins (Fig. 2B). The PBMs were conducted in collaboration with Martha Bulyk's laboratory where array analysis was done by Trevor Siggers, PhD. Both the 4 and 6 zinc Fingers of CLAMP directly bound the MRE sequences. The PBM results were significant because they revealed for the first time in the *Drosophila* Dosage Compensation field the previously missing factor that can directly bind to the MRE sequences that are necessary for recruitment of the MSL complex that regulates Dosage Compensation. In addition, I performed CLAMP Chip in *Drosophila* larvae that had been genetically engineered with MRE sequences in which the 8bp core had been mutated. The 8bp core of MREs is the most conserved portion of the MREs and critical for MSL recruitment. My CLAMP Chip results demonstrated for the first time that the wild type 8bp core of MREs is required for CLAMP recruitment *in vivo* as mutating this 8bp core lead to significant loss of CLAMP recruitment (Fig. S2D). These results provided *in vivo* validation for the direct binding of CLAMP to MREs.

**The CLAMP protein links the MSL complex to the X chromosome
during *Drosophila* dosage compensation**

Marcela M.L. Soruco*, **Jessica Chery***, Eric P. Bishop*, Trevor Siggers, Michael Y.
Tolstorukov, Alexander R. Leydon, Arthur U. Sugden, Karen Goebel, Jessica Feng,
Peng Xia, Anastasia Vedenko, Martha L. Bulyk, Peter J. Park, and Erica Larschan,

Genes & Dev. 2013.

Abstract

The *Drosophila* Male-Specific Lethal (MSL) dosage compensation complex increases transcript levels on the single male X-chromosome to equal the transcript levels in XX females. However, it is not known how the MSL complex is linked to its DNA recognition elements, the critical first step in dosage compensation. Here, we demonstrate that a previously uncharacterized zinc-finger protein, CLAMP (Chromatin-Linked Adaptor for MSL Proteins) functions as the first link between the MSL complex and the X-chromosome. CLAMP directly binds to the MSL complex DNA recognition elements and is required for the recruitment of MSL complex. The discovery of CLAMP identifies a key factor required for the chromosome-specific targeting of dosage compensation, providing new insights into how sub-nuclear domains of coordinate gene regulation are formed within metazoan genomes.

Introduction

Diverse mechanisms have evolved to correct for dosage imbalances between genes on the X-chromosomes and autosomes in mammals, *C. elegans*, and *Drosophila* (Gelbart and Kuroda 2009). In mammals, dosage compensation involves inactivation of one of the two X-chromosomes in females (Nguyen and Disteché 2006; Payer and Lee 2008; Meyer 2010). In *C. elegans*, gene expression along each of the two hermaphroditic X-chromosomes is reduced (Meyer 2010). In *Drosophila*, dosage compensation increases the transcript levels from most active genes along the length of the single male X-chromosome by at most two-fold to equal expression from the two female X-chromosomes (Belote and Lucchesi 1980b; Hamada et al. 2005; Larschan et al. 2007). A unifying feature of these dosage compensation mechanisms is that they all involve distinguishing the X-chromosome from autosomes to coordinately regulate a large number of genes along the length of the X-chromosome(s). However, the mechanisms used to generate the precise chromosome-specificity of dosage compensation are not well understood.

The key regulator of dosage compensation in *Drosophila* is the Male Specific Lethal (MSL) complex, which is expressed specifically in males and distinguishes the X-chromosome from the autosomes (Belote and Lucchesi 1980b; Gelbart and Kuroda 2009). It has been hypothesized that the MSL complex identifies its X-chromosome target genes in a three-stage process. First, the MSL complex recognizes the sites of transcription of its *roX1* and *roX2* long non-coding RNA components, which are

encoded on the X-chromosome, and co-transcriptionally incorporates them (Kelley et al. 1999; Kelley et al. 2008). Second, the MSL complex recognizes 150-300 additional genomic loci distributed across the X-chromosome called “Chromatin Entry Sites” (CES). CES were defined by their high levels of MSL complex occupancy and retention of the core MSL components in the absence of MSL3 (Kelley et al. 1999; Alekseyenko et al. 2008). Most CES contain one or multiple 21-bp GA-rich *cis*-acting DNA sequences called MSL Recognition Elements (MREs) that are required for MSL complex recruitment, but there are no known sequence-specific contacts between MSL complex and MREs (Alekseyenko et al. 2008; Straub et al. 2008; Fauth et al. 2010). Third, MSL complex spreads from CES to the bodies of active genes on the X-chromosome by recognizing histone modifications associated with transcriptional activation, including H3K36me3 (Larschan et al. 2007; Bell et al. 2008). Once active genes are recognized, their transcript levels are increased by the MSL complex (Hamada et al. 2005; Larschan et al. 2011b).

While both the *roX* non-coding RNAs and MREs contribute to the recognition of the male X-chromosome, they are individually insufficient to generate X-chromosome specificity (Meller and Rattner 2002b; Park et al. 2002; Alekseyenko et al. 2008; Straub et al. 2008). Therefore, we hypothesized that additional factors are required to tether and enrich the MSL complex at its target sites on the male X-chromosome. Thus, we performed a cell-based genome-wide RNAi screen, allowing for the possibility that MSL regulators might have genome-wide roles in both males and females in addition to X-specific roles in males (Larschan et al. 2012). Regulators

with non-sex specific functions as well as male-specific functions would not have been recovered from the powerful male-specific lethal screens that identified all of the MSL complex components (Belote and Lucchesi 1980b). Such non-sex specific regulators provide a key opportunity to understand the first steps in X-identification including recognition of highly-conserved MREs within CES (Alekseyenko et al 2013).

To define new regulators of MSL complex targeting or function, our genome-wide RNAi screen identified proteins that specifically modulate the activity of a luciferase reporter gene fused to the CES adjacent to *roX2* (Larschan et al. 2012). Among many other candidate positive regulators, we identified the conserved non-sex specific CG1832 protein that has a glutamine-rich N-terminus and a C-terminal domain containing seven C2H2 zinc fingers (Larschan et al. 2012). We found that CG1832 is enriched at CES, and that CG1832 RNAi strongly reduces MSL complex recruitment at five CES tested in male tissue culture cells (Larschan et al. 2012). However, the key questions remained: 1) Does CG1832 provide a direct link between the MSL complex and the X-chromosome? 2) Does CG1832 regulate MSL complex recruitment to the entire X-chromosome in the fly? 3) Is CG1832 enriched on the X-chromosome independent of the MSL complex to facilitate the discrimination of the X-chromosome from autosomes?

Here, we have named CG1832: CLAMP (Chromatin-Linked Aaptor for MSL Proteins). We identify CLAMP as the previously unknown link between the MSL

complex and the X-chromosome. First, we show that CLAMP directly recognizes MREs and is required for recruitment of the MSL complex to the entire X-chromosome. Second, we demonstrate that CLAMP is enriched at key CES independent of MSL complex. Third, we show that the MSL complex and CLAMP have a synergistic occupancy relationship that increases the X-enrichment of both factors at critical CES such as those at the *roX* loci. Our results support a mechanism whereby the MSL complex uses the X-enriched CLAMP protein to mark CES, followed by further enrichment from synergistic interactions between the two factors.

Results and Discussion

CLAMP directly recognizes MREs to tether MSL complex to the X-chromosome

To determine whether CLAMP and the MSL complex co-localize at CES *in vivo*, we performed CLAMP ChIP-seq experiments in male SL2 cells and compared CLAMP occupancy profiles with available MSL complex occupancy profiles (Fig. 1) (Alekseyenko et al. 2008; Larschan et al. 2012). We found CLAMP occupancy at many loci throughout the genome (X: 2695; autosomes: 10009) and strong colocalization with MRE sequences (Figs. 1A,B Fig. S1). Moreover, CLAMP and MSL complex exhibit their highest occupancy levels when colocalized (Figs. 1A; Fig. S2A,B). Non-overlapping CLAMP occupancy sites occur at 5' ends of active genes on all chromosomes (Fig. S2C) and non-overlapping MSL binding sites are localized

over the bodies of active X-linked genes (Fig. S2B). Furthermore, CLAMP sites occur precisely over MREs within CES compared with the broader domains of MSL complex occupancy (Figs. 1A,B, 4A). Together, these patterns are consistent with a role for CLAMP in recruiting MSL complex to CES followed by spreading of MSL complex to the bodies of active genes and additional non-sex specific roles for CLAMP at 5' ends of active genes.

Next, we used the MEME software package (Bailey et al. 2009) to generate a position weight matrix (PWM) representing the most enriched sequence within 500 regions surrounding CLAMP occupancy sites (2 kb each) at the 5' and 3' ends of genes genome-wide. All CLAMP localization motifs on the X-chromosome and on autosomes shared an 8-bp GA-rich core region that is very similar to the previously identified MRE consensus sequence (Fig. 2A). Therefore, we hypothesized that CLAMP functions as an MRE recognition factor.

To test this hypothesis, we characterized the *in vitro* DNA-binding specificity of CLAMP. We used protein-binding microarrays (PBMs) to assay CLAMP binding to all possible 10-bp sequences (Berger et al. 2006). Universal PBMs have been used to define binding specificities for hundreds of proteins from diverse classes of sequence-specific DNA binding transcription factors and from a wide range of species (Robasky and Bulyk 2011). Each PBM is a double-stranded DNA array that contains all possible 10-bp sequences embedded within variable flanking sequences. The CLAMP zinc finger domain contains seven C2H2 zinc fingers. After testing the

complete seven-finger construct and many additional combinations of zinc fingers, we were able to express soluble protein for the C-terminal 4-finger and 6-finger zinc-finger regions of CLAMP (see Methods). The *in vitro* PBM analysis for these proteins yielded statistically significant motifs with high PBM enrichment (6-finger motif: seed 8-bp PBM E-score = 0.497; 4-finger motif: seed 8-bp PBM E-score = 0.498) in excellent agreement with the CLAMP *in vivo* binding site motif and the previously defined MRE consensus (Fig. 2B). Therefore, CLAMP binds directly to the MRE motif *in vitro*.

Next, we examined whether MREs are required for the recruitment of CLAMP *in vivo* in male and female larvae using available stocks with three intact or mutated MREs embedded within a minimal 150 bp fragment of CES5C2 inserted site-specifically into an autosomal location (Alekseyenko et al., 2008) (Fig. S2D). Occupancy of both MSL complex and the CLAMP protein are both greater on the X-chromosome compared with the autosomal insertion, likely due to the more open chromatin environment on the X-chromosome compared with autosomes (Alekseyenko et al., 2008) (Fig. S2D). The autosomal insertion with three intact MREs shows greater occupancy of CLAMP than that in which the 8-bp core of each of the three MREs is mutated. Therefore, MREs are required for CLAMP recruitment in both males and females, as expected for a non-sex specific factor. In males, the occupancy of CLAMP at the insertion of CES5C2 with intact MREs is greater than that in females, consistent with synergistic interactions between CLAMP and the MSL complex that are described in more detail below.

To determine whether enrichment at CES is specific to CLAMP or is a common feature shared with other factors that can recognize MRE-like sequences, we examined a well-characterized *Drosophila* DNA-binding protein, the GAGA factor (GAF) (Lee et al. 2008). In contrast to CLAMP, which is strongly enriched at CES, GAF occupancy was lower at CES compared with other GAF binding sites throughout the genome (Figs. S2E, F). These data suggest that CLAMP, but not GAF, directly recognizes MRE elements within CES. In addition, CLAMP was recently shown to physically associate with the MSL complex on chromatin using an approach that combined chromatin immunoprecipitation with mass spectrometry (Wang et al. 2013b). Therefore, CLAMP directly recognizes MREs and physically associates with the MSL complex, providing the first link between the MSL complex and the X-chromosome. Due to the high degree of conservation of both CLAMP (www.flybase.org) and the MRE motif (Alekseyenko et al. 2013) across distantly related *Drosophila* species, it is likely that the CLAMP-MRE interaction is highly conserved.

CLAMP is required for MSL complex recruitment to the male X-chromosome *in vivo*

We previously demonstrated that CLAMP is required for recruitment of the MSL complex to 5 CES in SL2 male tissue culture cells (Larschan et al. 2012). To determine whether CLAMP is required for MSL complex recruitment to the entire X-chromosome in flies, we used a publicly available RNAi transgenic *Drosophila* line to reduce CLAMP mRNA levels (see Methods; www.flyrnai.org). Because CLAMP is highly expressed in all tissues, we expressed the RNAi construct ubiquitously (see Methods; www.flybase.org). Although the CLAMP RNAi construct did not completely eliminate CLAMP mRNA levels (Fig. S3A), immunostaining of polytene chromosomes for CLAMP (Figs. 3A, S3D) and CLAMP ChIP occupancy (Figs. S3B,C) were strongly reduced in this line. CLAMP RNAi causes a pupal lethal phenotype in both males and females (Larschan et al. 2012) and no deficiencies are available due to pericentric location of the CLAMP gene (www.flybase.org).

Next, we used ChIP for the MSL2 core component to examine MSL complex recruitment to five CES in the presence and absence of CLAMP RNAi treatment (Fig. S3F). Our results showed that CLAMP RNAi strongly reduced the occupancy of the MSL complex in male larvae (Fig. S3F). As a control, we measured H3 occupancy by ChIP to determine whether CLAMP RNAi globally disrupts chromatin. In contrast to the strong effects on MSL complex occupancy, we found that CLAMP RNAi induced only modest effects on H3 ChIP occupancy at CES (Fig. S3G). We also performed

polytene staining for the MSL3 protein after CLAMP RNAi and observed no detectable MSL3 staining, revealing that CLAMP is required for MSL complex recruitment to the entire X-chromosome (Fig. S3B). Thus, CLAMP promotes MSL complex localization to the entire X-chromosome in flies.

It was previously shown that loss of targeting of the MSL complex reduces complex stability, likely due to ubiquitylation of complex members by MSL2 (Copps et al. 1998; Villa et al. 2012). Therefore, measuring MSL protein levels after CLAMP RNAi would not distinguish between effects on complex stability and targeting to the X-chromosome. However, we previously demonstrated that CLAMP RNAi does not affect mRNA levels of MSL complex components (Larschan et al. 2012) and CLAMP is strongly enriched at 92% of CES (Fig. 1, S1), suggesting that indirect effects on MSL complex stability are unlikely (Larschan et al. 2012).

The MSL complex is highly enriched on the X-chromosome compared to autosomal controls (Fig. S3B) (Alekseyenko et al. 2008), but it mediates at most a two-fold increase in transcription. Furthermore, only a 1.4-fold decrease in expression of X-linked genes is seen after MSL2 RNAi in male SL2 cells (Hamada et al. 2005). To determine how CLAMP regulates gene expression globally, we conducted mRNA-seq experiments in male SL2 cells before and after CLAMP RNAi treatment because tissue culture cells are a uniform cell population, unlike whole animals. As expected for a non-sex specific factor present on both the X-chromosome and autosome (Fig. S2, S3B), CLAMP RNAi caused changes in gene expression genome-wide. However,

CLAMP RNAi caused a greater decrease in the gene expression of X-linked genes compared to autosomal genes ($p < 9.1E-5$), consistent with a function in MSL complex targeting. Consistent with its localization throughout the genome and enrichment on the X-chromosome (Fig. S1), CLAMP RNAi affects gene expression on both the X-chromosome and autosomes with a bias towards X-linked genes.

In contrast to average X-linked genes, the genes that encode the *roX* non-coding RNA components are highly dependent on the MSL complex (Bai et al. 2004). Therefore, we measured the levels of *roX* RNAs before and after CLAMP RNAi treatment by qRT-PCR in male and female third-instar larvae (Fig. S3A). In males, we found that *roX2* levels were reduced 4-fold after CLAMP RNAi, indicating that CLAMP serves as an activator of *roX2* expression. In contrast, *roX1* levels were largely unchanged in males after CLAMP RNAi. In females, *roX1* was significantly increased after CLAMP RNAi treatment (6.5-fold) and *roX2* levels remained largely unchanged. These patterns of CLAMP-mediated regulation agree with the established differential regulation of *roX1* and *roX2* and recent genetic evidence implicating *roX2* as the first site of X-identification (Meller and Rattner 2002b; Bai et al. 2004; Lim and Kelley 2012).

CLAMP is enriched on the X-chromosome independent of MSL complex

Identifying CLAMP as the protein that directly recognizes the X-chromosome during *Drosophila* dosage compensation provides a key opportunity to reveal new insight

into the process of X-identification. We hypothesized that X-enrichment of CLAMP independent of the MSL complex would indicate that CLAMP could function as a beacon on the X-chromosome to distinguish it from autosomes.

To determine how the MSL complex regulates CLAMP occupancy, we compared CLAMP occupancy at MREs on the X-chromosome and on autosomes in male cells before and after MSL2 RNAi, and in female cells that lack MSL complex. Overall, we found that the percentage of MREs occupied by CLAMP was significantly greater on the X-chromosome when compared with autosomes in both male cells (50% on X-chromosome vs. 32% on autosomes) and female cells (62% on X-chromosome vs. 38% on autosomes) (Figs. S4 A,B,C). Therefore, CLAMP is enriched at X-chromosome MREs compared with autosomal MREs, independent of the MSL complex.

Next, we determined how CLAMP occupancy varies in its dependence on the MSL complex at MREs within CES. To define the total number of likely CES for subsequent analysis, we determined how the largest class of previously characterized CES (306 loci) (Sural et al. 2008) overlapped with CLAMP sites present in male cells and found 265 common sites (87% overlap). Then, we clustered these sites based on their overall binding pattern and occupancy level of CLAMP in male cells, male cells after hypomorphic MSL2 RNAi (Hamada et al. 2005) and female cells (see Supplemental Methods) (Fig. 3A). We also plotted the average occupancy at each of these classes of CES compared with the average

CLAMP site on the X-chromosome and the autosome (Fig. 3B). In this way, three distinct classes of CES were defined (Figs. 3A, B): **Group A**) Largely MSL-dependent CES exhibit a strongly decreased CLAMP occupancy in the absence of MSL complex and their occupancy in female cells is much below that of the average CLAMP site (177 sites); **Group B**) Partially MSL-dependent CES, including both *roX* loci, exhibit higher than average levels of CLAMP occupancy in the absence of MSL complex and further increase in their occupancy and breadth in the presence of MSL complex (43 sites); **Group C**) Largely MSL-independent CES have higher than average CLAMP occupancy in both male and female cell lines (45 sites). We then validated the genome-wide analysis by measuring occupancy of CLAMP at members of each of these three classes of CES by ChIP-qPCR in male and female cells and larvae (Fig. S4D).

Thus, Group B and C CES, including both *roX* loci, have higher than average levels of MSL-independent CLAMP occupancy compared with Group A sites that have lower than average MSL-independent CLAMP occupancy. These data suggest that Group B and C CES may function as beacons to promote X-identification because they are occupied by CLAMP independent of MSL complex. In addition, CLAMP is required for MSL complex recruitment all along the X-chromosome and the MSL complex is required to increase CLAMP occupancy at Group A and Group B sites (Fig. 3; Fig. S3B, F). Therefore, synergistic interactions between CLAMP and the MSL complex at Group A and Group B sites likely increase the occupancy of both factors.

To define additional common features other than CLAMP occupancy that distinguish each subclass of CES, we compared the following properties: 1) the number of tandem MREs within each CES; 2) the two-dimensional clustering of CES along the X-chromosome; 3) the average nucleosome occupancy in the absence of the MSL complex (Fig. 4, S4). First, we found a strong increase in the percentage of Group B and Group C sites with three or more tandem MREs present within each CES locus compared with Group A sites (Fig. 4A). Second, Group B sites are clustered in two dimensions along the length of the X within a 5 Mb region surrounding *roX2* ($p < 0.027$), with the majority of sites between the *roX* genes (Fig. 4B). In contrast, Group A sites are uniformly distributed and Group C sites exhibit minimal clustering distal to the Group B sites. Third, we examined chromatin organization at each subclass of sites in the absence of MSL complex using publicly available female cell data sets (Fig. S4E: (www.modencode.org) (Henikoff et al. 2009). Under high salt extraction conditions, which are likely to provide the most accurate nucleosome occupancy profiles over CES, the decreased nucleosome density at Group B and Group C loci in female cells resembles that previously reported for male cells (Fig. S4E) (Alekseyenko et al. 2012). Therefore, a chromatin feature distinguishes Group B and C CES from those in Group A, even in the absence of MSL complex.

In summary, the number of tandem MREs within each CES locus (Fig. 4A), two dimensional clustering along the length of the X-chromosome (Fig. 4B), and MSL-independent nucleosome occupancy (Fig. S4E) are features that distinguish subclasses of CES and were not previously identified by analysis of the average

properties of all CES (Alekseyenko et al. 2008; Alekseyenko et al. 2012). By defining three distinct subclasses of CES, we provide a new tool for examining how the MSL complex identifies the X-chromosome.

Our identification of CLAMP as the previously unknown link that tethers MSL complex to the X-chromosome and three new subclasses of CES supports the following mechanism for X-identification (Fig. 4C): 1) Prior to MSL complex targeting, maternally-loaded CLAMP binds directly to MREs within Group B and C CES including *roX* loci due to clustering of tandem MREs within these loci (Fig. 4A). 2) During early embryogenesis, MSL-independent transcription of *roX* genes is stimulated (Meller and Rattner 2002b) (Kelley et al. 1997) and is likely to facilitate co-transcriptional MSL complex assembly (Lim and Kelley 2012). CLAMP is enriched at the CES adjacent to *roX* loci independent of the MSL complex and tethers the MSL complex to the X-chromosome (Figs. 1-3). 3) The MSL complex catalyzes H4K16ac (Smith et al. 2001), thereby opening chromatin (Shogren-Knaak et al. 2006) and increasing the accessibility of MRE sequences for recognition by CLAMP. Synergistic interactions between the MSL complex and CLAMP (Figs. 3A, S3B,C,F) are likely to contribute to the high occupancy of both factors at the *roX* loci, followed by spreading to additional CES (Figs. 3A, S3F, S4). Through this proposed mechanism, we suggest that CLAMP functions together with *roX* RNAs to target the MSL complex to the X-chromosome. It is also possible that additional factors function with CLAMP during MSL complex recruitment. In the future, new methods will be developed to define this process in real-time in the early embryo. The high

degree of conservation of both CLAMP and MREs across *Drosophila* species (Alekseyenko et al. 2013) suggests that they are key components of a conserved mechanism for establishing an active chromatin domain within a metazoan genome.

Acknowledgements

We thank Mitzi Kuroda, Fred Winston, Judith Bender, and Tricia Serio for critical reading of the manuscript. We are grateful to Mitzi Kuroda for the gift of the MSL3 and MSL2 antibodies. We thank Timothy Hughes for helpful advice. This work was supported by the following grants to E.L: NIH GM098461-01, a Pew Biomedical Scholars Program grant, a NSF-ADVANCE pilot project grant, a RI-INBRE pilot project grant, a Rhode Island Foundation grant, and the Brown University Salomon grant. M.M.L.S. is supported by the NIH F31 GM099399-01. M.L.B is supported by NIH R01 HG005287 and P.J.P is supported by NIH U01HG004258. All of the Illumina sequencing in this study was conducted at the Brown University Center for Genomics and Proteomics and was supported by Grant Number P30RR031153 from the National Center for Research Resources.

Author contributions

M.M.L.S., J.C, J.F., P.X., K.G., A.L., A.S., E.L., and A.V. performed experiments. E.B, M.Y.T., and T.S. performed computational analysis. E.L., M.L.B., and P.J.P. designed and consulted on experiments. E.L., and M.M.L.S prepared the manuscript.

Methods

Flies and crosses

The CLAMP RNAi fly line was obtained from the TRiP project (www.flyrnai.org). See supplemental methods for details.

Chromatin immunoprecipitation (ChIP) from third instar larvae

A protocol was adapted from Larschan et al., 2007. See supplemental methods for details.

Cell culture and RNAi

Standard cell culture approaches were used to maintain *Drosophila* SL2 and Kc cells and RNAi treatment conditions are described at www.flyrnai.org. See supplemental methods for details.

mRNA sample preparation

A protocol was adapted from Gelbart et al., 2009. See supplemental methods for details.

Preparation of Illumina sequencing libraries for ChIP-seq and mRNA-seq

A protocol was adapted from Alekseyenko et al., 2008. See supplemental methods for details.

Bioinformatics

See supplemental methods for details.

***in vitro* protein expression for protein binding microarrays (PBMs) and PBM**

A protocol was adapted from Berger et al., 2006. See supplemental methods for details.

Immunostaining of polytene chromosomes

A protocol was adapted from Kelley et al., 1999. See supplemental methods for details.

References

- Alekseyenko AA, Ellison CE, Gorchakov AA, Zhou Q, Kaiser VB, Toda N, Walton Z, Peng S, Park PJ, Bachtrog D et al. 2013. Conservation and de novo acquisition of dosage compensation on newly evolved sex chromosomes in *Drosophila*. *Genes Dev* **27**: 853-858.
- Alekseyenko AA, Ho JW, Peng S, Gelbart M, Tolstorukov MY, Plachetka A, Kharchenko PV, Jung YL, Gorchakov AA, Larschan E et al. 2012. Sequence-specific targeting of dosage compensation in *Drosophila* favors an active chromatin context. *PLoS Genet* **8**: e1002646.
- Alekseyenko AA, Peng S, Larschan E, Gorchakov AA, Lee OK, Kharchenko P, McGrath SD, Wang CI, Mardis ER, Park PJ et al. 2008. A sequence motif within chromatin entry sites directs MSL establishment on the *Drosophila* X chromosome. *Cell* **134**: 599-609.
- Bai X, Alekseyenko AA, Kuroda MI. 2004. Sequence-specific targeting of MSL complex regulates transcription of the roX RNA genes. *Embo J* **23**: 2853-2861.
- Bailey TL, Boden M, Buske FA, Frith M, Grant CE, Clementi L, Ren J, Li WW, Noble WS. 2009. MEME SUITE: tools for motif discovery and searching. *Nucleic Acids Res* **37**: W202-208.
- Bell O, Conrad T, Kind J, Wirbelauer C, Akhtar A, Schubeler D. 2008. Transcription-coupled methylation of histone H3 at lysine 36 regulates dosage compensation by enhancing recruitment of the MSL complex in *Drosophila melanogaster*. *Mol Cell Biol* **28**: 3401-3409.

- Belote JM, Lucchesi JC. 1980. Male-specific lethal mutations of *Drosophila melanogaster*. *Genetics* **96**: 165-186.
- Copps K, Richman R, Lyman LM, Chang KA, Rampersad-Ammons J, Kuroda MI. 1998. Complex formation by the *Drosophila* MSL proteins: role of the MSL2 RING finger in protein complex assembly. *Embo J* **17**: 5409-5417.
- Fauth T, Muller-Planitz F, Konig C, Straub T, Becker PB. 2010. The DNA binding CXC domain of MSL2 is required for faithful targeting the Dosage Compensation Complex to the X chromosome. *Nucleic Acids Res* **38**: 3209-3221.
- Gelbart ME, Kuroda MI. 2009. *Drosophila* dosage compensation: a complex voyage to the X chromosome. *Development* **136**: 1399-1410.
- Hamada FN, Park PJ, Gordadze PR, Kuroda MI. 2005. Global regulation of X chromosomal genes by the MSL complex in *Drosophila melanogaster*. *Genes Dev* **19**: 2289-2294.
- Henikoff S, Henikoff JG, Sakai A, Loeb GB, Ahmad K. 2009. Genome-wide profiling of salt fractions maps physical properties of chromatin. *Genome Res* **19**: 460-469.
- Kelley RL, Lee OK, Shim YK. 2008. Transcription rate of noncoding roX1 RNA controls local spreading of the *Drosophila* MSL chromatin remodeling complex. *Mech Dev* **125**: 1009-1019.
- Kelley RL, Meller VH, Gordadze PR, Roman G, Davis RL, Kuroda MI. 1999. Epigenetic spreading of the *Drosophila* dosage compensation complex from roX RNA genes into flanking chromatin. *Cell* **98**: 513-522.

- Kelley RL, Wang J, Bell L, Kuroda MI. 1997. Sex lethal controls dosage compensation in *Drosophila* by a non-splicing mechanism. *Nature* **387**: 195-199.
- Larschan E, Alekseyenko AA, Gortchakov AA, Peng S, Li B, Yang P, Workman JL, Park PJ, Kuroda MI. 2007. MSL complex is attracted to genes marked by H3K36 trimethylation using a sequence-independent mechanism. *Mol Cell* **28**: 121-133.
- Larschan E, Bishop EP, Kharchenko PV, Core LJ, Lis JT, Park PJ, Kuroda MI. 2011. X chromosome dosage compensation via enhanced transcriptional elongation in *Drosophila*. *Nature* **471**: 115-118.
- Larschan E, Soruco MM, Lee OK, Peng S, Bishop E, Chery J, Goebel K, Feng J, Park PJ, Kuroda MI. 2012. Identification of chromatin-associated regulators of MSL complex targeting in *Drosophila* dosage compensation. *PLoS Genet* **8**: e1002830.
- Lee C, Li X, Hechmer A, Eisen M, Biggin MD, Venters BJ, Jiang C, Li J, Pugh BF, Gilmour DS. 2008. NELF and GAGA factor are linked to promoter-proximal pausing at many genes in *Drosophila*. *Mol Cell Biol* **28**: 3290-3300.
- Lim CK, Kelley RL. 2012. Autoregulation of the *Drosophila* Noncoding roX1 RNA Gene. *PLoS Genet* **8**: e1002564.
- Meller VH, Rattner BP. 2002. The roX genes encode redundant male-specific lethal transcripts required for targeting of the MSL complex. *EMBO J* **21**: 1084-1091.
- Meyer BJ. 2010. Targeting X chromosomes for repression. *Curr Opin Genet Dev* **20**: 179-189.

- Nguyen DK, Disteche CM. 2006. Dosage compensation of the active X chromosome in mammals. *Nat Genet* **38**: 47-53.
- Park Y, Kelley RL, Oh H, Kuroda MI, Meller VH. 2002. Extent of chromatin spreading determined by roX RNA recruitment of MSL proteins. *Science* **298**: 1620-1623.
- Payer B, Lee JT. 2008. X chromosome dosage compensation: how mammals keep the balance. *Annu Rev Genet* **42**: 733-772.
- Robasky K, Bulyk ML. 2011. UniPROBE, update 2011: expanded content and search tools in the online database of protein-binding microarray data on protein-DNA interactions. *Nucleic Acids Res* **39**: D124-128.
- Shogren-Knaak M, Ishii H, Sun JM, Pazin MJ, Davie JR, Peterson CL. 2006. Histone H4-K16 acetylation controls chromatin structure and protein interactions. *Science* **311**: 844-847.
- Smith ER, Allis CD, Lucchesi JC. 2001. Linking global histone acetylation to the transcription enhancement of X-chromosomal genes in *Drosophila* males. *J Biol Chem* **276**: 31483-31486.
- Straub T, Grimaud C, Gilfillan GD, Mitterweger A, Becker PB. 2008. The chromosomal high-affinity binding sites for the *Drosophila* dosage compensation complex. *PLoS Genet* **4**: e1000302.
- Sural TH, Peng S, Li B, Workman JL, Park PJ, Kuroda MI. 2008. The MSL3 chromodomain directs a key targeting step for dosage compensation of the *Drosophila melanogaster* X chromosome. *Nat Struct Mol Biol* **15**: 1318-1325.

- Villa R, Forne I, Muller M, Imhof A, Straub T, Becker PB. 2012. MSL2 Combines Sensor and Effector Functions in Homeostatic Control of the Drosophila Dosage Compensation Machinery. *Mol Cell* **48**: 647-654.
- Wang CI, Alekseyenko AA, Leroy G, Elia AE, Gorchakov AA, Britton LM, Elledge SJ, Kharchenko PV, Garcia BA, Kuroda MI. 2013. Chromatin proteins captured by ChIP-mass spectrometry are linked to dosage compensation in Drosophila. *Nat Struct Mol Biol*.

Figure 1. CLAMP localizes precisely at MREs at high levels when co-localized with MSL complex.

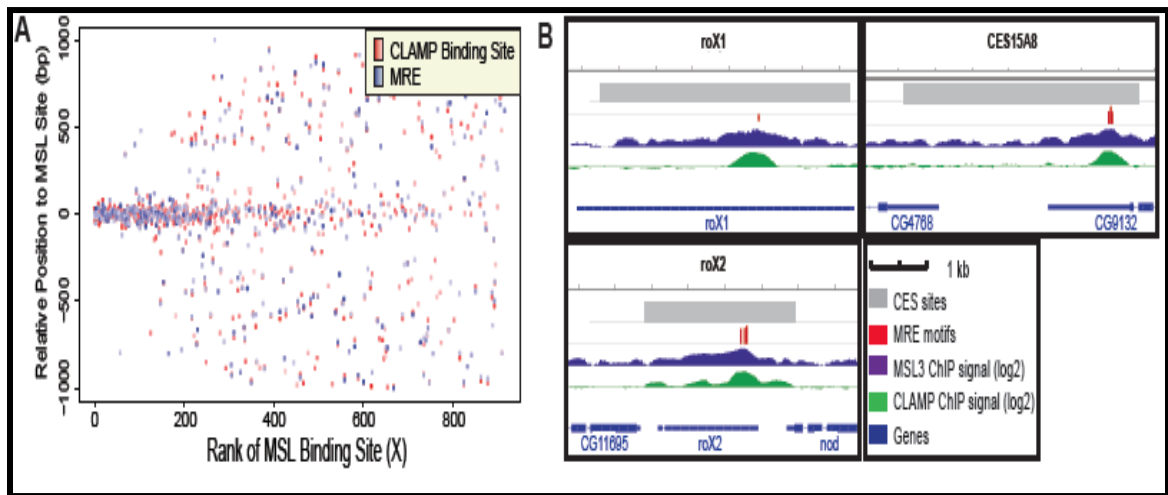


Figure 1. CLAMP localizes precisely at MREs at high levels when colocalized with MSL complex. **A)** Occupancy relationships were calculated based on input-normalized ChIP-seq experiments that were performed on CLAMP from male SL2 cells (Larschan et al. 2012) and previously defined CES (Alekseyenko et al. 2008). Two independent ChIP-seq libraries were prepared from biological replicate chromatin samples that correlated strongly with each other (see Methods). MSL complex binding sites were ranked based on MSL complex occupancy from ChIP-seq read counts (*x*-axis) (strongest on the left and weakest on the right) (Alekseyenko et al. 2008). The *y*-axis indicates the position relative to the center of the MSL complex sites. Each red dot represents the position of a CLAMP binding site that is within 1 kb of an MSL complex binding site. The closer the red dot is located to the zero position on the *y*-axis, the closer the CLAMP site is to the center of the MSL complex binding site. The intensity of the red color is proportional to CLAMP occupancy (see Methods). The intensity of the blue color is proportional to how close the individual MRE sequence is to the consensus sequence and does not correlate with MSL complex occupancy, in contrast to CLAMP occupancy (see Supplemental Methods). **B)** Examples of CLAMP and MSL complex occupancy at three well-characterized CES are shown from SL2 cells. Input normalized read counts are shown on a log scale for CLAMP and MSL complex.

Figure 2. CLAMP directly binds to MRE sequences.

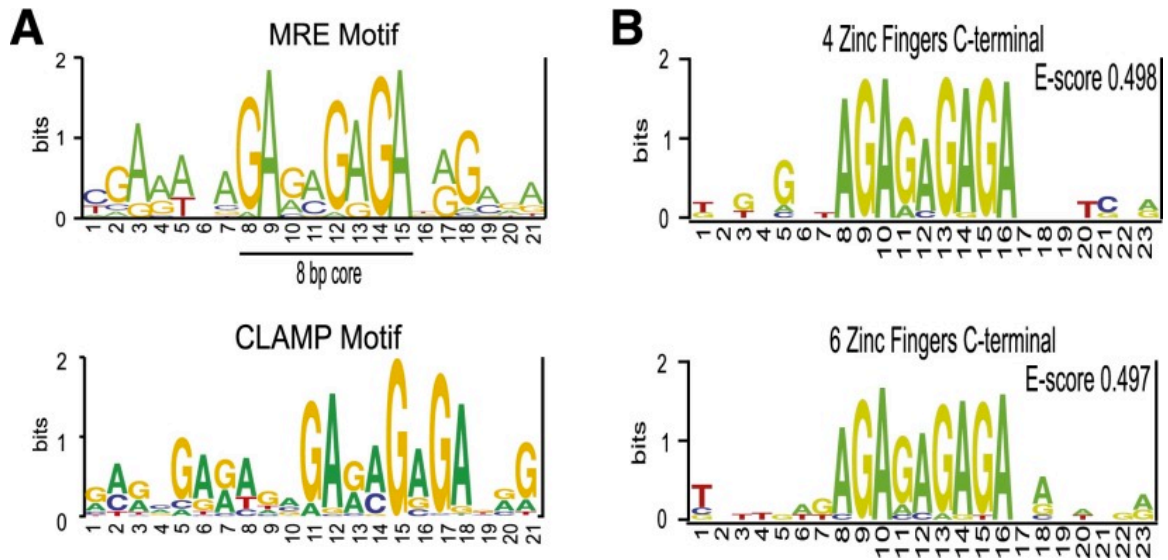


Figure 2. CLAMP directly binds to MRE sequences. **A)** (top) Position weight matrix (PWM) for the original MRE motif (Alekseyenko et al. 2008). (bottom) PWMs similar to MREs were derived from CLAMP ChIP-seq data from SL2 cells. MEME was used to analyze 200-bp regions flanking 500 CLAMP occupancy sites that were located 3' ends of genes on the X-chromosome. **B)** Protein binding microarray analysis resulted in a motif similar to the MRE for the direct *in vitro* binding specificities of the C-terminal CLAMP four (top) and six zinc-finger fragments (bottom).

Figure 3. CES have three classes of CLAMP occupancy patterns

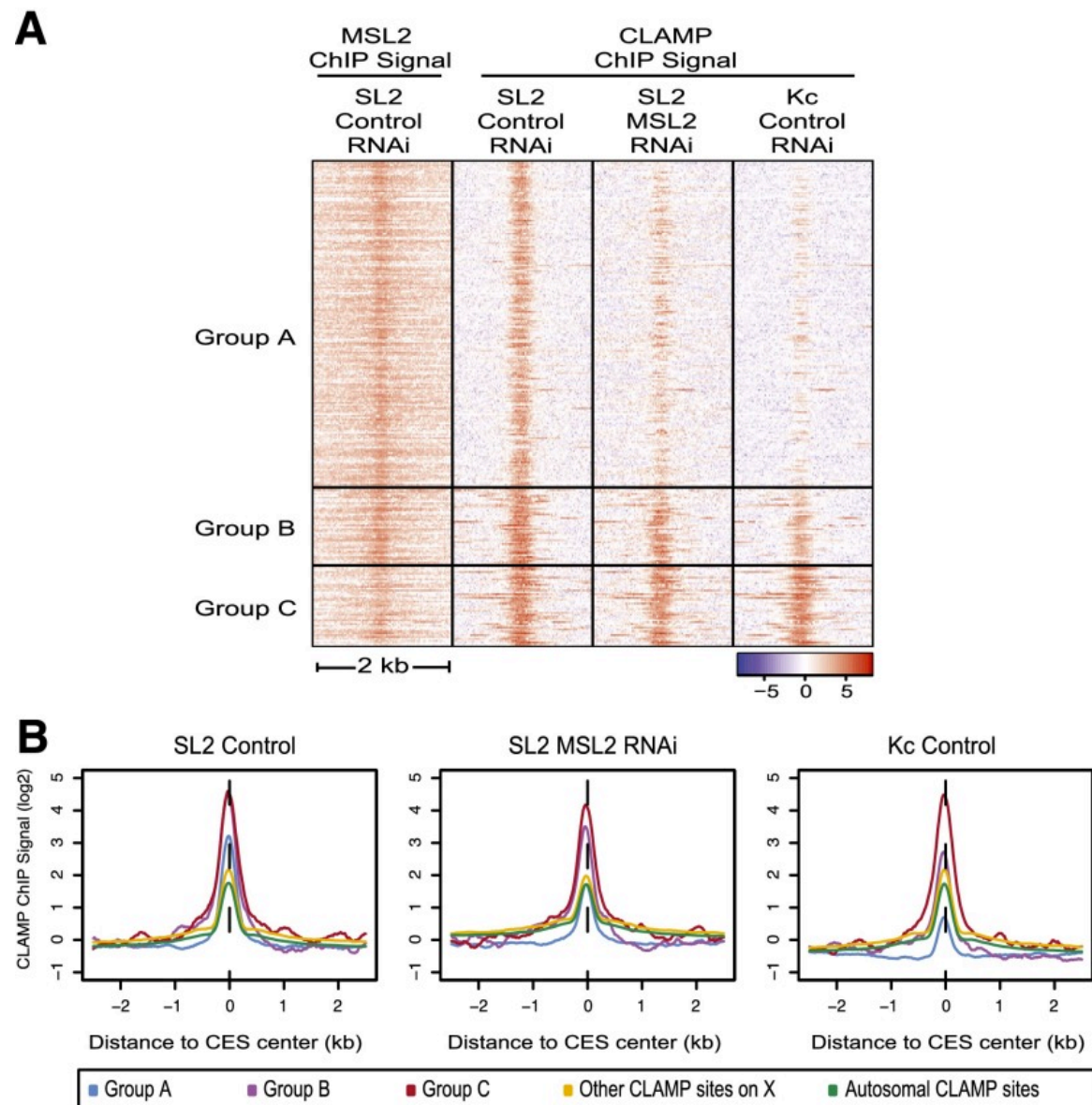


Figure 3. CES have three classes of CLAMP occupancy patterns. A) CLAMP or MSL2 occupancy patterns (reads per kilobase per million mapped reads [RPKM] normalized to Input) are shown within a 2 kb window centered at the MRE for each member of each subclass of CES (265 total CES). 1) Group A sites are highly MSL-dependent. 2) Group B sites are partially MSL-dependent. 3) Group C sites are independent of MSL complex. **B)** Average profiles of CLAMP occupancy (RPKM) normalized to Input are centered at group A, group B, or C CES peaks and are shown compared with average occupancy at all CLAMP sites on the X-chromosome or autosomes (log 2 scale).

Figure 4. Three subclasses of CES differ in their number of tandem MREs, two-dimensional clustering along the X-chromosome.

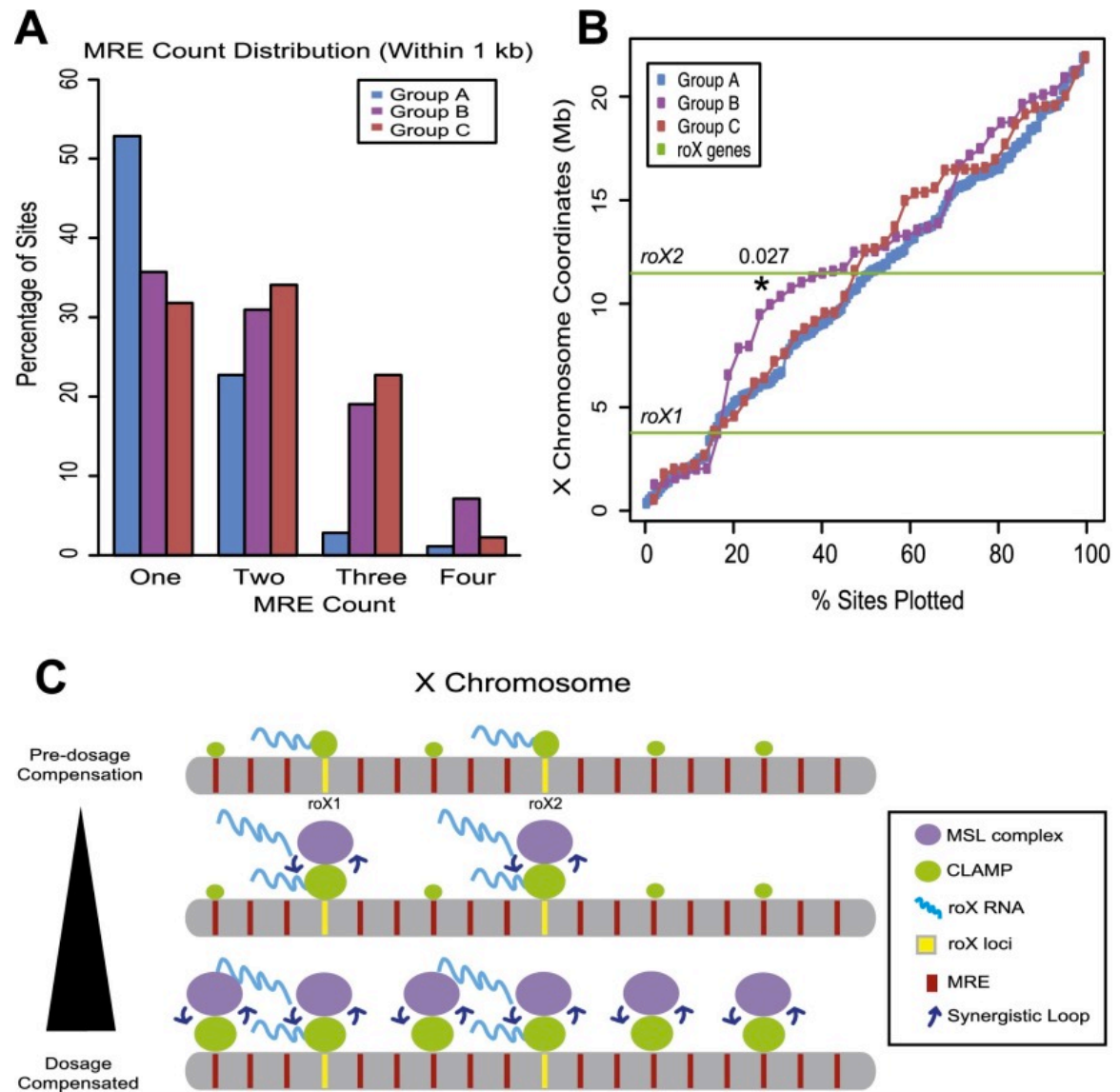


Figure 4. Three subclasses of CES differ in their number of tandem MREs, two-dimensional clustering along the X-chromosome. A) Analysis of the number of tandem MREs within each subclass of CESs. **B)** The locations of CES from each subclass of sites is plotted to allow their relative positions along the length of the X-chromosome to be determined. The *y*-axis is the position along the length of the X-chromosome and the *x*-axis shows the % of sites plotted. The positions of the *roX* loci are shown as green lines. **C)** Model for CLAMP and *roX* RNAs functioning together during X-identification: **1)** CLAMP binds to MREs within Group B and Group C independent of MSL complex and *roX* RNAs are transcribed independent of MSL complex recruitment. **2)** MSL complex is likely to be cotranscriptionally assembled at *roX* loci. CLAMP tethers MSL complex to CES including *roX* loci. MSL complex catalyzes H4K16ac, thereby opening chromatin and increasing the accessibility of MRE sequences for CLAMP recognition. Next, increased levels of CLAMP associate with CES and thereby recruit additional MSL complexes via synergistic interactions. **3)** The MSL complex spreads to additional CES, potentially stabilized by CLAMP, thereby initiating additional synergistic interactions.

Supplemental Methods

Flies and crosses

Fly stocks were provided by the Bloomington stock center and raised on yeast, sugar, cornmeal-based media at 25 °C. The CLAMP RNAi experiment was performed by crossing the CG1832 RNAi stock #27080 ($y[1] \ v[1]; \ P\{y[+t7.7] \ v[+t1.8]=TRiP.JF02426\}attP2$) with a driver da-Gal4 stock Bloomington #8641 ($w[1118]; \ P\{da-GAL4.w[-]\}3$) and the matched control experiment was performed by crossing the CG1832 RNAi stock to a ($y[1] \ w[1]$) control stock.

Chromatin immunoprecipitation (ChIP) from third instar larvae

Chromatin Preparation

The following protocol is modified from Larschan et al., 2007. Chromatin was prepared from 150 sexed third instar larvae ground into a powder in a mortar and pestle while frozen in liquid nitrogen. Larvae were homogenized in a 40 mL Dounce homogenizer in 40 mLs of PBS containing a protease inhibitor cocktail (PI) (pancrease-extract 0.002 mg/ml, thermolysin 0.0005 mg/ml, chymotrypsin 0.002 mg/ml, trypsin 0.02 mg/ml, papain 0.33 mg/ml) (Roche). Next, the extract was crosslinked with 1% formaldehyde. Samples were rotated for 20 min at room temperature followed by quenching with 2 mLs 2.5 M glycine for 5 minutes at room temperature. The extract was chilled on ice and centrifuged for 5 minutes at 1500 g at 4 °C. The pellet was resuspended in 40 mLs PBS containing 1x PI and 0.2 mM

PMSF followed by centrifugation for 5 minutes at 1500 g at 4 °C. Centrifugation and resuspension were repeated as described above with the following buffers: 1) 10 mLs Wash A (10 mM HEPES pH 7.6, 10 mM EDTA pH 8.0, 0.5 mM EGTA pH 8.0, 0.25% Triton-X 100, 1x PI, , 0.2 mM PMSF); 2) 10 mLs Wash B (10 mM HEPES pH 7.6, 10 mM NaCl, 10 mM EDTA pH 8.0, 0.5 mM EGTA pH 8.0, 0.25% Triton-X 100, 1x PI, 0.2 mM PMSF); 3) 10 mLs TE Wash (10 mM Tris-HCL pH 8.0, 1 mM EDTA pH 8.0, 0.01% SDS, 1x PI, 0.2 mM PMSF); 4) 10 mLs TE Wash with SDS (10 mM Tris-HCL pH 8.0, 1 mM EDTA pH 8.0, 1% SDS, 1x PI, 0.2 mM PMSF), and 5) 2 washes with 10 mLs TE Wash. Next, the pellet was resuspended in 2 mLs Pre-RIPA (0.1% SDS, 10 mM Tris-HCL pH 8.0, 1 mM EDTA pH 8.0, 1x PI, 0.2 mM PMSF). A Misonix Q125 tip sonicator was used to fragment 400 µL aliquots of chromatin for 1 min total sonication time (20 % amplitude, 6 intervals of 10 sec). The aliquots were then pooled together and Triton-X 100 was added to a final concentration of 1% followed by rotation at 4 °C for 1 minute. Next, deoxycholate was added to a 0.1% final concentration followed by rotation at 4 °C for 1 minute. Then, NaCl was added to a final concentration of 140 mM followed by rotation at 4 °C for 5 minutes.

Next, 200 µL of the chromatin preparation was saved for the input control sample. To reverse the crosslinks on the input samples, the following reagents were added: 1) 200 µL of RIPA 150 (50 mM Tris-HCl pH 8.0, 1% NP-40, 2 mM EDTA pH 8.0, 0.1% sodium deoxycholate, 0.1% SDS, 150 mM NaCl, 1x PI, 0.2 mM PMSF, 1 mM DTT); 2) SDS to a final concentration of 1%; 3) 15 µL 5M NaCl; 4) 1 µL of DNAase-free RNase

A (Roche). Samples were incubated at 65 °C overnight. To digest proteins and RNA, the following reagents were added: 1) an additional 1 µL of RNase A followed by incubation at 37 °C for 30 minutes; 2) 20 µL 1M Tris-HCl pH 6.8; 3) 10 µL 0.5M EDTA pH 8.0; and 4) 3 µL 20 mg/mL proteinase K (Invitrogen) followed by incubation at 42 °C for 90 minutes. To remove digested protein and enzymes, phenol-chloroform-isoamyl alcohol (Invitrogen) extraction was performed followed by chloroform extraction using Phase-Lock tubes (Eppendorf). DNA was precipitated by adding 10 µL glycogen (Ambion), 1/10 volume of 3 M NaAc, and 2 volumes of 100% EtOH followed by incubation at -80 °C overnight. Samples were centrifuged at 20,000 g at 4 °C for 10 minutes followed by washing the pellet with 1 mL 70% ethanol and resuspension in ddH₂O. Concentrations were obtained using the Qubit DNA High Sensitivity reagents (Invitrogen) and fluorimeter. 500 ng of each sample was run on a 2% agarose gel to determine that chromatin fragmentation produced fragments from 200-500 bp in length. The remainder of the chromatin preparation was resuspended in RIPA 150 buffer to a final concentration of 10 µg/mL.

Immunoprecipitation

The immunoprecipitation protocol was modified from Alekseyenko et al., 2006. Immunoprecipitation was performed using 1 mL of the appropriate chromatin preparation combined with antibody at 4 °C overnight. Anti-MSL2 antibodies were

described previously (Kelley et al. 1999). The anti-CLAMP antibody was described previously (Larschan et al. 2012) and the anti-H3 was obtained from Abcam. 60 μ L of pre-blocked Protein A agarose beads (Millipore) were added to each IP followed by rotation at 4 °C for 2 hrs. Each IP was then centrifuged at 1000 g at 4 °C for 3 minutes. Beads were resuspended in cold RIPA 150 buffer followed by rotation at 4 °C for 5 minutes and centrifugation at 1000 g at 4 °C for 3 minutes. Washes were repeated with RIPA 150, RIPA 300 (50 mM Tris-HCl pH 8.0, 1% NP-40, 2 mM EDTA pH 8.0, 0.1% sodium deoxycholate, 0.1% SDS, 300 mM NaCl), LiCl/TE (0.25 M LiCl, 1% NP-40, 1% sodium deoxycholate, 10 mM Tris-HCl pH 8.0, 1 mM EDTA pH 8.0) and two times with TE buffer (10 mM Tris-HCl pH 8.0, 1 mM EDTA pH 8.0). The samples were eluted off the beads twice with 250 μ L Sodium Bicarbonate Buffer (1% SDS, 0.1 M sodium bicarbonate) by rotating at room temperature for 15 minutes followed by centrifugation at 1000 g for 3 minutes.

Quantitative Real-Time PCR

Primer sequences are available upon request and were previously reported (Alekseyenko et al. 2008). Briefly, all primers were tested to assure that they have 100% efficiency and good linearity prior to use with an ABI 7000 real time PCR machine. The average relative quantitation (ABI User Bulletin #2) of three replicate reactions was determined for all samples. Platinum SYBR master mix w/ROX (Invitrogen) was used. Two independent chromatin preparations were conducted for each experiment and standard deviations are reported. Dilutions of input and IP

DNA were conducted to assure that samples were in the linear range for quantitative PCR. To assess the overall ChIP enrichment, normalization was conducted using primers within either the CG15570 or PKA internal control genes that span genomic loci that are unbound by MSL complex or CLAMP as assessed by ChIP-seq. Normalization to input DNA was also conducted for ChIP experiments. After normalization to input, ChIP data were normalized within each biological replicate to data from the primer pair with the highest enrichment value. Subsequently, biological replicates were averaged and standard deviations calculated.

Cell culture and RNAi

dsRNA to target the CLAMP gene (CG1832) was generated as described in detail at www.flyrnai.org. The DRSC03718 and DRSC29935 constructs were synthesized using the T7 megascript kit (Ambion) and purified using the RNAeasy kit (Qiagen). Both dsRNAs yielded very similar results and neither had specificity for additional genes as off-targets using a 19 bp criterion (www.flyrnai.org; Larschan et al. 2012). All experiments shown used the DRSC03718 construct but the DRSC29935 construct yielded very similar results (data not shown). GFP dsRNA used as a control and the MSL2 dsRNA construct were synthesized as previously described (Hamada et al. 2005; Larschan et al. 2012). Primer sequences are available on request and at www.flyrnai.org.

RNAi treatment of SL2 cells

RNAi treatment was conducted in T225 flasks as described at www.flyrnai.org. SL2 cells were diluted to 1.5×10^6 cells/ml and 10 mLs of cells were added to each T225 flask with 225 µg of dsRNA and 5 mL of water. Six days after dsRNA treatment, chromatin preparation (as described above) and mRNA preparation using the RNAeasy Plus kit (Qiagen) were performed. SL2 cells were maintained in Gibco Schneider's medium with 10% heat inactivated FBS (JRH).

mRNA sample preparation

Larval mRNA preparation

From the crosses generated for ChIP experiments described above, 10 sexed third instar larvae were collected, flash frozen in liquid nitrogen and homogenized using a Tissue Lyser (Invitrogen) and stainless steel beads. The total RNA was extracted using Trizol Reagent as previously described (Bai et al. 2004). Further purification of total RNA and removal of genomic DNA was conducted using the RNAeasy Plus kit (Qiagen).

cDNA synthesis and Quantitative real time RT-PCR

Synthesis of cDNA was conducted as described in Alekseyenko et al. 2006 using 1 µg of total RNA, Vilo mix (random hexamer and oligo dT) (Invitrogen), and Super Script III (Invitrogen). cDNA was heated for 5 minutes at 85 °C to inactivate enzymes prior to real time PCR with the ABI 7000 machine. Primer sequences are available upon request. cDNA was diluted prior to qPCR to ensure values were in linear range. Values were normalized to PKA, a highly-expressed gene.

Preparation of Illumina sequencing libraries

ChIP-seq

Chromatin preparation and immunoprecipitation were conducted as described in Larschan et al., 2012. ChIP-seq libraries were generated using a protocol adapted from the Illumina TruSeq sample preparation guide. Each reaction was followed by purification using AMPure XP magnetic Beads (Beckman Coulter). The fragmented DNA was 3' end repaired and 5' phosphorylated using the End-it kit (Epicenter). A 3' adenosine was added using Klenow Fragment (3'->5' exo) (NEB). The indexed annealed adapters were ligated in a 1:50 dilution of the 100 µM stock using T4 DNA Ligase Quick Ligase Kit (Enzymatics). The libraries were then PCR amplified using Phusion high fidelity polymerase (NEB) for 15 cycles. Next, the libraries were run on a 2% agarose gel using the E-Gel system to select a size range from 200 bp to 500

bp followed by concentration on a MinElute column (Qiagen). Six libraries were multiplexed per flow cell lane. Sequencing was performed on the Illumina Hi-Seq platform.

mRNA-seq

mRNA-seq libraries were prepared using the Illumina TrueSeq Kit reagents following the sample preparation protocol provided.

Bioinformatics

Analysis of ChIP-seq data

Alignment and relationships between MSL complex and CLAMP ChIP-seq data

Reads were aligned to the dm3 assembly of the *D. melanogaster* genome using the Bowtie aligner (Langmead et al. 2009). Reads that aligned with more than one mismatch or did not align uniquely were discarded. Peaks were called using the SPP software package (Kharchenko et al. 2008) with an FDR threshold of 0.05. Read counts from both input and ChIP samples were calculated within 20 bp bins across the genome. A pseudocount of one was added to both input and ChIP counts to prevent undefined ratios. Counts were further normalized by the total number of reads in the sample. Enrichment within each bin was then calculated as the $\log(2)$

ratio of sample-size normalized ChIP read counts to sample-size normalized input read counts. All data are publicly available at

<http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE39271>.

Data from two independent biological replicates of each ChIP-seq experiment were combined after we verified that correlation coefficients between the replicates were high. Furthermore, all bioinformatic analyses described below were conducted on each independent replicate and very similar results were obtained (data not shown).

The sequencing statistics are listed below in addition to Pearson R coefficients:

CLAMP ChIP-seq from Kc cells (control RNAi) (read number):

Replicate 1 Input: 74673445 raw, 44755545 aligned (59.9%)

Replicate 1 ChIP: 79102347 raw, 50139342 aligned (63.4%)

Replicate 2 Input: 26546524 raw, 15479624 aligned (58.3%)

Replicate 2 ChIP: 35921511 raw, 17536896 aligned (48.8%)

Pearson R for CLAMP ChIP replicates, 1 kb bins: 0.96

CLAMP ChIP-seq from SL2 cells (control RNAi) (read number):

Replicate 1 Input: 29495696 raw, 17591247 aligned (59.6%)

Replicate 1 ChIP: 27820142 raw, 19210177 aligned (69.1%)

Replicate 2 Input: 22498098 raw, 13510882 aligned (60.1%)

Replicate 2 ChIP: 28231400 raw, 15194724 aligned (53.8%)

Pearson R for CLAMP ChIP replicates, 1 kb bins: 0.95

CLAMP ChIP-seq from SL2 cells (*msl2* RNAi) (read number):

Replicate 1 Input: 69003930 raw, 39942739 aligned (57.9%)

Replicate 1 ChIP: 64882446 raw, 38190623 aligned (58.9%)

Replicate 2 Input: 43707761 raw, 22989749 aligned (52.6%)

Replicate 2 ChIP: 37850575 raw, 18060728 aligned (47.7%)

Pearson R for CLAMP ChIP replicates, 1kb bins: 0.96

For direct comparison between MSL complex and CLAMP, MSL3 ChIP-seq data from Alekseyenko et al. 2008 were re-processed using the same method described above for the CLAMP ChIP-seq data. In the scatter plots (Figs. 1A, S1A), sites of MSL complex recruitment were sorted from strongest to weakest (*x*-axis) and the distance to the nearest CLAMP binding site for each MSL site was plotted on the *y*-axis. Red color intensity is proportional to the log of the CLAMP binding site strength calculated with the SPP software. For Fig. 1A, locations of MREs on the X-chromosome identified by Alekseyenko et al (2008) were plotted relative to MSL binding sites in an analogous manner to the CLAMP binding sites. Instead of binding site strength, color intensity of the blue dots representing MREs is proportional to the negative log of the p-value that measures how well a particular sequence matches the canonical MRE motif.

Identification of sequence motifs within CLAMP binding sites

To define the sequence motifs that were associated with CLAMP binding, MEME (Bailey et al. 2009) was used to search for a motif within the DNA sequence surrounding each CLAMP binding site (± 1 kb). Because performing MEME analysis on all 12,704 identified binding sites was computationally prohibitive, MEME was performed on 500 sequences and their reverse complements, chosen randomly from the entire dataset, and configured to report the top 3 motifs identified. Multiple runs produced similar results. In all cases, the top three reported motifs were GA-rich. To examine whether CLAMP associates with a different sequence in the absence of MSL complex, four groups of binding sites were considered: 1) sites ± 1 kb of autosomal TSS (Transcription Start Sites); 2) sites ± 1 kb of autosomal TTS (Transcription Termination Sites); 3) sites ± 1 kb of X-chromosome TSS loci; 4) sites ± 1 kb of X-chromosome TTS loci. MEME analysis was performed separately on a subset of each of these groups of binding sites containing 500 sequences. The resulting identified motifs were similar in all cases and Fig. 2B shows the sites ± 1 kb of X-chromosome TTS loci (Fig. 2B).

Generating average gene profiles

To generate average enrichment profiles, ChIP and input datasets were each binned into 20-bp bins after normalization (Figs. 4, S1-4). Enrichment in each bin was calculated as a $\log_2$ ratio of the read counts in the ChIP and the corresponding input

bin. A pseudocount of 1 was then added to each bin to avoid undefined ratios. Regions of a consistent length (e.g. 2 kb or 5 kb) were extracted around a feature of interest (e.g. TSS, binding peak center) and the enrichments were averaged.

Defining CES subclasses

A feature vector was constructed for each of the CES consisting of all three ChIP datasets and clustered by hierarchical clustering using *Ward's* minimum variance criterion and the Euclidean distance metric. From this clustering, two distinct groups of sites were identified (Group 1 and Group 2). To determine the MSL-dependence of these sites in an unbiased way, we further classified these sites into three groups based on the ratio of CLAMP occupancy in S2 cells (+MSL complex) compared with Kc cells (no MSL complex). This analysis subdivided the original Group 2 sites into two groups: Group B and Group C. Group 1 remained intact and was renamed as Group A. qPCR data validated that the classification of CES loci into three groups was more accurate than the original two groups (Fig. S4D).

In Fig. 4A, MSL complex ChIP enrichment profiles computed from the Alekseyenko et al. 2008 ChIP-seq dataset are also plotted. In Figure 5A, the number of MRE motifs within 1 kb of the center of each CES was counted. The percentage of CES in each of the two groups with one, two, three or four sites was then computed. In Fig. 5B, the location of each CES along the length of the X-chromosome was plotted (y-axis) in relation to the percentage of sites plotted (x-axis).

Analysis of nucleosome occupancy at CES prior to MSL complex recruitment

Profiled chromatin properties at CES in female Kc cells using publicly available data include: 1) nucleosome density without salt extraction (green); 2) nucleosome density after high salt extraction (600 mM NaCl) of MNase-treated chromatin (blue), and 3) the insoluble pellet (red) (www.modencode.org) (Mito et al. 2007; Henikoff et al. 2009). High salt extraction of the chromatin produces the most accurate profiles of CES because they are present at 3' ends that are not easily extracted like the 5' TSS regions. The salt-insoluble pellet fraction likely contains chromatin-associated protein complexes including a CLAMP-chromatin complex. The profiles of nucleosome density surrounding Group A, B and C CES loci are shown in Fig. S4E.

mRNA-seq analysis

The expression of each gene is quantified in terms of RPKM (reads per million reads per kilobase). The distribution of gene expression data was assessed and a cut-off of RPKM=3 was determined to be a good threshold to separate active from inactive genes.

***in vitro* protein expression for protein binding microarrays (PBMs)**

The pDEST15 Gateway vector (Invitrogen) was used to generate the expression plasmid for the C-terminal CLAMP four and six zinc-finger constructs. For the six-

finger construct (amino acid residues 350-561), the following primers were used for cloning: 5'ACTGTTAAACTGAAAATCCGTCCAA3' and 3'CTATAACCCACCGATAATCACTTCCTGT5'.

For the four-finger construct, (amino acid residues 412-561), the following primers were used for cloning: 5'GGCGAGAAACCCTATAAATGCCAAAC 3' and 3'CTATAACCCACCGATAATCACTTCCTGT 5'. Cloning was performed using standard Gateway Technology (Invitrogen). Proteins were expressed using the TnT Quick-coupled Transcription/Translation Systems Kit (Promega). *In vitro* transcription/translation reactions were carried out according to standard protocols with the following adaptation: zinc acetate was added to a final concentration of 50 μ M.

Protein Binding Microarrays (PBMs)

In vitro produced C-terminal 4 and 6 zinc-finger regions of CLAMP were fused to GST at their N-termini and used in PBM experiments. Proteins were quantified using standard Western blotting protocols with a GST dilution series as a standard. The concentrations of the proteins used for PBMs were as follows: 1) C-terminal 4 zinc-finger protein (24.3 nM); 2) C-terminal 6 zinc-finger protein (13.9 nM). Protein binding to microarrays was assayed as previously described (Bulyk, 2009) using a custom-designed 4x44K format oligonucleotide array (Agilent Technologies, Inc., AMADID #015681), that was converted to double-stranded DNA by primer extension as described previously (Berger et al. 2006). All protein binding

microarrays were performed in duplicate using independently prepared protein samples and the same motifs were obtained. Zinc acetate to 50 μ M was added to each sample before binding to the PBMs. PBM data were quantified, normalized, and analyzed using the Universal PBM Data Analysis Suite and the Seed-and-Wobble motif derivation algorithm (Berger and Bulyk 2009). An E-score of 0.45 is significant for our PBM experiments (Badis et al., 2009). We found that $E > 0.37$ corresponds to a false discovery rate (Q-value) of 0.001 by conducting a series of negative control PBM experiments (Badis et al., 2009).

Immunostaining of polytene chromosomes

Polytene staining was performed as described previously (Kelley et al. 1999). A goat anti-MSL3 antibody was used at 1:100 and the rabbit anti-CLAMP antibody was used at 1:50. Alexafluor 488 (anti-goat) and 594 (anti-rabbit) secondary antibodies (Invitrogen) were used for immuno-detection at a 1:200 dilution.

Supplemental Figures

Figure S1. Global comparisons of CLAMP and MSL complex occupancy at MREs.

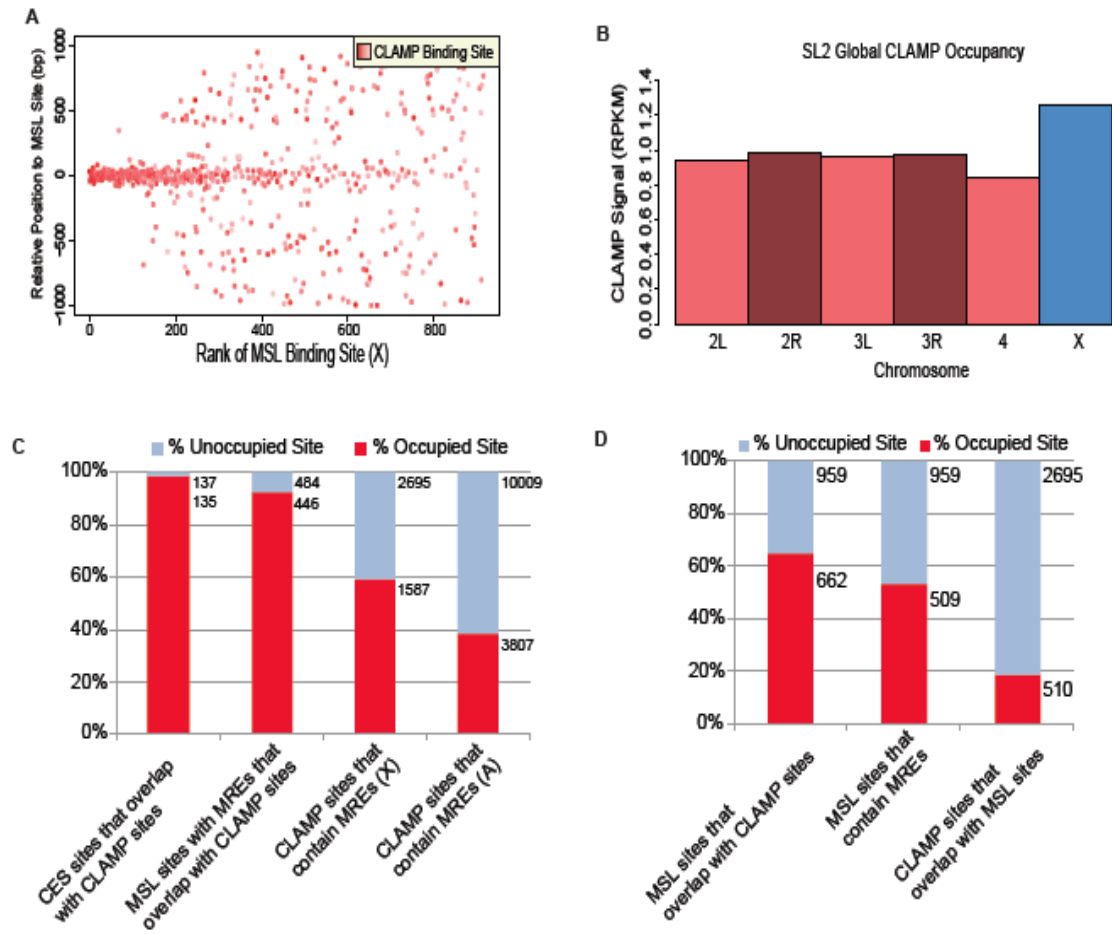


Figure S1. Global comparisons of CLAMP and MSL complex occupancy at MREs.

A) Occupancy relationships were calculated based on input-normalized ChIP-seq experiments that were performed on CLAMP from male SL2 cells (Larschan et al. 2012) and previously defined CES (Alekseyenko et al. 2008). Two independent ChIP-seq libraries were prepared from biological replicate chromatin samples that correlated strongly with each other (see Methods). MSL complex binding sites were ranked based on MSL complex occupancy from ChIP-seq read counts (x -axis) (strongest on the left and weakest on the right) (Alekseyenko et al. 2008). The y -axis indicates the position relative to the center of the MSL complex sites. Each red dot represents the position of a CLAMP binding site that is within 1 kb of an MSL complex binding site. The closer the red dot is located to the zero position on the y -axis, the closer the CLAMP site is to the center of the MSL complex binding site. The intensity of the red color is proportional to CLAMP occupancy (see Methods). **B)** The total read count per kb per million is plotted for CLAMP ChIP-seq data for each chromosome and normalized to copy number (RPKM) (see Methods for details). **C and D)** Occupancy relationships among CLAMP binding sites, MSL binding sites, and MREs.

Figure S2. CLAMP and MSL complex are highly enriched when co-localized and CLAMP also localizes to the 5' ends of active genes genome-wide.

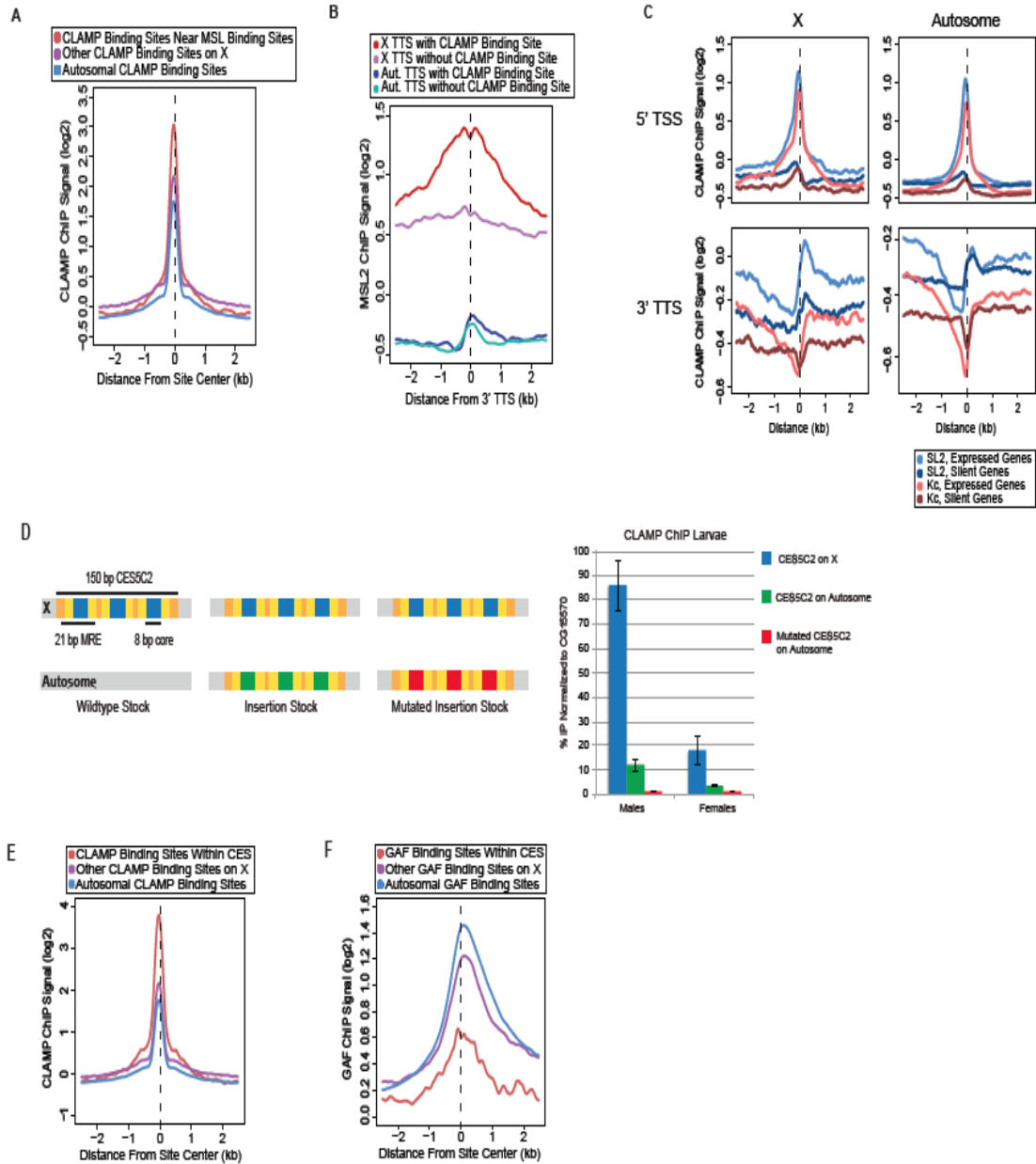


Figure S2. CLAMP and MSL complex are highly enriched when colocalized and CLAMP also localizes to the 5' ends of active genes genome-wide. **A)** CLAMP occupancy sites that co-localize with a MSL complex binding site have enhanced occupancy from SL2 ChIP-seq analysis compared with other CLAMP occupied sites on the X-chromosome and autosomes. **B)** Average gene profiles of MSL complex ChIP-seq data centered at the transcription termination sites (TTS) indicates that MSL complex is enriched near 3' ends of genes in the presence of CLAMP. **C)** Average profiles generated from CLAMP ChIP-seq data are centered around the transcription start site (TSS) or transcription termination site (TTS). CLAMP is highly enriched at TSS regions of active genes on all chromosomes in both SL2 and Kc cells. Enrichment near the TTS is specific to the male X chromosome and corresponds with the location of CES near these regions. **D)** Wild type and mutant MREs were inserted site-specifically into an autosomal location as previously described (Alekseyenko et al. 2008). The minimal 150-bp fragment of the CES5C2 high affinity site (yellow) was inserted with either three MREs (green) with wild type GA-rich 8-bp cores (blue) or with three mutated 8-bp cores (red). ChIP-qPCR data are shown from third instar larvae for the CLAMP protein. Primers within the attP landing site adjacent to the insertion were used as previously described to distinguish between the autosomal insertion and the endogenous sequence on the X-chromosome (Alekseyenko et al. 2008). **E and F)** CLAMP is enriched at CES (**E**) *in vivo* in SL2 cells while the GAF protein is enriched at sites outside of CES (**F**) in SL2 cells as shown by average gene profiles surrounding their respective binding sites (CLAMP (ChIP-seq) and GAF (ChIP-chip)).

Figure S3. CLAMP is required for *roX* gene transcription and MSL localization to the X.

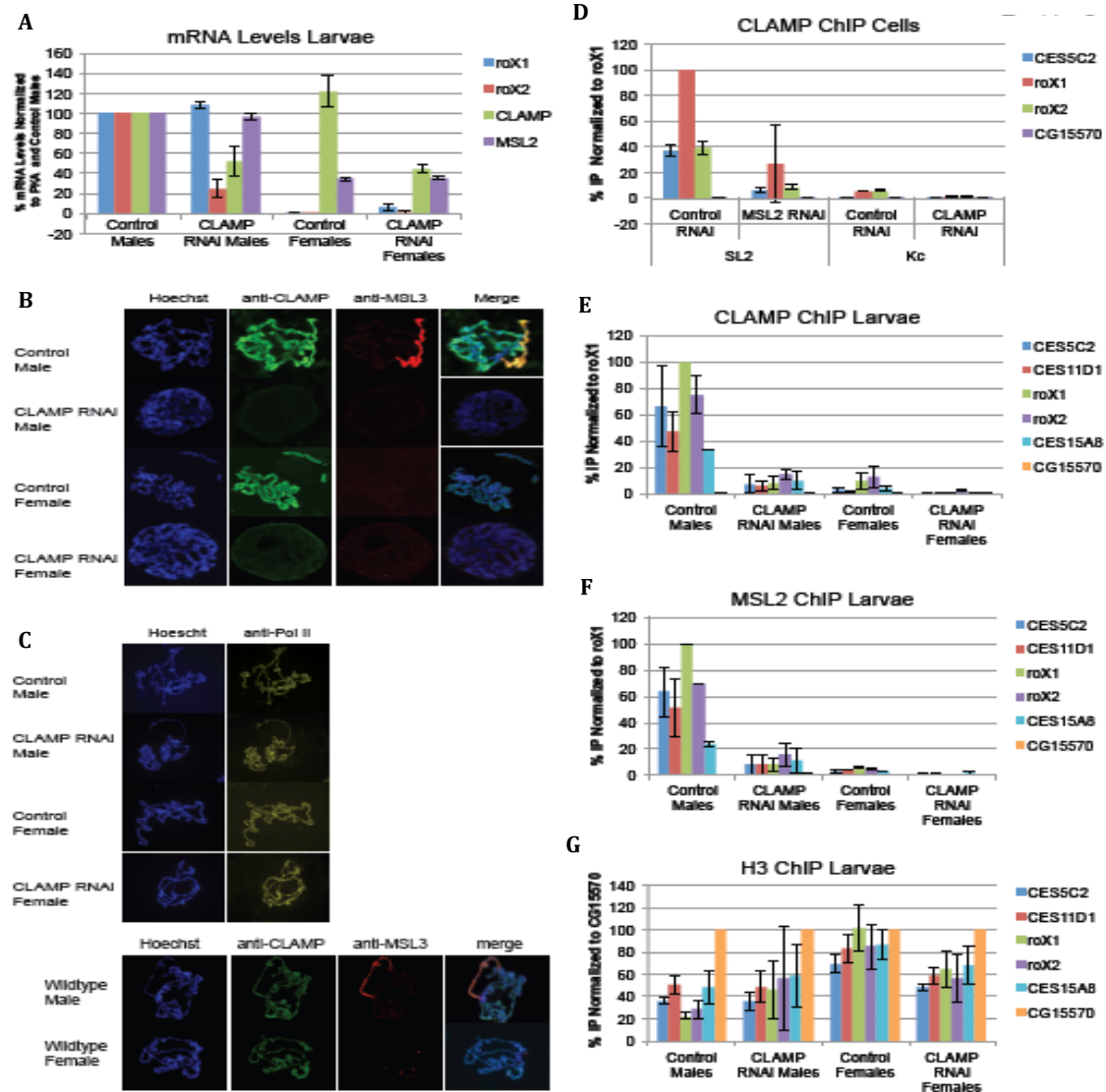


Figure S3. CLAMP is required for *roX* gene transcription and MSL localization to the X. **A)** qRT-PCR analysis of transcription of *roX* genes, CLAMP mRNA, and MSL2 mRNA after CLAMP RNAi in male and female *Drosophila* larvae. **B)** The CLAMP RNAi fly line was crossed a control stock or the *daughterless* Gal4 driver followed by immunostaining with the anti-CLAMP and anti-MSL3 antibodies. Crosses were performed at 25 °C to assure that RNAi is effectively induced by the Gal4 driver. However, chromosome morphology was poor due to the high temperature. “Control” refers to progeny of the cross between the CLAMP RNAi stock and a control y[1] w[1] stock that does not contain the *daughterless* Gal4 driver. The following false colors were used: **1)** Hoecht stain for DNA (blue); **2)** CLAMP immunostaining (green); **3)** MSL3 immunostaining (red); **4)** overlay of CLAMP and MSL3 (yellow). **C)** The following false colors were used: **1)** Hoecht stain for DNA (blue); **2)** Pol II CTD immunostaining (yellow). “Wild type” refers to a y[1] w[1] stock that was grown at 18°C to improve chromosome morphology. The following false colors were used: 1) Hoecht stain for DNA (blue); 2) CLAMP immunostaining (green); 3) MSL3 immunostaining (red); 4) overlay of CLAMP and MSL3 (yellow). **D and E)** ChIP-qPCR data for CLAMP ChIP conducted in tissue culture male (SL2) and female (Kc) cells (**D**) and in sexed third instar larvae before and after CLAMP RNAi at 5 CES (**E**), as assayed by ChIP-qPCR indicate that MSL complex is required to enrich CLAMP levels at CES. First, normalization was conducted to the input sample. Subsequently, internal normalization was performed within each experiment to facilitate comparisons between biological replicates. The value for the highest locus within each sample, *roX1*, was set to 100%.

The occupancy at *CG15570*, an untranscribed gene, is shown as a negative control. Results are the average and standard deviation of at least two replicates. **F)** ChIP for MSL2 before and after CLAMP RNAi at 5 CES as assayed by ChIP-qPCR in sexed third instar larvae. Normalization was conducted as described in **D)**. **G)** H3 ChIP at 5 CES loci in male and female larvae before and after CLAMP RNAi is shown. Normalization was conducted as described in **D)** except the *CG15570* gene was used for normalization instead of *roX1* because it exhibits the highest H3 occupancy levels.

Figure S4. Comparison of CLAMP binding characteristics in the presence and absence of MSL complex

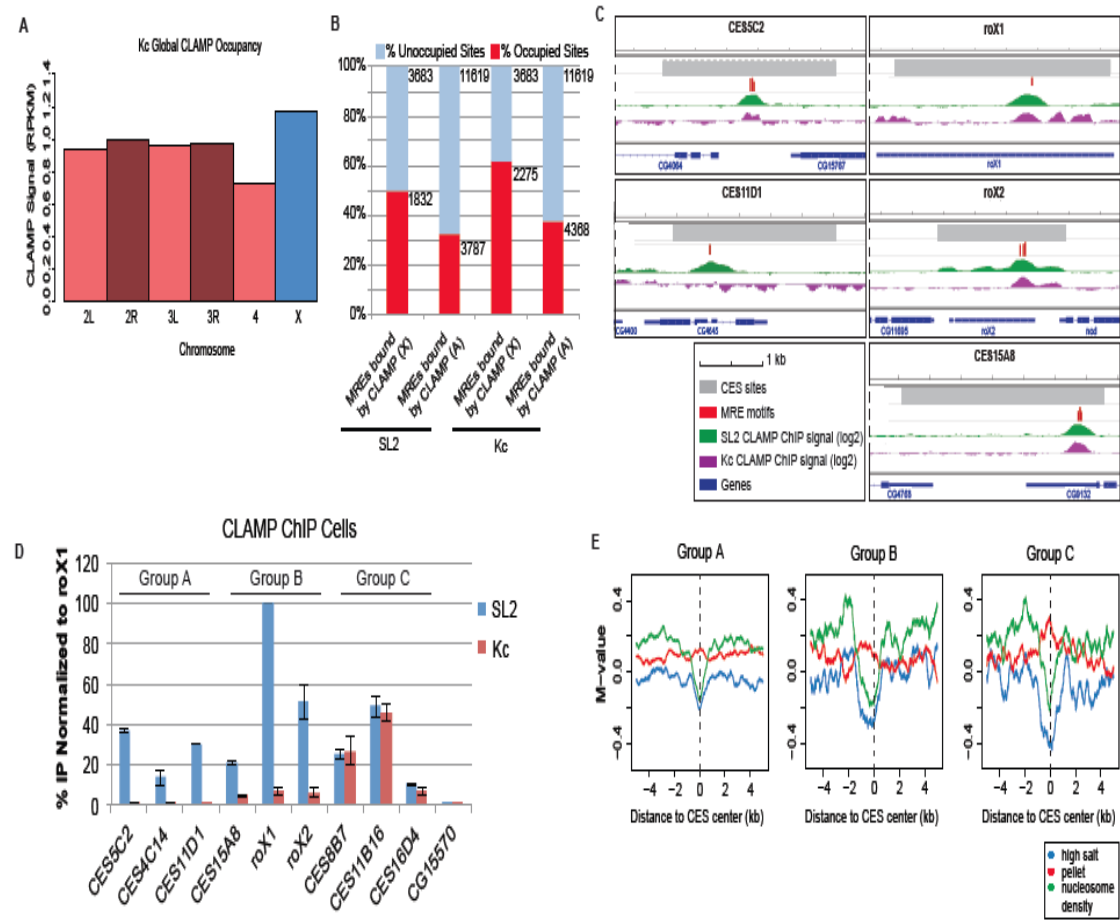


Figure S4. Comparison of CLAMP binding characteristics in the presence and absence of MSL complex. **A)** Global X-enrichment of CLAMP in female (Kc) cells independent of MSL complex as measured by total read-counts on each chromosome arm. **B)** CLAMP occupancy at MREs in male (SL2) and female (Kc) cells on the X-chromosome and autosomes. **C)** Localization of CLAMP and MSL complex near MREs in male and female cells is shown from ChIP-seq data at 5 CES.. **D)** CLAMP ChIP-qPCR validation of three CES within each subclass of sites in both SL2 and Kc cells. The % IP was calculated followed by normalization to the highest value within each replicate (*roX1*) that was set to 100%. The occupancy of CLAMP at the *CG15570* untranscribed gene is shown as a negative control. Data are the average and standard deviation of two biological replicate chromatin preparations that are independent from each other and from those used for ChIP-seq analysis. **E)** Chromatin properties of subclasses of CES are profiled in the absence of MSL complex (female Kc cells) (www.modENCODE.org). Properties are: 1) total nucleosome density (green); 2) high salt extracted chromatin (blue, 600 mM NaCl); 3) the insoluble pellet (red, salt insoluble; large protein-DNA aggregates). The dashed black lines on each plot indicate the location of the binding site centers.

**Chapter 4: The *Drosophila* CLAMP zinc-finger protein recognizes GA
dinucleotide expansions that target dosage compensation to the male X-
chromosome**

Chapter 4 of my thesis dissertation is my first-author manuscript in preparation for submission for publication. In consultation with Dr. Erica Larschan, Trevor Siggers, PhD and I designed the custom protein binding microarrays. I performed the protein binding microarray (PBM) experiments in collaboration with Dr. Martha Bulyk's laboratory at Harvard Medical School. Analysis of PBM data was done initially by Trevor Siggers, PhD and subsequently duplicated at higher stringency by Jason Dobson, PhD. I performed the Chromatin Immunoprecipitation (ChIP) followed by qPCR of *Drosophila* larvae and Electrophoretic Mobility Shift Assays that provided both the *in vivo* and additional *in vitro* validation of the PBM observations. A collaborator Lin Yang from Dr. Remo Rohs' laboratory performed DNA Shape analysis of data generated from my PBM experiments. From this analysis Lin Yang generated Figures 5B and SF3. Bioinformatic analysis of previously existing CLAMP ChIP-Seq data from Marcela Soruco and publicly available MODENCODE data was done by Jason Dobson, PhD. Using analysis of data from my PBM experiments and ChIP-seq data, Jason Dobson generated Figures 2, 3, 4B, 5A, 6B and SF1. Immunostains for the polytene chromosome figure, Figure 1, were performed by Leila Reider, PhD. I, Jessica Chery, generated all other figures. Liu Wei, PhD from Dr. Gerwald Jogl's laboratory assisted me with protein purification.

The *Drosophila* CLAMP zinc-finger protein recognizes GA dinucleotide expansions that target dosage compensation to the male X-chromosome

Abstract

Dosage compensation is an essential process that equalizes gene dose of X-linked genes between sexes in heterogametic species by forming a domain of coordinate gene expression. GA dinucleotide repeats are critical for the identification of the *Drosophila* X-chromosome, but the mechanism by which these repeats are recognized to mediate the dosage compensation process remains unknown. CLAMP, a zinc finger protein, links the dosage compensation complex to the X-chromosome. However, CLAMP occupies GA-rich sequences genome-wide. Here, we demonstrate that CLAMP occupancy on the X differs from autosomes due to altered genomic architecture that includes changes in the number of contiguous GA dinucleotide repeats. Increasing GA repeats increases binding of CLAMP and recruitment of the MSL complex. Increased recruitment of CLAMP to the X-chromosome is directly related to a greater number of GA-repeats present on the X-chromosome. In this way, we propose a mechanism for how the rapid evolution of additional GA-repeats on the X-chromosome increased the occupancy of the dosage compensation complex. Overall, we provide key insight into how subtle changes in genomic architecture can nucleate the formation of a domain of coordinate gene expression.

Introduction

Changes in primary DNA sequence that occur over evolutionary time alter transcription factor occupancy on DNA (Paris et al. 2013). Furthermore, transcription factors can recruit cofactors that alter chromatin (Voss and Hager 2014). However, it is not known how transcription factors identify their binding sites among thousands of potential genomic locations with similar DNA sequences to precisely modulate chromatin and gene expression. Sex chromosome dosage compensation in heterogametic species is a useful model system with which to study this process because a large number of genes on a single chromosome are co-regulated and share a common chromatin modification, acetylation of histone H4 at lysine 16 (H4K16ac) (Gelbart et al. 2009).

The conserved step of the sex-chromosome dosage compensation process across species is the increase in transcript levels of genes along the single active X-chromosome that is mediated by the H4K16ac chromatin mark (Deng et al. 2011; Deng et al. 2013). The H4K16ac chromatin mark physically opens chromatin by disrupting inter-nucleosomal contacts (Shogren-Knaak et al. 2006). In this way, an open chromosome conformation is likely to result in increased transcription levels (Hilfiker et al. 1997; Gu et al. 1998a; Lucchesi et al. 2005; Alekseyenko et al. 2006; Larschan et al. 2011b).

In *Drosophila* dosage compensation is mediated by the MSL complex that contains the MOF histone acetyltransferase that deposits H4K16ac (Alekseyenko et al. 2008). The MSL complex is recruited by 21-bp GA-repeat DNA elements called MREs (MSL Recognition Elements) (Alekseyenko et al. 2008). MREs that recruit MSL complex are embedded within larger 1.5 kb sequences called Chromatin Entry Sites (CES) defined by the highest levels of MSL complex occupancy (Alekseyenko et al. 2008). MREs have an 8-bp core GA-repeat element and additional flanking sequences required for high levels of MSL complex recruitment (Alekseyenko et al. 2008). There is a 1.8-fold enrichment of MRE sequences found on the X-chromosome when compared with autosomes. One of the ways in which the X-chromosome is likely to have evolved an increased number of MREs is via expansion of GA dinucleotide sequences that may have originated as transposon insertions or de-novo via replication errors (Alekseyenko et al. 2013; Ellison and Bachtrog 2013). Extended repeat sequences increase MSL complex occupancy (Ellison and Bachtrog 2013). However, the precise mechanism by which extended GA-repeats increase MSL recruitment remains unknown.

We recently demonstrated that the CLAMP C₂H₂ zinc finger protein directly recognizes MRE sequences and is required for MSL complex recruitment (Soruco et al. 2013). Furthermore, ChIP-mass spectrometry experiments demonstrated that CLAMP and the MSL complex physically interact (Wang et al. 2013a). Additionally, immunofluorescence revealed co-localization of CLAMP and the MSL complex, which contains MOF that deposits the H4K16ac mark (Soruco et al. 2013; Wang et

al. 2013a). The H4K16ac mark is thought to open chromatin structure, allowing CLAMP to identify its binding sites more efficiently. We therefore hypothesized that the CLAMP protein facilitates X-chromosome identification by promoting targeting of MSL complex specifically to the X-chromosome. However, the CLAMP protein occupies MREs genome-wide (Sorucu et al. 2013); therefore it was not understood how the pattern of CLAMP enrichment differs on the X from autosomes to promote X-chromosome identification.

Here, we integrate *in vivo* and *in vitro* approaches to determine how CLAMP localization differs on the X-chromosome from autosomes, independent of MSL complex. First, there are more CLAMP binding sites on the X-chromosome leading to greater CLAMP occupancy on the X-chromosome than autosomes. Second, the distribution of CLAMP motifs specific to the X-chromosome promotes co-localization of the CLAMP protein with 3' ends of genes, which is also the location of high affinity MSL complex sites. Third, CLAMP occupancy increases as the number of GA-repeats increases both *in vivo* and *in vitro*. Integrating these results we provide a mechanism by which expansion of GA-repeats that occurred over evolutionary time increases MSL complex occupancy on the X-chromosome via CLAMP binding.

Results and Discussion

The occupancy of the CLAMP protein is increased on the X-chromosome compared to autosomes due to increased numbers of motifs

The CLAMP protein is necessary for MSL complex recruitment to the X-chromosome (Soruco et al. 2013). However, CLAMP occupies sites all over the genome (Fig. 1). Therefore, we used genomic approaches to compare the occupancy pattern of CLAMP on the X-chromosome and autosomes. Next, we sought to determine how this pattern influences differential MSL complex recruitment. It was previously determined that recruitment of MSL complex to the X-chromosome in males dramatically changes the chromatin environment by de-compacting chromatin and increasing CLAMP occupancy (Shogren-Knaak et al. 2006; Shogren-Knaak and Peterson 2006; Kharchenko et al. 2011; Alekseyenko et al. 2012; Soruco et al. 2013). To determine how CLAMP occupancy on the X-chromosome differs from autosomes in the absence of MSL complex, we used a female tissue culture system (Kc cells) in which MSL complex is not expressed; MSL complex is translationally repressed by sex-lethal in females (Bashaw and Baker 1997; Soruco et al. 2013).

We determined that CLAMP occupancy is increased on the X-chromosome compared with autosomes (data not shown). Next, we derived the CLAMP recognition element (CRE) using MEME (Fig 2). The number of CREs on the X-chromosome is increased compared to autosomes similarly to the MREs themselves

(approximately 1.8 fold) (Sorucu et al. 2013). The percentage of sites occupied by CLAMP is increased approximately two-fold on the X-chromosome versus autosomes in both males and females (Sorucu et al. 2013). However, the occupancy per site does not differ between X-chromosome and autosomes (data not shown). Extensive analysis of chromatin marks data from modENCODE revealed that there were no significant differences between CLAMP binding sites on X versus autosomes (data not shown). No chromatin marks appeared to distinguish CLAMP binding sites on X-chromosome from autosomes. Therefore, the greater CLAMP occupancy on the X-chromosome independent of MSL complex results primarily from an increased number of binding sites and an increased percentage of sites that are bound by CLAMP.

CLAMP occupancy and motif distribution relative to genes differs 3' versus 5' on X-chromosome versus autosomes

The overall magnitude of the chromosome-wide X-enrichment of CLAMP is modest (1.8 fold) and therefore we hypothesized that additional factors such as the position of CLAMP relative to gene bodies differ between the X-chromosome and autosomes. MSL complex is recruited to gene bodies and transcription termination sites (TTS) (Alekseyenko et al. 2006) and this is the location of the majority of the CES. Thus, we hypothesized that CLAMP occupancy may differ on the X-chromosome via recruitment to sites proximal to gene bodies, and in this way contribute to the specific localization of MSL complex to its targets.

To test this hypothesis, we performed contingency table analysis to determine how CLAMP motifs and occupancy differ in their position on the X-chromosome compared with autosomes. On both the X-chromosome and autosomes, we found over-representation of the CLAMP motifs and CLAMP occupancy at the transcription site/promoters similar to many other transcription factors (Fig 3A, 3B). However, on the X-chromosome CLAMP motifs and occupancy are enriched over gene bodies and towards the 3' ends unlike other transcription factors (Fig 3A, 3B). We compared the distribution of CLAMP motifs and occupancy to other transcription factors, such as GAF. The GAF transcription factor is also altered in its occupancy on the X-chromosome over gene bodies but not at TTS. The higher CLAMP occupancy in gene bodies on the X-chromosome compared to autosomes correlates with CLAMP recruitment of MSL complex to gene bodies on the X-chromosome.

Sequences flanking the core 8bp motif within the MRE are required for CLAMP recruitment

Next, we wanted to determine whether the length of repetitive GA motifs influenced CLAMP occupancy and the subsequent recruitment of the MSL complex. Using the protein binding microarray (PBM) approach (Berger and Bulyk 2009), we assessed the direct DNA binding of CLAMP to thousands of genomic sequences. Our PBM was customized with probes designed to contain endogenous genomic sequences from CLAMP ChIP-seq data and synthetic sequences designed to identify critical regions

of the MRE. There were four major categories of probes on the array (Fig 4A). The first type of probe contained the conserved 8bp MRE core motif. Within this type of probe there were both sequences occupied by CLAMP *in vivo* (ChIP Positive PBM-MRE Positive) and not occupied by CLAMP *in vivo* (ChIP Negative PBM-MRE Positive). There was also a category of probes with sequences that did not contain the MRE 8bp core motif but that were occupied by CLAMP *in vivo* (ChIP Positive PBM-MRE Negative). In addition, probes in which the MRE 8bp core motif was intact but the flanking DNA was mutated to an artificial constant (non-endogenous) sequence (PBM-MRE Positive Constant Flank) were included on the array.

Using this custom array, we defined a number of sequence features that contribute to CLAMP binding. First, sequences flanking the 8-bp core of the MRE are critical for high affinity CLAMP binding (Fig 4B, Supplemental Fig 1). To provide additional *in vitro* testing of this result, we performed electrophoretic mobility assays (EMSAs) with purified CLAMP protein and synthetically designed sequences from the PBM. Prior PBMs revealed that CLAMP binds dsDNA via a 6-Zinc finger domain (Soruco et al. 2013). For the purposes of our EMSAs, the 6-Zinc finger domain of CLAMP was fused to MBP (Maltose Binding Protein) tag. The MBP tag was not cleaved as this rendered CLAMP insoluble. Functional binding of the CLAMP 6-Zinc finger protein was tested with a MRE sequence previously identified by PBM (Supplemental Fig 2A, 2B). As a negative control CLAMP protein was also incubated with the motif for the transcription factor *Dissatisfaction (Dsf)*, a nuclear hormone receptor. Additional controls included incubating purified *Dsf* with a CLAMP motif or with a

Dsf motif. CLAMP did not bind the *Dsf* motif (Supplemental Fig 2A). Further, purified *Dsf* bound the *Dsf* motif but did not bind the CLAMP motif (data not shown). In addition, MBP protein alone was incubated with CLAMP binding motifs to ensure that the binding we observed was due to CLAMP and not the MBP tag. MBP alone did not bind to the CLAMP motif (Supplemental Fig 2A). Together these results indicate that CLAMP 6-Zinc finger binds specifically to the PBM-identified motif in the EMSA assay.

We compared CLAMP binding to a sequence that we determined was a high affinity CLAMP site on the X-chromosome versus a sequence that had constant non-endogenous DNA flanking the 8bp core motif. The high affinity status of the CLAMP binding site was defined using probes with the highest positive fluorescence signal value from the CLAMP PBMs. Using EMSAs, we tried to compete away CLAMP binding by adding unlabeled probe to a reaction containing biotin labeled probe and CLAMP protein. Sequences with constant (non-endogenous) flank did not compete well with CLAMP binding compared to sequences identical to the high affinity X-chromosome site (Fig 4C). In addition, these constant flank probes when incubated with increasing concentrations of CLAMP protein did not bind CLAMP *in vitro* (data not shown). We therefore conclude that the flanking DNA outside the 8bp core is critical for direct CLAMP binding.

We also performed ChIP on male and female larvae that had been genetically engineered by autosomal site-specific insertion of a tandem array of three MREs in a

150bp window (Alekseyenko et al. 2008). In one type of larvae (Scrambled (ART)), the inserted sequences had the 8bp core intact and all adjacent nucleotides scrambled. In a second type of larvae (Control (SYA)), the entire 21bp MRE sequence was intact but the bases immediately adjacent to the 21bp MREs were scrambled. Mutating the regions immediately adjacent to the 8bp core led to significant loss in CLAMP occupancy at these sites (Fig 4D), revealing that sequences outside the 8bp core increase CLAMP binding *in vivo*. We therefore hypothesized that there must be specific nucleotides within these flanking DNA that are critical to generating a high affinity CLAMP binding sequence.

Using the custom PBM data signal intensities, we tested how flanking DNA contributed to binding. We ranked all probes based on their CLAMP binding signal into three classes: top 5%, top 25%, and top 50% probes. Next, we defined the CLAMP motifs using a z-score: the top 5% motif had the best z-score (Fig 5A). We found that the highest ranked motifs had the best z-scores and also had more GA dinucleotide repeats than sequences exhibiting modest binding. Taken together, these observations indicate that additional GA repeats are strong predictors of the strongest CLAMP binding sites.

DNA Shape analysis reveals GA repeats as critical for CLAMP binding

We also performed DNA shape analysis on the strongest and weakest X versus A sequences bound by CLAMP to determine whether the additional GA repeat effect

was due to sequences found on the X-chromosome versus autosomes. Using a Mann-Whitney test, we found modest overall increase in binding of sequences from the X-chromosome versus autosomes (Supplemental Fig 3A). When we made a histogram of $\log_2(\text{signal_intensity})$ to detect the root cause of this difference in binding, we observed two Gaussian distributions (Supplemental Fig 3B). Fitting a 2-component Gaussian model to the data, we separated the probes into weak, medium, and strong (Supplemental Fig 3C). Weak and strong probes were defined by signal intensity from PBM. When we compared CLAMP binding intensity within the classes, we determined that the increased binding of X-chromosome probes was due to effects on binding to weak X-chromosome probes as opposed to strong probes (Supplemental Fig 3D, 3E).

We then analyzed four DNA shape features: Minor Groove Width (MGW), inter-bp Roll, inter-bp Helix twist (Helt), and intra-bp Propeller twist (ProT) using the established DNAShape method for predicting DNA structural features (Zhou et al. 2013). The difference in shape features was more striking when comparing X vs A probes for the weak rather than strong probes (Supplemental Fig 3F). However, when we did multiple linear regression for the strong versus weak sites to assess how predictive these DNA shape differences were for CLAMP binding, the improvement in binding predictions was not significant $p\text{-value} < 0.05$. Therefore we concluded that DNA shape was not a primary predictor of CLAMP binding.

However, when we did a direct comparison of binding effects due to sequence only versus sequence plus shape based on a Mann-Whitney test, we found that sequence alone produced a significant effect (Supplemental Fig 3G). Given the MRE motif is constituted of GA repeats, we hypothesized that dinucleotide content within CLAMP binding sites could be an indicator of increased CLAMP binding. Performing multiple linear regression based on dinucleotide content revealed that GA repeats, not DNA shape, is a main predictor of CLAMP occupancy (Figure 5B).

Increased number of dinucleotide repeats increases CLAMP binding to DNA *in vitro* and *in vivo*

We further tested the effects of additional GA repeats on CLAMP binding by comparing *in vitro* binding of two CLAMP zinc finger domains. We generated a CLAMP truncated protein with 4 Zinc fingers instead of 6 Zinc fingers and compared their *in vitro* binding on a PBM. The CLAMP 6-Zinc finger protein had increased binding compared to the CLAMP 4-Zinc finger protein (data not shown). Together, these observations suggest that additional zinc fingers help stabilize the binding.

The PBM classification system revealed that the top CLAMP motif has more GA repeats than lower motifs (Fig. 5A) and that more zinc fingers increase probability of binding. We therefore wanted to test how an increased number of repeats alters CLAMP binding *in vivo* and *in vitro*. We used RPKM (Reads per Kilobase transcript per Million reads) analysis for CLAMP ChIP-seq data over regions with different

numbers of GA repeats and assessed the number of GA repeats at sites occupied by CLAMP *in vivo*. We found that increased numbers of GA repeats increase CLAMP occupancy *in vivo* (data not shown). Gel-shifts with probes of identical base-pair (bp) length that encoded an increasing number of GA repeats within the same flanking DNA were used to determine the number of GA repeats at which CLAMP *in vitro* binding is acquired. CLAMP did not bind *in vitro* to probes with 10bp or 14bp GA repeats but bound to probes with 26bp, 30bp, or 58bp GA repeats (Fig 6A). We also used RPKM analysis to measure the number of GA repeats on the X-chromosome versus autosomes. We found that the X-chromosome has a greater number of GA repeats relative to autosomes (Fig 6B). The expanded number of GA repeats on the X-chromosome correlates with the observed increase in CLAMP binding and occupancy on the X-chromosome relative to autosomes.

We propose that from the DNA sequence based perspective, increasing GA repeat number is key for high affinity CLAMP binding. Increasing GA dinucleotides is a strong driver of CLAMP binding and as more CLAMP binds more MSL complex is recruited to sites of active transcription marked by H3K36me3 (Fig 7: Model). This new finding reveals the significant effects DNA sequence alone independent of chromatin status can have on co-factor recruitment. However, we also recognize that CLAMP binding due to GA-repeat numbers does not occur in a chromatin vacuum physiologically. CLAMP must work in concert with active chromatin features such as H3K36me3 that initially flags the functional CLAMP sequences to be bound based on their accessibility. This concept is supported by previous studies

demonstrating that active chromatin features, such as H3K36me3 and increased GC content in the DNA flanking MREs, distinguishes MREs on the X-chromosome targeted in dosage compensation independent of MSL binding (Alekseyenko et al. 2012).

Conclusion

Our findings support a model whereby subtle differences in specific DNA motif distribution that alter the binding of one protein facilitate the discrimination of the X-chromosome from autosomes. In the case of the transcription factor CLAMP, higher numbers of GA repeats on the X-chromosome relative to autosomes increases CLAMP occupancy on the X-chromosome. Increased CLAMP occupancy elevates MSL complex recruitment, and MSL complex can release paused polymerase to upregulate transcription of the X-chromosome for dosage compensation (Larschan et al. 2011b). While DNA sequence is a powerful driver of CLAMP targeting, roX RNAs and K36me3 also contribute. In the future we hope to determine how these three different components (DNA, RNA, and chromatin marks) work together to modulate co-factor targeting. It is possible that JIL-1 kinase which associates with MSL complex and loosens chromatin by activating PARP is a first step in making CLAMP binding motifs accessible to CLAMP protein (Thomas et al. 2014). After the first wave of CLAMP protein binds to its target sequences, a symphony of interactions between roX2 RNAs and H3K36me3 further mark the functional sequences to be targeted for dosage compensation. roX2 RNAs and

H3K36me3 have been previously demonstrated to have participatory roles in helping distinguish sites of active transcription (Meller 2003; Deng and Meller 2006; Bell et al. 2008; Alekseyenko et al. 2012; Ilik et al. 2013).

CLAMP binding to GA repeats helps provide a mechanism to partially explain the X-chromosome specific targeting of MSL complex. As a GA repeat binding protein, CLAMP fits into the ranks of other repeat binding proteins involved in transcriptional regulation: SatB1 and bric à brac (Bab) proteins. Bab proteins are necessary for the development of certain organs in *Drosophila melanogaster* (Lours et al. 2003). The BAB proteins, Bab1 and Bab2, do not contain a zinc finger domain but share a conserved domain (BabCD) that binds to A/T rich sequences (Lours et al. 2003). TA or TAA repeats constitute the best BabCD binding sequences (Lours et al. 2003). SatB1, a chromatin remodeler, specifically binds AT-rich sequences that contain ATC repeats (Dickinson et al. 1992). Mutations in the AT sequences decrease SatB1 binding (Dickinson et al. 1992). Mutations altering the winding of DNA, such as mutations disrupting shape of the DNA, also reduce SatB1 binding (Dickinson et al. 1992).

Like the SatB1 repeat binder, CLAMP is associated with factors involved in chromatin remodeling, such as JIL-1 kinase, H4K16 acetylation, and H3K36me3 (Dickinson et al. 1992; Soruco et al. 2013; Wang et al. 2013a). CLAMP's ability to functionally bind GA repeats provides a potentially broader function for CLAMP in helping regulate gene expression genome wide. The enrichment of GA repeats on

the X-chromosome helps explain why we observe modest enrichment of CLAMP protein on the X-chromosome relative to autosomes although CLAMP binds genome-wide.

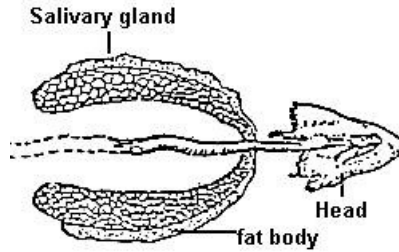
CLAMP binding to GA repeats also highlights its role in the initial steps of establishing a domain of coordinate gene regulation. Chromatin entry sites that recruit MSL complex to the X-chromosome for dosage compensation were recently demonstrated to have evolved through transposon element mediated GA-repeat expansion (Alekseyenko et al. 2013; Ellison and Bachtrög 2013). This was found to be a common mechanism for the acquisition of X-chromosome dosage compensation across a variety of *Drosophila* species, including *D. miranda*, *D. subobscura*, and *D. pseudoobscura* (Alekseyenko et al. 2013; Ellison and Bachtrög 2013). Future studies analyzing CLAMP occupancy in these different species could provide insight into the role of CLAMP in the evolution of dosage compensation. Our work here indicates that CLAMP's direct binding of GA repeats is the missing factor that targets MSL complex directly to these CES in the evolution of dosage compensation.

Materials and Methods

Immunostaining of polytene chromosomes

Polytene staining was performed as described below. The rabbit anti-CLAMP antibody was used at 1:50. Alexafluor 594 (anti-rabbit) secondary antibodies (Invitrogen) were used for immuno-detection at a 1:200 dilution.

1. Grow and collect fat, healthy third instar larvae from 18°C. Third instar crawl up the vial sides prior to pupation. Vials should not be too crowded, and the first wave of wanderers is the best for polytene chromosome preparation.
2. Sex larvae, if desired. Females have better polytenes (and twice as much X chromosome DNA). Sex larvae in ice-cold 0.7% NaCl or water. Look for presence or absence of testes. Keep collection dish on ice to paralyze larvae.
3. Siliconize 22mm square coverslips with RainEX and then wipe with a wet kimwipe. Keep covered to avoid dust.
4. Extract the salivary glands in 0.7% NaCl solution (hypotonic, so makes the tissues swell) or 1X PBS. Use #5 tweezers to dissect the fat body (more opaque) without puncturing or splitting the glands.



5. Collect a few glands in RT 1X PBS.
6. Add ~200ul **Fix1** to a well slide. Move several glands to this slide for **60 seconds**. Do not overfix.
7. Next, add ~200ul **Fix2** to another well slide and fix glands for **2 minutes**.
8. Put ~15ul **1:2:3 fixative** solution on each siliconized coverslip. Using tweezers, carefully transfer 2-3 glands to each (#1) coverslip droplet. Allow to fix for **5 minutes**.
9. Touch the fixative droplet to a slide labeled in pencil. Quickly flip over. Press with a kimwipe on one side of the coverslip so it doesn't move laterally. Using a dissecting needle, tap on the coverslip over the glands to break apart the nuclei. Using medium pressure, draw 2-3 times over the coverslip. Blot fixative with a kimwipe.
10. Examine the chromosomes under the phase contrast microscope for sufficient spreading. Do this quickly, as the chromosomes will firm up. Tap more with the needle if necessary. Score with a diamond scribe.
11. Store slides in a humid chamber (pipette box with wet paper towel below the holed insert) until you have at least 5.

12. Freeze slides in liquid nitrogen. Immediately flick off the cover slip with a razor blade.

13. Store slides in 100% EtOH at -20°C for up to a week.

Immunostaining:

1. Wash 15 min in 1X PBS.
2. Wash 10 min in 1% Triton X-100 (in 4°C) in 1X PBS.
3. Block 1 hour in blocking solution (0.5% BSA in 1X PBS).
 - a. After 40 min, start preparing primary antibody.
4. Incubate primary antibody: put 80ul on each cover slip, touch each to the appropriate slide, and incubate overnight in a **dark** humid chamber (wrap in foil) at 4°C (or 2 hours at RT). For Immunostaining of proteins use dilutions specific to each antibody and protein.
5. Shake off the cover slips and wash X3 15 min in 1X PBS.
6. Incubate secondary antibody: to detect use a secondary AlexaFluor (1:1000 dilution), 0.5ul/500ul blocking solution, 80ul per cover slip. Incubate 2 hrs at RT in a **dark** humid chamber.
7. Wash X3 5 min in 1X PBS. Pull Prolong Gold medium out of -20°C to thaw.
8. Counterstain with Hoechst: 10ul in 100mL 1X PBS. Pass slides through Hoechst for 10 sec at RT.
9. Wash X3 10 min in 1X PBS.

10. Mount slides: dry slides as much as possible. On each (#1.5) cover slip put a drop of Prolong Gold Anti-fade mounting medium. Touch drop to slide and flip over.
11. Tack corners of cover slip with nail polish and dry for 10 minutes.
12. Analyze under a fluorescent microscope

0.7% NaCl:
0.7g NaCl
100 mL H₂O

5X formaldehyde stock (700uL aliquots, store unused solution at -20°C):
0.74g paraformaldehyde
4.0mL dH₂O
28ul 1N KOH

*Fix1:
167ul 10X PBS
16.7ul Triton X-100 (stored at 4°C)
1150ul dH₂O
333ul 5X formaldehyde stock

*Fix2:
500ul dH₂O
833ul glacial acetic acid
333ul 5X formaldehyde stock

*1:2:3 fixative solution:
1 part lactic acid
2 parts water
3 parts acetic acid
(by volume. Make ~2mL at most)

20X SSC:
175.3g NaCl (3M)
88.2g Sodium Citrate (0.3M)
fill to 1L with H₂O

10X PBS:
80g NaCl
2g KCl
14.4g Na₂HPO₄
2.4g KH₂PO₄
fill to 1L with H₂O

1% Triton X-100 in 1X PBS:
2.5 mL Triton X-100
247.5 mL 1X PBS

0.5% BSA in 1X PBS: (blocking)
0.5 g BSA

Chromatin immunoprecipitation (ChIP-qPCR) from third instar larvae

Chromatin Preparation

The following protocol is modified from Larschan et al., 2007. Chromatin was prepared from 150 sexed third instar larvae ground into a powder in a mortar and pestle while frozen in liquid nitrogen. Larvae were homogenized in a 40 mL Dounce homogenizer in 40 mLs of PBS containing a protease inhibitor cocktail (PI) (pancrease-extract 0.002 mg/ml, thermolysin 0.0005 mg/ml, chymotrypsin 0.002 mg/ml, trypsin 0.02 mg/ml, papain 0.33 mg/ml) (Roche). Next, the extract was crosslinked with 1% formaldehyde. Samples were rotated for 20 min at room temperature followed by quenching with 2 mLs 2.5 M glycine for 5 minutes at room temperature. The extract was chilled on ice and centrifuged for 5 minutes at 1500 g at 4 °C. The pellet was resuspended in 40 mLs PBS containing 1x PI and 0.2 mM PMSF followed by centrifugation for 5 minutes at 1500 g at 4 °C. Centrifugation and resuspension were repeated as described above with the following buffers: 1) 10 mLs Wash A (10 mM HEPES pH 7.6, 10 mM EDTA pH 8.0, 0.5 mM EGTA pH 8.0, 0.25% Triton-X 100, 1x PI, 0.2 mM PMSF); 2) 10 mLs Wash B (10 mM HEPES pH 7.6, 10 mM NaCl, 10 mM EDTA pH 8.0, 0.5 mM EGTA pH 8.0, 0.25% Triton-X 100, 1x PI, 0.2 mM PMSF); 3) 10 mLs TE Wash (10 mM Tris-HCL pH 8.0, 1 mM EDTA pH 8.0, 0.01% SDS, 1x PI, 0.2 mM PMSF); 4) 10 mLs TE Wash with SDS (10 mM Tris-HCL pH 8.0, 1 mM EDTA pH 8.0, 1% SDS, 1x PI, 0.2 mM PMSF), and 5) 2 washes with 10 mLs TE Wash. Next, the pellet was resuspended in 2 mLs Pre-RIPA (0.1% SDS, 10 mM

Tris-HCL pH 8.0, 1 mM EDTA pH 8.0, 1x PI, 0.2 mM PMSF). A Misonix Q125 tip sonicator was used to fragment 400 μ L aliquots of chromatin for 1 min total sonication time (20 % amplitude, 6 intervals of 10 sec). The aliquots were then pooled together and Triton-X 100 was added to a final concentration of 1% followed by rotation at 4 °C for 1 minute. Next, deoxycholate was added to a 0.1% final concentration followed by rotation at 4 °C for 1 minute. Then, NaCl was added to a final concentration of 140 mM followed by rotation at 4 °C for 5 minutes.

Next, 200 μ L of the chromatin preparation was saved for the input control sample. To reverse the crosslinks on the input samples, the following reagents were added: 1) 200 μ L of RIPA 150 (50 mM Tris-HCl pH 8.0, 1% NP-40, 2 mM EDTA pH 8.0, 0.1% sodium deoxycholate, 0.1% SDS, 150 mM NaCl, 1x PI, 0.2 mM PMSF, 1 mM DTT); 2) SDS to a final concentration of 1%; 3) 15 μ L 5M NaCl; 4) 1 μ L of DNAase-free RNase A (Roche). Samples were incubated at 65 °C overnight. To digest proteins and RNA, the following reagents were added: 1) an additional 1 μ L of RNase A followed by incubation at 37 °C for 30 minutes; 2) 20 μ L 1M Tris-HCl pH 6.8; 3) 10 μ L 0.5M EDTA pH 8.0; and 4) 3 μ L 20 mg/mL proteinase K (Invitrogen) followed by incubation at 42 °C for 90 minutes. To remove digested protein and enzymes, phenol-chloroform-isoamyl alcohol (Invitrogen) extraction was performed followed by chloroform extraction using Phase-Lock tubes (Eppendorf). DNA was precipitated by adding 10 μ L glycogen (Ambion), 1/10 volume of 3 M NaAc, and 2 volumes of 100% EtOH followed by incubation at -80 °C overnight. Samples were

centrifuged at 20,000 g at 4 °C for 10 minutes followed by washing the pellet with 1 mL 70% ethanol and resuspension in ddH₂O. Concentrations were obtained using the Qubit DNA High Sensitivity reagents (Invitrogen) and fluorimeter. 500 ng of each sample was run on a 2% agarose gel to determine that chromatin fragmentation produced fragments from 200-500 bp in length. The remainder of the chromatin preparation was resuspended in RIPA 150 buffer to a final concentration of 10 µg/mL.

Immunoprecipitation

The immunoprecipitation protocol was modified from Alekseyenko et al., 2006. Immunoprecipitation was performed using 1 mL of the appropriate chromatin preparation combined with antibody at 4 °C overnight. Anti-MSL2 antibodies were described previously (Kelley et al. 1999). The anti-CLAMP antibody was described previously (Larschan et al. 2012) and the anti-H3 was obtained from Abcam. 60 µL of pre-blocked Protein A agarose beads (Millipore) were added to each IP followed by rotation at 4 °C for 2 hrs. Each IP was then centrifuged at 1000 g at 4 °C for 3 minutes. Beads were resuspended in cold RIPA 150 buffer followed by rotation at 4 °C for 5 minutes and centrifugation at 1000 g at 4 °C for 3 minutes. Washes were repeated with RIPA 150, RIPA 300 (50 mM Tris-HCl pH 8.0, 1% NP-40, 2 mM EDTA pH 8.0, 0.1% sodium deoxycholate, 0.1% SDS, 300 mM NaCl), LiCl/TE (0.25 M LiCl, 1% NP-40, 1% sodium deoxycholate, 10 mM Tris-HCl pH 8.0, 1 mM EDTA pH 8.0) and two times with TE buffer (10 mM Tris-HCl pH 8.0, 1 mM EDTA pH 8.0). The

samples were eluted off the beads twice with 250 μ l Sodium Bicarbonate Buffer (1% SDS, 0.1 M sodium bicarbonate) by rotating at room temperature for 15 minutes followed by centrifugation at 1000 g for 3 minutes.

Quantitative Real-Time PCR

Primer sequences are available upon request and were previously reported (Alekseyenko et al. 2008). Briefly, all primers were tested to assure that they have 100% efficiency and good linearity prior to use with an ABI 7000 real time PCR machine. The average relative quantitation (ABI User Bulletin #2) of three replicate reactions was determined for all samples. Platinum SYBR master mix w/ROX (Invitrogen) was used. Two independent chromatin preparations were conducted for each experiment and standard deviations are reported. Dilutions of input and IP DNA were conducted to assure that samples were in the linear range for quantitative PCR. To assess the overall ChIP enrichment, normalization was conducted using primers within either the CG15570 or PKA internal control genes that span genomic loci that are unbound by MSL complex or CLAMP as assessed by ChIP-seq. Normalization to input DNA was also conducted for ChIP experiments. After normalization to input, ChIP data were normalized within each biological replicate to data from the primer pair with the highest enrichment value. Subsequently, biological replicates were averaged and standard deviations calculated.

***in vitro* protein expression for protein binding microarrays (PBMs)**

The pDEST15 Gateway vector (Invitrogen) was used to generate the expression plasmid for the C-terminal CLAMP four and six zinc-finger constructs. For the six-finger construct (amino acid residues 350-561), the following primers were used for cloning:

5' ACTGTTAAACTGAAAATCCGTCCAA 3' and
3' CTATAACCCACCGATAATCACTTCCTGT 5'.

For the four-finger construct, (amino acid. residues 412-561), the following primers were used for cloning: 5' GGCGAGAAACCCTATAAATGCCAAAC 3' and 3' CTATAACCCACCGATAATCACTTCCTGT 5'. Cloning was performed using standard Gateway Technology (Invitrogen). Proteins were expressed using the TnT Quick-coupled Transcription/Translation Systems Kit (Promega). *In vitro* transcription/translation reactions were carried out according to standard protocols with the following adaptation: zinc acetate was added to a final concentration of 50 μ M.

Protein Binding Microarrays (PBMs)

In vitro produced C-terminal 4 and 6 zinc-finger regions of CLAMP were fused to GST at their N-termini and used in PBM experiments. Proteins were quantified using standard Western blotting protocols with a GST dilution series as a standard. The concentrations of the proteins used for PBMs were as follows: 1) C-terminal 4 zinc-finger protein (24.3 nM); 2) C-terminal 6 zinc-finger protein (13.9 nM).

Protein binding to microarrays was assayed as previously described (Bulyk, 2009) using a custom-designed 4x44K format oligonucleotide array (Agilent Technologies, Inc., AMADID #015681) that was converted to double-stranded DNA by primer extension as described previously (Berger et al. 2006). All protein binding microarrays were performed in duplicate using independently prepared protein samples and the same motifs were obtained. Zinc acetate to 50 μ M was added to each sample before binding to the PBMs. PBM data were quantified, normalized, and analyzed using the Universal PBM Data Analysis Suite and the Seed-and-Wobble motif derivation algorithm (Berger and Bulyk 2009). An E-score of 0.45 is significant for our PBM experiments (Badis et al., 2009). We found that $E > 0.37$ corresponds to a false discovery rate (Q-value) of 0.001 by conducting a series of negative control PBM experiments (Badis et al., 2009).

Protein expression from *E.coli* for EMSAs

The pMAL-c5X Vector containing the MBP (Maltose Binding Tag) was used to express MBP fusion CLAMP 6 Zinc finger protein using the Nde1 and Sbf1 restriction enzyme sites. The pMAL protein fusion and purification system from NEB (Catalog # E8200S) was used. The protein was purified using a maltose column. The following primers were used for cloning the 6 Zinc finger domain of CLAMP:

Forward

5' GGTAGCGTTAAACAGAGTGTTACCGT 3'

Reverse

3' TTACAGACCGCCAATAATAACTTCTT 5'

Electrophoretic Mobility Shift Assay

EMSAs were performed using the Chemiluminescent Nucleic Acid Detection Module (Thermo Scientific Cat#89880)/LightShift Chemiluminescent EMSA Kit (Thermo Scientific Cat#20148). DNA binding conditions for CLAMP protein were previously optimized with 1 ug/uL Poly (dI•dC, Cat#20148E) and Elution Buffer: 5mM Tris-HCL, pH 8.5 (AE Macherey-Nagel Cat#740588.10).

DNA labeling

The CLAMP binding sequences were end labeled with Biotin using Biotin 3'End Labeling Kit (Thermo Scientific Pierce Cat#89818) with the following adaptations.

Each strand of DNA was Biotin labeled independently then annealed using a slow cool down cycle: 95°C 10 mins, 72°C 2 hours, 52°C 2 hours, 37°C 2 hours, 4°C using a PCR machine.

Single strand DNA digestion of annealed dsDNA was performed for 10 mins at 37°C with Mung Bean Nuclease 2,000 U/uL (New England Biolabs Cat#M0250S) followed by PCR Clean-up with Qiagen Kit: Qiaquick PCR Purification Kit Cat# 28104.

Figures

Figure 1. CLAMP occupies sites across the genome

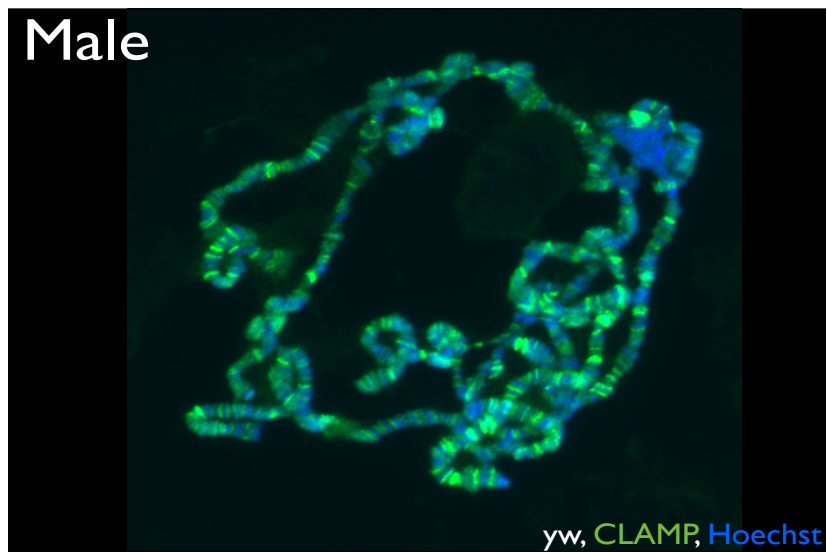
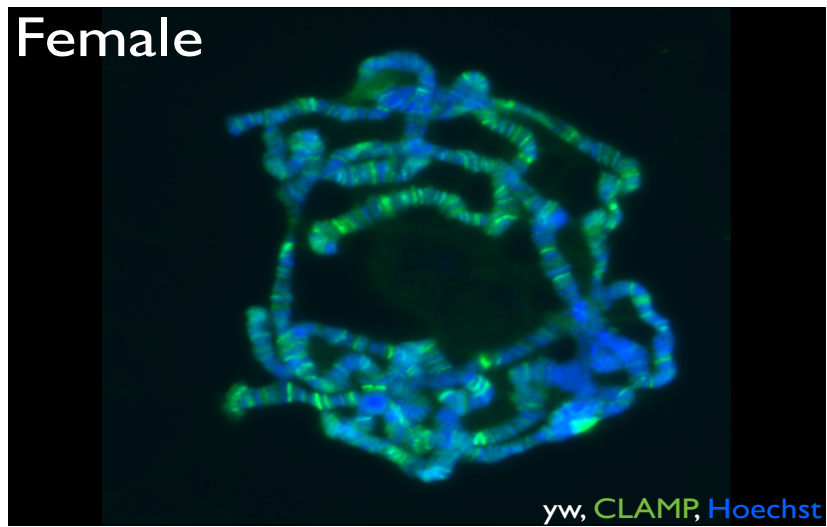


Figure 1. Clamp occupies sites across the genome. Polytene chromosomes from the salivary glands of third instar wildtype *Drosophila* male and female larvae were stained for CLAMP binding. The CLAMP protein was detected with CLAMP antibody and fluorescently labeled secondary antibody. Genomic DNA was stained with Hoechst fluorescent dye. Yw indicates the wildtype state of the larvae used for the polytene chromosomes.

Figure 2. CLAMP binding sequences are enriched on the X-chromosome relative to autosomes

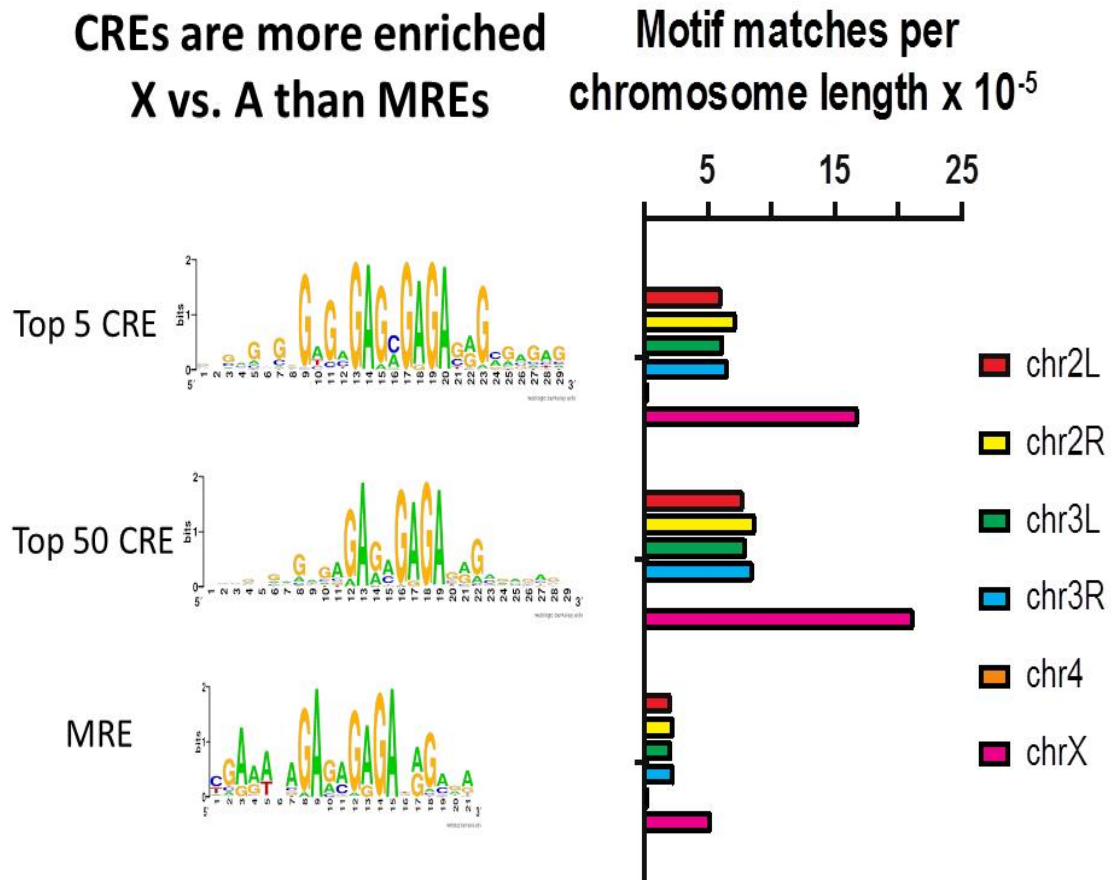


Figure 2. Using MEME (Multiple EM for Motif Elicitation) analysis, CLAMP recognition element (CRE) is the sequence discovered in the top 5% and top 50% of sites bound by CLAMP. This CLAMP binding sequence was enriched on the X-chromosome versus autosomes.

Figure 3A. CLAMP motifs are enriched in gene bodies and towards 3'ends on the X-chromosome

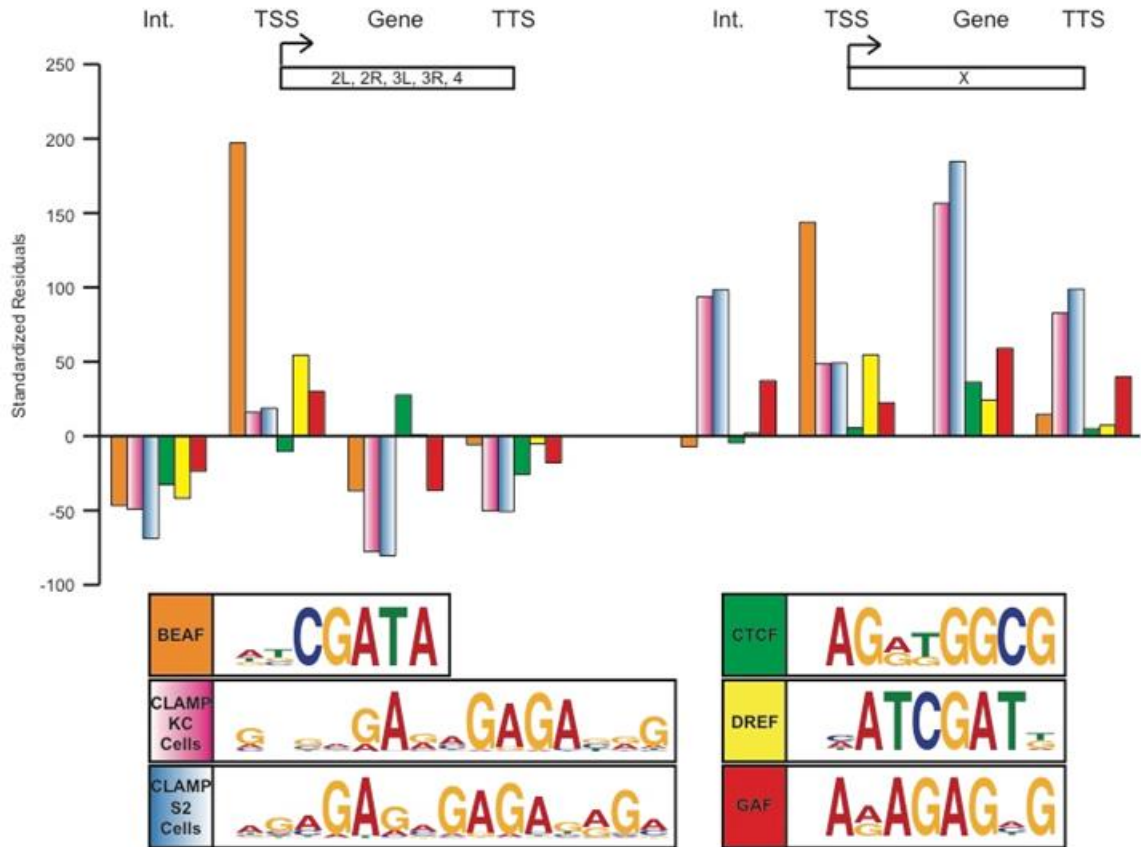


Figure 3A. The position of the binding motifs of various transcription factors (GAF, DREF, CTCF, CP190, BEAF) and CLAMP were compared across the *Drosophila melanogaster* genome. Unlike all other factors tested, the CLAMP binding motif was found enriched at 3' ends of genes and within gene bodies on the X-chromosome but not on autosomes. The motifs of each transcription factor tested was color-coded: orange = BEAF motif, pink = CLAMP motif in female (Kc) cells, blue = CLAMP motif in male (SL2) cells, green = CTCF motif, yellow = DREF motif, and red = GAF motif.

Figure 3B. On the X-chromosome, CLAMP occupancy is greater over gene bodies and towards 3' ends.

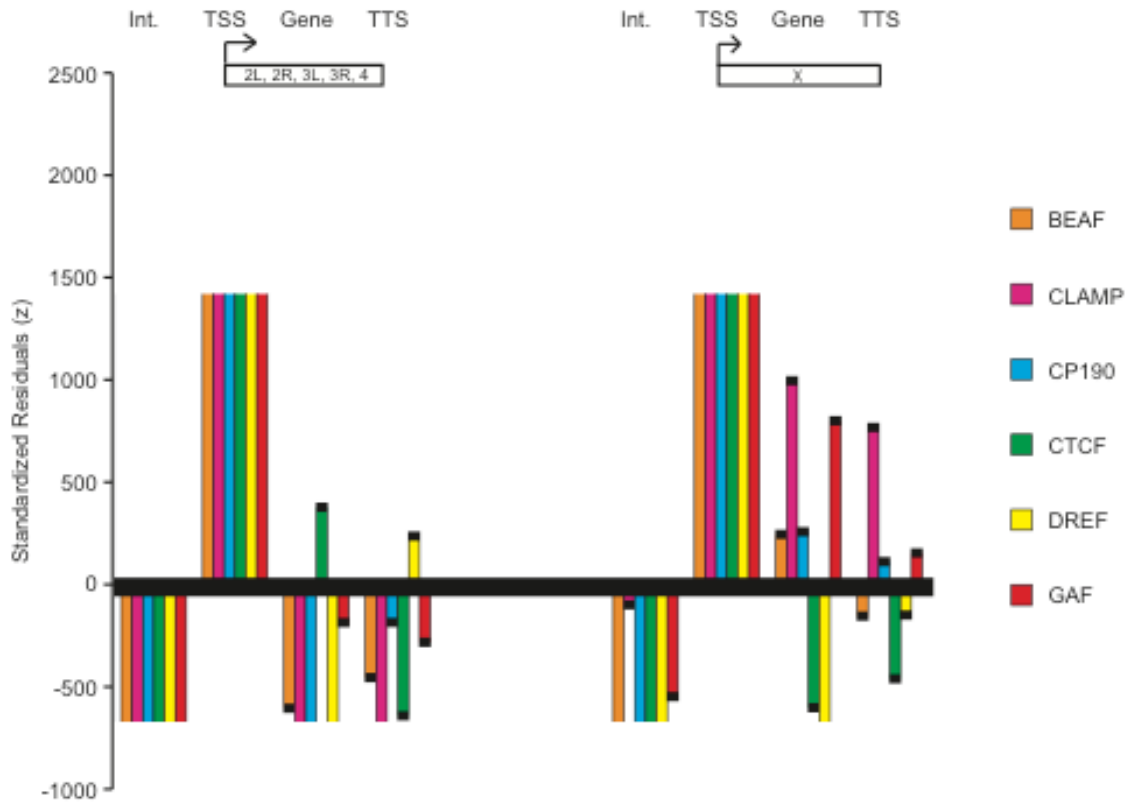
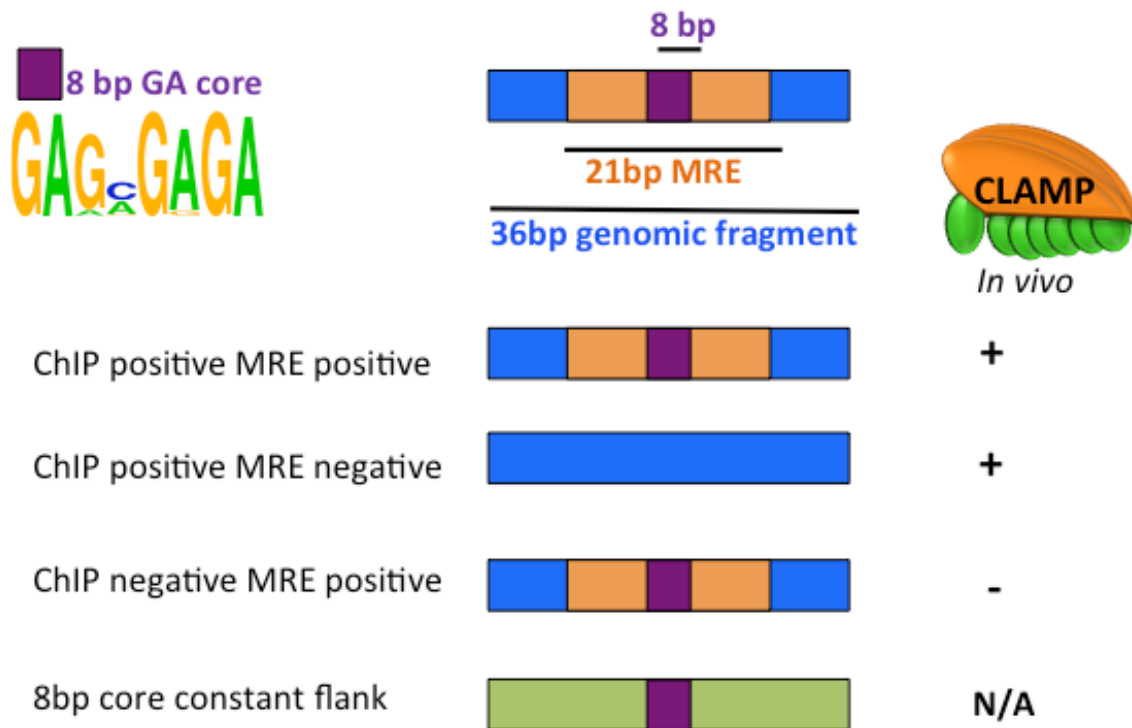


Figure 3B. RPKM (Reads Per Kilobase transcript per Million reads) analysis of CLAMP ChIP-seq reads in comparison to published data sets for BEAF, CP190, CTCF, DREF, and GAF reveal enriched CLAMP occupancy in gene bodies and 3'ends of genes on the X-chromosome but not autosomes. Motifs are similarly color-coded as in Figure 3A.

Figure 4A. Probes used for customized Protein Binding Microarray



The Clamp 8bp core *in vitro* binding sequence was previously determined using protein binding microarrays (Soruco et al. 2013). Probes were designed using genomic sequences with or without this 8bp core present to identify CLAMP's *in vitro* binding requirements beyond the 8bp core.

Figure 4B. Sequences flanking the core 8-bp region sequences increase CLAMP binding

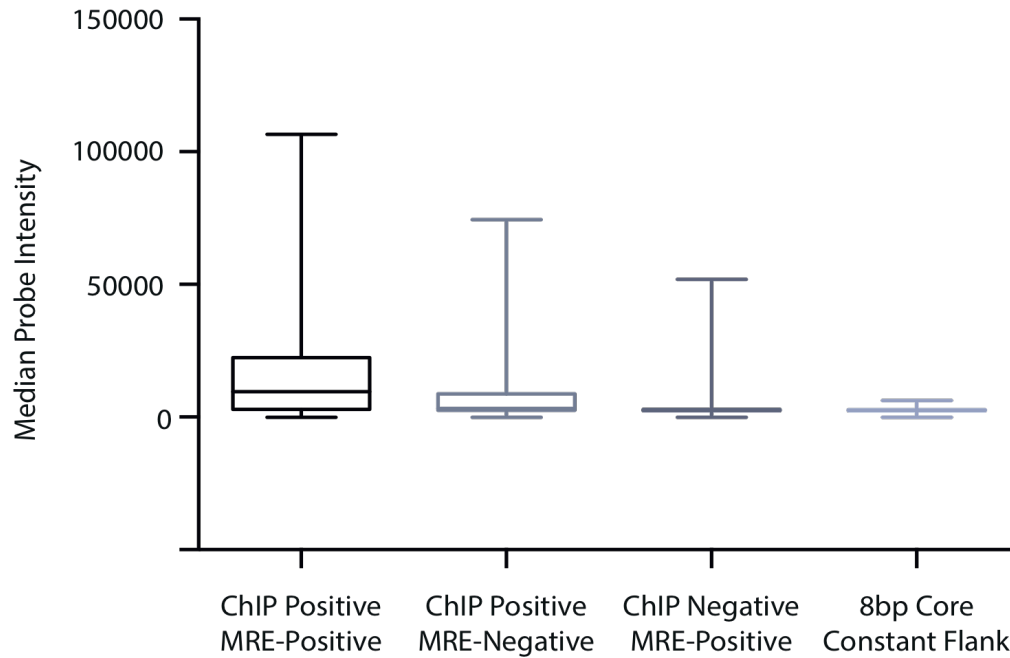


Figure 4B. The CLAMP *in vitro* binding intensities from PBM depicted higher signal for probes that contain the 8bp within particular flanking DNA. Probes with endogenous DNA flanking the 8bp core of genomic sites occupied by CLAMP *in vivo* are more highly bound *in vitro* than those without the MRE present. Probes with DNA flanking 8bp altered to an artificial non-endogenous constant flank are significantly less bound *in vitro* by CLAMP.

Figure 4C. Constant flanking DNA does not compete well with wild-type flank for binding

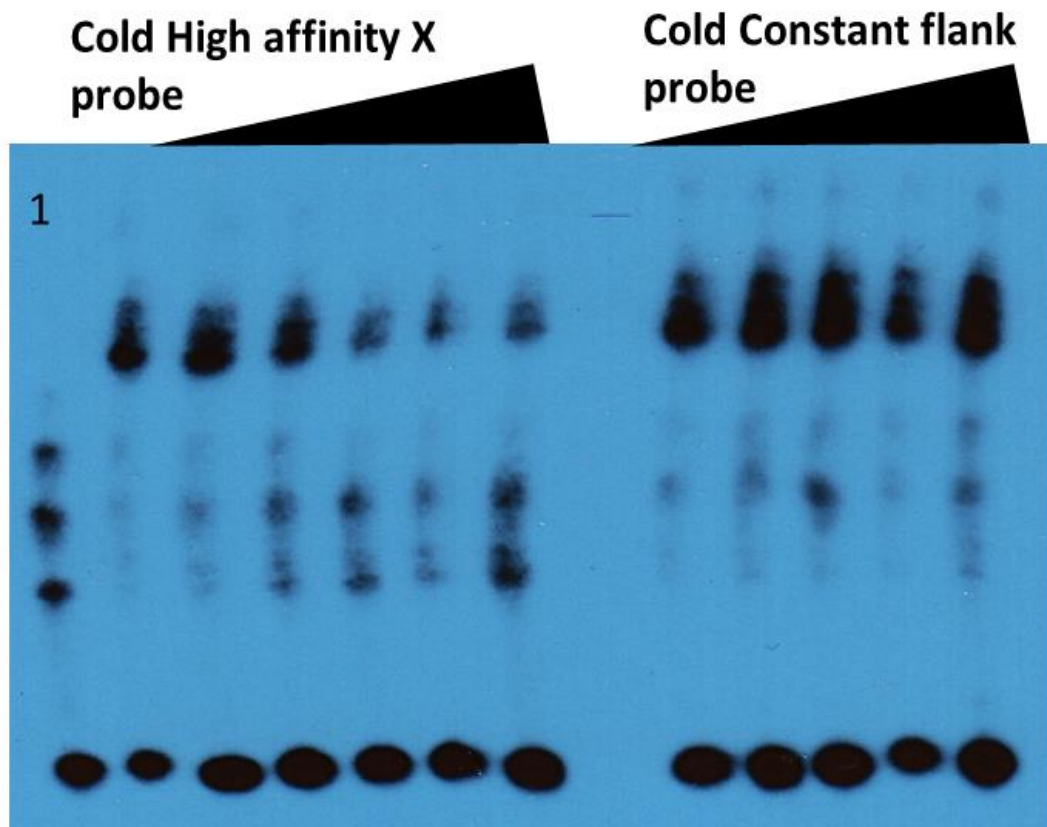


Figure 4C. Constant DNA flanking the 8bp core does not compete well for binding of CLAMP compared to a wild-type flanking sequence. Lane 1 contains Biotin labeled dsDNA with a sequence matching a high affinity CLAMP binding site on the X-chromosome. Lane 1 contains DNA alone with no protein. All other lanes contain protein, 0.11 ug/ul, with a constant concentration of labeled probe, 3.12 ng/ul. Increasing concentrations of unlabelled competitor with identical sequence to the labeled probe were added to each lane from left to right. In the cold constant flank probe gel on the right, cold competitor containing 8bp core intact with constant/non-endogenous flanking DNA sequence was added at increasing concentration. All probes used in the EMSAs were double-stranded.

High affinity X-chromosome sequence:

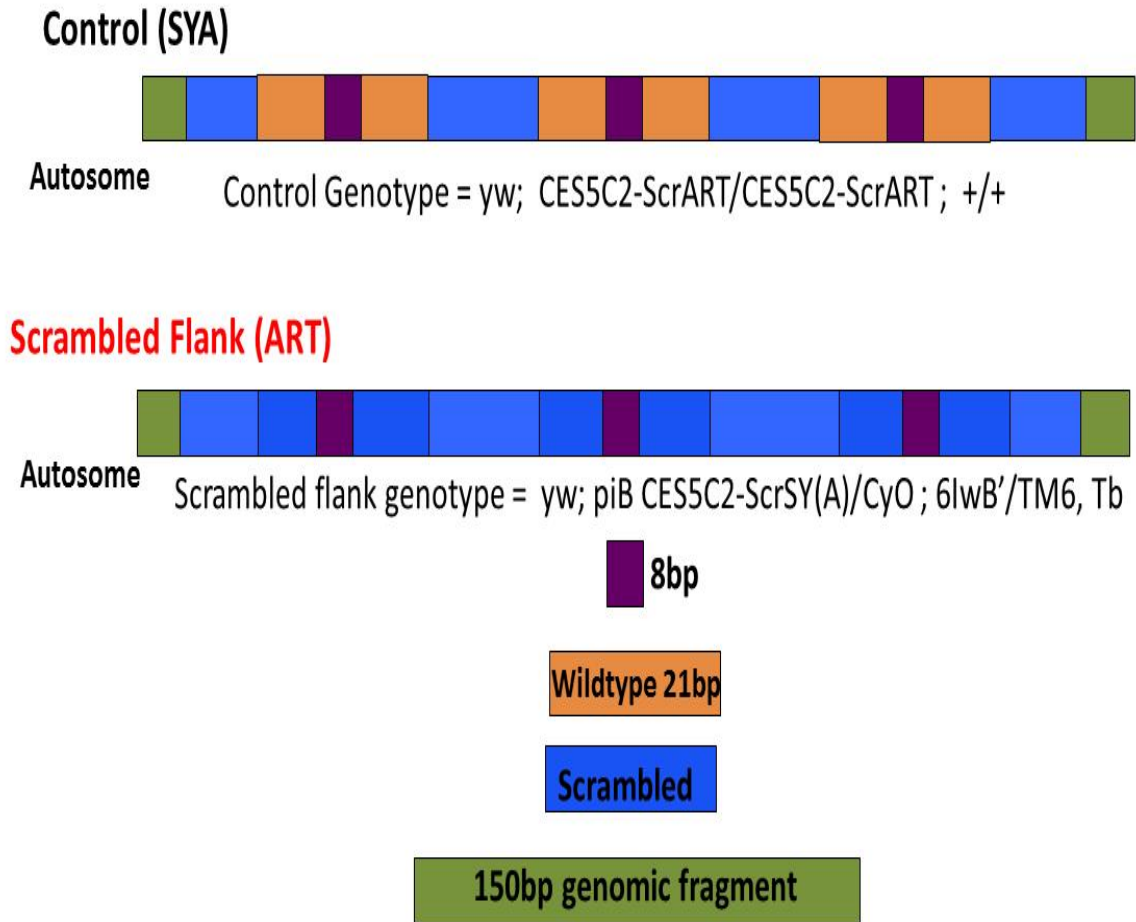
5' CTGCATGCTCTCTCTCTCCCCCTCTCGCTCACTCGC 3'

Constant sequence:

5' CTACTATAGCAATGGGAGCGAGAAGTATCAGTCAGT 3'

Figure 4D. Sequences flank the core 8 bp region increase binding *in vivo*

Drosophila melanogaster strains previously genomically engineered (Alekseyenko et al. 2008) used for CLAMP ChIP



Sex larvae and collect → Chromatin Immunoprecipitation → qPCR

Figure 4D. Sequences flanking the core 8 bp region increase binding *in vivo*

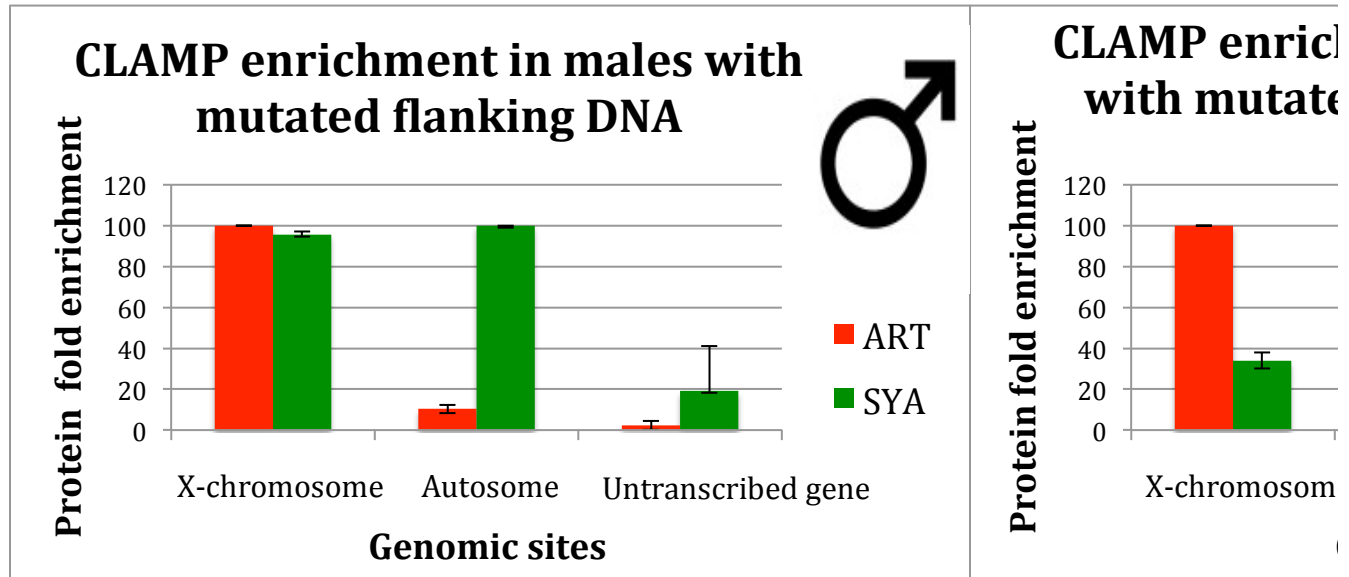


Figure 4D. *Drosophila* strains were engineered with altered genomic fragments site-specifically inserted onto autosomes as previously described (Alekseyenko et al. 2008). Larvae of these strains were used to perform CLAMP ChIP using the published method (Soruco et al. 2013). Autosome genomic sites contain inserted fragments with flanking DNA mutated. CLAMP ChIP at X-chromosome sites with the wild-type version of the inserted fragment on the autosome was also measured as a positive control. CLAMP ChIP at an untranscribed gene (CG18734) was measured as negative control.

Figure 5. The strongest CLAMP binding sites have additional GA-repeats

Figure 5A. Motifs defined from custom PBMS reveal that the top motif with the best z-score (5%) has more GAs than modestly bound motif (50%).

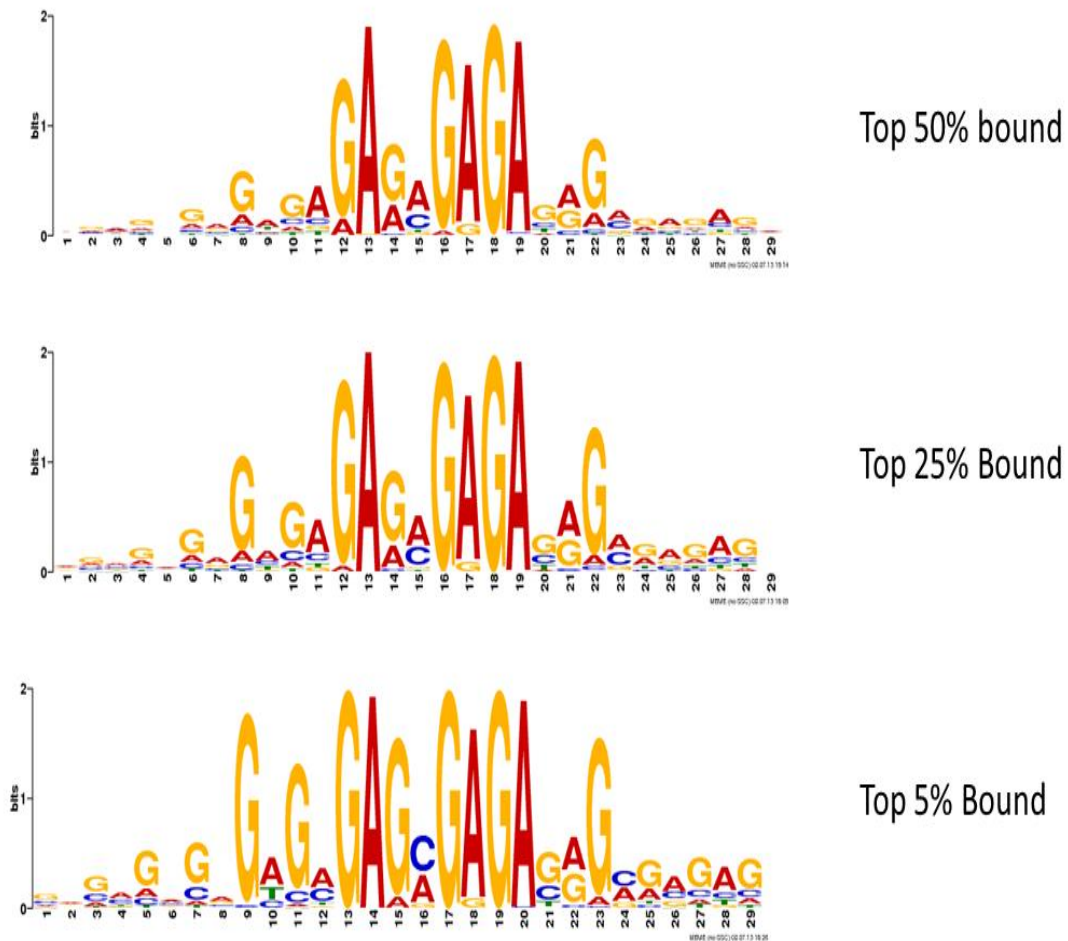


Figure 5A. MEME analysis of custom PBM data for CLAMP was done to discover motifs enriched for CLAMP binding.

Figure 5. The strongest CLAMP binding sites have additional GA-repeats

Figure 5B. Dinucleotides are the strongest predictors of CLAMP occupancy

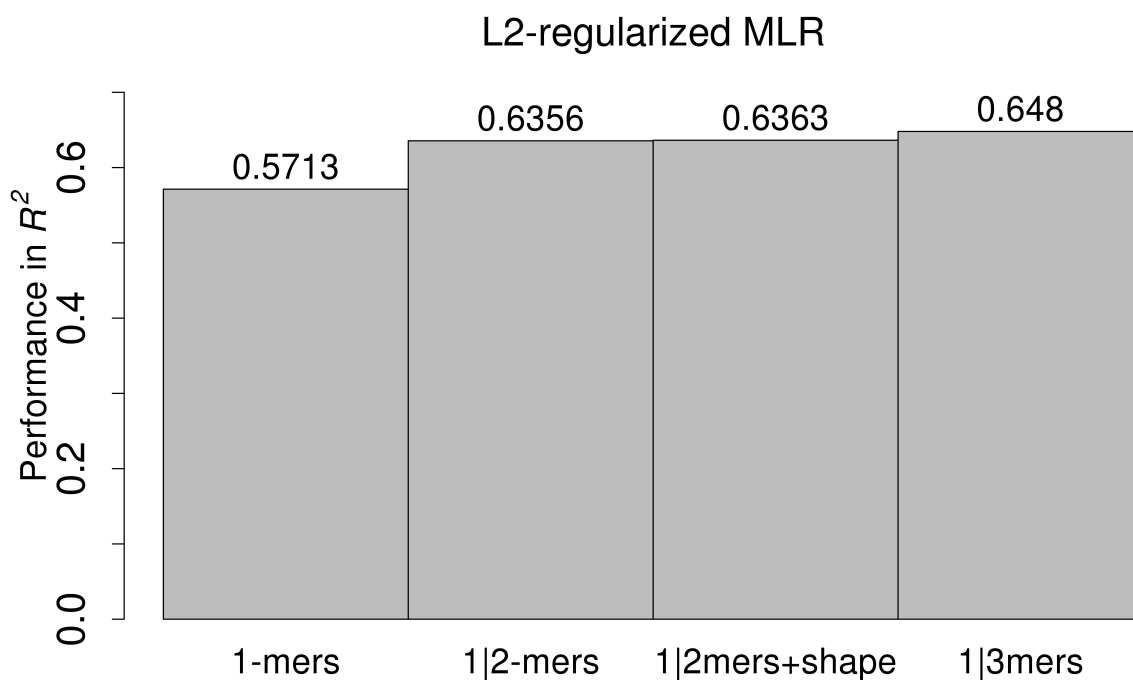
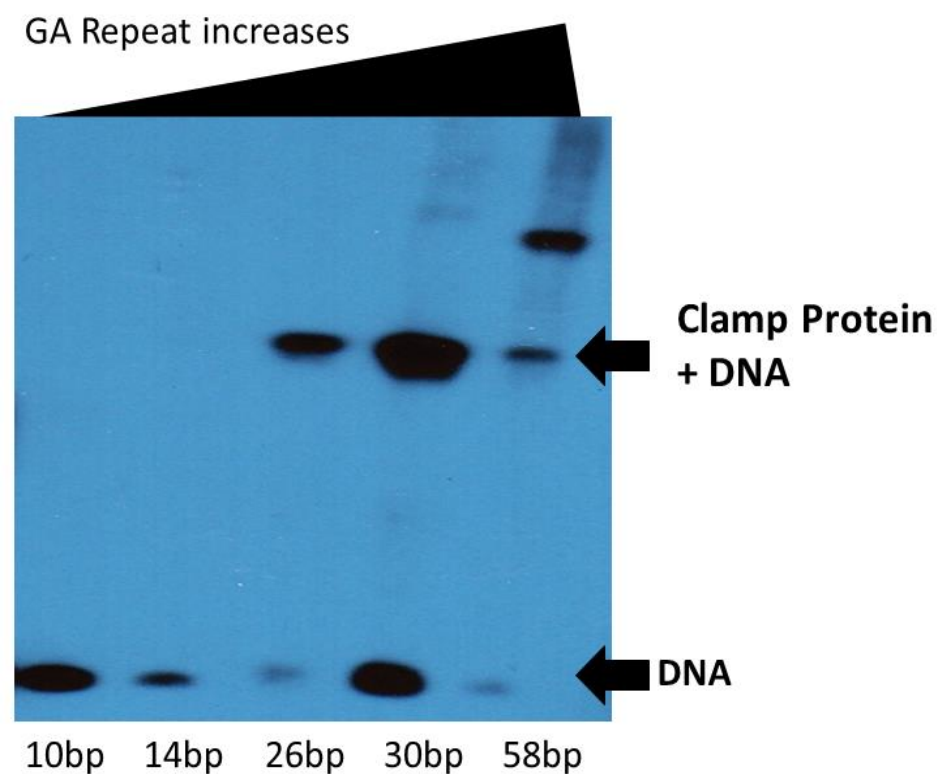


Figure 5B. Multiple linear regression modeling reveals that dinucleotide repeat, not DNA shape, is strongest predictor of CLAMP binding. Differences in the predictive powers of a single basepair (1-mers, $R^2 = 0.5713$) versus two (1|2-mers, $R^2 = 0.6356$) or three basepairs (1|3 mers, $R^2=0.648$) were compared. Dinucleotides showed significant increase in ability to predict increased CLAMP binding: 0.6356 versus 0.5713. The predictive power of dinucleotides plus shape (1|2mers+shape, $R^2 = 0.6363$) did not increase relative to dinucleotides alone (1|2-mers, $R^2 = 0.6356$) indicating that repeats alone, not DNA shape, were the strongest predictor of CLAMP occupancy.

Figure 6: Increased number of GA-repeats increases CLAMP occupancy

Figure 6A. Optimal GA repeat number for CLAMP binding *in vitro* is 26bp GA sequence



Probe Length	Probe Sequence
10bp	TCTACTTACCTCTACTACTTCTACGAGAGAGAGACTTCTACT ACTTCTACTTCTCCCAAT
14bp	TCTACTTACCTCTACTACTTCTGAGAGAGAGAGAGATCTAC TACTTCTACTTCTCCCAAT
26bp	TCTACTTACCTCTACTGAGAGAGAGAGAGAGAGAGAGAGAG AGAACTTCTACTTCTCCCAAT
30bp	TCTACTTACCTCTAGAGAGAGAGAGAGAGAGAGAGAGAGAG AGAGATTCTACTTCTCCCAA
58bp	CGA GAGAGAGAGAGAGAGAGAGAGAC

Figure 6A. The number of GA-repeats required for CLAMP *in vitro* binding was tested using probes with various lengths of GA-repeats. While the number of GA repeats was varied, all stretches of GA-repeats were embedded within the same artificial (non-genomic) flanking sequence. CLAMP did not bind to 10, 14 bp GA-repeats, but did bind to 26, 30, and 58bp GA-repeats. The supershifted band observed with the 58bp probe suggests the binding of multiple CLAMP molecules/protein to the 58bp GA-repeats probe.

Figure 6: Increased number of GA-repeats increases CLAMP occupancy

Figure 6B. X-chromosome has more GA-repeats than autosomes

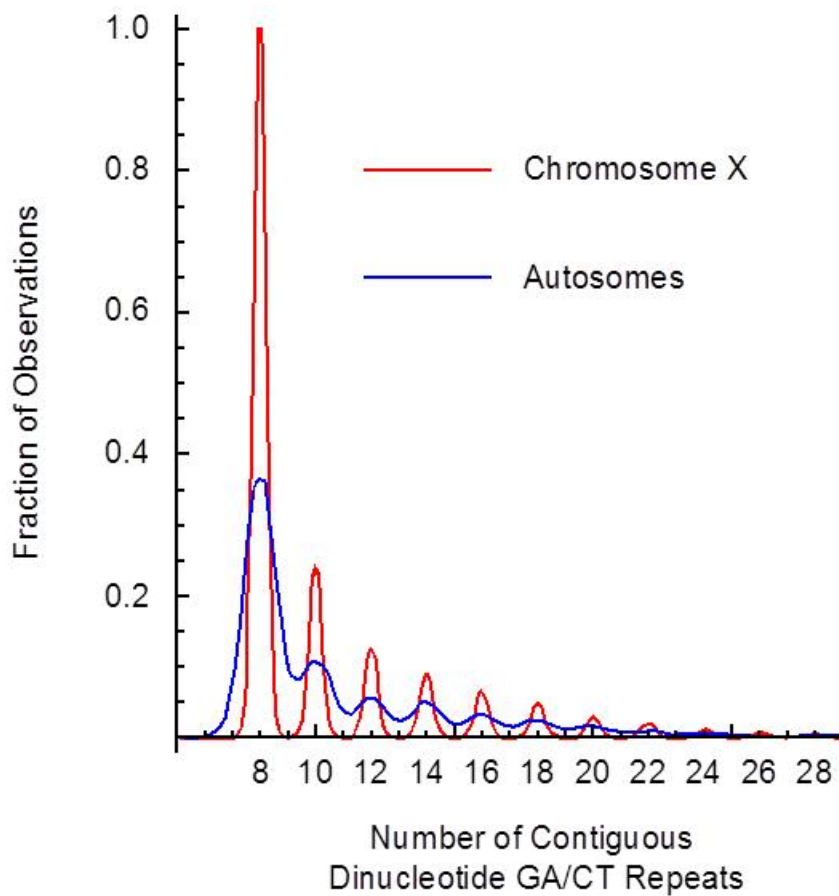
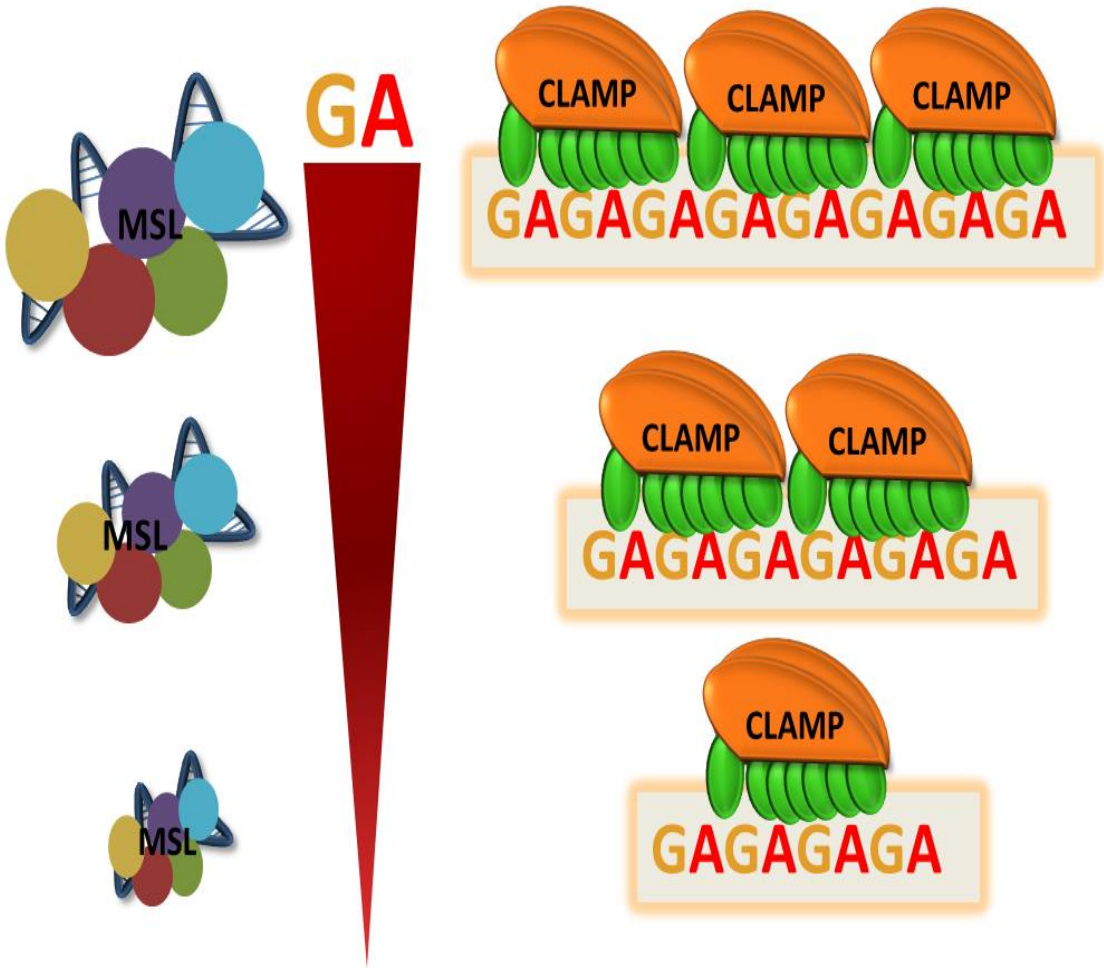
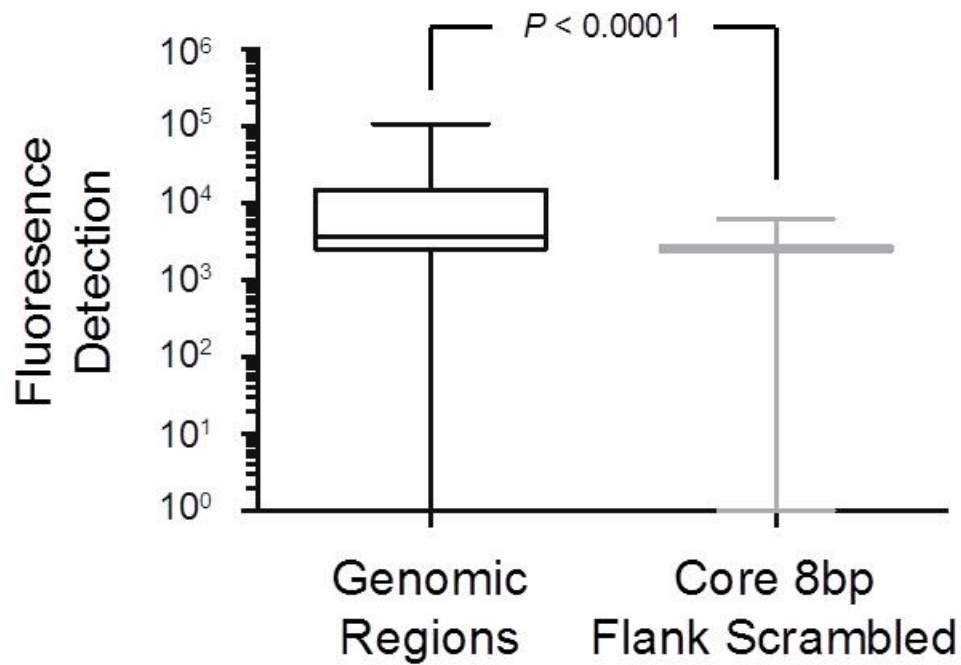


Figure 6B. GA-repeat stretches were compared between X-chromosome and autosomes. Out of the total observations of GA-repeats of particular lengths (8, 10, 12, 14, etc) the X-chromosome (red) had a higher number of GA repeats than autosomes (blue).



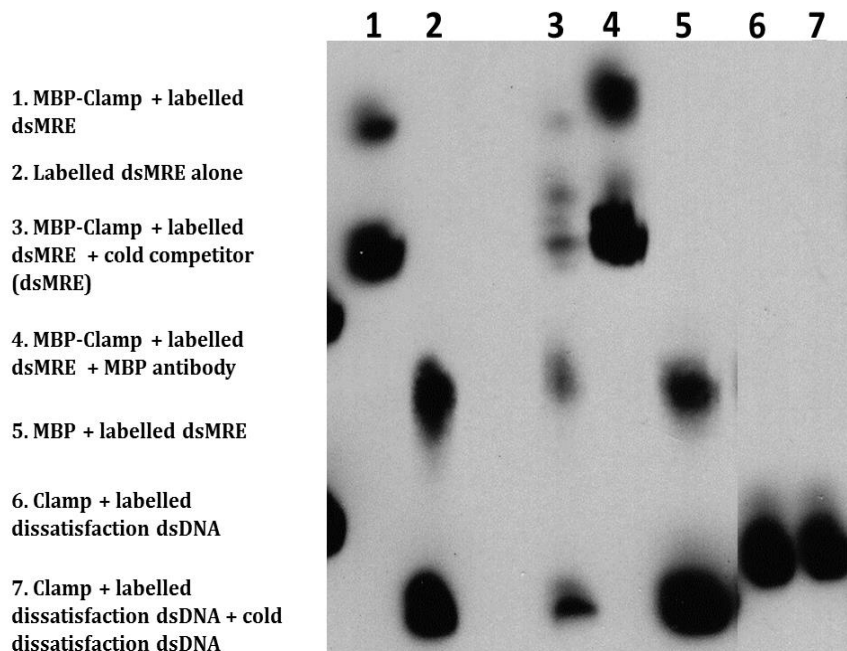
Supplemental Figures

Supplemental Figure 1. Sequences flanking the 8bp core of the CLAMP motif significantly contribute to CLAMP binding *in vitro* (PBM)



Supplemental Figure 1. Comparing the binding intensities of CLAMP protein on the PBM revealed that probes with genomic flanking DNA were more highly bound than probes with 8bp core intact and flanking DNA scrambled or artificially generated.

Supplemental Figure 2A. MBP-CLAMP binding is specific for CLAMP motif

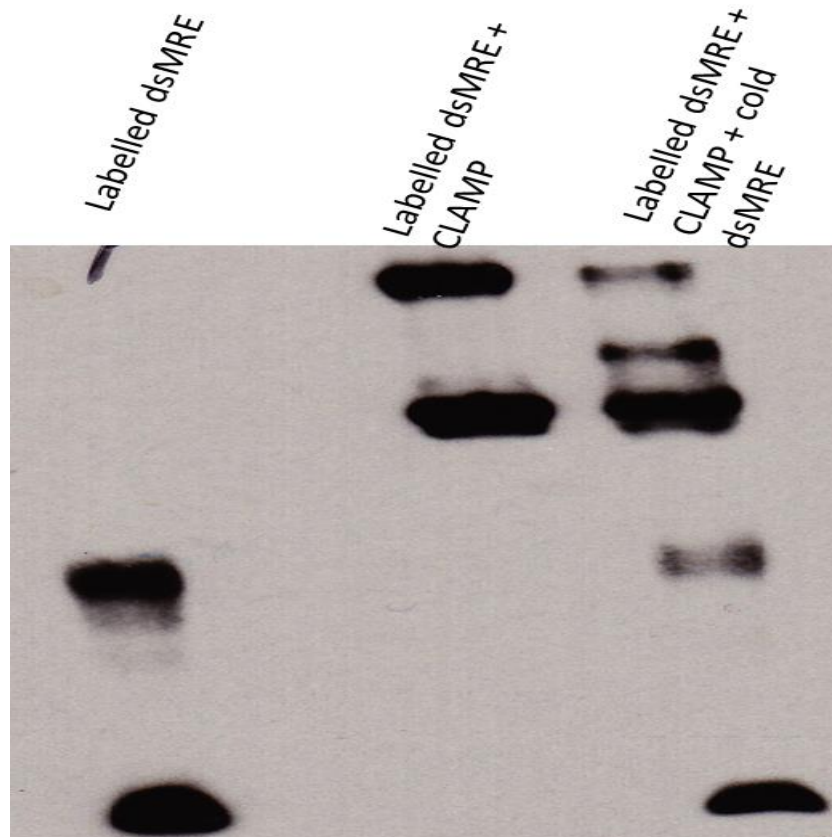


Dissatisfaction (dsf) motif: ACGATGTTGTCCTCAaagtcagcaagtcaGCGCCGTAGTCCCAA

dsMRE: 5' CGCAGACGCGGGTGCGAGCGAGACGGCCAGATGCGG 3'

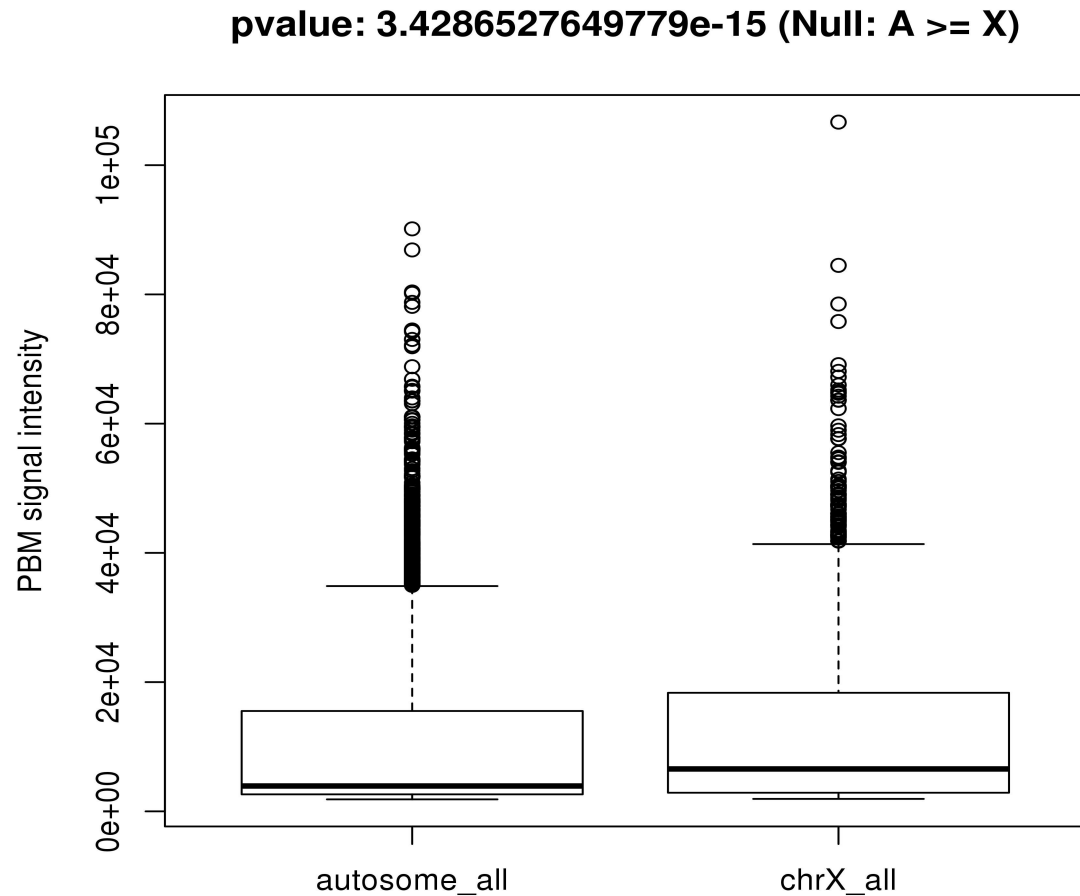
Supplemental Figure 2A. MBP-CLAMP 6 Zinc finger was expressed and purified from *E.coli*. Electrophoretic Mobility Shift Assays (EMSAs) were used to demonstrate the specificity of this purified MBP-CLAMP to a dsMRE probe from the Protein Binding Microarrays (PBMs). We interpret the top band in lane 2 (probe alone) as a secondary structure adopted by the probe. This is supported by the previously published findings that G-rich probes have a tendency to form secondary structures, referred to as G-quads (Huppert 2010).

Supplemental Figure 2B. MBP-CLAMP binding is specific for CLAMP motif



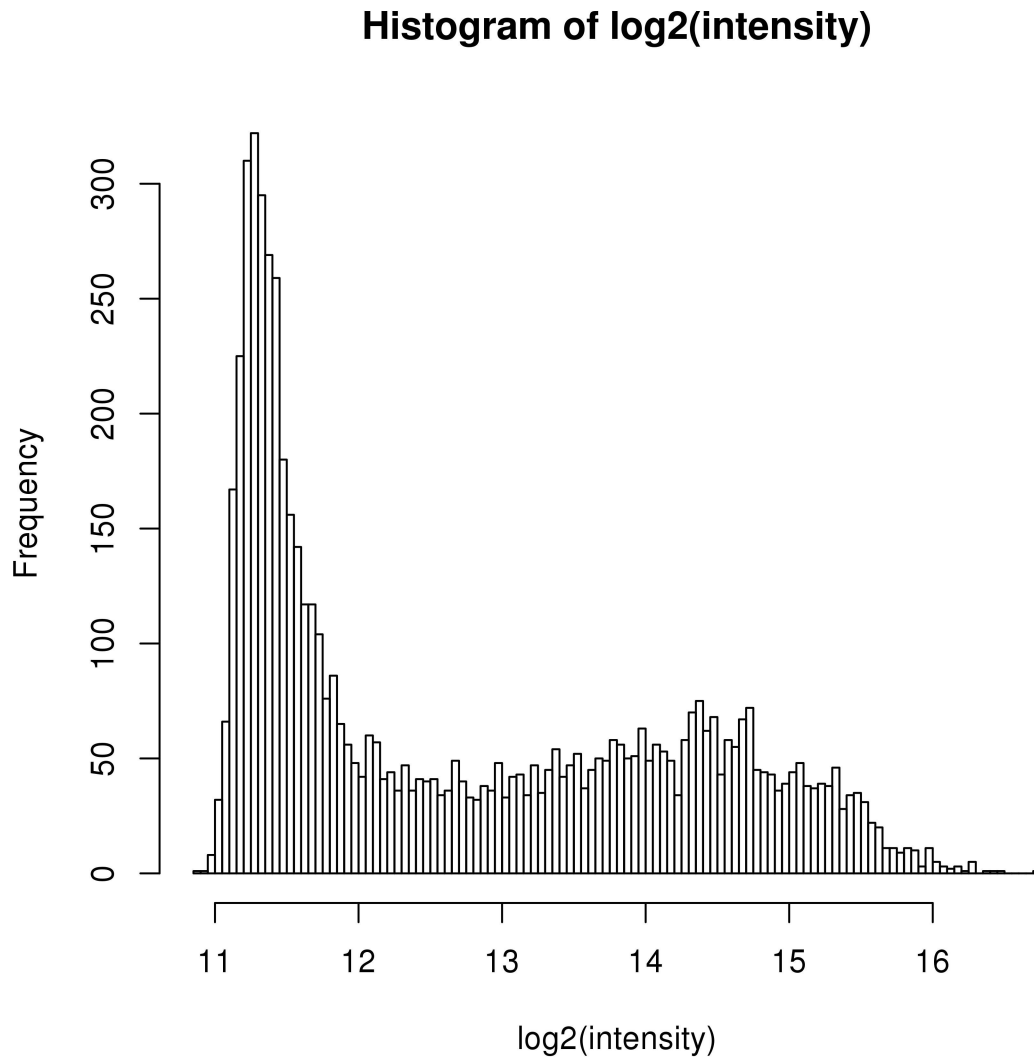
Supplemental Figures 2B. Electrophoretic Mobility Shift Assays (EMSAs) with purified MBP tagged CLAMP and a Biotin labeled dsMRE probe demonstrates purified CLAMP protein's binding specificity for a GA-rich MRE.

Supplemental Figure 3A. PBM probes from the X-chromosome have greater CLAMP binding than autosomal probes



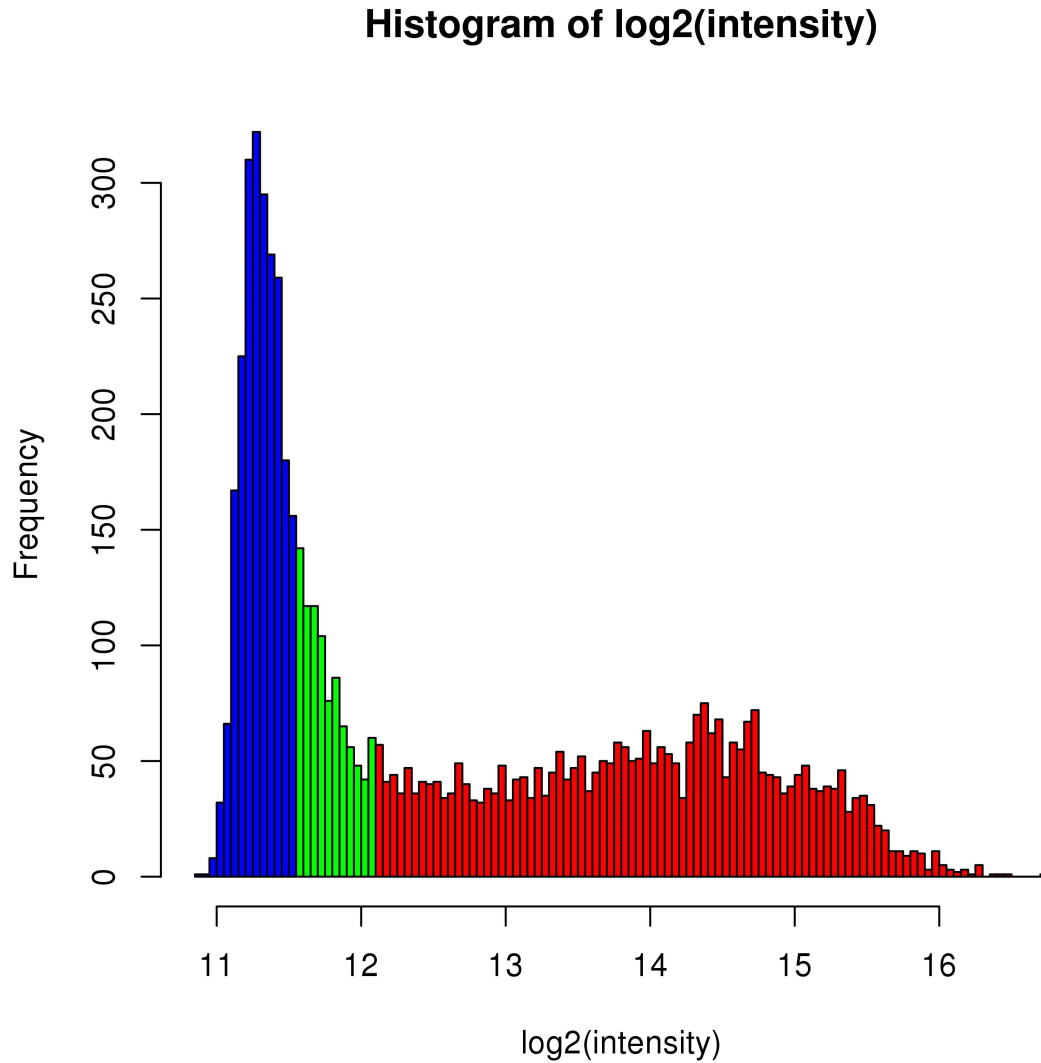
Supplemental Figure 3A. A Mann-Whitney test was used to test the null hypothesis that CLAMP binding to PBM probes with autosomal versus X-chromosome sequences displays no difference in binding intensity. The null hypothesis was rejected revealing a modest increase in CLAMP binding to X-chromosome versus autosome sequences.

Supplemental Figure 3B. Histogram indicates mixed Gaussian in probe signal intensity



Supplemental Figure 3B. The difference in binding intensity of probes from the X-chromosome versus autosome was further analyzed by plotting CLAMP binding signal to all probes (both X and A) as a histogram. This revealed that the probes are largely either strongly or weakly bound by CLAMP.

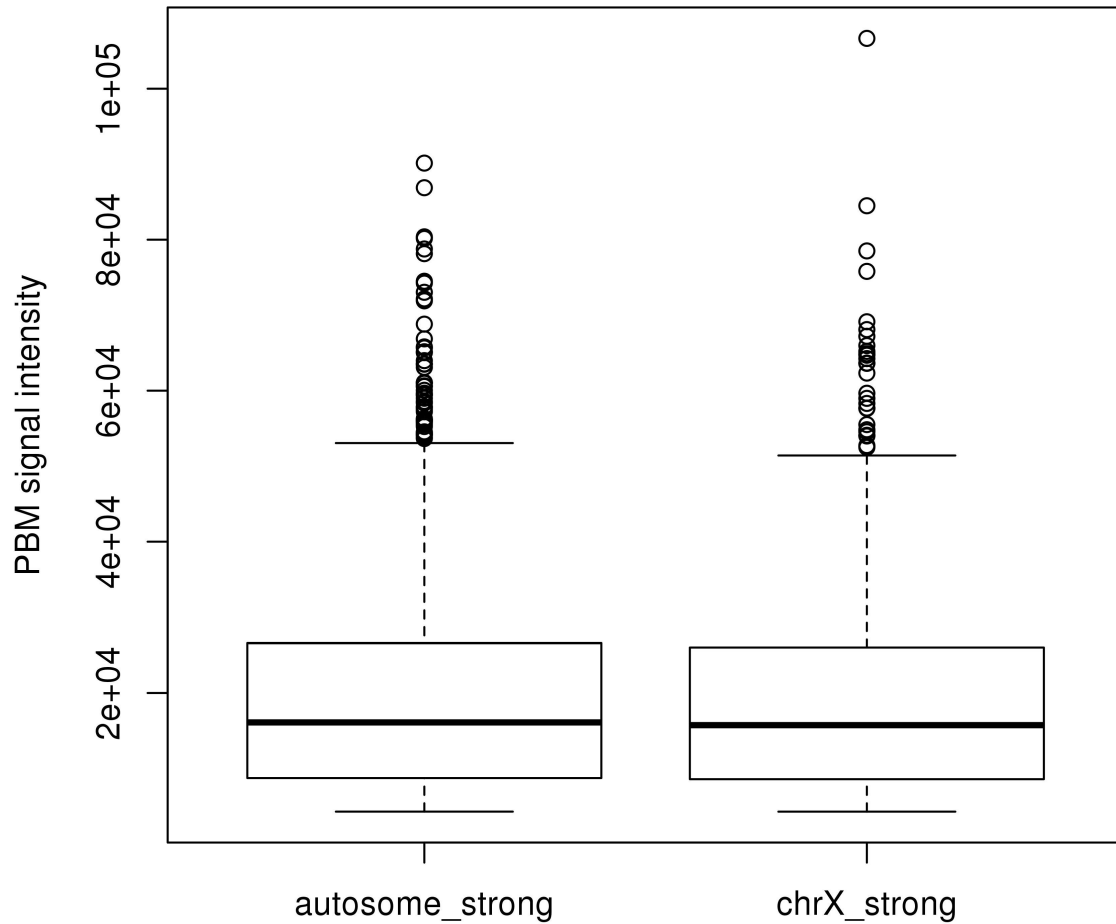
Supplemental Figure 3C. Classification of probes into weak (blue), medium (green), and strong (red) probes



Supplemental Figure 3C. Based on the two Gaussian distribution observed in the histogram of CLAMP binding signal intensity, probes were classified as weak, medium, and strong.

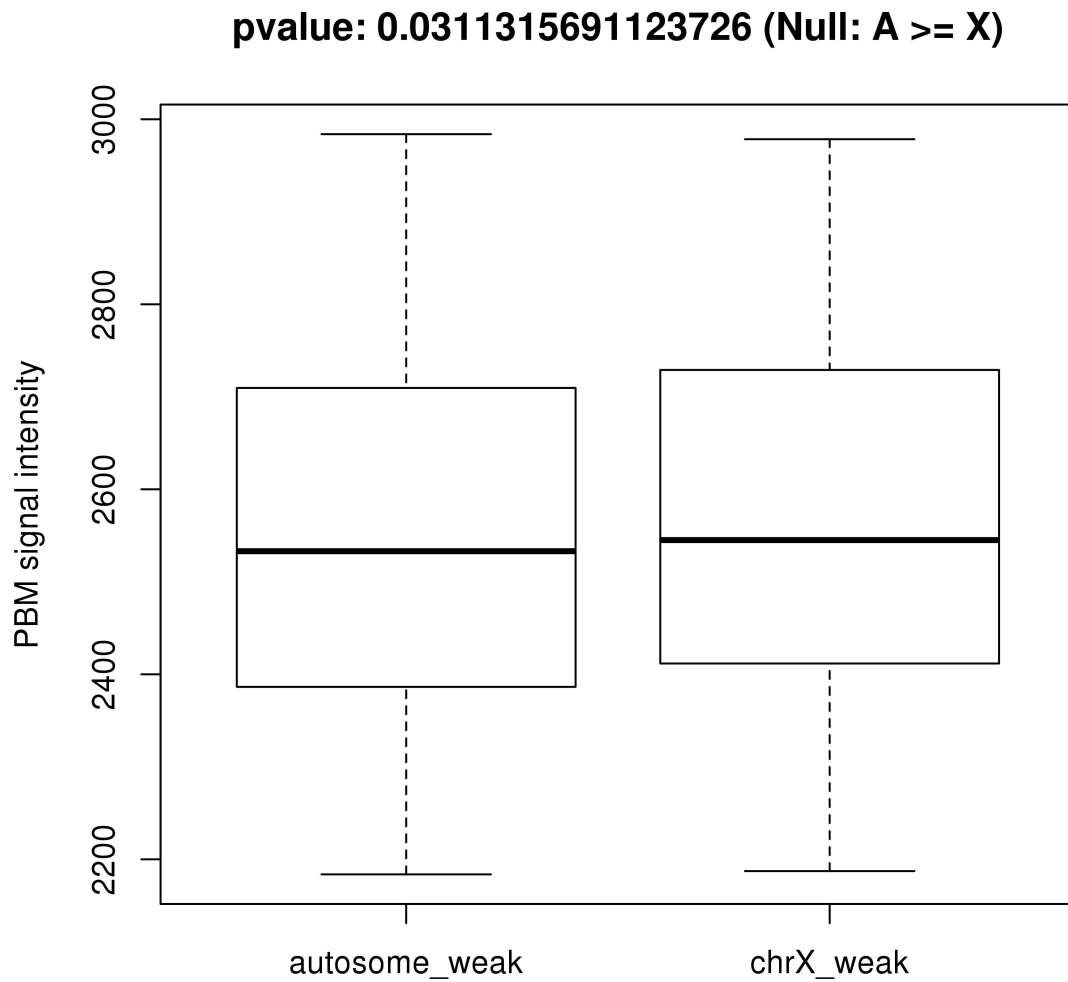
Supplemental Figure 3D. Comparison of A vs X for probes of strong signal

pvalue: 0.690800631647755 (Null: A $\sim$ X)



Supplemental Figure 3D. A Mann-Whitney test was used to compare the binding intensities of strong probes with sequences from autosomes versus X-chromosome. No significant preference was observed for strongly bound probes from X-chromosome.

Supplemental Figure 3E. Comparison of A vs X probes for weak signal intensity. Weak probes on X are more strongly bound



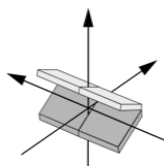
Supplemental Figure 3E. A Mann-Whitney test was used to determine if there was any difference in CLAMP binding due to weakly bound X versus autosome probes. Weakly bound X probes show modest increase in CLAMP binding.

Supplemental Figure 3F. Difference in shape features is more striking for weak probes instead of strong probes

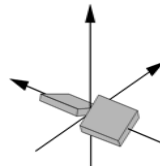
DNA shape features: Minor Groove Width (MGW), inter-bp Roll, inter-bp Helix twist (Helt), and intra-bp Propeller twist (ProT)



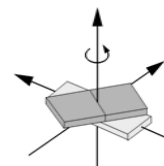
MGW



Roll

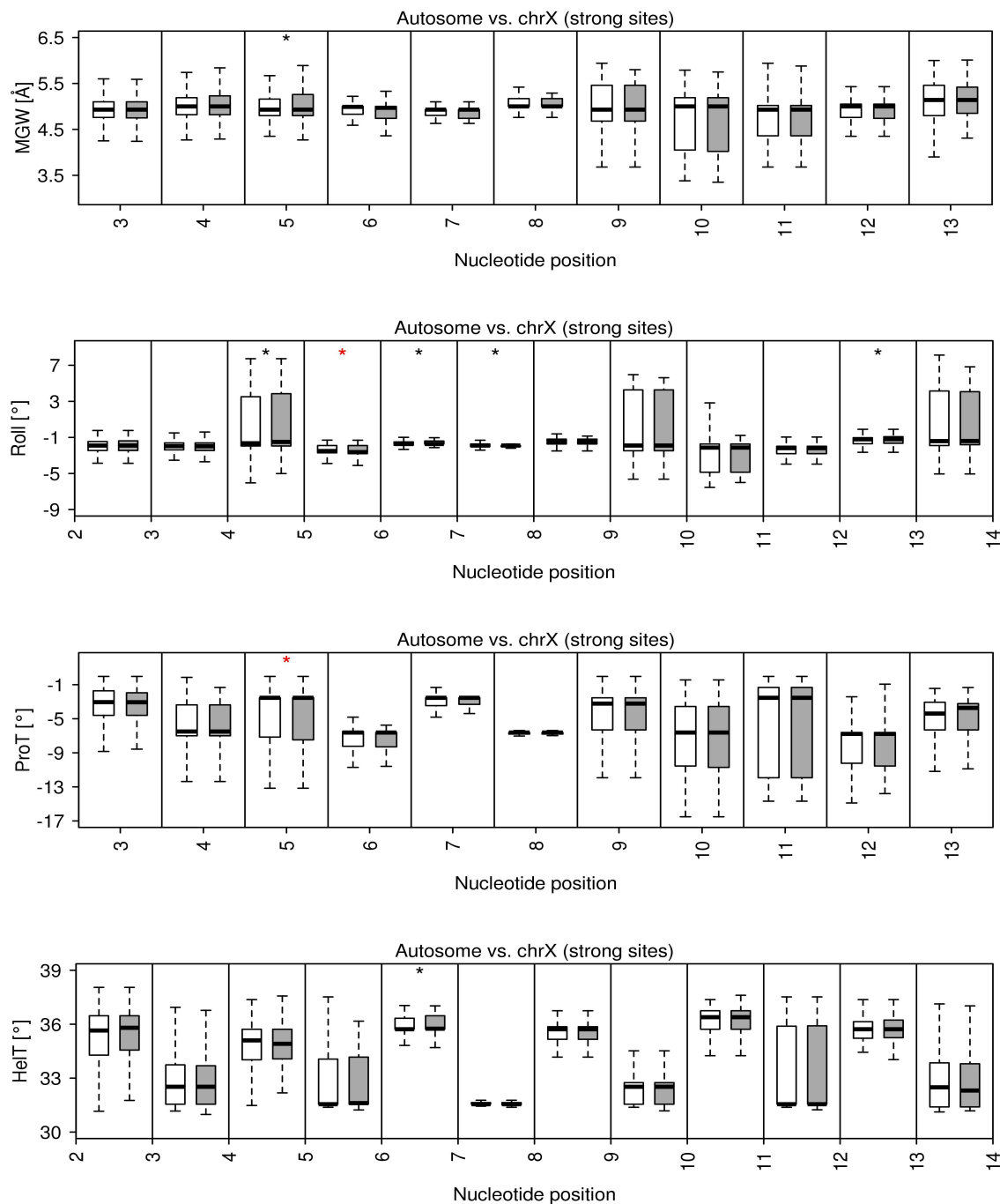


ProT



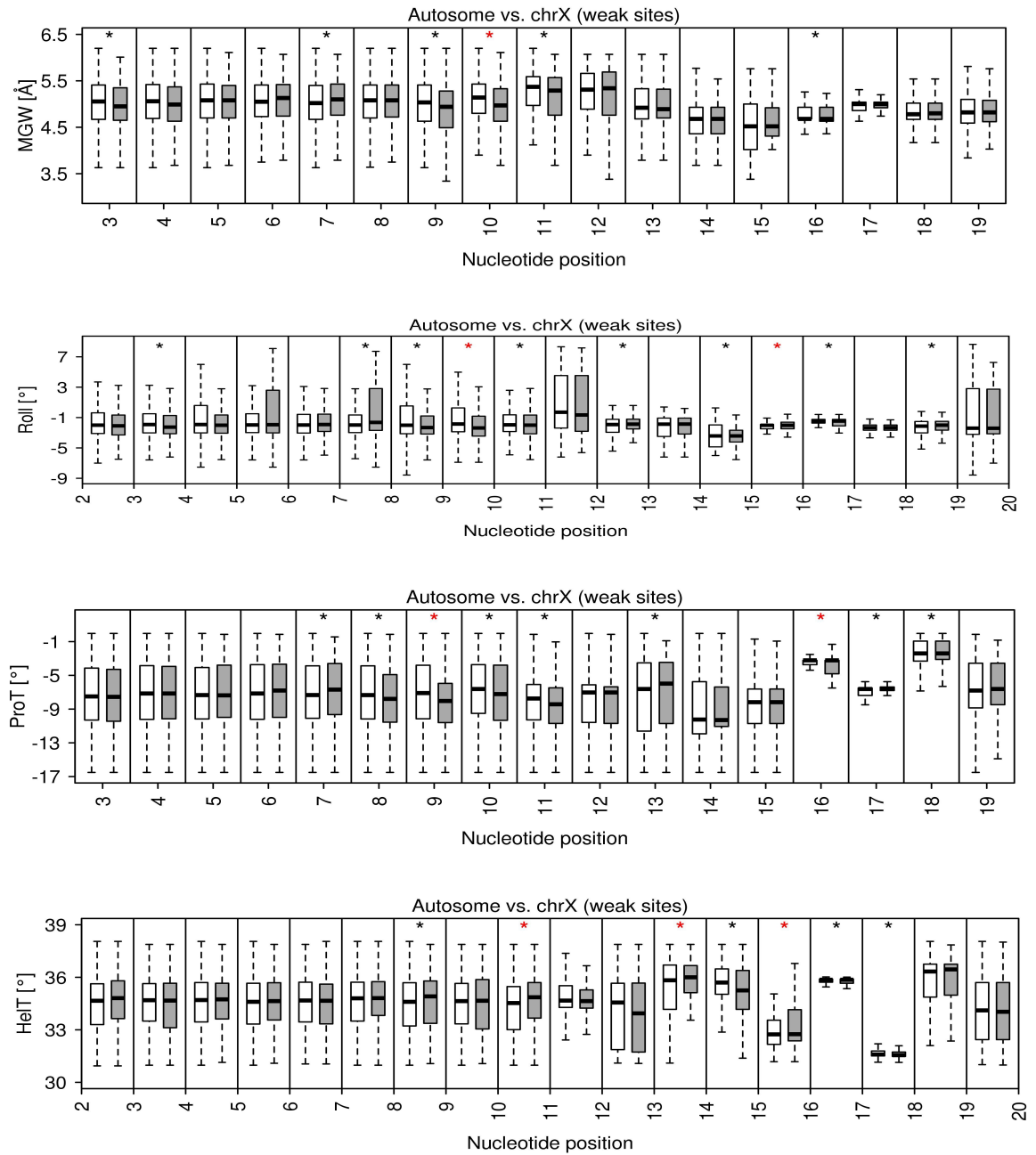
Helt

Supplemental Figure 3F. Difference in shape features is more striking for weak probes instead of strong probes



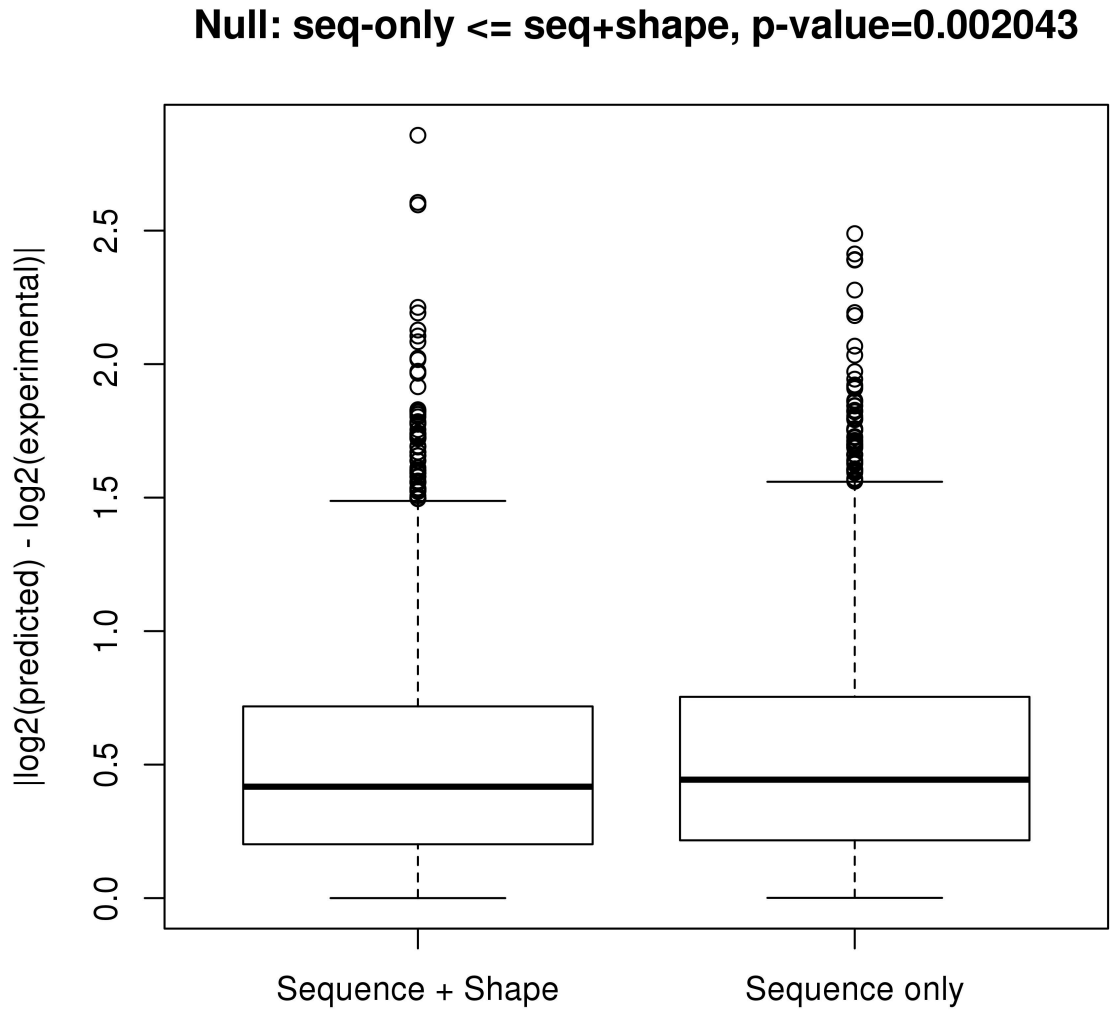
Supplemental Figure 3F. KS-test between A and X binding site DNA shape profiles in the strong group. Black stars indicate significant difference with p-value < 0.05 , and red stars indicate significance different with pvalue < 0.001 .

Supplemental Figure 3F. Difference in shape features is more striking for weak probes instead of strong probes



Supplemental Figure 3F. KS-test between A and X binding site DNA shape profiles in the weak group. Black stars indicate significant difference with p-value < 0.05, and red stars indicate significant difference with pvalue < 0.001.

Supplemental Figure 3G. Mann-Whitney test of sequence +shape versus sequence only model shows sequence only has greater effect on binding



Supplemental Figure 3G. $R^2 \sim 0.1$ between observed and predicted signal intensity for strong probes and decreased performance (0.296 to 0.274) for weak probes revealed that DNA shape was not a primary driver of CLAMP binding intensity.

References

- Alekseyenko AA, Ellison CE, Gorchakov AA, Zhou Q, Kaiser VB, Toda N, Walton Z, Peng S, Park PJ, Bachtrog D et al. 2013. Conservation and de novo acquisition of dosage compensation on newly evolved sex chromosomes in *Drosophila*. *Genes & development* **27**: 853-858.
- Alekseyenko AA, Ho JW, Peng S, Gelbart M, Tolstorukov MY, Plachetka A, Kharchenko PV, Jung YL, Gorchakov AA, Larschan E et al. 2012. Sequence-specific targeting of dosage compensation in *Drosophila* favors an active chromatin context. *PLoS genetics* **8**: e1002646.
- Alekseyenko AA, Larschan E, Lai WR, Park PJ, Kuroda MI. 2006. High-resolution ChIP-chip analysis reveals that the *Drosophila* MSL complex selectively identifies active genes on the male X chromosome. *Genes & development* **20**: 848-857.
- Alekseyenko AA, Peng S, Larschan E, Gorchakov AA, Lee OK, Kharchenko P, McGrath SD, Wang CI, Mardis ER, Park PJ et al. 2008. A sequence motif within chromatin entry sites directs MSL establishment on the *Drosophila* X chromosome. *Cell* **134**: 599-609.
- Bashaw GJ, Baker BS. 1997. The regulation of the *Drosophila* msl-2 gene reveals a function for Sex-lethal in translational control. *Cell* **89**: 789-798.
- Bell O, Conrad T, Kind J, Wirbelauer C, Akhtar A, Schubeler D. 2008. Transcription-coupled methylation of histone H3 at lysine 36 regulates dosage compensation by enhancing recruitment of the MSL complex in *Drosophila melanogaster*. *Molecular and cellular biology* **28**: 3401-3409.
- Berger MF, Bulyk ML. 2009. Universal protein-binding microarrays for the comprehensive characterization of the DNA-binding specificities of transcription factors. *Nature protocols* **4**: 393-411.
- Berger MF, Philippakis AA, Qureshi AM, He FS, Estep PW, 3rd, Bulyk ML. 2006. Compact, universal DNA microarrays to comprehensively determine transcription-factor binding site specificities. *Nat Biotechnol* **24**: 1429-1435.
- Deng X, Berletch JB, Ma W, Nguyen DK, Hiatt JB, Noble WS, Shendure J, Disteche CM. 2013. Mammalian X upregulation is associated with enhanced transcription initiation, RNA half-life, and MOF-mediated H4K16 acetylation. *Dev Cell* **25**: 55-68.
- Deng X, Hiatt JB, Nguyen DK, Ercan S, Sturgill D, Hillier LW, Schlesinger F, Davis CA, Reinke VJ, Gingeras TR et al. 2011. Evidence for compensatory upregulation of expressed X-linked genes in mammals, *Caenorhabditis elegans* and *Drosophila melanogaster*. *Nature genetics* **43**: 1179-1185.
- Deng X, Meller VH. 2006. roX RNAs are required for increased expression of X-linked genes in *Drosophila melanogaster* males. *Genetics* **174**: 1859-1866.
- Dickinson LA, Joh T, Kohwi Y, Kohwi-Shigematsu T. 1992. A tissue-specific MAR/SAR DNA-binding protein with unusual binding site recognition. *Cell* **70**: 631-645.
- Ellison CE, Bachtrog D. 2013. Dosage compensation via transposable element mediated rewiring of a regulatory network. *Science* **342**: 846-850.

- Gelbart ME, Larschan E, Peng S, Park PJ, Kuroda MI. 2009. Drosophila MSL complex globally acetylates H4K16 on the male X chromosome for dosage compensation. *Nature structural & molecular biology* **16**: 825-832.
- Gu W, Szauter P, Lucchesi JC. 1998. Targeting of MOF, a putative histone acetyl transferase, to the X chromosome of Drosophila melanogaster. *Developmental genetics* **22**: 56-64.
- Hilfiker A, Hilfiker-Kleiner D, Pannuti A, Lucchesi JC. 1997. mof, a putative acetyl transferase gene related to the Tip60 and MOZ human genes and to the SAS genes of yeast, is required for dosage compensation in Drosophila. *The EMBO journal* **16**: 2054-2060.
- Ilik IA, Quinn JJ, Georgiev P, Tavares-Cadete F, Maticzka D, Toscano S, Wan Y, Spitale RC, Luscombe N, Backofen R et al. 2013. Tandem stem-loops in roX RNAs act together to mediate X chromosome dosage compensation in Drosophila. *Molecular cell* **51**: 156-173.
- Kharchenko PV, Alekseyenko AA, Schwartz YB, Minoda A, Riddle NC, Ernst J, Sabo PJ, Larschan E, Gorchakov AA, Gu T et al. 2011. Comprehensive analysis of the chromatin landscape in Drosophila melanogaster. *Nature* **471**: 480-485.
- Larschan E, Bishop EP, Kharchenko PV, Core LJ, Lis JT, Park PJ, Kuroda MI. 2011. X chromosome dosage compensation via enhanced transcriptional elongation in Drosophila. *Nature* **471**: 115-118.
- Lours C, Bardot O, Godt D, Laski FA, Couderc JL. 2003. The Drosophila melanogaster BTB proteins bric a brac bind DNA through a composite DNA binding domain containing a pipsqueak and an AT-Hook motif. *Nucleic acids research* **31**: 5389-5398.
- Lucchesi JC, Kelly WG, Panning B. 2005. Chromatin remodeling in dosage compensation. *Annu Rev Genet* **39**: 615-651.
- Meller VH. 2003. Initiation of dosage compensation in Drosophila embryos depends on expression of the roX RNAs. *Mechanisms of development* **120**: 759-767.
- Paris M, Kaplan T, Li XY, Villalta JE, Lott SE, Eisen MB. 2013. Extensive divergence of transcription factor binding in Drosophila embryos with highly conserved gene expression. *PLoS genetics* **9**: e1003748.
- Shogren-Knaak M, Ishii H, Sun JM, Pazin MJ, Davie JR, Peterson CL. 2006. Histone H4-K16 acetylation controls chromatin structure and protein interactions. *Science* **311**: 844-847.
- Shogren-Knaak M, Peterson CL. 2006. Switching on chromatin: mechanistic role of histone H4-K16 acetylation. *Cell Cycle* **5**: 1361-1365.
- Soruco MM, Chery J, Bishop EP, Siggers T, Tolstorukov MY, Leydon AR, Sugden AU, Goebel K, Feng J, Xia P et al. 2013. The CLAMP protein links the MSL complex to the X chromosome during Drosophila dosage compensation. *Genes & development* **27**: 1551-1556.
- Thomas CJ, Kotova E, Andrade M, Adolf-Bryfogle J, Glaser R, Regnard C, Tulin AV. 2014. Kinase-Mediated Changes in Nucleosome Conformation Trigger Chromatin Decondensation via Poly(ADP-Ribosyl)ation. *Molecular cell* **53**: 831-842.

- Voss TC, Hager GL. 2014. Dynamic regulation of transcriptional states by chromatin and transcription factors. *Nature reviews Genetics* **15**: 69-81.
- Wang CI, Alekseyenko AA, LeRoy G, Elia AE, Gorchakov AA, Britton LM, Elledge SJ, Kharchenko PV, Garcia BA, Kuroda MI. 2013. Chromatin proteins captured by ChIP-mass spectrometry are linked to dosage compensation in *Drosophila*. *Nature structural & molecular biology* **20**: 202-209.
- Zhou T, Yang L, Lu Y, Dror I, Dantas Machado AC, Ghane T, Di Felice R, Rohs R. 2013. DNASHape: a method for the high-throughput prediction of DNA structural features on a genomic scale. *Nucleic acids research* **41**: W56-62.

Conclusions and Future Directions: Implications for the role of CLAMP in genomic targeting

Conclusions and Future Directions: Implications for the role of CLAMP in genomic targeting

Transcriptional regulation is a major area of interest especially in disease and therapeutics research. A number of transcription factors have been directly implicated in many diseases, such as cancer (James E. Darnell 2002). Zinc finger proteins form a large fraction of transcription factors and are widespread across nature. For example ~2% of proteins encoded in the human genome contain zinc fingers (Sunde 2002). This prevalence of zinc fingers can in part be explained by their functional versatility: zinc fingers can bind DNA, RNA, and protein, and dimerize with themselves (Sunde 2002). Zinc fingers can also be acetylated or phosphorylated, and can distinguish between methylated and unmethylated DNA sites to activate or repress gene expression (Shiro Iuchi 2005).

While there are varying types of zinc fingers such as CCHC, C₂C₂, and C₂CH that differ based on the secondary structures/folds they facilitate and the spatial arrangement of their cystidines and histidines, the most predominant form of zinc fingers is the C₂H₂ type (Sunde 2002; Shiro Iuchi 2005; Enuameh et al. 2013). C₂H₂ zinc fingers - the first zinc finger structures to be identified - are found in 3% of genes in the human genome and are involved in diverse cellular events related to development, differentiation, and tumor suppression (Shiro Iuchi 2005). When a census was generated to identify a full repertoire of transcription factors (TFs) in the human genome based on the ability to bind DNA sequence specifically, C₂H₂ was one of the

three types of TFs that dominated the human genome (Vaquerizas et al. 2009). This census determined that C₂H₂ zinc fingers, found in 675 transcription factors, composed one of the three types of TFs that accounted for 80% of the repertoire (Vaquerizas et al. 2009).

My thesis work has focused on characterizing the role of the transcription factor CLAMP, a C₂H₂ zinc finger protein, in targeting the MSL complex specifically to the X-chromosome for *Drosophila* dosage compensation. We are using the *Drosophila* dosage compensation model as a paradigm to understand the larger mystery of genomic targeting of transcription factors for spatial and temporal regulation of gene expression. Furthermore, it is highly likely that CLAMP has additional roles in gene expression beyond its function in dosage compensation due to its targeting to approximately 30% of active promoters.

CLAMP in *Drosophila* Dosage Compensation

In addition to targeting MSL complex to the X-chromosome by directly and sequence specifically binding to GA-rich sites on the X-chromosome, we speculate that CLAMP could be acetylated by MSL complex as a means of regulating CLAMP levels. MSL complex functions as a histone-acetyltransferase and zinc finger proteins can be acetylated. For example, the highly conserved C₂H₂ zinc finger transcription factor YY1 can be acetylated at its central region and C-terminal DNA binding domain (Lucchesi et al. 2005). YY1, which is required for wild-type development of

mammalian embryos, has a role in activating and repressing genes during cell growth and differentiation (Lucchesi et al. 2005). Acetylation of YY1's central domain was necessary for YY1's repressor function and also served as a target for deacetylation. However acetylation of the C-terminal DNA binding domain of YY1 decreased DNA binding and was irreversible, interacting instead with histone deacetylases (HDACs) to provide stable contact between YY1 and HDACs (Lucchesi et al. 2005).

Acetylation of CLAMP protein left over from maternal loading could serve as a means to control CLAMP's autoregulation. We have observed that CLAMP binds to its own promoter (data not shown), and we hypothesize this promoter binding could facilitate increased CLAMP expression. Acetylation of the CLAMP protein could interfere with CLAMP's DNA binding ability, thereby inhibiting CLAMP binding to its promoter to auto-regulate leading to increased CLAMP expression.

As the H4K16 acetylation mark deposited by MSL complex induces remodeling of the X-chromosome into a more open chromatin state, more CLAMP protein binds to MREs. This model correlates with our previously published findings of dramatically elevated CLAMP levels at high affinity sites for MSL complex that are highly acetylated at H4K16 (Soruco et al. 2013). In order to test these models, future experiments could use mass spectrometry to detect acetylated CLAMP in cell or tissue lysates in the presence and absence of the MOF histone acetyltransferase. Also, CLAMP binding to MREs embedded within *in vitro* assembled nucleosome

arrays that are acetylated or unacetylated could be used to further test these models of acetylation of CLAMP.

Is CLAMP an activator or repressor?

While our work has begun to elucidate the role of CLAMP in facilitating *Drosophila* male dosage compensation that occurs through hyper-transcription of the X-chromosome, we have yet to determine whether CLAMP functions as an activator or repressor. Given that CLAMP is expressed in both male and female *Drosophila*, we hypothesize that there are multiple functions of CLAMP that can be mediated through interactions with different protein complexes. For example, we found that CLAMP is able to identify its binding sites within the heterochromatin that is interspersed with euchromatic regions (data not shown). Upon first glance this would appear to contradict its role in associating with the activating complex MSL. MSL increases transcription of genes on the X-chromosome in males by releasing paused polymerase into active elongation (Larschan et al. 2011b). Heterochromatin involves packed chromatin, which is repressive to gene expression. However, CLAMP was co-purified with the *Drosophila* Myb-E2F2/RBF transcriptional repressor complex (Keene et al. 2006). Myb-MuvB complex was found at promoters repressing transcription of developmentally regulated genes. We have also observed that CLAMP appears to repress developmentally regulated genes (Sorucu et al, unpublished data).

In addition, the Myb-MuvB complex associates with members of the NURF (Nucleosome Remodeling Factor) complex (Keene et al. 2006). NURF facilitates ATP-dependent nucleosome sliding, thereby disrupting nucleosome periodicity and allowing transcription activation via chromatin remodeling (Gu et al. 1998a). NURF has been demonstrated to repress *roX* transcription in *Drosophila* females (Hilfiker et al. 1997). We have also observed that CLAMP appears to repress *roX1* (Soruco et al., unpublished data) indicating a possible role for CLAMP-mediated NURF repression via altering nucleosome occupancy.

In addition, the zinc finger protein GAF, similarly to CLAMP, binds to GAG sequences across the genome. In fact, CLAMP and GAF appear to colocalize at a large number of GAGA sequences (data not shown). GAF acts in concert with NURF to create nucleosome free regions and counteracts repressors to facilitate RNA Pol II transcription (Larschan et al. 2006; Alekseyenko et al. 2013; Ellison and Bachtrog 2013). NURF plays a key role in GAF factor transcriptional activation; GAF and NURF are both involved in recruiting RNA Pol II (Larschan et al. 2006; Alekseyenko et al. 2013; Ellison and Bachtrog 2013). Therefore, it is possible that CLAMP functions similarly to GAF by recruiting NURF for transcriptional activation. This is supported by the finding that NURF mutants lead to puffing of the X-chromosome configuration, and wildtype NURF is required for normal X-chromosome morphology in males (Hilfiker et al. 1997).

Given that components of the NURF complex associate with the Myb-MuvB repressor complex and CLAMP interacts with the Myb-MuvB repressor complex, it is possible CLAMP's interaction with NURF or Myb-MuvB complex components allows CLAMP to function either as an activator or repressor. This line of study will be further pursued by others using CLAMP mass-spectrometry to identify CLAMP's protein partners.

The histone deacetylase Rpd3 was also found to interact with the Myb-MuvB repressor complex which CLAMP co-purified with. Given CLAMP's association with a histone acetyltransferase complex (MSL) and a repressor complex containing a histone deacetylase (Rpd3), it is possible that CLAMP functions as a targeting factor for different chromatin remodeling complexes. CLAMP targeting of these factors could also in turn be regulated by acetylation and deacetylation of CLAMP.

The ability of CLAMP to interact both with an activating complex (MSL) and a repressor (Myb-E2F2/RBF) indicates the multidimensional interaction potential of this new zinc finger protein (Keene et al. 2006; Wang et al. 2013a). In other words, CLAMP may function both in an activating or repressive context depending on the group of proteins it associates with. Different complexes may co-opt CLAMP because of its ability to sequence specifically bind to DNA thereby allowing precise genomic targeting facilitating transcriptional regulation.

Future Directions

Identification of CLAMP interacting partners will provide key insight into how domains of coordinate gene regulation are established by precisely targeting transcription factors. It is possible that these interactions will define a new paradigm for how transcriptional regulators establish sub-nuclear domains across species and could potentially provide new therapeutic targets for diverse diseases including cancers and neurological disorders. For example, zinc fingers are a promising new gene therapy approach for HIV treatment and potentially Huntington's Disease. In a mouse model for Huntington's Disease, a synthetic zinc finger repressor was used to reduce expression of the mutant form of the huntington gene that is primarily causative of this neurodegenerative disorder (Garriga-Canut et al. 2012). A zinc finger nuclease was used to eliminate CCR5 expression in mice to generate an HIV resistant immune system (Hofer et al. 2013). Both of these advancements demonstrate the therapeutic power of zinc finger proteins to manipulate transcriptional regulation in disease. By identifying how the zinc finger protein CLAMP tandemly modulates three different forms of interaction with DNA, RNA, and proteins for coordinate gene expression, we could provide a powerful new multi-faceted tool for genome editing.

References

- Adelman K, Wei W, Ardehali MB, Werner J, Zhu B, Reinberg D, Lis JT. 2006. Drosophila Paf1 modulates chromatin structure at actively transcribed genes. *Mol Cell Biol* **26**: 250-260.
- Adkins NL, Hagerman TA, Georgel P. 2006. GAGA protein: a multi-faceted transcription factor. *Biochem Cell Biol* **84**: 559-567.
- Akhtar A, Becker PB. 2000. Activation of transcription through histone H4 acetylation by MOF, an acetyltransferase essential for dosage compensation in Drosophila. *Mol Cell* **5**: 367-375.
- Akhtar A, Zink D, Becker PB. 2000. Chromodomains are protein-RNA interaction modules. *Nature* **407**: 405-409.
- Alekseyenko AA, Ellison CE, Gorchakov AA, Zhou Q, Kaiser VB, Toda N, Walton Z, Peng S, Park PJ, Bachtrog D et al. 2013. Conservation and de novo acquisition of dosage compensation on newly evolved sex chromosomes in Drosophila. *Genes & development* **27**: 853-858.
- Alekseyenko AA HJ, Peng S, Gelbart M, Tolstorukov M, Plachetka A, Kharchenko PV, Jung YL, Gorchakov AA, Larschan E, Gu T, Minoda A, Riddle NC, Schwartz YB, Elgin SCR, Karpen GH, Pirrotta V, Kuroda MI**, Park PJ**. . 2012. Sequence-specific targeting of dosage compensation in Drosophila favors an active chromatin context. *In press PLoS Genetics*.
- Alekseyenko AA, Ho JW, Peng S, Gelbart M, Tolstorukov MY, Plachetka A, Kharchenko PV, Jung YL, Gorchakov AA, Larschan E et al. 2012. Sequence-specific targeting of dosage compensation in Drosophila favors an active chromatin context. *PLoS genetics* **8**: e1002646.
- Alekseyenko AA, Larschan E, Lai WR, Park PJ, Kuroda MI. 2006. High-resolution ChIP-chip analysis reveals that the Drosophila MSL complex selectively identifies active genes on the male X chromosome. *Genes & development* **20**: 848-857.
- Alekseyenko AA, Peng S, Larschan E, Gorchakov AA, Lee OK, Kharchenko P, McGrath SD, Wang CI, Mardis ER, Park PJ et al. 2008. A sequence motif within chromatin entry sites directs MSL establishment on the Drosophila X chromosome. *Cell* **134**: 599-609.
- Amrein H, Axel R. 1997. Genes expressed in neurons of adult male Drosophila. *Cell* **88**: 459-469.
- Atanesyan L, Gunther V, Dichtl B, Georgiev O, Schaffner W. 2012. Polyglutamine tracts as modulators of transcriptional activation from yeast to mammals. *Biological chemistry* **393**: 63-70.
- Bai X, Alekseyenko AA, Kuroda MI. 2004. Sequence-specific targeting of MSL complex regulates transcription of the roX RNA genes. *The EMBO journal* **23**: 2853-2861.
- Bai X, Larschan E, Kwon SY, Badenhorst P, Kuroda MI. 2007. Regional control of chromatin organization by noncoding roX RNAs and the NURF remodeling complex in Drosophila melanogaster. *Genetics* **176**: 1491-1499.

- Bailey TL, Boden M, Buske FA, Frith M, Grant CE, Clementi L, Ren J, Li WW, Noble WS. 2009. MEME SUITE: tools for motif discovery and searching. *Nucleic Acids Res* **37**: W202-208.
- Bashaw GJ, Baker BS. 1997. The regulation of the *Drosophila* msl-2 gene reveals a function for Sex-lethal in translational control. *Cell* **89**: 789-798.
- Batista PJ, Chang HY. 2013a. Cytotopic localization by long noncoding RNAs. *Current opinion in cell biology* **25**: 195-199.
- . 2013b. Long Noncoding RNAs: Cellular Address Codes in Development and Disease. *Cell* **152**: 1298-1307.
- Beatriz Vicoso DB. 2009. Progress and prospects toward our understanding of the evolution of dosage compensation.
- Bell O, Conrad T, Kind J, Wirbelauer C, Akhtar A, Schubeler D. 2008. Transcription-coupled methylation of histone H3 at lysine 36 regulates dosage compensation by enhancing recruitment of the MSL complex in *Drosophila melanogaster*. *Molecular and cellular biology* **28**: 3401-3409.
- Belote JM, Lucchesi JC. 1980a. Control of X chromosome transcription by the maleless gene in *Drosophila*. *Nature* **285**: 573-575.
- . 1980b. Male-specific lethal mutations of *Drosophila melanogaster*. *Genetics* **96**: 165-186.
- Berger MF, Bulyk ML. 2009. Universal protein-binding microarrays for the comprehensive characterization of the DNA-binding specificities of transcription factors. *Nature protocols* **4**: 393-411.
- Berger MF, Philippakis AA, Qureshi AM, He FS, Estep PW, 3rd, Bulyk ML. 2006. Compact, universal DNA microarrays to comprehensively determine transcription-factor binding site specificities. *Nat Biotechnol* **24**: 1429-1435.
- Bryson TD, Weber CM, Henikoff S. 2010. Baculovirus-encoded protein expression for epigenomic profiling in *Drosophila* cells. *Fly (Austin)* **4**: 258-265.
- Buscaino A, Legube G, Akhtar A. 2006. X-chromosome targeting and dosage compensation are mediated by distinct domains in MSL-3. *EMBO reports* **7**: 531-538.
- Carrozza MJ, Li B, Florens L, Suganuma T, Swanson SK, Lee KK, Shia WJ, Anderson S, Yates J, Washburn MP et al. 2005. Histone H3 methylation by Set2 directs deacetylation of coding regions by Rpd3S to suppress spurious intragenic transcription. *Cell* **123**: 581-592.
- Chow JC, Ciaudo C, Fazzari MJ, Mise N, Servant N, Glass JL, Attreed M, Avner P, Wutz A, Barillot E et al. 2010. LINE-1 activity in facultative heterochromatin formation during X chromosome inactivation. *Cell* **141**: 956-969.
- Chu C, Qu K, Zhong FL, Artandi SE, Chang HY. 2011. Genomic maps of long noncoding RNA occupancy reveal principles of RNA-chromatin interactions. *Molecular cell* **44**: 667-678.
- Chuang PT, Albertson DG, Meyer BJ. 1994. DPY-27: a chromosome condensation protein homolog that regulates *C. elegans* dosage compensation through association with the X chromosome. *Cell* **79**: 459-474.

- Colgan J, Ashali H, Manley JL. 1995. A direct interaction between a glutamine-rich activator and the N terminus of TFIIB can mediate transcriptional activation in vivo. *Mol Cell Biol* **15**: 2311-2320.
- Conrad T, Cavalli FM, Holz H, Hallacli E, Kind J, Ilik I, Vaquerizas JM, Luscombe NM, Akhtar A. 2012. The MOF chromobarrel domain controls genome-wide H4K16 acetylation and spreading of the MSL complex. *Dev Cell* **22**: 610-624.
- Copps K, Richman R, Lyman LM, Chang KA, Rampersad-Ammons J, Kuroda MI. 1998. Complex formation by the Drosophila MSL proteins: role of the MSL2 RING finger in protein complex assembly. *Embo J* **17**: 5409-5417.
- Csankovszki G, McDonel P, Meyer BJ. 2004. Recruitment and spreading of the C. elegans dosage compensation complex along X chromosomes. *Science* **303**: 1182-1185.
- Dahlsveen IK, Gilfillan GD, Shelest VI, Lamm R, Becker PB. 2006. Targeting determinants of dosage compensation in Drosophila. *PLoS genetics* **2**: e5.
- Davis TL, Meyer BJ. 1997. SDC-3 coordinates the assembly of a dosage compensation complex on the nematode X chromosome. *Development* **124**: 1019-1031.
- Deng X, Berletch JB, Ma W, Nguyen DK, Hiatt JB, Noble WS, Shendure J, Disteche CM. 2013. Mammalian X upregulation is associated with enhanced transcription initiation, RNA half-life, and MOF-mediated H4K16 acetylation. *Dev Cell* **25**: 55-68.
- Deng X BJ, Ma W, Nguyen DK, Hiatt JB, Noble WS, Shendure J, Disteche CM. 2013. Mammalian X Upregulation Is Associated with Enhanced Transcription Initiation, RNA Half-Life, and MOF-Mediated H4K16 Acetylation. *Developmental Cell* **25**: 1-14.
- Deng X, Hiatt JB, Nguyen DK, Ercan S, Sturgill D, Hillier LW, Schlesinger F, Davis CA, Reinke VJ, Gingeras TR et al. 2011. Evidence for compensatory upregulation of expressed X-linked genes in mammals, Caenorhabditis elegans and Drosophila melanogaster. *Nature genetics* **43**: 1179-1185.
- Deng X, Meller VH. 2006. roX RNAs are required for increased expression of X-linked genes in Drosophila melanogaster males. *Genetics* **174**: 1859-1866.
- Derrien T, Johnson R, Bussotti G, Tanzer A, Djebali S, Tilgner H, Guernec G, Martin D, Merkel A, Knowles DG et al. 2012. The GENCODE v7 catalog of human long noncoding RNAs: analysis of their gene structure, evolution, and expression. *Genome research* **22**: 1775-1789.
- Deuring R, Fanti L, Armstrong JA, Sarte M, Papoulas O, Prestel M, Daubresse G, Verardo M, Moseley SL, Berloco M et al. 2000. The ISWI chromatin-remodeling protein is required for gene expression and the maintenance of higher order chromatin structure in vivo. *Mol Cell* **5**: 355-365.
- Dickinson LA, Joh T, Kohwi Y, Kohwi-Shigematsu T. 1992. A tissue-specific MAR/SAR DNA-binding protein with unusual binding site recognition. *Cell* **70**: 631-645.
- Disteche CM. 2012. Dosage compensation of the sex chromosomes. *Annu Rev Genet* **46**: 537-560.
- Dundr M, Misteli T. 2010. Biogenesis of nuclear bodies. *Cold Spring Harbor perspectives in biology* **2**: a000711.

- Dunlap D, Yokoyama R, Ling H, Sun HY, McGill K, Cugusi S, Lucchesi JC. 2012. Distinct contributions of MSL complex subunits to the transcriptional enhancement responsible for dosage compensation in *Drosophila*. *Nucleic acids research* **40**: 11281-11291.
- Ellison CE, Bachtrog D. 2013. Dosage compensation via transposable element mediated rewiring of a regulatory network. *Science* **342**: 846-850.
- Engreitz JM, Pandya-Jones A, McDonel P, Shishkin A, Sirokman K, Surka C, Kadri S, Xing J, Goren A, Lander ES et al. 2013. The Xist lncRNA exploits three-dimensional genome architecture to spread across the X chromosome. *Science* **341**: 1237973.
- Enuameh MS, Asriyan Y, Richards A, Christensen RG, Hall VL, Kazemian M, Zhu C, Pham H, Cheng Q, Blatti C et al. 2013. Global analysis of *Drosophila* Cys(2)-His(2) zinc finger proteins reveals a multitude of novel recognition motifs and binding determinants. *Genome research* **23**: 928-940.
- Fauth T, Muller-Planitz F, Konig C, Straub T, Becker PB. 2010. The DNA binding CXC domain of MSL2 is required for faithful targeting the Dosage Compensation Complex to the X chromosome. *Nucleic acids research* **38**: 3209-3221.
- Fisher WW, Li JJ, Hammonds AS, Brown JB, Pfeiffer BD, Weiszmann R, MacArthur S, Thomas S, Stamatoyannopoulos JA, Eisen MB et al. 2012. DNA regions bound at low occupancy by transcription factors do not drive patterned reporter gene expression in *Drosophila*. *Proceedings of the National Academy of Sciences of the United States of America* **109**: 21330-21335.
- Furuhashi H, Nakajima M, Hirose S. 2006. DNA supercoiling factor contributes to dosage compensation in *Drosophila*. *Development* **133**: 4475-4483.
- Gallach M, Arnau V, Aldecoa R, Marin I. 2010. A sequence motif enriched in regions bound by the *Drosophila* dosage compensation complex. *BMC genomics* **11**: 169.
- Gamsjaeger R, Liew CK, Loughlin FE, Crossley M, Mackay JP. 2007. Sticky fingers: zinc-fingers as protein-recognition motifs. *Trends in biochemical sciences* **32**: 63-70.
- Garriga-Canut M, Agustin-Pavon C, Herrmann F, Sanchez A, Dierssen M, Fillat C, Isalan M. 2012. Synthetic zinc finger repressors reduce mutant huntingtin expression in the brain of R6/2 mice. *Proceedings of the National Academy of Sciences of the United States of America* **109**: E3136-3145.
- Gelbart ME, Kuroda MI. 2009. *Drosophila* dosage compensation: a complex voyage to the X chromosome. *Development* **136**: 1399-1410.
- Gelbart ME, Larschan E, Peng S, Park PJ, Kuroda MI. 2009. *Drosophila* MSL complex globally acetylates H4K16 on the male X chromosome for dosage compensation. *Nature structural & molecular biology* **16**: 825-832.
- Gilfillan GD, Straub T, de Wit E, Greil F, Lamm R, van Steensel B, Becker PB. 2006. Chromosome-wide gene-specific targeting of the *Drosophila* dosage compensation complex. *Genes Dev* **20**: 858-870.
- Girirajan S, Campbell CD, Eichler EE. 2011. Human copy number variation and complex genetic disease. *Annu Rev Genet* **45**: 203-226.

- Gontan C, Achame EM, Demmers J, Barakat TS, Rentmeester E, van IW, Grootegeod JA, Gribnau J. 2012. RNF12 initiates X-chromosome inactivation by targeting REX1 for degradation. *Nature* **485**: 386-390.
- Gu W, Szauter P, Lucchesi JC. 1998a. Targeting of MOF, a putative histone acetyl transferase, to the X chromosome of *Drosophila melanogaster*. *Developmental genetics* **22**: 56-64.
- Gu WG, Szauter P, Lucchesi JC. 1998b. Targeting of MOF, a putative histone acetyl transferase, to the X chromosome of *Drosophila melanogaster*. *Dev Genet* **22**: 56-64.
- Guertin MJ, Lis JT. 2012. Mechanisms by which transcription factors gain access to target sequence elements in chromatin. *Current opinion in genetics & development*.
- Hallacli E, Lipp M, Georgiev P, Spielman C, Cusack S, Akhtar A, Kadlec J. 2012. Msl1-mediated dimerization of the dosage compensation complex is essential for male X-chromosome regulation in *Drosophila*. *Molecular cell* **48**: 587-600.
- Hamada FN, Park PJ, Gordadze PR, Kuroda MI. 2005. Global regulation of X chromosomal genes by the MSL complex in *Drosophila melanogaster*. *Genes Dev* **19**: 2289-2294.
- Hattori M, Fujiyama A, Taylor TD, Watanabe H, Yada T, Park HS, Toyoda A, Ishii K, Totoki Y, Choi DK et al. 2000. The DNA sequence of human chromosome 21. *Nature* **405**: 311-319.
- Heard E, Turner J. 2011. Function of the sex chromosomes in mammalian fertility. *Cold Spring Harbor perspectives in biology* **3**: a002675.
- Henikoff S, Henikoff JG, Sakai A, Loeb GB, Ahmad K. 2009. Genome-wide profiling of salt fractions maps physical properties of chromatin. *Genome Res* **19**: 460-469.
- Henry KW, Wyce A, Lo WS, Duggan LJ, Emre NC, Kao CF, Pillus L, Shilatifard A, Osley MA, Berger SL. 2003. Transcriptional activation via sequential histone H2B ubiquitylation and deubiquitylation, mediated by SAGA-associated Ubp8. *Genes Dev* **17**: 2648-2663.
- Herpin A, Scharf M. 2011. Sex determination: switch and suppress. *Current biology : CB* **21**: R656-659.
- Hilfiker A, Hilfiker-Kleiner D, Pannuti A, Lucchesi JC. 1997. mof, a putative acetyl transferase gene related to the Tip60 and MOZ human genes and to the SAS genes of yeast, is required for dosage compensation in *Drosophila*. *The EMBO journal* **16**: 2054-2060.
- Hofer U, Henley JE, Exline CM, Mulhern O, Lopez E, Cannon PM. 2013. Pre-clinical modeling of CCR5 knockout in human hematopoietic stem cells by zinc finger nucleases using humanized mice. *The Journal of infectious diseases* **208 Suppl 2**: S160-164.
- Hsu DR, Meyer BJ. 1994. The dpy-30 gene encodes an essential component of the *Caenorhabditis elegans* dosage compensation machinery. *Genetics* **137**: 999-1018.
- Huppert JL. 2010. Structure, location and interactions of G-quadruplexes. *The FEBS journal* **277**: 3452-3458.

- Ilik IA, Quinn JJ, Georgiev P, Tavares-Cadete F, Maticzka D, Toscano S, Wan Y, Spitale RC, Luscombe N, Backofen R et al. 2013. Tandem stem-loops in roX RNAs act together to mediate X chromosome dosage compensation in *Drosophila*. *Molecular cell* **51**: 156-173.
- Jacobs PA, Baikie AG, Court Brown WM, Strong JA. 1959. The somatic chromosomes in mongolism. *Lancet* **1**: 710.
- James E, Darnell J. 2002. TRANSCRIPTION FACTORS AS TARGETS FOR CANCER THERAPY. *Nature Reviews Cancer* **2**.
- Jans J, Gladden JM, Ralston EJ, Pickle CS, Michel AH, Pferdehirt RR, Eisen MB, Meyer BJ. 2009. A condensin-like dosage compensation complex acts at a distance to control expression throughout the genome. *Genes & development* **23**: 602-618.
- Jeon Y, Lee JT. 2011. YY1 tethers Xist RNA to the inactive X nucleation center. *Cell* **146**: 119-133.
- Jin Y, Wang Y, Johansen J, Johansen KM. 2000. JIL-1, a chromosomal kinase implicated in regulation of chromatin structure, associates with the male specific lethal (MSL) dosage compensation complex. *J Cell Biol* **149**: 1005-1010.
- Johnson NA, Lachance J. 2012. The genetics of sex chromosomes: evolution and implications for hybrid incompatibility. *Annals of the New York Academy of Sciences* **1256**: E1-22.
- Kadlec J, Hallacli E, Lipp M, Holz H, Sanchez-Weatherby J, Cusack S, Akhtar A. 2011. Structural basis for MOF and MSL3 recruitment into the dosage compensation complex by MSL1. *Nat Struct Mol Biol* **18**: 142-149.
- Keene JD, Komisarow JM, Friedersdorf MB. 2006. RIP-Chip: the isolation and identification of mRNAs, microRNAs and protein components of ribonucleoprotein complexes from cell extracts. *Nature protocols* **1**: 302-307.
- Kelley RL, Lee OK, Shim YK. 2008. Transcription rate of noncoding roX1 RNA controls local spreading of the *Drosophila* MSL chromatin remodeling complex. *Mechanisms of development* **125**: 1009-1019.
- Kelley RL, Meller VH, Gordadze PR, Roman G, Davis RL, Kuroda MI. 1999. Epigenetic spreading of the *Drosophila* dosage compensation complex from roX RNA genes into flanking chromatin. *Cell* **98**: 513-522.
- Kelley RL, Wang J, Bell L, Kuroda MI. 1997. Sex lethal controls dosage compensation in *Drosophila* by a non-splicing mechanism. *Nature* **387**: 195-199.
- Keogh MC, Kurdistani SK, Morris SA, Ahn SH, Podolny V, Collins SR, Schuldiner M, Chin K, Punna T, Thompson NJ et al. 2005. Cotranscriptional set2 methylation of histone H3 lysine 36 recruits a repressive Rpd3 complex. *Cell* **123**: 593-605.
- Kharchenko PV, Alekseyenko AA, Schwartz YB, Minoda A, Riddle NC, Ernst J, Sabo PJ, Larschan E, Gorchakov AA, Gu T et al. 2011. Comprehensive analysis of the chromatin landscape in *Drosophila melanogaster*. *Nature* **471**: 480-485.
- Kharchenko PV, Tolstorukov MY, Park PJ. 2008. Design and analysis of ChIP-seq experiments for DNA-binding proteins. *Nat Biotechnol* **26**: 1351-1359.

- Kind J, Vaquerizas JM, Gebhardt P, Gentzel M, Luscombe NM, Bertone P, Akhtar A. 2008. Genome-wide analysis reveals MOF as a key regulator of dosage compensation and gene expression in *Drosophila*. *Cell* **133**: 813-828.
- Klug A. 2010. The discovery of zinc fingers and their development for practical applications in gene regulation and genome manipulation. *Q Rev Biophys* **43**: 1-21.
- Klug A, Schwabe JW. 1995. Protein motifs 5. Zinc fingers. *FASEB journal : official publication of the Federation of American Societies for Experimental Biology* **9**: 597-604.
- Ko-Hsuan Hung YW, Jin Crystal Zhao. 2013. Regulation of Mammalian Gene Dosage by Long Non-coding RNAs. *Biomolecules* **3(1)**: 124-142.
- Kulkarni MM, Booker M, Silver SJ, Friedman A, Hong P, Perrimon N, Mathey-Prevot B. 2006. Evidence of off-target effects associated with long dsRNAs in *Drosophila melanogaster* cell-based assays. *Nat Methods* **3**: 833-838.
- Lai F, Orom UA, Cesaroni M, Beringer M, Taatjes DJ, Blobel GA, Shiekhata R. 2013. Activating RNAs associate with Mediator to enhance chromatin architecture and transcription. *Nature* **494**: 497-501.
- Langmead B, Trapnell C, Pop M, Salzberg SL. 2009. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol* **10**: R25.
- Larschan E, Alekseyenko AA, Gortchakov AA, Peng S, Li B, Yang P, Workman JL, Park PJ, Kuroda MI. 2007. MSL complex is attracted to genes marked by H3K36 trimethylation using a sequence-independent mechanism. *Mol Cell* **28**: 121-133.
- Larschan E, Alekseyenko AA, Lai WR, Park PJ, Kuroda MI. 2006. MSL complex associates with clusters of actively transcribed genes along the *Drosophila* male X chromosome. *Cold Spring Harbor symposia on quantitative biology* **71**: 385-394.
- Larschan E, Bishop E, Kharchenko P, Park P, Kuroda M. 2011a. X chromosome dosage compensation via enhanced transcription elongation in *Drosophila* males. *Nature* **In press**.
- Larschan E, Bishop EP, Kharchenko PV, Core LJ, Lis JT, Park PJ, Kuroda MI. 2011b. X chromosome dosage compensation via enhanced transcriptional elongation in *Drosophila*. *Nature* **471**: 115-118.
- Larschan E, Soruco MM, Lee OK, Peng S, Bishop E, Chery J, Goebel K, Feng J, Park PJ, Kuroda MI. 2012. Identification of chromatin-associated regulators of MSL complex targeting in *Drosophila* dosage compensation. *PLoS Genet* **8**: e1002830.
- Lee C, Li X, Hechmer A, Eisen M, Biggin MD, Venters BJ, Jiang C, Li J, Pugh BF, Gilmour DS. 2008. NELF and GAGA factor are linked to promoter-proximal pausing at many genes in *Drosophila*. *Mol Cell Biol* **28**: 3290-3300.
- Lee JT. 2012. Epigenetic regulation by long noncoding RNAs. *Science* **338**: 1435-1439.
- Lee JT, Bartolomei MS. 2013. X-Inactivation, Imprinting, and Long Noncoding RNAs in Health and Disease. *Cell* **152**: 1308-1323.

- Lejeune J, Gautier M, Turpin R. 1959. [Study of somatic chromosomes from 9 mongoloid children]. *Comptes rendus hebdomadaires des seances de l'Academie des sciences* **248**: 1721-1722.
- Leppig KA, Disteche CM. 2001. Ring X and other structural X chromosome abnormalities: X inactivation and phenotype. *Seminars in reproductive medicine* **19**: 147-157.
- Li F, Parry DA, Scott MJ. 2005. The amino-terminal region of Drosophila MSL1 contains basic, glycine-rich, and leucine zipper-like motifs that promote X chromosome binding, self-association, and MSL2 binding, respectively. *Molecular and cellular biology* **25**: 8913-8924.
- Li F, Schiemann AH, Scott MJ. 2008. Incorporation of the noncoding roX RNAs alters the chromatin-binding specificity of the Drosophila MSL1/MSL2 complex. *Molecular and cellular biology* **28**: 1252-1264.
- Lim CK, Kelley RL. 2012. Autoregulation of the Drosophila Noncoding roX1 RNA Gene. *PLoS Genet* **8**: e1002564.
- Lipeng Wu BMZ, Yanming Wang, Benjamin A. Garcia, Yali Dou. 2011. The RING Finger Protein MSL2 in the MOF Complex Is an E3 Ubiquitin Ligase for H2B K34 and Is Involved in Crosstalk with H3 K4 and K79 Methylation. *Molecular cell* **43**: 132-144.
- Little M, Holmes G, Walsh P. 1999. WT1: what has the last decade told us? *BioEssays : news and reviews in molecular, cellular and developmental biology* **21**: 191-202.
- Lours C, Bardot O, Godt D, Laski FA, Couderc JL. 2003. The Drosophila melanogaster BTB proteins bric a brac bind DNA through a composite DNA binding domain containing a pipsqueak and an AT-Hook motif. *Nucleic acids research* **31**: 5389-5398.
- Lucchesi JC. 1998. Dosage compensation in flies and worms: the ups and downs of X-chromosome regulation. *Current opinion in genetics & development* **8**: 179-184.
- Lucchesi JC, Kelly WG, Panning B. 2005. Chromatin remodeling in dosage compensation. *Annu Rev Genet* **39**: 615-651.
- Lucchesi JMBJC. 1980. Control of X Chromosome transcription by the maleless gene in *Drosophila*. *Nature* **285**: 573-575.
- Lusk RW, Eisen MB. 2013. Spatial promoter recognition signatures may enhance transcription factor specificity in yeast. *PloS one* **8**: e53778.
- Lyman LM, Copps K, Rastelli L, Kelley RL, Kuroda MI. 1997. Drosophila male-specific lethal-2 protein: Structure/function analysis and dependence on MSL-1 for chromosome association. *Genetics* **147**: 1743-1753.
- Maenner S, Muller M, Frohlich J, Langer D, Becker PB. 2013. ATP-dependent roX RNA remodeling by the helicase maleless enables specific association of MSL proteins. *Molecular cell* **51**: 174-184.
- Mank JE. 2009. The W, X, Y and Z of sex-chromosome dosage compensation. *Trends in genetics : TIG* **25**: 226-233.

- McDonel P, Jans J, Peterson BK, Meyer BJ. 2006. Clustered DNA motifs mark X chromosomes for repression by a dosage compensation complex. *Nature* **444**: 614-618.
- Meller VH. 2003. Initiation of dosage compensation in *Drosophila* embryos depends on expression of the roX RNAs. *Mechanisms of development* **120**: 759-767.
- Meller VH, Kuroda MI. 2002. Sex and the single chromosome. *Advances in genetics* **46**: 1-24.
- Meller VH, Rattner BP. 2002a. The roX genes encode redundant male-specific lethal transcripts required for targeting of the MSL complex. *The EMBO journal* **21**: 1084-1091.
- . 2002b. The roX genes encode redundant male-specific lethal transcripts required for targeting of the MSL complex. *Embo J* **21**: 1084-1091.
- Meller VH, Wu KH, Roman G, Kuroda MI, Davis RL. 1997. roX1 RNA paints the X chromosome of male *Drosophila* and is regulated by the dosage compensation system. *Cell* **88**: 445-457.
- Mendjan S, Taipale M, Kind J, Holz H, Gebhardt P, Schelder M, Vermeulen M, Buscaino A, Duncan K, Mueller J et al. 2006. Nuclear pore components are involved in the transcriptional regulation of dosage compensation in *Drosophila*. *Molecular cell* **21**: 811-823.
- Meyer BJ. 1997. Sex Determination and X Chromosome Dosage Compensation.
- . 2010. Targeting X chromosomes for repression. *Current opinion in genetics & development* **20**: 179-189.
- Mito Y, Henikoff JG, Henikoff S. 2007. Histone replacement marks the boundaries of cis-regulatory domains. *Science* **315**: 1408-1411.
- Morra R, Smith ER, Yokoyama R, Lucchesi JC. 2008. The MLE subunit of the *Drosophila* MSL complex uses its ATPase activity for dosage compensation and its helicase activity for targeting. *Molecular and cellular biology* **28**: 958-966.
- Nagaike T, Logan C, Hotta I, Rozenblatt-Rosen O, Meyerson M, Manley JL. 2011. Transcriptional activators enhance polyadenylation of mRNA precursors. *Mol Cell* **41**: 409-418.
- Nguyen DK, Disteche CM. 2006. Dosage compensation of the active X chromosome in mammals. *Nat Genet* **38**: 47-53.
- Nikiforov MA, Chandriani S, Park J, Kotenko I, Matheos D, Johnsson A, McMahon SB, Cole MD. 2002. TRRAP-dependent and TRRAP-independent transcriptional activation by Myc family oncoproteins. *Mol Cell Biol* **22**: 5054-5063.
- Nora EP, Lajoie BR, Schulz EG, Giorgetti L, Okamoto I, Servant N, Piolot T, van Berkum NL, Meisig J, Sedat J et al. 2012. Spatial partitioning of the regulatory landscape of the X-inactivation centre. *Nature* **485**: 381-385.
- Nybakken K, Perrimon N. 2002. Hedgehog signal transduction: recent findings. *Curr Opin Genet Dev* **12**: 503-511.
- Oh H, Park Y, Kuroda MI. 2003. Local spreading of MSL complexes from roX genes on the *Drosophila* X chromosome. *Genes & development* **17**: 1334-1339.

- Onodera O, Roses AD, Tsuji S, Vance JM, Strittmatter WJ, Burke JR. 1996. Toxicity of expanded polyglutamine-domain proteins in *Escherichia coli*. *FEBS letters* **399**: 135-139.
- Orom UA, Derrien T, Beringer M, Gumireddy K, Gardini A, Bussotti G, Lai F, Zytnicki M, Notredame C, Huang Q et al. 2010. Long noncoding RNAs with enhancer-like function in human cells. *Cell* **143**: 46-58.
- Palmer MJ, Richman R, Richter L, Kuroda MI. 1994. Sex-Specific Regulation of the Male-Specific Lethal-1 Dosage Compensation Gene in *Drosophila*. *Gene Dev* **8**: 698-706.
- Paris M, Kaplan T, Li XY, Villalta JE, Lott SE, Eisen MB. 2013. Extensive divergence of transcription factor binding in *Drosophila* embryos with highly conserved gene expression. *PLoS genetics* **9**: e1003748.
- Park SW, Oh H, Lin YR, Park Y. 2010. MSL cis-spreading from roX gene up-regulates the neighboring genes. *Biochemical and biophysical research communications* **399**: 227-231.
- Park Y, Kelley RL, Oh H, Kuroda MI, Meller VH. 2002. Extent of chromatin spreading determined by roX RNA recruitment of MSL proteins. *Science* **298**: 1620-1623.
- Park Y, Mengus G, Bai X, Kageyama Y, Meller VH, Becker PB, Kuroda MI. 2003. Sequence-specific targeting of *Drosophila* roX genes by the MSL dosage compensation complex. *Molecular cell* **11**: 977-986.
- Payer B, Lee JT. 2008. X chromosome dosage compensation: how mammals keep the balance. *Annu Rev Genet* **42**: 733-772.
- Pelassa I, Cora D, Cesano F, Monje FJ, Montarolo PG, Fiumara F. 2014. Association of polyalanine and polyglutamine coiled coils mediates expansion disease-related protein aggregation and dysfunction. *Hum Mol Genet*.
- Pferdehirt RR, Kruesi WS, Meyer BJ. 2011. An MLL/COMPASS subunit functions in the *C. elegans* dosage compensation complex to target X chromosomes for transcriptional regulation of gene expression. *Genes & development* **25**: 499-515.
- Pinter SF, Sadreyev RI, Yildirim E, Jeon Y, Ohsumi TK, Borowsky M, Lee JT. 2012. Spreading of X chromosome inactivation via a hierarchy of defined Polycomb stations. *Genome research* **22**: 1864-1876.
- Prestel M, Feller C, Becker PB. 2010a. Dosage compensation and the global rebalancing of aneuploid genomes. *Genome biology* **11**: 216.
- Prestel M, Feller C, Straub T, Mitlohner H, Becker PB. 2010b. The activation potential of MOF is constrained for dosage compensation. *Molecular cell* **38**: 815-826.
- Raja SJ, Charapitsa I, Conrad T, Vaquerizas JM, Gebhardt P, Holz H, Kadlec J, Fraterman S, Luscombe NM, Akhtar A. 2010. The nonspecific lethal complex is a transcriptional regulator in *Drosophila*. *Mol Cell* **38**: 827-841.
- Ramadan N, Flockhart I, Booker M, Perrimon N, Mathey-Prevot B. 2007. Design and implementation of high-throughput RNAi screens in cultured *Drosophila* cells. *Nat Protoc* **2**: 2245-2264.
- Rinn JL, Chang HY. 2012. Genome regulation by long noncoding RNAs. *Annual review of biochemistry* **81**: 145-166.

- Robasky K, Bulyk ML. 2011. UniPROBE, update 2011: expanded content and search tools in the online database of protein-binding microarray data on protein-DNA interactions. *Nucleic Acids Res* **39**: D124-128.
- Salz HK, Erickson JW. 2010. Sex determination in *Drosophila*: The view from the top. *Fly (Austin)* **4**: 60-70.
- Schulz EG, Heard E. 2013. Role and control of X chromosome dosage in mammalian development. *Current opinion in genetics & development*.
- Sekido R, Lovell-Badge R. 2009. Sex determination and SRY: down to a wink and a nudge? *Trends in genetics : TIG* **25**: 19-29.
- Shi X, Finkelstein A, Wolf AJ, Wade PA, Burton ZF, Jaehning JA. 1996. Paf1p, an RNA polymerase II-associated factor in *Saccharomyces cerevisiae*, may have both positive and negative roles in transcription. *Mol Cell Biol* **16**: 669-676.
- Shiro Iuchi NK. 2005. *Zinc Finger Proteins: From Atomic Contact to Cellular Function*. Landes Bioscience / Eureka.com, Kluwer Academic / Plenum Publishers 233 Spring Street New York, New York U.S.A. 100013.
- Shogren-Knaak M, Ishii H, Sun JM, Pazin MJ, Davie JR, Peterson CL. 2006. Histone H4-K16 acetylation controls chromatin structure and protein interactions. *Science* **311**: 844-847.
- Shogren-Knaak M, Peterson CL. 2006. Switching on chromatin: mechanistic role of histone H4-K16 acetylation. *Cell Cycle* **5**: 1361-1365.
- Simon MD, Pinter SF, Fang R, Sarma K, Rutenberg-Schoenberg M, Bowman SK, Kesner BA, Maier VK, Kingston RE, Lee JT. 2013. High-resolution Xist binding maps reveal two-step spreading during X-chromosome inactivation. *Nature*.
- Smith ER, Allis CD, Lucchesi JC. 2001. Linking global histone acetylation to the transcription enhancement of X-chromosomal genes in *Drosophila* males. *J Biol Chem* **276**: 31483-31486.
- Smith ER, Pannuti A, Gu W, Steurnagel A, Cook RG, Allis CD, Lucchesi JC. 2000. The *drosophila* MSL complex acetylates histone H4 at lysine 16, a chromatin modification linked to dosage compensation. *Molecular and cellular biology* **20**: 312-318.
- Soruco MM, Chery J, Bishop EP, Siggers T, Tolstorukov MY, Leydon AR, Sugden AU, Goebel K, Feng J, Xia P et al. 2013. The CLAMP protein links the MSL complex to the X chromosome during *Drosophila* dosage compensation. *Genes & development* **27**: 1551-1556.
- Spierer A, Seum C, Delattre M, Spierer P. 2005. Loss of the modifiers of variegation Su(var)3-7 or HP1 impacts male X polytene chromosome morphology and dosage compensation. *J Cell Sci* **118**: 5047-5057.
- Straub T, Grimaud C, Gilfillan GD, Mitterweger A, Becker PB. 2008. The chromosomal high-affinity binding sites for the *Drosophila* dosage compensation complex. *PLoS genetics* **4**: e1000302.
- Sunde JMMaM. 2002. Zinc Fingers--Folds for Many Occasions. *IUBMB Life* **54**: 351-355.

- Sural TH, Peng S, Li B, Workman JL, Park PJ, Kuroda MI. 2008. The MSL3 chromodomain directs a key targeting step for dosage compensation of the *Drosophila melanogaster* X chromosome. *Nature structural & molecular biology* **15**: 1318-1325.
- Tanny JC, Erdjument-Bromage H, Tempst P, Allis CD. 2007. Ubiquitylation of histone H2B controls RNA polymerase II transcription elongation independently of histone H3 methylation. *Genes Dev* **21**: 835-847.
- Thomas CJ, Kotova E, Andrade M, Adolf-Bryfogle J, Glaser R, Regnard C, Tulin AV. 2014. Kinase-Mediated Changes in Nucleosome Conformation Trigger Chromatin Decondensation via Poly(ADP-Ribosyl)ation. *Molecular cell* **53**: 831-842.
- Thomas Cremer MC. 2010. Chromosome Territories. *Cold Spring Harbor perspectives in biology*.
- Thomas S, Li XY, Sabo PJ, Sandstrom R, Thurman RE, Canfield TK, Giste E, Fisher W, Hammonds A, Celniker SE et al. 2011. Dynamic reprogramming of chromatin accessibility during *Drosophila* embryo development. *Genome biology* **12**: R43.
- Vaquerizas JM, Kummerfeld SK, Teichmann SA, Luscombe NM. 2009. A census of human transcription factors: function, expression and evolution. *Nature reviews Genetics* **10**: 252-263.
- Vaquero A, Blanch M, Espinas ML, Bernues J. 2008a. Activation properties of GAGA transcription factor. *Biochimica et biophysica acta* **1779**: 312-317.
- . 2008b. Activation properties of GAGA transcription factor. *Bba-Gene Regul Mech* **1779**: 312-317.
- Villa R, Forne I, Muller M, Imhof A, Straub T, Becker PB. 2012. MSL2 combines sensor and effector functions in homeostatic control of the *Drosophila* dosage compensation machinery. *Molecular cell* **48**: 647-654.
- Voss TC, Hager GL. 2014. Dynamic regulation of transcriptional states by chromatin and transcription factors. *Nature reviews Genetics* **15**: 69-81.
- Wang CI, Alekseyenko AA, LeRoy G, Elia AE, Gorchakov AA, Britton LM, Elledge SJ, Kharchenko PV, Garcia BA, Kuroda MI. 2013a. Chromatin proteins captured by ChIP-mass spectrometry are linked to dosage compensation in *Drosophila*. *Nature structural & molecular biology* **20**: 202-209.
- . 2013b. Chromatin proteins captured by ChIP-mass spectrometry are linked to dosage compensation in *Drosophila*. *Nat Struct Mol Biol*.
- Wells MB, Snyder MJ, Custer LM, Csankovszki G. 2012. *Caenorhabditis elegans* dosage compensation regulates histone H4 chromatin state on X chromosomes. *Molecular and cellular biology* **32**: 1710-1719.
- Worringer KA, Panning B. 2007. Zinc finger protein Zn72D promotes productive splicing of the maleless transcript. *Mol Cell Biol* **27**: 8760-8769.
- Wu L, Zee BM, Wang Y, Garcia BA, Dou Y. 2011. The RING Finger Protein MSL2 in the MOF Complex Is an E3 Ubiquitin Ligase for H2B K34 and Is Involved in Crosstalk with H3 K4 and K79 Methylation. *Mol Cell* **43**: 132-144.

- Xiao-Yong Li ST, Peter J Sabo, Michael B Eisen, John A Stamatoyannopoulos and Mark D Biggin. 2011. The role of chromatin accessibility in directing the widespread, overlapping patterns of *Drosophila* transcription factor binding. *Genome biology* **12**.
- Yonker SA, Meyer BJ. 2003. Recruitment of *C. elegans* dosage compensation proteins for gene-specific versus chromosome-wide repression. *Development* **130**: 6519-6532.
- Zhao J, Sun BK, Erwin JA, Song JJ, Lee JT. 2008. Polycomb proteins targeted by a short repeat RNA to the mouse X chromosome. *Science* **322**: 750-756.
- Zhou K, Kuo WH, Fillingham J, Greenblatt JF. 2009. Control of transcriptional elongation and cotranscriptional histone modification by the yeast BUR kinase substrate Spt5. *Proc Natl Acad Sci U S A* **106**: 6956-6961.
- Zhou T, Yang L, Lu Y, Dror I, Dantas Machado AC, Ghane T, Di Felice R, Rohs R. 2013. DNASHape: a method for the high-throughput prediction of DNA structural features on a genomic scale. *Nucleic acids research* **41**: W56-62.

**Appendix 1: The glutamine domain of CLAMP is necessary for MSL complex
recruitment**

Jessica Chery

Appendix 1: The glutamine domain of CLAMP is necessary for MSL complex recruitment

Jessica Chery

Introduction

Dosage compensation equalizes X-linked gene expression between males and females and with autosomes. In the model organisms mammals, *C.elegans*, and *Drosophila*, an imbalanced gene dose can occur because females have two X-chromosomes while males have a single X-chromosome. However, the *Drosophila* dosage compensation mechanism allows males (XY) to equalize expression of their X-linked genes with females (XX) by increasing their expression two-fold (Lucchesi 1980; Lucchesi et al. 2005). A key regulator of dosage compensation is the conserved Male-Specific Lethal (MSL) complex, which associates specifically with the male X-chromosome and is present only in males (Kelley et al. 1999).

The MSL complex is a ribonucleoprotein complex composed of five proteins and two noncoding RNAs required for dosage compensation (Gelbart and Kuroda 2009). The proteins include MSL1, MSL2, MSL3, MLE (maleless), and MOF (males absent on the first) (Palmer et al. 1994; Lyman et al. 1997; Gu et al. 1998b). The RNA components of the MSL complex are two noncoding RNAs, roX1 and roX2 (RNA on X) that are functionally redundant although structurally different (Meller and Rattner 2002a). They are encoded on the X and themselves serve as sites of initial MSL recruitment to then be incorporated into the fully functional MSL complex that

then mediates dosage compensation (Meller et al. 1997; Kelley et al. 1999). The complex primarily functions as a histone acetyltransferase responsible for modifying histone H4 at lysine 16 (H4K16). H4K16ac is the chromatin mark that stimulates transcription by promoting global chromatin de-condensation and activating elongation of paused polymerase into active transcription (Akhtar and Becker 2000; Larschan et al. 2011b). However, none of the components of the MSL complex sequence specifically bind to DNA, hence the requirement for CLAMP to link MSL complex directly to the MRE sequences necessary for MSL complex recruitment.

CLAMP contains a seven C2H2 zinc finger domain that binds directly to MRE sequences (Soruco et al. 2013). Furthermore, the CLAMP protein associates with MSL complex *in vivo* as identified from a crosslinking mass spectrometry approach (Wang et al. 2013a). Additionally, CLAMP has a glutamine-rich N-terminus similar to other transcription factor domains that recruit co-activators, such as GAGA factor (GAF). Glutamine-rich domains have been demonstrated to interact with each other, implicating the ability for glutamine domains to recruit other factors with glutamine-rich domains (Atanesyan et al. 2012). The glutamine-rich domain of GAF recruits co-activators and general transcription factors and mediates protein-protein interactions (Adkins et al. 2006; Vaquero et al. 2008a). In addition, the DNA binding domain of GAF contains a single zinc finger flanked by three basic regions giving it specificity for a single GAG sequence of DNA (Adkins et al. 2006). GAF is also involved in keeping chromatin open for transcription and assisting in

chromatin remodeling via its protein-protein interactions with the NURF chromatin remodeling complex (Adkins et al. 2006).

Despite these intriguing observations, the domains of CLAMP that are sufficient for MSL recruitment are undefined. Many highly conserved transcriptional activators contain glutamine-rich regions that directly recruit co-activators and general transcription factors (Colgan et al. 1995). Therefore, we hypothesized that CLAMP recruits the MSL complex via its glutamine-rich N-terminus, similarly to how other zinc-finger domain containing proteins interact with co-activator complexes (Vaquero et al. 2008b). To test this hypothesis, we performed structure-function analysis using a series of deletions within the glutamine-rich region of CLAMP to determine which segment is required for MSL complex recruitment *in vivo*.

Results and Discussion

CLAMP is a 562 amino acid protein. The glutamine-rich domain is contained within the first 318 amino acids, and the remaining 244 amino acids compose the seven C₂H₂ zinc fingers of CLAMP (Fig 2). Although the first zinc finger in the C-terminal 7 C₂H₂ series does not occur until amino acid 319, the 7 zinc finger domain of CLAMP was designed to include amino acids 288 – 562 to enable protein expression (Fig 2).

We fused Tandem affinity purification (TAP) and the HA (hemagglutinin) tags to the CLAMP constructs. The TAP and HA tagged truncated proteins of CLAMP were expressed in male (SL2) *Drosophila* cells and used in the previously described MSL dependent luciferase reporter assay to determine effects on MSL complex

recruitment (Larschan et al. 2012). We successfully expressed the TAP-tagged version of full-length CLAMP in male *Drosophila* cells and confirmed specificity with Western blot analysis (Fig 1). The C-terminal TAP tagged construct was under the control of a copper inducible promoter and constructed by a genome-wide epitope-tagging project (www.flybase.org). The copper induction system was used instead of the endogenous promoter to allow protein levels to be titrated and because the putative endogenous CLAMP promoter is very large and uncharacterized (Bryson et al. 2010). The C-terminal TAP-tagged CLAMP construct did not associate with the normal endogenous CLAMP binding sites as evaluated by chromatin immunoprecipitation qPCR (data not shown) and therefore was not used in subsequent studies. We suggest this lack of appropriate targeting of TAP-CLAMP may be due to the interference of the TAP tag with CLAMP DNA specificity as the TAP tag is fused to CLAMP's DNA binding domain.

Therefore, we next used a small N-terminal HA epitope tag to avoid disrupting the DNA binding domain. We expressed six HA-tagged CLAMP truncations from an actin promoter in male cells (Fig 2). The actin promoter was used because the CLAMP promoter is very long and uncharacterized. Expression was confirmed through Western blot analysis with an anti-HA antibody (Fig 2B). To reduce levels of the endogenous CLAMP protein, dsRNA targeting the 3'UTR of CLAMP was designed, allowing us to specifically define the effects of the HA-tagged truncations in an available MSL-dependent luciferase assay (Larschan et al. 2012). We initially designed three dsRNA constructs to target three distinct regions of the 3'UTR of

CLAMP. Reduction in CLAMP mRNA levels was assessed in male cells by measuring CLAMP mRNA levels and luciferase activity after transfection with dsRNA before transfecting in HA truncations (Fig 3, 4). The two dsRNA sequences causing greatest CLAMP reduction (dsRNA #2 and #3) were chosen for subsequent experiments. dsRNA targeting the endogenous CLAMP mRNA and HA-tagged CLAMP truncations were then concurrently transfected into male cells and their effects on MSL complex recruitment assessed through the established luciferase reporter assay that identified the CLAMP gene (Larschan et al. 2012) (Dissertation Chapter 2).

Unfortunately, full length HA-tagged CLAMP protein could not be expressed in the SL2 cells after transfection (data not shown). Therefore, we compared the shorter constructs to the longest construct that we were able to express, which contained amino acids 161-562 (missing N-terminal 160 amino acids). None of the constructs could rescue the decrease in levels of activity after RNAi treatment of CLAMP. This observation may be due to the overexpression of the constructs using a strong promoter. In the absence of high levels of endogenous CLAMP protein, the constructs that contained only the 6 zinc finger (amino acids 350-562) or 7 zinc finger domain (amino acids 288-562) of CLAMP displayed the greatest loss of luciferase activity (Fig 5). This loss of activity was greater than that seen with vector alone suggesting a dominant negative activity of these constructs in the absence of normal levels of endogenous CLAMP. These data suggest that the glutamine-rich domain of CLAMP is likely to be important for MSL complex recruitment.

In contrast, the two constructs that contain portions of the glutamine-rich domain and the full zinc finger domain (161-562 and 213-562) both caused reductions in levels of activity in the presence of wild type CLAMP and in its absence (Fig 5). The glutamine-rich domain on its own does not have any activity (1-260 or 1-287). These data suggest the potential for overexpressed CLAMP that contains 161-562 to compete with the endogenous protein and prevent MSL-dependent activity in a dominant negative activity. There are two possible explanations: 1) It is possible that the N-terminal 160 amino acids are important for MSL complex recruitment; or 2) overexpression of the 161-562 region prevents normal MSL-dependent luciferase activity. We previously demonstrated that the zinc fingers of CLAMP can directly bind to MREs and that MREs are required for CLAMP recruitment (Dissertation Chapter 3) (Sorucu et al. 2013). Therefore, our luciferase results support a model by which expression of excess zinc fingers leads to greatly elevated binding to the MREs located in the 3' end of the luciferase reporter (Fig 5). In the absence of CLAMP's glutamine domain to interact with MSL complex, MSL complex is not targeted to the MREs concurrently with CLAMP zinc finger binding. GFP RNAi was used as negative control in all experiments and positive controls included dsRNA targeting the protein-coding region of CLAMP. This RNAi was previously demonstrated to reduce endogenous CLAMP levels by 70 to 80%.

Conclusion

Expression of amino acids 161-562 of CLAMP displayed the lowest levels of luminescence relative to our control sample in which no CLAMP truncations were transfected (Fig 6). This construct of CLAMP retained the most glutamine stretches of all constructs tested in the luciferase assay. We were unable to express full-length recombinant CLAMP protein as the cells died after transfection of full-length CLAMP. We observed similar effects when we attempted to express the full-length GST fusion protein in *Escherichia coli* (*E.coli*). These effects are in line with previous observations that long polyglutamine domains expressed as GST-fusions in *E.coli* impede growth (Onodera et al. 1996). In addition, polyglutamine domains in some proteins promote aggregation, leading to toxicity (Pelassa et al. 2014). The CLAMP construct with the most glutamine stretches (amino acids 161-562) reduced reporter transcription (Fig 6), supporting the hypothesis that excess expression of glutamine domains potentially leads to aggregation of CLAMP with itself and possible sequestration of MSL complex away from the luciferase reporter. Aggregation of CLAMP could conceal the zinc finger domain, preventing binding to MREs and abrogating targeting of MSL complex to the MREs. CLAMP aggregates would therefore result in a loss of activation of the rox2 promoter that controls luciferase gene expression. In addition, these HA-tagged constructs are under actin promoter and are therefore most likely over-expressed in the cells. Further studies are needed to map the direct interaction between CLAMP and MSL complex.

Future work

In the future, we will map the endogenous promoter of CLAMP which can then be used to express the HA CLAMP constructs at physiologically relevant levels. It will be necessary to produce full length CLAMP protein for comparison with other truncations. Furthermore, once a CLAMP mutant in the fly is generated, constructs can be tested *in vivo*. CRISPR and TALEN approaches for mutant generation are currently underway and a promising insertion within the CLAMP 5' UTR is now available that reduces CLAMP mRNA levels to 20% of endogenous levels. Collectively, these experiments have allowed us to determine that the glutamine-rich portion of CLAMP is required for high levels of MSL-dependent luciferase activity. It is possible that this domain directly contacts components of MSL complex or that another protein mediates this interaction.

Materials and Methods

dsRNA synthesis

CLAMP (CG1832) has two isoforms: CG1832-RA and CG1832-RB. The CG1832-RA sequence was used to design the 3'UTR dsRNA constructs.

Three regions of the 3'UTR of CLAMP were targeted with dsRNA.

Construct	3'UTR Region	Number of off-targets
dsRNA#1	1-300 bp	2
dsRNA#2	1-470 bp	3*
dsRNA#3	1-500 bp	4*

*Off-targets include critical genes involved in mitosis, chromosome segregation, phagocytosis, dorsal closure, neuroblast proliferation.

These 3'UTR regions were chosen in part to help keep off-target sites under five and avoid essential genes.

RNAi Constructs' Primers

The following primers were designed to amplify the three target regions from the SD16766 plasmid that contains the cDNA of CLAMP.

3'UTR Forward Primer (gene specific region = 20bp)

TAATACGACTCACTATAGGGATAGTTTAAATGCAGAGGTG

RNAi#1 (300bp) Construct Reverse Primer

TAATACGACTCACTATAGGGCAAATAAATTAGGTTTCATTT

RNAi#2 (470bp) Construct Reverse Primer

TAATACGACTCACTATAGGGTTACCTGTGCATATTCATAC

RNAi#3(500bp) Construct Reverse Primer

TAATACGACTCACTATAGGGTACAACTTTCAACTTATATG

T7 promoter = TAATACGACTCACTATAGGG

T7 promoter was included in the dsRNA constructs to for *in vitro* transcription in the Ambion MEGAScript T7 system. The same forward primer was used in each dsRNA reaction (#1, #2, #3) with the respective reverse primer to amplify target regions.

PCR Reaction Mix:

dH₂O
10X Buffer
10mM dNTP
50mM MgCl₂
10 uM Forward Primer
10uM Reverse Primer
DNA template (SD16766, 121.1 ng/ul)
Platinum Taq

dsRNA was synthesized using Ambion MEGAScript kit (Invitrogen Catalog # AM1334).

dsRNA synthesis using Ambion MEGAscript kit

1) Make up a master mix by multiplying the following by the number of reactions required for each PCR product

ATP	2 ul
CTP	2 ul
GTP	2 ul
UTP	2 ul
H ₂ O	3 ul
10x Buffer	2 ul
Enzyme Mix	2 ul
Template (PCR)	5 ul

Assemble mix on ice and mix by pipeting

2) Incubate at 37 in the heat block for 4-5 hours

3) Add 1 ul DNase I/reaction and incubate at 37 for 15 minutes.

4) Purify using the RNeasy mini cleanup protocol (see manual) eluting 2x 50 ul and using 1 column for every 2 reactions

RNeasy mini kit: Qiagen, Catalog #: 74104

dsRNA were transfected into cells using 15ug of dsRNA per well in a 6 well plate of *Drosophila* male cells then incubated for 6 days at 25°C. GFP RNAi was used as a negative control and dsRNA targeting the protein-coding region of CLAMP as a positive control. Knockdown of endogenous CLAMP protein assayed using standard Western analysis with WesternBreeze Chemiluminescent Kit (Novex Catalog #WB7106), qPCR of CLAMP mRNA levels, and luciferase activity.

qPCR of CLAMP mRNA levels

For qPCR assessment, RNA was extracted from each well using RNAeasy Plus Kit Protocol (Qiagen, Catalog #: 74104). RNA was then reverse transcribed to synthesize cDNA using SuperScript vilo cDNA synthesis kit (Invitrogen, Catalog # 11754-050).

The following primers were used to determine expression levels via qPCR of cDNA.

Factor	Primer Name	Primer Sequence
CLAMP	CG1832 mRNA F	CTGACCACGGCAATAGAA
	CG1832 mRNA R	TCCCTCATTTGGCACTTC
PKA	PKA 2	CAATCAGCAGATTCTCCGGCT
	PKA 3	AGCCGCACTCGCGCTTCTAC
MSL2	MSL2 mRNA F	GCCCTGTCCGTATGAATG
	MSL2 mRNA R	CAACATGGGTAAACAACC
Rox2	roX2-2 F	AGCTCGGATGGCCATCGAAA
	roX2-11 R	GCCAGCAGCAGCATCTCCA

CLAMP Glutamine truncations

Recombinant CLAMP HA fusion truncations were made using standard cloning procedure with the Gateway technology (Invitrogen, Catalog #11824-026).

Luciferase Reporter Assay

Three constructs were transfected into male cells: CLAMP 3'UTR RNAi (RNAi#1,#2,#2), HA-tagged Glutamine truncations and the luciferase reporter.

Luciferase Reporter Assay with Glutamine Truncations, 96-well

1. Cells were counted and diluted to 0.7×10^6 cells/ml with media. 100 ul cells per well is needed.
2. RNAi constructs were diluted to 0.015 ug/ul
3. Glutamine truncations concentrations were calculated to be constant for transfection.
4. A master mix of luciferase reporters was made.
 - a. The following was added per well.
 - 0.25 ul Firefly
 - 0.25 ul Renilla
 - 8.5 ul EC Buffer
 - b. 1.25 ul Enhancer/per well was added and mix vortex 2-3 seconds.
 - c. The mix was incubated 2- 3 minutes at room temperature
 - d. 0.375 ul of Effectene/per well was added and the mix vortex for 10 seconds.
5. RNAi constructs were added to the plate.
6. Mastermix of glutamine truncations was made.

- a. Per well: 1 ul DNA construct + 8 ul EC Buffer
 - b. Added 1.25 ul Enhancer per well.
 - c. Vortex 2-3 seconds.
 - d. Incubated 2-3 minutes at room temperature.
 - e. Added 0.375 ul Effectene per well. Vortex 10 seconds.
7. 10.625 ul of both Firefly reporter mix and glutamine constructs mixes were added to each well.
8. The plate was incubated 4-7 minutes and then spun at 1000 RPM for 1 minute.
9. 100 ul of SL2 cells were added to each well.
10. The plate was incubated at 37°C for 5 days.

Figures

Figure 1. Western blot depicting Copper inducible expression of TAP-tagged CLAMP

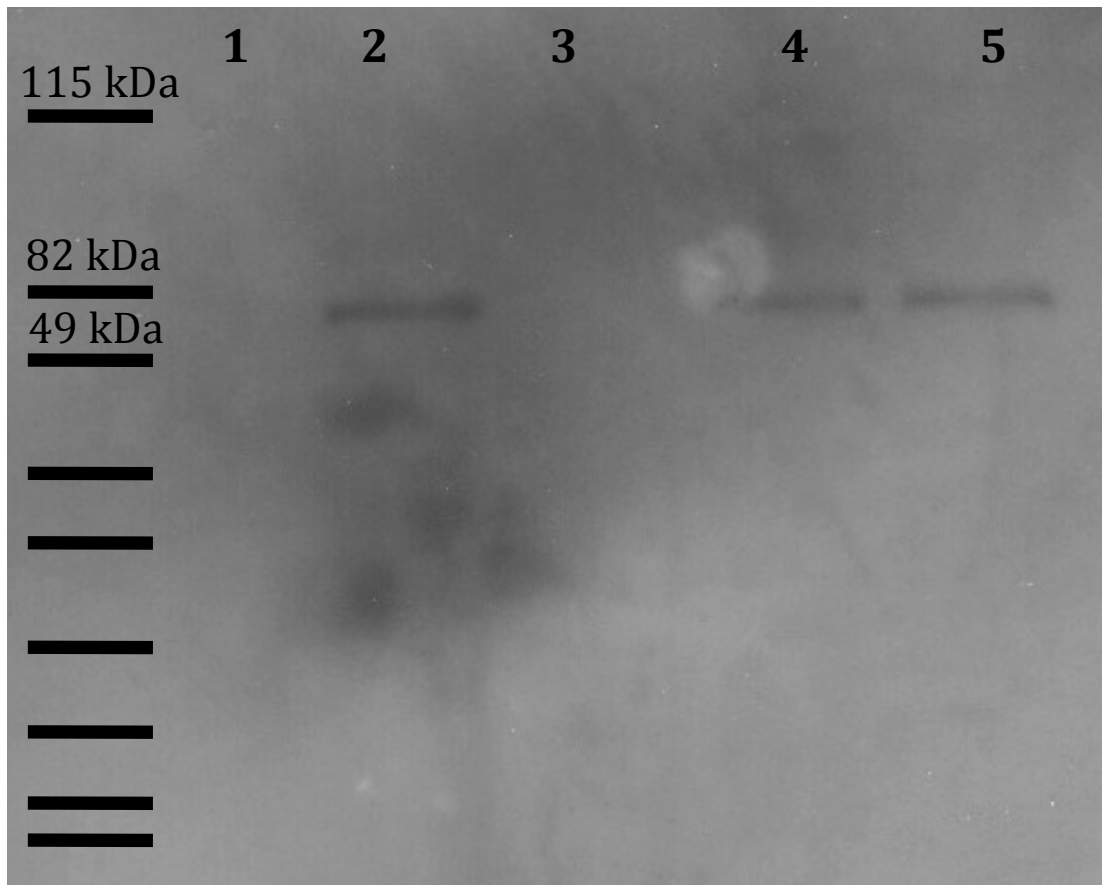


Figure 1. TAP-tagged CLAMP is only expressed upon induction with Copper.

(1) Water, Negative Control **(2)** Clamp + Copper **(3)** Clamp - Copper
(4) Clamp + Copper **(5)** Clamp + Copper

Figure 2A. HA tagged CLAMP truncations were expressed from an actin promoter in male cells

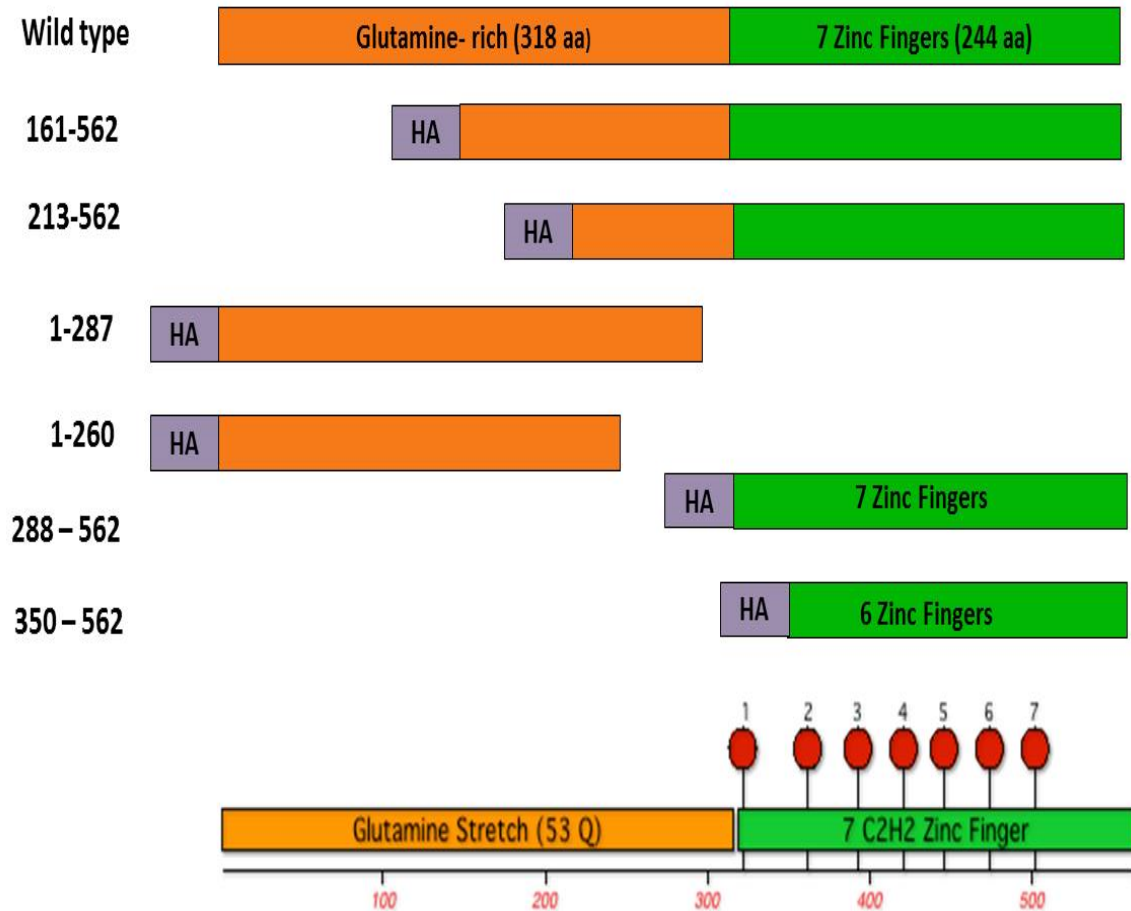


Figure 2B. HA tagged CLAMP truncations were expressed from an actin promoter in male cells

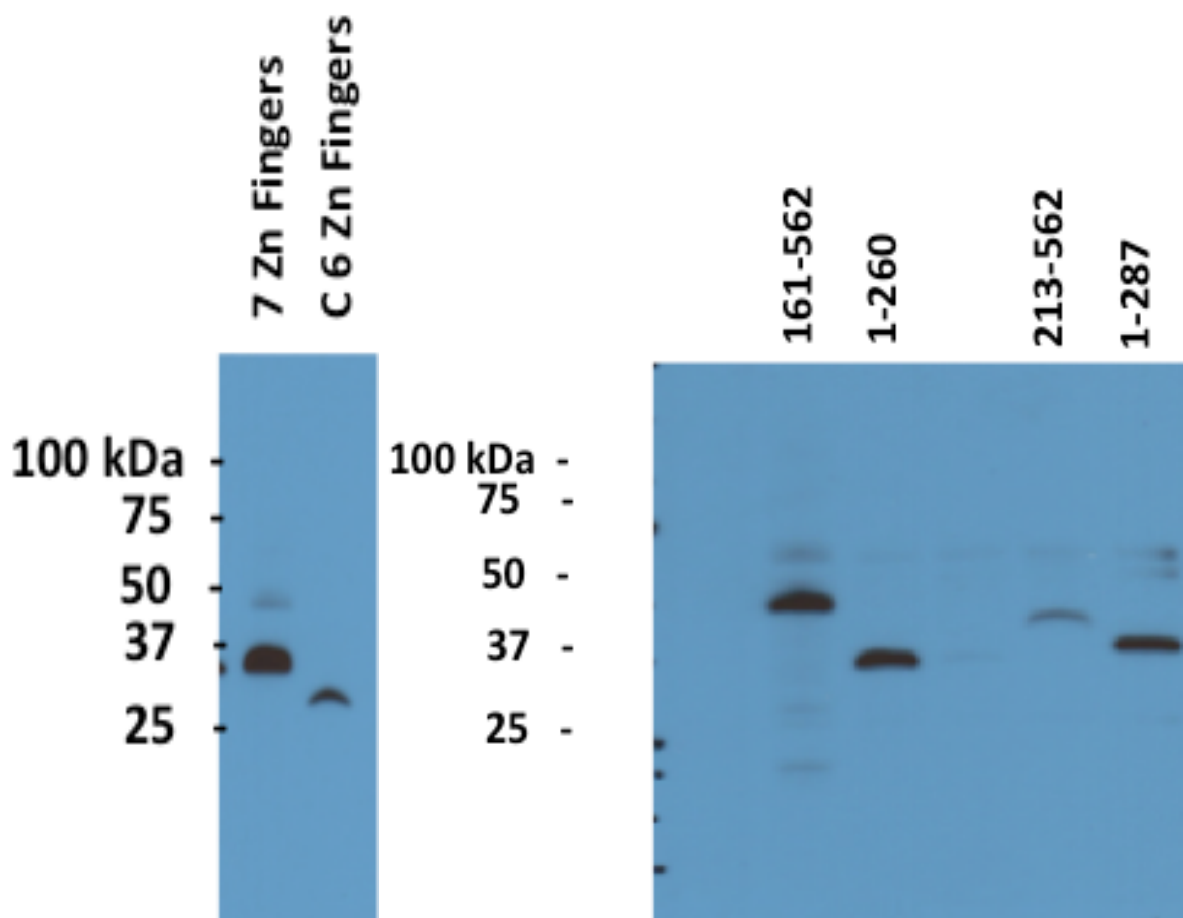


Figure 2B. Truncated versions of full-length CLAMP protein were expressed in *Drosophila* SL2 (male) cells from the actin promoter. Expression of the constructs was assayed using Western blots. Each construct ran at its predicted size: 161-562: 49.17 kDa, 1-260: 33.55 kDa, 213-562 43.45 kDa, 1-287 36.52 kDa, 7 Zn Fingers: 33 kDa, 6 Zn Fingers: 26 kDa kDa. Therefore the SL2 cells were able to successfully express these CLAMP domains. HA antibody was used for detection on the western blot. SL2 protein nuclear extractions from cells that had not been transfected with a CLAMP truncation were run as a negative control (data not shown).

Figure 3. CLAMP mRNA levels after transfection with dsRNA before transfecting in HA truncations

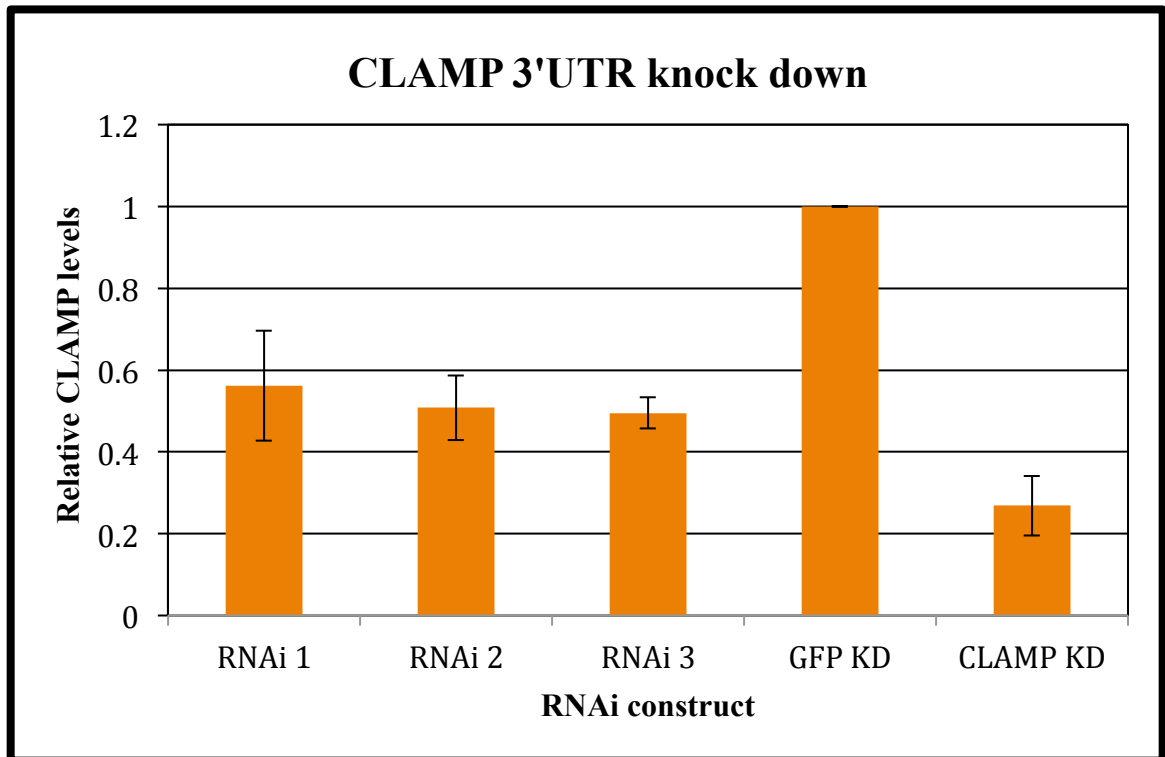


Figure 3. Three different RNAi constructs were designed to target the 3'UTR of CLAMP gene. All three constructs displayed similar efficiency of knockdown. GFP knockdown was used as a negative control and RNAi targeting the protein-coding region of CLAMP was used as a positive control. Similar to what we had observed previously, CLAMP protein expression levels were reduced about 70% in the positive control.

Figure 4. Luciferase activity after transfection with dsRNA before transfecting in HA truncations

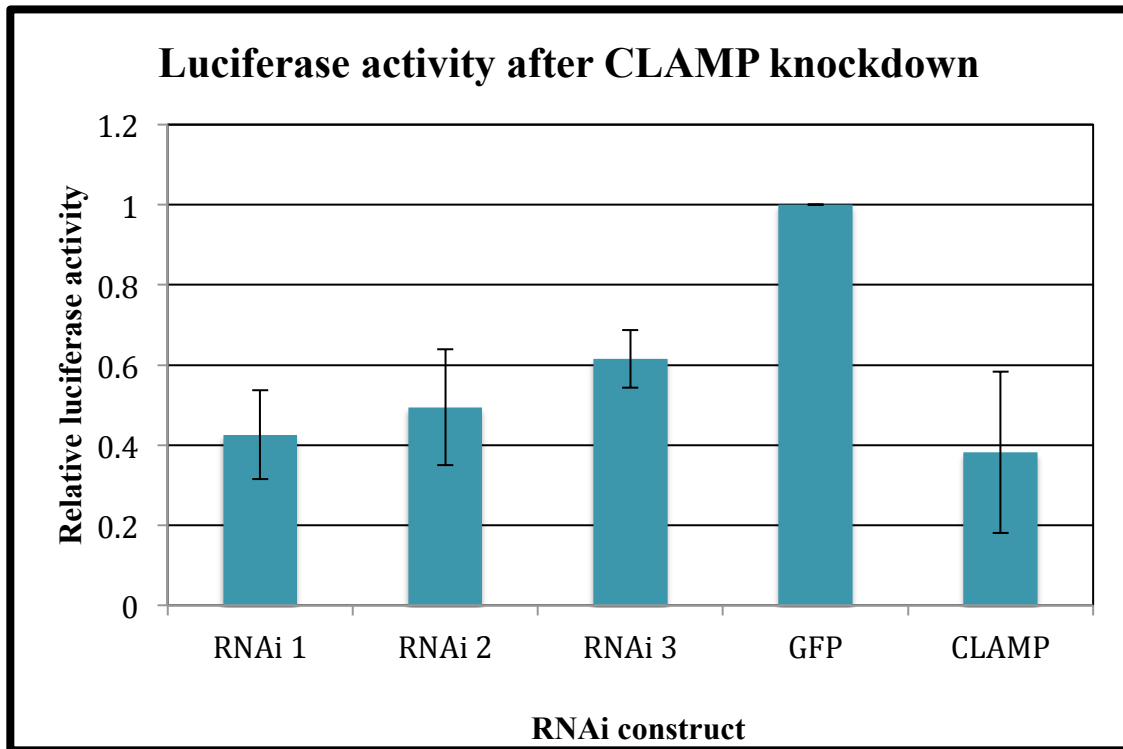


Figure 4. Luciferase activity was measured after CLAMP RNAi to evaluate how effective these constructs are with the luciferase reporter assay. All three RNAi constructs displayed similarly reduced luciferase gene expression, indicating that CLAMP had been effectively reduced in the cells. GFP was used as a negative control and CLAMP protein-coding RNAi as a positive control for CLAMP reduction.

Figure 5. Amino acids 161-562 and the zinc finger domain of CLAMP display the greatest loss in luminescence after dsRNA2 and dsRNA3 treatments

Figure 5A. Luciferase activity of amino acids 161-562 and 213-562 of CLAMP

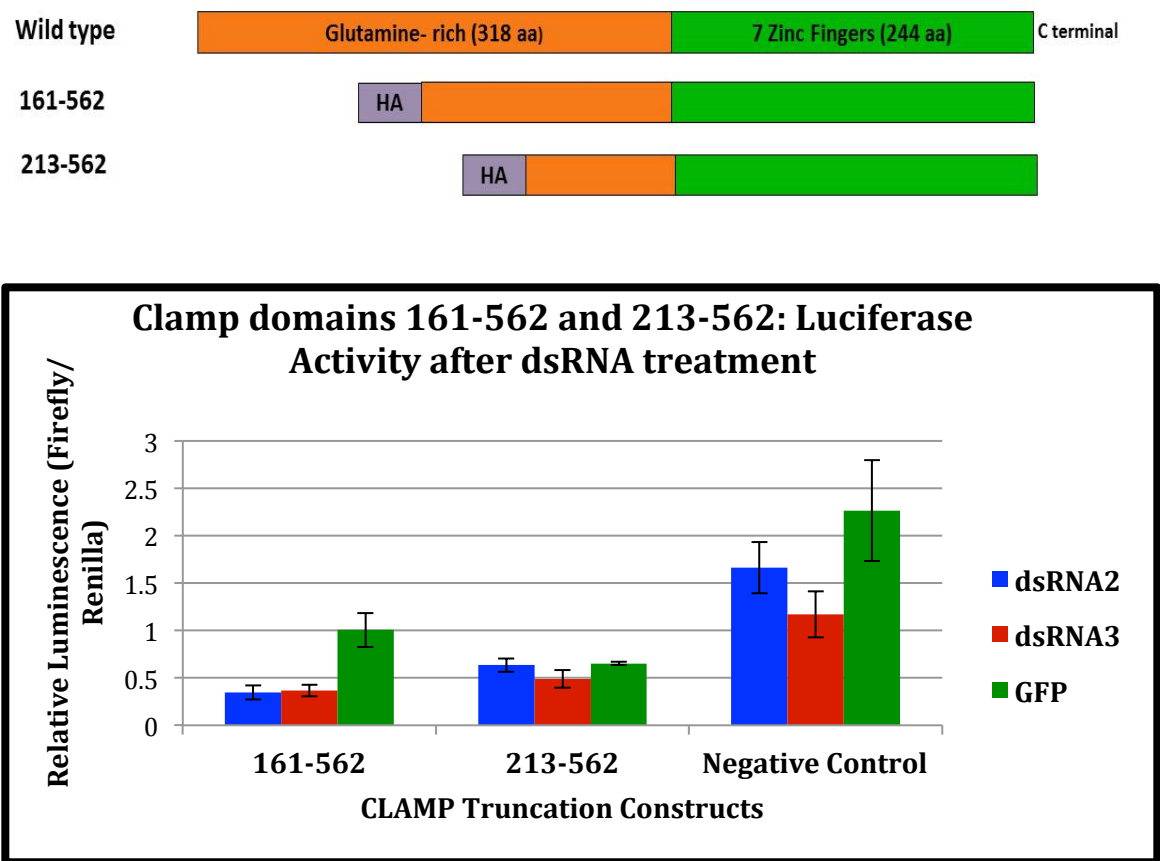


Figure 5A. In the presence and absence of endogenous CLAMP protein, a MSL dependent luciferase reporter shows decreased activity as measured by relative luminescence after transfection with both the 161-562 and 213-562 truncations of CLAMP. An empty vector without any CLAMP truncations was used as a negative control.

Figure 5. Amino acids 161-562 and the zinc finger domain of CLAMP display the greatest loss in luminescence after dsRNA2 and dsRNA3 treatments

Figure 5B. Luciferase activity of amino acids 1-287 and 1-260 of CLAMP

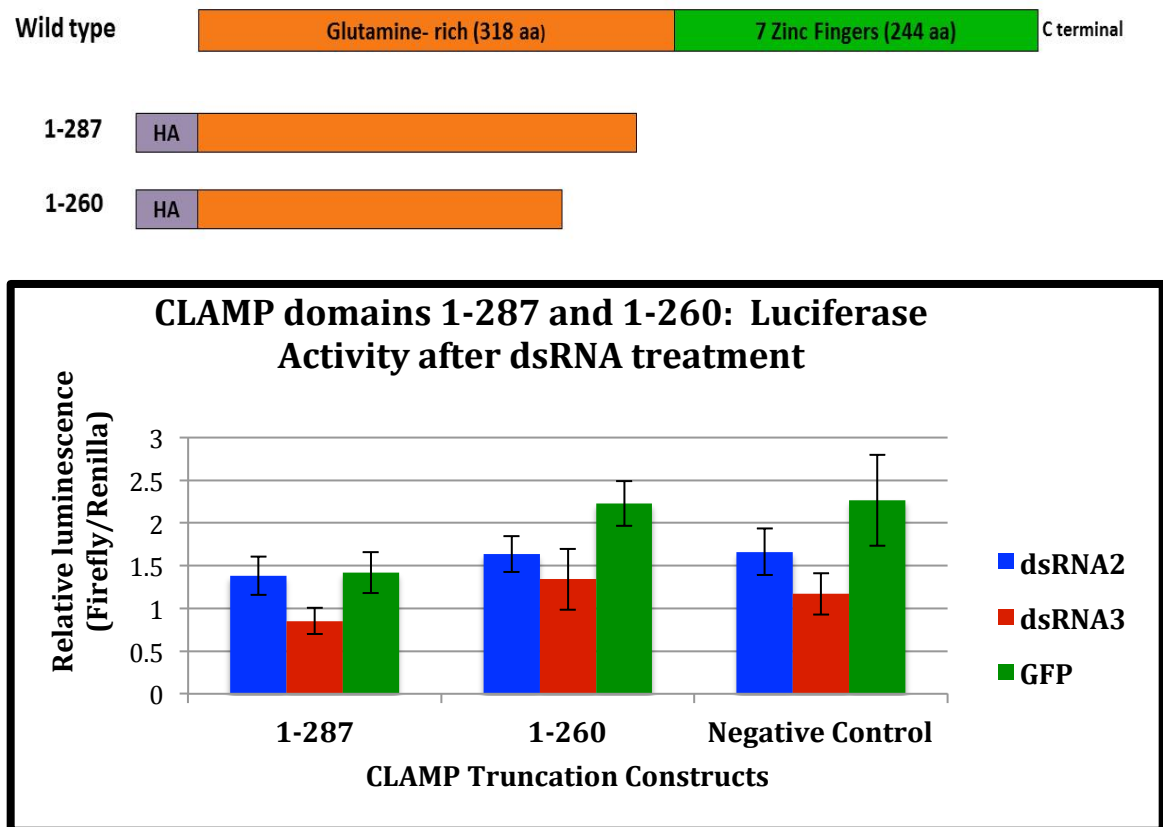


Figure 5B. In the presence and absence of endogenous CLAMP protein, a MSL dependent luciferase reporter shows little change in activity as measured by relative luminescence after transfection with both the 1-287 and 1-260 truncations of CLAMP. An empty vector without any CLAMP truncations was used as a negative control.

Figure 5. Amino acids 161-562 and the zinc finger domain of CLAMP display the greatest loss in luminescence after dsRNA2 and dsRNA3 treatments

Figure 5C. Luciferase activity of zinc finger domains of CLAMP

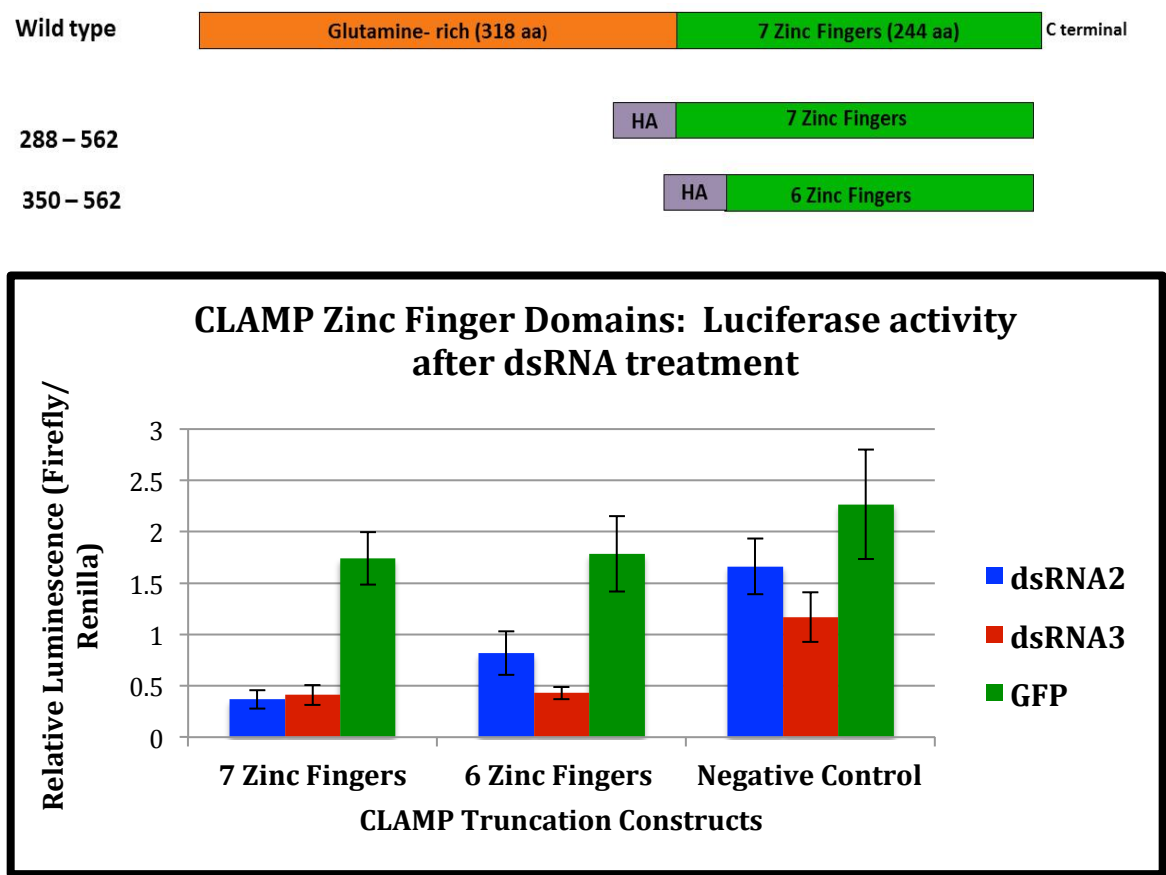


Figure 5C. In the absence of endogenous CLAMP protein, a MSL dependent luciferase reporter shows decreased activity as measured by relative luminescence after transfection with both the 6- and 7-Zinc finger truncations of CLAMP. No loss of activity is observed in the presence of endogenous CLAMP plus these zinc finger truncations. An empty vector without any CLAMP truncations was used as a negative control.

Figure 6. Working Model

Figure 6a. Luciferase reporter used to assay MSL complex recruitment.

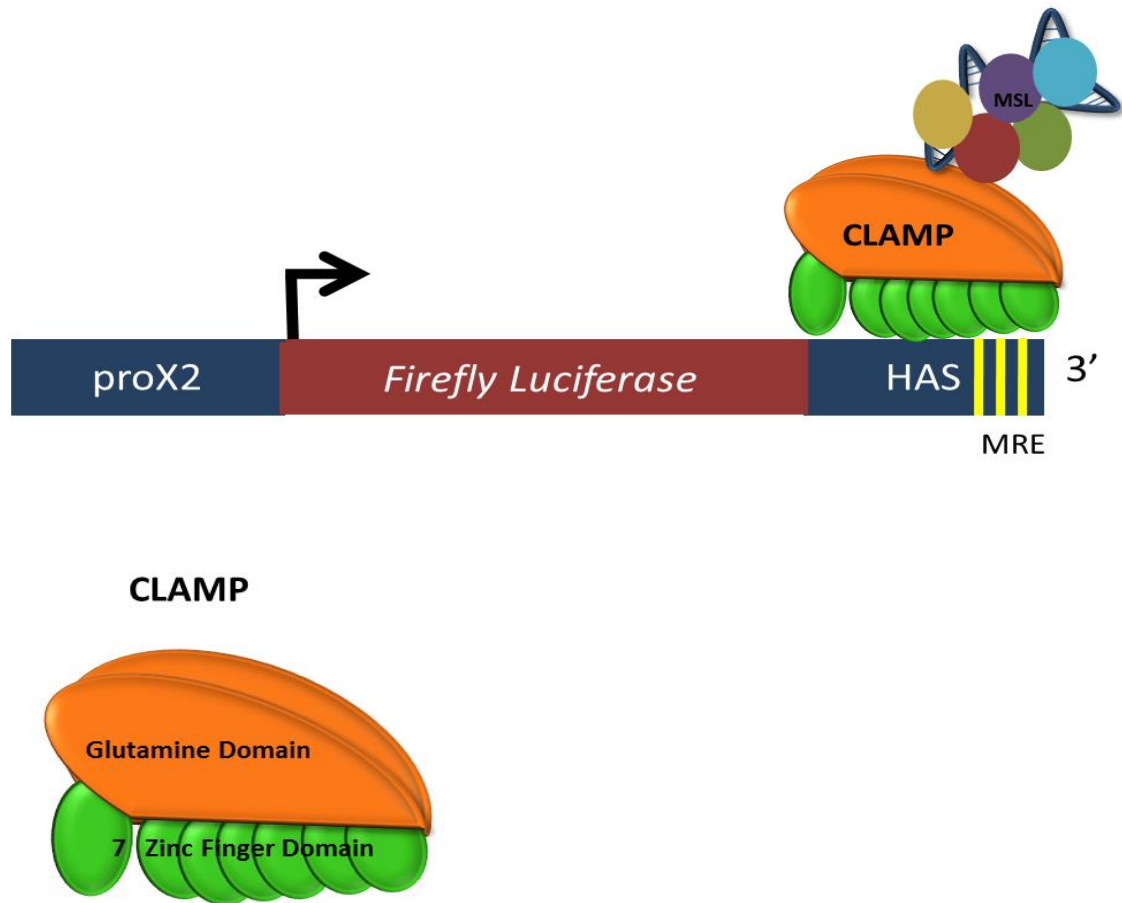


Figure 6a. A MSL dependent luciferase reporter was used to assay MSL complex recruitment activity. MREs located 3' of the luciferase gene recruit MSL complex using CLAMP. MSL complex recruitment activates roX2 promoter leading to luciferase gene expression.

Figure 6. Working Model

Figure 6b. Expression of excess CLAMP zinc finger domain prohibits MSL complex recruitment

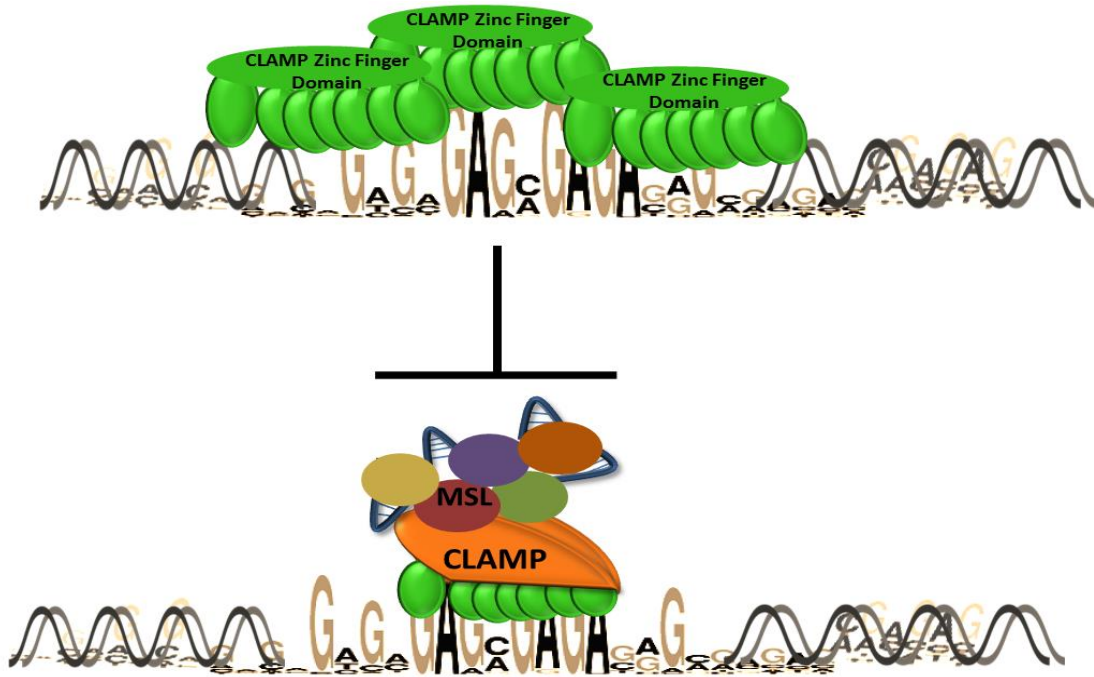


Figure 6b. Expression of excess CLAMP zinc finger domain prohibits MSL complex recruitment by binding to MREs, possibly impeding low levels of endogenous full-length CLAMP from binding to DNA to target MSL complex to MREs.

Figure 6. Working Model

Figure 6c. Excess levels of 161-562 domain of CLAMP sequester MSL complex

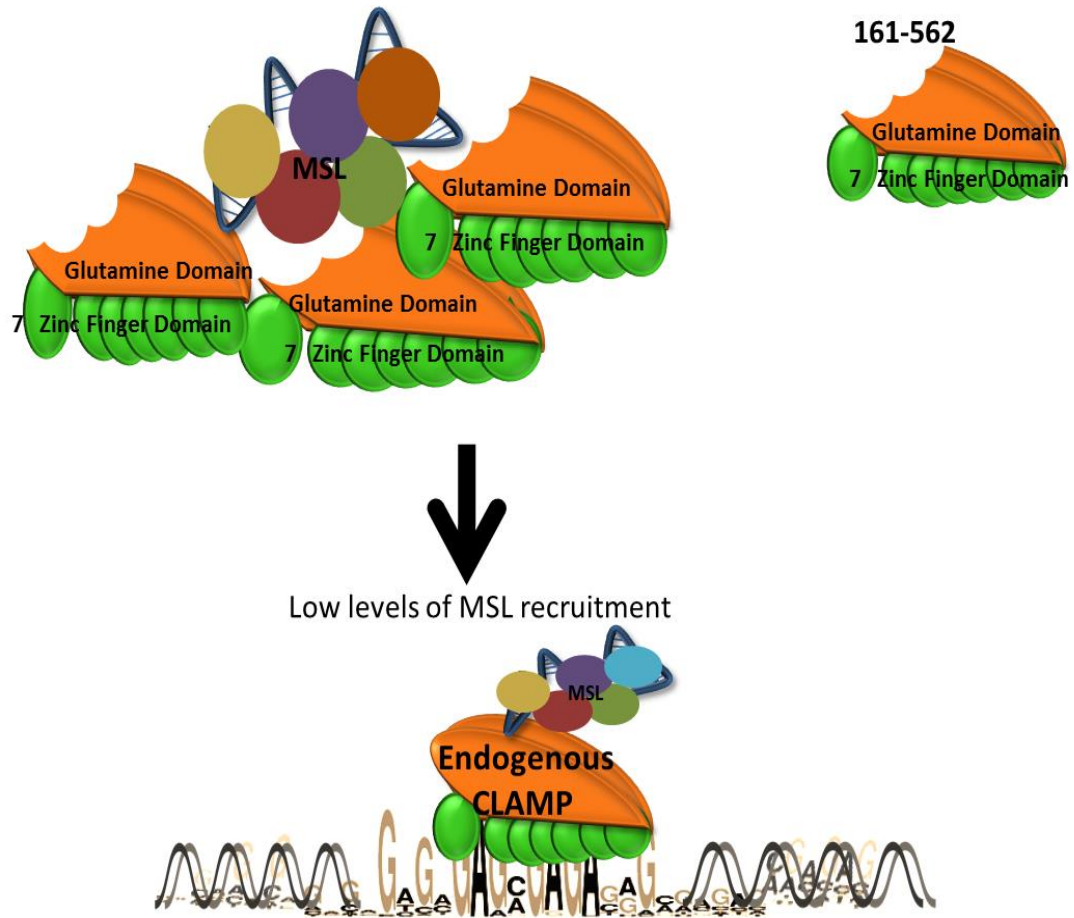


Figure 6c. Excess levels of 161-562 domain of CLAMP may sequester MSL complex, decreasing MSL complex recruitment to MREs.

References

- Adelman K, Wei W, Ardehali MB, Werner J, Zhu B, Reinberg D, Lis JT. 2006. Drosophila Paf1 modulates chromatin structure at actively transcribed genes. *Mol Cell Biol* **26**: 250-260.
- Adkins NL, Hagerman TA, Georgel P. 2006. GAGA protein: a multi-faceted transcription factor. *Biochem Cell Biol* **84**: 559-567.
- Akhtar A, Becker PB. 2000. Activation of transcription through histone H4 acetylation by MOF, an acetyltransferase essential for dosage compensation in Drosophila. *Mol Cell* **5**: 367-375.
- Akhtar A, Zink D, Becker PB. 2000. Chromodomains are protein-RNA interaction modules. *Nature* **407**: 405-409.
- Alekseyenko AA, Ellison CE, Gorchakov AA, Zhou Q, Kaiser VB, Toda N, Walton Z, Peng S, Park PJ, Bachtrog D et al. 2013. Conservation and de novo acquisition of dosage compensation on newly evolved sex chromosomes in Drosophila. *Genes & development* **27**: 853-858.
- Alekseyenko AA HJ, Peng S, Gelbart M, Tolstorukov M, Plachetka A, Kharchenko PV, Jung YL, Gorchakov AA, Larschan E, Gu T, Minoda A, Riddle NC, Schwartz YB, Elgin SCR, Karpen GH, Pirrotta V, Kuroda MI**, Park PJ**. 2012. Sequence-specific targeting of dosage compensation in Drosophila favors an active chromatin context. *In press PLoS Genetics*.
- Alekseyenko AA, Ho JW, Peng S, Gelbart M, Tolstorukov MY, Plachetka A, Kharchenko PV, Jung YL, Gorchakov AA, Larschan E et al. 2012. Sequence-specific targeting of dosage compensation in Drosophila favors an active chromatin context. *PLoS genetics* **8**: e1002646.
- Alekseyenko AA, Larschan E, Lai WR, Park PJ, Kuroda MI. 2006. High-resolution ChIP-chip analysis reveals that the Drosophila MSL complex selectively identifies active genes on the male X chromosome. *Genes & development* **20**: 848-857.
- Alekseyenko AA, Peng S, Larschan E, Gorchakov AA, Lee OK, Kharchenko P, McGrath SD, Wang CI, Mardis ER, Park PJ et al. 2008. A sequence motif within chromatin entry sites directs MSL establishment on the Drosophila X chromosome. *Cell* **134**: 599-609.
- Amrein H, Axel R. 1997. Genes expressed in neurons of adult male Drosophila. *Cell* **88**: 459-469.
- Atanesyan L, Gunther V, Dichtl B, Georgiev O, Schaffner W. 2012. Polyglutamine tracts as modulators of transcriptional activation from yeast to mammals. *Biological chemistry* **393**: 63-70.
- Bai X, Alekseyenko AA, Kuroda MI. 2004. Sequence-specific targeting of MSL complex regulates transcription of the roX RNA genes. *The EMBO journal* **23**: 2853-2861.
- Bai X, Larschan E, Kwon SY, Badenhorst P, Kuroda MI. 2007. Regional control of chromatin organization by noncoding roX RNAs and the NURF remodeling complex in Drosophila melanogaster. *Genetics* **176**: 1491-1499.

- Bailey TL, Boden M, Buske FA, Frith M, Grant CE, Clementi L, Ren J, Li WW, Noble WS. 2009. MEME SUITE: tools for motif discovery and searching. *Nucleic Acids Res* **37**: W202-208.
- Bashaw GJ, Baker BS. 1997. The regulation of the *Drosophila* msl-2 gene reveals a function for Sex-lethal in translational control. *Cell* **89**: 789-798.
- Batista PJ, Chang HY. 2013a. Cytotopic localization by long noncoding RNAs. *Current opinion in cell biology* **25**: 195-199.
- . 2013b. Long Noncoding RNAs: Cellular Address Codes in Development and Disease. *Cell* **152**: 1298-1307.
- Beatriz Vicoso DB. 2009. Progress and prospects toward our understanding of the evolution of dosage compensation.
- Bell O, Conrad T, Kind J, Wirbelauer C, Akhtar A, Schubeler D. 2008. Transcription-coupled methylation of histone H3 at lysine 36 regulates dosage compensation by enhancing recruitment of the MSL complex in *Drosophila melanogaster*. *Molecular and cellular biology* **28**: 3401-3409.
- Belote JM, Lucchesi JC. 1980a. Control of X chromosome transcription by the maleless gene in *Drosophila*. *Nature* **285**: 573-575.
- . 1980b. Male-specific lethal mutations of *Drosophila melanogaster*. *Genetics* **96**: 165-186.
- Berger MF, Bulyk ML. 2009. Universal protein-binding microarrays for the comprehensive characterization of the DNA-binding specificities of transcription factors. *Nature protocols* **4**: 393-411.
- Berger MF, Philippakis AA, Qureshi AM, He FS, Estep PW, 3rd, Bulyk ML. 2006. Compact, universal DNA microarrays to comprehensively determine transcription-factor binding site specificities. *Nat Biotechnol* **24**: 1429-1435.
- Bryson TD, Weber CM, Henikoff S. 2010. Baculovirus-encoded protein expression for epigenomic profiling in *Drosophila* cells. *Fly (Austin)* **4**: 258-265.
- Buscaino A, Legube G, Akhtar A. 2006. X-chromosome targeting and dosage compensation are mediated by distinct domains in MSL-3. *EMBO reports* **7**: 531-538.
- Carrozza MJ, Li B, Florens L, Suganuma T, Swanson SK, Lee KK, Shia WJ, Anderson S, Yates J, Washburn MP et al. 2005. Histone H3 methylation by Set2 directs deacetylation of coding regions by Rpd3S to suppress spurious intragenic transcription. *Cell* **123**: 581-592.
- Chow JC, Ciaudo C, Fazzari MJ, Mise N, Servant N, Glass JL, Attreed M, Avner P, Wutz A, Barillot E et al. 2010. LINE-1 activity in facultative heterochromatin formation during X chromosome inactivation. *Cell* **141**: 956-969.
- Chu C, Qu K, Zhong FL, Artandi SE, Chang HY. 2011. Genomic maps of long noncoding RNA occupancy reveal principles of RNA-chromatin interactions. *Molecular cell* **44**: 667-678.
- Chuang PT, Albertson DG, Meyer BJ. 1994. DPY-27: a chromosome condensation protein homolog that regulates *C. elegans* dosage compensation through association with the X chromosome. *Cell* **79**: 459-474.

- Colgan J, Ashali H, Manley JL. 1995. A direct interaction between a glutamine-rich activator and the N terminus of TFIIB can mediate transcriptional activation in vivo. *Mol Cell Biol* **15**: 2311-2320.
- Conrad T, Cavalli FM, Holz H, Hallaceli E, Kind J, Ilik I, Vaquerizas JM, Luscombe NM, Akhtar A. 2012. The MOF chromobarrel domain controls genome-wide H4K16 acetylation and spreading of the MSL complex. *Dev Cell* **22**: 610-624.
- Copps K, Richman R, Lyman LM, Chang KA, Rampersad-Ammons J, Kuroda MI. 1998. Complex formation by the Drosophila MSL proteins: role of the MSL2 RING finger in protein complex assembly. *Embo J* **17**: 5409-5417.
- Csankovszki G, McDonel P, Meyer BJ. 2004. Recruitment and spreading of the C. elegans dosage compensation complex along X chromosomes. *Science* **303**: 1182-1185.
- Dahlsveen IK, Gilfillan GD, Shelest VI, Lamm R, Becker PB. 2006. Targeting determinants of dosage compensation in Drosophila. *PLoS genetics* **2**: e5.
- Davis TL, Meyer BJ. 1997. SDC-3 coordinates the assembly of a dosage compensation complex on the nematode X chromosome. *Development* **124**: 1019-1031.
- Deng X, Berletch JB, Ma W, Nguyen DK, Hiatt JB, Noble WS, Shendure J, Disteche CM. 2013. Mammalian X upregulation is associated with enhanced transcription initiation, RNA half-life, and MOF-mediated H4K16 acetylation. *Dev Cell* **25**: 55-68.
- Deng X BJ, Ma W, Nguyen DK, Hiatt JB, Noble WS, Shendure J, Disteche CM. 2013. Mammalian X Upregulation Is Associated with Enhanced Transcription Initiation, RNA Half-Life, and MOF-Mediated H4K16 Acetylation. *Developmental Cell* **25**: 1-14.
- Deng X, Hiatt JB, Nguyen DK, Ercan S, Sturgill D, Hillier LW, Schlesinger F, Davis CA, Reinke VJ, Gingeras TR et al. 2011. Evidence for compensatory upregulation of expressed X-linked genes in mammals, Caenorhabditis elegans and Drosophila melanogaster. *Nature genetics* **43**: 1179-1185.
- Deng X, Meller VH. 2006. roX RNAs are required for increased expression of X-linked genes in Drosophila melanogaster males. *Genetics* **174**: 1859-1866.
- Derrien T, Johnson R, Bussotti G, Tanzer A, Djebali S, Tilgner H, Guernec G, Martin D, Merkel A, Knowles DG et al. 2012. The GENCODE v7 catalog of human long noncoding RNAs: analysis of their gene structure, evolution, and expression. *Genome research* **22**: 1775-1789.
- Deuring R, Fanti L, Armstrong JA, Sarte M, Papoulas O, Prestel M, Daubresse G, Verardo M, Moseley SL, Berloco M et al. 2000. The ISWI chromatin-remodeling protein is required for gene expression and the maintenance of higher order chromatin structure in vivo. *Mol Cell* **5**: 355-365.
- Dickinson LA, Joh T, Kohwi Y, Kohwi-Shigematsu T. 1992. A tissue-specific MAR/SAR DNA-binding protein with unusual binding site recognition. *Cell* **70**: 631-645.
- Disteche CM. 2012. Dosage compensation of the sex chromosomes. *Annu Rev Genet* **46**: 537-560.
- Dundr M, Misteli T. 2010. Biogenesis of nuclear bodies. *Cold Spring Harbor perspectives in biology* **2**: a000711.

- Dunlap D, Yokoyama R, Ling H, Sun HY, McGill K, Cugusi S, Lucchesi JC. 2012. Distinct contributions of MSL complex subunits to the transcriptional enhancement responsible for dosage compensation in *Drosophila*. *Nucleic acids research* **40**: 11281-11291.
- Ellison CE, Bachtrog D. 2013. Dosage compensation via transposable element mediated rewiring of a regulatory network. *Science* **342**: 846-850.
- Engreitz JM, Pandya-Jones A, McDonel P, Shishkin A, Sirokman K, Surka C, Kadri S, Xing J, Goren A, Lander ES et al. 2013. The Xist lncRNA exploits three-dimensional genome architecture to spread across the X chromosome. *Science* **341**: 1237973.
- Enuameh MS, Asriyan Y, Richards A, Christensen RG, Hall VL, Kazemian M, Zhu C, Pham H, Cheng Q, Blatti C et al. 2013. Global analysis of *Drosophila* Cys(2)-His(2) zinc finger proteins reveals a multitude of novel recognition motifs and binding determinants. *Genome research* **23**: 928-940.
- Fauth T, Muller-Planitz F, Konig C, Straub T, Becker PB. 2010. The DNA binding CXC domain of MSL2 is required for faithful targeting the Dosage Compensation Complex to the X chromosome. *Nucleic Acids Res* **38**: 3209-3221.
- Fisher WW, Li JJ, Hammonds AS, Brown JB, Pfeiffer BD, Weiszmman R, MacArthur S, Thomas S, Stamatoyannopoulos JA, Eisen MB et al. 2012. DNA regions bound at low occupancy by transcription factors do not drive patterned reporter gene expression in *Drosophila*. *Proceedings of the National Academy of Sciences of the United States of America* **109**: 21330-21335.
- Furuhashi H, Nakajima M, Hirose S. 2006. DNA supercoiling factor contributes to dosage compensation in *Drosophila*. *Development* **133**: 4475-4483.
- Gallach M, Arnau V, Aldecoa R, Marin I. 2010. A sequence motif enriched in regions bound by the *Drosophila* dosage compensation complex. *BMC genomics* **11**: 169.
- Gamsjaeger R, Liew CK, Loughlin FE, Crossley M, Mackay JP. 2007. Sticky fingers: zinc-fingers as protein-recognition motifs. *Trends in biochemical sciences* **32**: 63-70.
- Garriga-Canut M, Agustin-Pavon C, Herrmann F, Sanchez A, Dierssen M, Fillat C, Isalan M. 2012. Synthetic zinc finger repressors reduce mutant huntingtin expression in the brain of R6/2 mice. *Proceedings of the National Academy of Sciences of the United States of America* **109**: E3136-3145.
- Gelbart ME, Kuroda MI. 2009. *Drosophila* dosage compensation: a complex voyage to the X chromosome. *Development* **136**: 1399-1410.
- Gelbart ME, Larschan E, Peng S, Park PJ, Kuroda MI. 2009. *Drosophila* MSL complex globally acetylates H4K16 on the male X chromosome for dosage compensation. *Nature structural & molecular biology* **16**: 825-832.
- Gilfillan GD, Straub T, de Wit E, Greil F, Lamm R, van Steensel B, Becker PB. 2006. Chromosome-wide gene-specific targeting of the *Drosophila* dosage compensation complex. *Genes Dev* **20**: 858-870.
- Girirajan S, Campbell CD, Eichler EE. 2011. Human copy number variation and complex genetic disease. *Annu Rev Genet* **45**: 203-226.

- Gontan C, Achame EM, Demmers J, Barakat TS, Rentmeester E, van IW, Grootegeod JA, Gribnau J. 2012. RNF12 initiates X-chromosome inactivation by targeting REX1 for degradation. *Nature* **485**: 386-390.
- Gu W, Szauter P, Lucchesi JC. 1998a. Targeting of MOF, a putative histone acetyl transferase, to the X chromosome of *Drosophila melanogaster*. *Developmental genetics* **22**: 56-64.
- Gu WG, Szauter P, Lucchesi JC. 1998b. Targeting of MOF, a putative histone acetyl transferase, to the X chromosome of *Drosophila melanogaster*. *Dev Genet* **22**: 56-64.
- Guertin MJ, Lis JT. 2012. Mechanisms by which transcription factors gain access to target sequence elements in chromatin. *Current opinion in genetics & development*.
- Hallacli E, Lipp M, Georgiev P, Spielman C, Cusack S, Akhtar A, Kadlec J. 2012. Msl1-mediated dimerization of the dosage compensation complex is essential for male X-chromosome regulation in *Drosophila*. *Molecular cell* **48**: 587-600.
- Hamada FN, Park PJ, Gordadze PR, Kuroda MI. 2005. Global regulation of X chromosomal genes by the MSL complex in *Drosophila melanogaster*. *Genes Dev* **19**: 2289-2294.
- Hattori M, Fujiyama A, Taylor TD, Watanabe H, Yada T, Park HS, Toyoda A, Ishii K, Totoki Y, Choi DK et al. 2000. The DNA sequence of human chromosome 21. *Nature* **405**: 311-319.
- Heard E, Turner J. 2011. Function of the sex chromosomes in mammalian fertility. *Cold Spring Harbor perspectives in biology* **3**: a002675.
- Henikoff S, Henikoff JG, Sakai A, Loeb GB, Ahmad K. 2009. Genome-wide profiling of salt fractions maps physical properties of chromatin. *Genome Res* **19**: 460-469.
- Henry KW, Wyce A, Lo WS, Duggan LJ, Emre NC, Kao CF, Pillus L, Shilatifard A, Osley MA, Berger SL. 2003. Transcriptional activation via sequential histone H2B ubiquitylation and deubiquitylation, mediated by SAGA-associated Ubp8. *Genes Dev* **17**: 2648-2663.
- Herpin A, Scharf M. 2011. Sex determination: switch and suppress. *Current biology : CB* **21**: R656-659.
- Hilfiker A, Hilfiker-Kleiner D, Pannuti A, Lucchesi JC. 1997. mof, a putative acetyl transferase gene related to the Tip60 and MOZ human genes and to the SAS genes of yeast, is required for dosage compensation in *Drosophila*. *The EMBO journal* **16**: 2054-2060.
- Hofer U, Henley JE, Exline CM, Mulhern O, Lopez E, Cannon PM. 2013. Pre-clinical modeling of CCR5 knockout in human hematopoietic stem cells by zinc finger nucleases using humanized mice. *The Journal of infectious diseases* **208 Suppl 2**: S160-164.
- Hsu DR, Meyer BJ. 1994. The dpy-30 gene encodes an essential component of the *Caenorhabditis elegans* dosage compensation machinery. *Genetics* **137**: 999-1018.
- Ilik IA, Quinn JJ, Georgiev P, Tavares-Cadete F, Maticzka D, Toscano S, Wan Y, Spitale RC, Luscombe N, Backofen R et al. 2013. Tandem stem-loops in roX RNAs act

- together to mediate X chromosome dosage compensation in *Drosophila*. *Molecular cell* **51**: 156-173.
- Jacobs PA, Baikie AG, Court Brown WM, Strong JA. 1959. The somatic chromosomes in mongolism. *Lancet* **1**: 710.
- James E. Darnell J. 2002. TRANSCRIPTION FACTORS AS TARGETS FOR CANCER THERAPY. *Nature Reviews Cancer* **2**.
- Jans J, Gladden JM, Ralston EJ, Pickle CS, Michel AH, Pferdehirt RR, Eisen MB, Meyer BJ. 2009. A condensin-like dosage compensation complex acts at a distance to control expression throughout the genome. *Genes & development* **23**: 602-618.
- Jeon Y, Lee JT. 2011. YY1 tethers Xist RNA to the inactive X nucleation center. *Cell* **146**: 119-133.
- Jin Y, Wang Y, Johansen J, Johansen KM. 2000. JIL-1, a chromosomal kinase implicated in regulation of chromatin structure, associates with the male specific lethal (MSL) dosage compensation complex. *J Cell Biol* **149**: 1005-1010.
- Johnson NA, Lachance J. 2012. The genetics of sex chromosomes: evolution and implications for hybrid incompatibility. *Annals of the New York Academy of Sciences* **1256**: E1-22.
- Kadlec J, Hallacli E, Lipp M, Holz H, Sanchez-Weatherby J, Cusack S, Akhtar A. 2011. Structural basis for MOF and MSL3 recruitment into the dosage compensation complex by MSL1. *Nat Struct Mol Biol* **18**: 142-149.
- Keene JD, Komisarow JM, Friedersdorf MB. 2006. RIP-Chip: the isolation and identification of mRNAs, microRNAs and protein components of ribonucleoprotein complexes from cell extracts. *Nature protocols* **1**: 302-307.
- Kelley RL, Lee OK, Shim YK. 2008. Transcription rate of noncoding roX1 RNA controls local spreading of the *Drosophila* MSL chromatin remodeling complex. *Mechanisms of development* **125**: 1009-1019.
- Kelley RL, Meller VH, Gordadze PR, Roman G, Davis RL, Kuroda MI. 1999. Epigenetic spreading of the *Drosophila* dosage compensation complex from roX RNA genes into flanking chromatin. *Cell* **98**: 513-522.
- Kelley RL, Wang J, Bell L, Kuroda MI. 1997. Sex lethal controls dosage compensation in *Drosophila* by a non-splicing mechanism. *Nature* **387**: 195-199.
- Keogh MC, Kurdistani SK, Morris SA, Ahn SH, Podolny V, Collins SR, Schuldiner M, Chin K, Punna T, Thompson NJ et al. 2005. Cotranscriptional set2 methylation of histone H3 lysine 36 recruits a repressive Rpd3 complex. *Cell* **123**: 593-605.
- Kharchenko PV, Alekseyenko AA, Schwartz YB, Minoda A, Riddle NC, Ernst J, Sabo PJ, Larschan E, Gorchakov AA, Gu T et al. 2011. Comprehensive analysis of the chromatin landscape in *Drosophila melanogaster*. *Nature* **471**: 480-485.
- Kharchenko PV, Tolstorukov MY, Park PJ. 2008. Design and analysis of ChIP-seq experiments for DNA-binding proteins. *Nat Biotechnol* **26**: 1351-1359.
- Kind J, Vaquerizas JM, Gebhardt P, Gentzel M, Luscombe NM, Bertone P, Akhtar A. 2008. Genome-wide analysis reveals MOF as a key regulator of dosage compensation and gene expression in *Drosophila*. *Cell* **133**: 813-828.

- Klug A. 2010. The discovery of zinc fingers and their development for practical applications in gene regulation and genome manipulation. *Q Rev Biophys* **43**: 1-21.
- Klug A, Schwabe JW. 1995. Protein motifs 5. Zinc fingers. *FASEB journal : official publication of the Federation of American Societies for Experimental Biology* **9**: 597-604.
- Ko-Hsuan Hung YW, Jin Crystal Zhao. 2013. Regulation of Mammalian Gene Dosage by Long Non-coding RNAs. *Biomolecules* **3**(1): 124-142.
- Kulkarni MM, Booker M, Silver SJ, Friedman A, Hong P, Perrimon N, Mathey-Prevot B. 2006. Evidence of off-target effects associated with long dsRNAs in *Drosophila melanogaster* cell-based assays. *Nat Methods* **3**: 833-838.
- Lai F, Orom UA, Cesaroni M, Beringer M, Taatjes DJ, Blobel GA, Shiekhattar R. 2013. Activating RNAs associate with Mediator to enhance chromatin architecture and transcription. *Nature* **494**: 497-501.
- Langmead B, Trapnell C, Pop M, Salzberg SL. 2009. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol* **10**: R25.
- Larschan E, Alekseyenko AA, Gortchakov AA, Peng S, Li B, Yang P, Workman JL, Park PJ, Kuroda MI. 2007. MSL complex is attracted to genes marked by H3K36 trimethylation using a sequence-independent mechanism. *Mol Cell* **28**: 121-133.
- Larschan E, Alekseyenko AA, Lai WR, Park PJ, Kuroda MI. 2006. MSL complex associates with clusters of actively transcribed genes along the *Drosophila* male X chromosome. *Cold Spring Harbor symposia on quantitative biology* **71**: 385-394.
- Larschan E, Bishop E, Kharchenko P, Park P, Kuroda M. 2011a. X chromosome dosage compensation via enhanced transcription elongation in *Drosophila* males. *Nature* **In press**.
- Larschan E, Bishop EP, Kharchenko PV, Core LJ, Lis JT, Park PJ, Kuroda MI. 2011b. X chromosome dosage compensation via enhanced transcriptional elongation in *Drosophila*. *Nature* **471**: 115-118.
- Larschan E, Soruco MM, Lee OK, Peng S, Bishop E, Chery J, Goebel K, Feng J, Park PJ, Kuroda MI. 2012. Identification of chromatin-associated regulators of MSL complex targeting in *Drosophila* dosage compensation. *PLoS Genet* **8**: e1002830.
- Lee C, Li X, Hechmer A, Eisen M, Biggin MD, Venters BJ, Jiang C, Li J, Pugh BF, Gilmour DS. 2008. NELF and GAGA factor are linked to promoter-proximal pausing at many genes in *Drosophila*. *Mol Cell Biol* **28**: 3290-3300.
- Lee JT. 2012. Epigenetic regulation by long noncoding RNAs. *Science* **338**: 1435-1439.
- Lee JT, Bartolomei MS. 2013. X-Inactivation, Imprinting, and Long Noncoding RNAs in Health and Disease. *Cell* **152**: 1308-1323.
- Lejeune J, Gautier M, Turpin R. 1959. [Study of somatic chromosomes from 9 mongoloid children]. *Comptes rendus hebdomadaires des seances de l'Academie des sciences* **248**: 1721-1722.

- Leppig KA, Disteché CM. 2001. Ring X and other structural X chromosome abnormalities: X inactivation and phenotype. *Seminars in reproductive medicine* **19**: 147-157.
- Li F, Parry DA, Scott MJ. 2005. The amino-terminal region of Drosophila MSL1 contains basic, glycine-rich, and leucine zipper-like motifs that promote X chromosome binding, self-association, and MSL2 binding, respectively. *Molecular and cellular biology* **25**: 8913-8924.
- Li F, Schiemann AH, Scott MJ. 2008. Incorporation of the noncoding roX RNAs alters the chromatin-binding specificity of the Drosophila MSL1/MSL2 complex. *Molecular and cellular biology* **28**: 1252-1264.
- Lim CK, Kelley RL. 2012. Autoregulation of the Drosophila Noncoding roX1 RNA Gene. *PLoS Genet* **8**: e1002564.
- Lipeng Wu BMZ, Yanming Wang, Benjamin A. Garcia, Yali Dou. 2011. The RING Finger Protein MSL2 in the MOF Complex Is an E3 Ubiquitin Ligase for H2B K34 and Is Involved in Crosstalk with H3 K4 and K79 Methylation. *Molecular cell* **43**: 132-144.
- Little M, Holmes G, Walsh P. 1999. WT1: what has the last decade told us? *BioEssays : news and reviews in molecular, cellular and developmental biology* **21**: 191-202.
- Lours C, Bardot O, Godt D, Laski FA, Couderc JL. 2003. The Drosophila melanogaster BTB proteins bric a brac bind DNA through a composite DNA binding domain containing a pipsqueak and an AT-Hook motif. *Nucleic acids research* **31**: 5389-5398.
- Lucchesi JC. 1998. Dosage compensation in flies and worms: the ups and downs of X-chromosome regulation. *Current opinion in genetics & development* **8**: 179-184.
- Lucchesi JC, Kelly WG, Panning B. 2005. Chromatin remodeling in dosage compensation. *Annu Rev Genet* **39**: 615-651.
- Lucchesi JMBJC. 1980. Control of X Chromosome transcription by the maleless gene in *Drosophila*. *Nature* **285**: 573-575.
- Lusk RW, Eisen MB. 2013. Spatial promoter recognition signatures may enhance transcription factor specificity in yeast. *PloS one* **8**: e53778.
- Lyman LM, Copps K, Rastelli L, Kelley RL, Kuroda MI. 1997. Drosophila male-specific lethal-2 protein: Structure/function analysis and dependence on MSL-1 for chromosome association. *Genetics* **147**: 1743-1753.
- Maenner S, Muller M, Frohlich J, Langer D, Becker PB. 2013. ATP-dependent roX RNA remodeling by the helicase maleless enables specific association of MSL proteins. *Molecular cell* **51**: 174-184.
- Mank JE. 2009. The W, X, Y and Z of sex-chromosome dosage compensation. *Trends in genetics : TIG* **25**: 226-233.
- McDonel P, Jans J, Peterson BK, Meyer BJ. 2006. Clustered DNA motifs mark X chromosomes for repression by a dosage compensation complex. *Nature* **444**: 614-618.
- Meller VH. 2003. Initiation of dosage compensation in Drosophila embryos depends on expression of the roX RNAs. *Mechanisms of development* **120**: 759-767.

- Meller VH, Kuroda MI. 2002. Sex and the single chromosome. *Advances in genetics* **46**: 1-24.
- Meller VH, Rattner BP. 2002a. The roX genes encode redundant male-specific lethal transcripts required for targeting of the MSL complex. *Embo J* **21**: 1084-1091.
- . 2002b. The roX genes encode redundant male-specific lethal transcripts required for targeting of the MSL complex. *The EMBO journal* **21**: 1084-1091.
- Meller VH, Wu KH, Roman G, Kuroda MI, Davis RL. 1997. roX1 RNA paints the X chromosome of male *Drosophila* and is regulated by the dosage compensation system. *Cell* **88**: 445-457.
- Mendjan S, Taipale M, Kind J, Holz H, Gebhardt P, Schelder M, Vermeulen M, Buscaino A, Duncan K, Mueller J et al. 2006. Nuclear pore components are involved in the transcriptional regulation of dosage compensation in *Drosophila*. *Molecular cell* **21**: 811-823.
- Meyer BJ. 1997. Sex Determination and X Chromosome Dosage Compensation.
- . 2010. Targeting X chromosomes for repression. *Current opinion in genetics & development* **20**: 179-189.
- Mito Y, Henikoff JG, Henikoff S. 2007. Histone replacement marks the boundaries of cis-regulatory domains. *Science* **315**: 1408-1411.
- Morra R, Smith ER, Yokoyama R, Lucchesi JC. 2008. The MLE subunit of the *Drosophila* MSL complex uses its ATPase activity for dosage compensation and its helicase activity for targeting. *Molecular and cellular biology* **28**: 958-966.
- Nagaike T, Logan C, Hotta I, Rozenblatt-Rosen O, Meyerson M, Manley JL. 2011. Transcriptional activators enhance polyadenylation of mRNA precursors. *Mol Cell* **41**: 409-418.
- Nguyen DK, Disteche CM. 2006. Dosage compensation of the active X chromosome in mammals. *Nat Genet* **38**: 47-53.
- Nikiforov MA, Chandriani S, Park J, Kotenko I, Matheos D, Johnsson A, McMahon SB, Cole MD. 2002. TRRAP-dependent and TRRAP-independent transcriptional activation by Myc family oncoproteins. *Mol Cell Biol* **22**: 5054-5063.
- Nora EP, Lajoie BR, Schulz EG, Giorgetti L, Okamoto I, Servant N, Piolot T, van Berkum NL, Meisig J, Sedat J et al. 2012. Spatial partitioning of the regulatory landscape of the X-inactivation centre. *Nature* **485**: 381-385.
- Nybakken K, Perrimon N. 2002. Hedgehog signal transduction: recent findings. *Curr Opin Genet Dev* **12**: 503-511.
- Oh H, Park Y, Kuroda MI. 2003. Local spreading of MSL complexes from roX genes on the *Drosophila* X chromosome. *Genes & development* **17**: 1334-1339.
- Onodera O, Roses AD, Tsuji S, Vance JM, Strittmatter WJ, Burke JR. 1996. Toxicity of expanded polyglutamine-domain proteins in *Escherichia coli*. *FEBS letters* **399**: 135-139.
- Orom UA, Derrien T, Beringer M, Gumireddy K, Gardini A, Bussotti G, Lai F, Zytnicki M, Notredame C, Huang Q et al. 2010. Long noncoding RNAs with enhancer-like function in human cells. *Cell* **143**: 46-58.

- Palmer MJ, Richman R, Richter L, Kuroda MI. 1994. Sex-Specific Regulation of the Male-Specific Lethal-1 Dosage Compensation Gene in *Drosophila*. *Gene Dev* **8**: 698-706.
- Paris M, Kaplan T, Li XY, Villalta JE, Lott SE, Eisen MB. 2013. Extensive divergence of transcription factor binding in *Drosophila* embryos with highly conserved gene expression. *PLoS genetics* **9**: e1003748.
- Park SW, Oh H, Lin YR, Park Y. 2010. MSL cis-spreading from roX gene up-regulates the neighboring genes. *Biochemical and biophysical research communications* **399**: 227-231.
- Park Y, Kelley RL, Oh H, Kuroda MI, Meller VH. 2002. Extent of chromatin spreading determined by roX RNA recruitment of MSL proteins. *Science* **298**: 1620-1623.
- Park Y, Mengus G, Bai X, Kageyama Y, Meller VH, Becker PB, Kuroda MI. 2003. Sequence-specific targeting of *Drosophila* roX genes by the MSL dosage compensation complex. *Molecular cell* **11**: 977-986.
- Payer B, Lee JT. 2008. X chromosome dosage compensation: how mammals keep the balance. *Annu Rev Genet* **42**: 733-772.
- Pelassa I, Cora D, Cesano F, Monje FJ, Montarolo PG, Fiumara F. 2014. Association of polyalanine and polyglutamine coiled coils mediates expansion disease-related protein aggregation and dysfunction. *Hum Mol Genet*.
- Pferdehirt RR, Kruesi WS, Meyer BJ. 2011. An MLL/COMPASS subunit functions in the *C. elegans* dosage compensation complex to target X chromosomes for transcriptional regulation of gene expression. *Genes & development* **25**: 499-515.
- Pinter SF, Sadreyev RI, Yildirim E, Jeon Y, Ohsumi TK, Borowsky M, Lee JT. 2012. Spreading of X chromosome inactivation via a hierarchy of defined Polycomb stations. *Genome research* **22**: 1864-1876.
- Prestel M, Feller C, Becker PB. 2010a. Dosage compensation and the global re-balancing of aneuploid genomes. *Genome biology* **11**: 216.
- Prestel M, Feller C, Straub T, Mitlohner H, Becker PB. 2010b. The activation potential of MOF is constrained for dosage compensation. *Molecular cell* **38**: 815-826.
- Raja SJ, Charapitsa I, Conrad T, Vaquerizas JM, Gebhardt P, Holz H, Kadlec J, Fraterman S, Luscombe NM, Akhtar A. 2010. The nonspecific lethal complex is a transcriptional regulator in *Drosophila*. *Mol Cell* **38**: 827-841.
- Ramadan N, Flockhart I, Booker M, Perrimon N, Mathey-Prevot B. 2007. Design and implementation of high-throughput RNAi screens in cultured *Drosophila* cells. *Nat Protoc* **2**: 2245-2264.
- Rinn JL, Chang HY. 2012. Genome regulation by long noncoding RNAs. *Annual review of biochemistry* **81**: 145-166.
- Robasky K, Bulyk ML. 2011. UniPROBE, update 2011: expanded content and search tools in the online database of protein-binding microarray data on protein-DNA interactions. *Nucleic Acids Res* **39**: D124-128.
- Salz HK, Erickson JW. 2010. Sex determination in *Drosophila*: The view from the top. *Fly (Austin)* **4**: 60-70.

- Schulz EG, Heard E. 2013. Role and control of X chromosome dosage in mammalian development. *Current opinion in genetics & development*.
- Sekido R, Lovell-Badge R. 2009. Sex determination and SRY: down to a wink and a nudge? *Trends in genetics : TIG* **25**: 19-29.
- Shi X, Finkelstein A, Wolf AJ, Wade PA, Burton ZF, Jaehning JA. 1996. Paf1p, an RNA polymerase II-associated factor in *Saccharomyces cerevisiae*, may have both positive and negative roles in transcription. *Mol Cell Biol* **16**: 669-676.
- Shiro Iuchi NK. 2005. *Zinc Finger Proteins: From Atomic Contact to Cellular Function*. Landes Bioscience / Eureka.com, Kluwer Academic / Plenum Publishers 233 Spring Street New York, New York U.S.A. 100013.
- Shogren-Knaak M, Ishii H, Sun JM, Pazin MJ, Davie JR, Peterson CL. 2006. Histone H4-K16 acetylation controls chromatin structure and protein interactions. *Science* **311**: 844-847.
- Shogren-Knaak M, Peterson CL. 2006. Switching on chromatin: mechanistic role of histone H4-K16 acetylation. *Cell Cycle* **5**: 1361-1365.
- Simon MD, Pinter SF, Fang R, Sarma K, Rutenberg-Schoenberg M, Bowman SK, Kesner BA, Maier VK, Kingston RE, Lee JT. 2013. High-resolution Xist binding maps reveal two-step spreading during X-chromosome inactivation. *Nature*.
- Smith ER, Allis CD, Lucchesi JC. 2001. Linking global histone acetylation to the transcription enhancement of X-chromosomal genes in *Drosophila* males. *J Biol Chem* **276**: 31483-31486.
- Smith ER, Pannuti A, Gu W, Steurnagel A, Cook RG, Allis CD, Lucchesi JC. 2000. The *drosophila* MSL complex acetylates histone H4 at lysine 16, a chromatin modification linked to dosage compensation. *Molecular and cellular biology* **20**: 312-318.
- Soruco MM, Chery J, Bishop EP, Siggers T, Tolstorukov MY, Leydon AR, Sugden AU, Goebel K, Feng J, Xia P et al. 2013. The CLAMP protein links the MSL complex to the X chromosome during *Drosophila* dosage compensation. *Genes & development* **27**: 1551-1556.
- Spierer A, Seum C, Delattre M, Spierer P. 2005. Loss of the modifiers of variegation Su(var)3-7 or HP1 impacts male X polytene chromosome morphology and dosage compensation. *J Cell Sci* **118**: 5047-5057.
- Straub T, Grimaud C, Gilfillan GD, Mitterweger A, Becker PB. 2008. The chromosomal high-affinity binding sites for the *Drosophila* dosage compensation complex. *PLoS genetics* **4**: e1000302.
- Sunde JMMaM. 2002. Zinc Fingers--Folds for Many Occasions. *IUBMB Life* **54**: 351-355.
- Sural TH, Peng S, Li B, Workman JL, Park PJ, Kuroda MI. 2008. The MSL3 chromodomain directs a key targeting step for dosage compensation of the *Drosophila melanogaster* X chromosome. *Nature structural & molecular biology* **15**: 1318-1325.
- Tanny JC, Erdjument-Bromage H, Tempst P, Allis CD. 2007. Ubiquitylation of histone H2B controls RNA polymerase II transcription elongation independently of histone H3 methylation. *Genes Dev* **21**: 835-847.

- Thomas CJ, Kotova E, Andrade M, Adolf-Bryfogle J, Glaser R, Regnard C, Tulin AV. 2014. Kinase-Mediated Changes in Nucleosome Conformation Trigger Chromatin Decondensation via Poly(ADP-Ribosylation). *Molecular cell* **53**: 831-842.
- Thomas Cremer MC. 2010. Chromosome Territories. *Cold Spring Harbor perspectives in biology*.
- Thomas S, Li XY, Sabo PJ, Sandstrom R, Thurman RE, Canfield TK, Giste E, Fisher W, Hammonds A, Celniker SE et al. 2011. Dynamic reprogramming of chromatin accessibility during Drosophila embryo development. *Genome biology* **12**: R43.
- Vaquerizas JM, Kummerfeld SK, Teichmann SA, Luscombe NM. 2009. A census of human transcription factors: function, expression and evolution. *Nature reviews Genetics* **10**: 252-263.
- Vaquero A, Blanch M, Espinas ML, Bernues J. 2008a. Activation properties of GAGA transcription factor. *Biochimica et biophysica acta* **1779**: 312-317.
- . 2008b. Activation properties of GAGA transcription factor. *Bba-Gene Regul Mech* **1779**: 312-317.
- Villa R, Forne I, Muller M, Imhof A, Straub T, Becker PB. 2012. MSL2 combines sensor and effector functions in homeostatic control of the Drosophila dosage compensation machinery. *Molecular cell* **48**: 647-654.
- Voss TC, Hager GL. 2014. Dynamic regulation of transcriptional states by chromatin and transcription factors. *Nature reviews Genetics* **15**: 69-81.
- Wang CI, Alekseyenko AA, LeRoy G, Elia AE, Gorchakov AA, Britton LM, Elledge SJ, Kharchenko PV, Garcia BA, Kuroda MI. 2013a. Chromatin proteins captured by ChIP-mass spectrometry are linked to dosage compensation in Drosophila. *Nature structural & molecular biology* **20**: 202-209.
- . 2013b. Chromatin proteins captured by ChIP-mass spectrometry are linked to dosage compensation in Drosophila. *Nat Struct Mol Biol*.
- Wells MB, Snyder MJ, Custer LM, Csankovszki G. 2012. Caenorhabditis elegans dosage compensation regulates histone H4 chromatin state on X chromosomes. *Molecular and cellular biology* **32**: 1710-1719.
- Worringer KA, Panning B. 2007. Zinc finger protein Zn72D promotes productive splicing of the maleless transcript. *Mol Cell Biol* **27**: 8760-8769.
- Wu L, Zee BM, Wang Y, Garcia BA, Dou Y. 2011. The RING Finger Protein MSL2 in the MOF Complex Is an E3 Ubiquitin Ligase for H2B K34 and Is Involved in Crosstalk with H3 K4 and K79 Methylation. *Mol Cell* **43**: 132-144.
- Xiao-Yong Li ST, Peter J Sabo, Michael B Eisen, John A Stamatoyannopoulos and Mark D Biggin. 2011. The role of chromatin accessibility in directing the widespread, overlapping patterns of Drosophila transcription factor binding. *Genome biology* **12**.
- Yonker SA, Meyer BJ. 2003. Recruitment of C. elegans dosage compensation proteins for gene-specific versus chromosome-wide repression. *Development* **130**: 6519-6532.
- Zhao J, Sun BK, Erwin JA, Song JJ, Lee JT. 2008. Polycomb proteins targeted by a short repeat RNA to the mouse X chromosome. *Science* **322**: 750-756.

- Zhou K, Kuo WH, Fillingham J, Greenblatt JF. 2009. Control of transcriptional elongation and cotranscriptional histone modification by the yeast BUR kinase substrate Spt5. *Proc Natl Acad Sci U S A* **106**: 6956-6961.
- Zhou T, Yang L, Lu Y, Dror I, Dantas Machado AC, Ghane T, Di Felice R, Rohs R. 2013. DNASHape: a method for the high-throughput prediction of DNA structural features on a genomic scale. *Nucleic acids research* **41**: W56-62.

Appendix 2: CLAMP exhibits an *in vivo* interaction with the *roX2* RNA component of MSL complex

Jessica Chery

Appendix 2: CLAMP exhibits an *in vivo* interaction with the *roX2* RNA component of MSL complex

Jessica Chery

Introduction

The C₂H₂ or Cys₂His₂ zinc finger is comprised of a zinc ion ligated to a pair of cysteine and histidine residues confined within a hydrophobic core usually formed by amino acids Tyr6, Phe17, and Leu23 (Shiro Iuchi 2005). This structure is typically composed of twenty to thirty amino acids. C₂H₂ zinc fingers can bind specific DNA sequences via a Beta-Beta sheet-alpha (ββ α) helix structure formed by approximately thirty amino acid residues. This DNA binding specificity is achieved by contact between the major groove of DNA and the amino acids present on the α helix of the zinc finger and wrapping of the ββ α structure around the strands of DNA. DNA binding specificity and affinity is also strengthened by a well-conserved linker, TGEKP, between adjoining zinc fingers. Each finger can recognize 3bp of DNA. Having triple or multiple C₂H₂ fingers adjacently or separated in pairs was demonstrated to be able to bind to DNA using only three consecutive fingers without the need for other domains (Klug and Schwabe 1995; Shiro Iuchi 2005). These findings agree with our previously published results where we observed that both the 4- and 6- C-terminal C₂H₂ fingers of our seven C₂H₂ protein CLAMP were able to directly bind identical 8bp motifs (Soruco et al. 2013).

In addition to binding DNA, C₂H₂ zinc fingers can also bind RNA via amino acids of the α helix and multimerize with themselves thereby facilitating protein-protein

interactions. For example, in *Xenopus laevis* the 9 zinc finger transcription factor TFIIA binds both ribosomal RNA (rRNA) and DNA using distinct zinc fingers and structures (Shiro Iuchi 2005). Another classical zinc finger protein with dual DNA/RNA binding capacity is the Wilm's tumour protein 1 (WT1), which functions in transactivation and repression (Little et al. 1999; Sunde 2002). WT1 mutations and deletions are thought to be causal of an aggressive type of kidney tumor in children and also observed in acute myeloid leukemia (Little et al. 1999; Sunde 2002).

While multiple zinc fingers are often involved in conferring DNA sequence binding specificity, zinc finger proteins can interact with other proteins via single or multiple zinc finger domains. Two such examples include the Ikaros transcription factors and FOG-1 (Friend of GATA). The FOG-1 transcription factor contains nine zinc fingers, five of which can interact with proteins (Gamsjaeger et al. 2007). Of these five zinc fingers, four can interact with zinc fingers of the GATA protein independently of each other (Sunde 2002; Gamsjaeger et al. 2007). Ikaros contains two zinc finger domains: four zinc fingers at the N-terminus bind DNA and the two zinc fingers at the C-terminus are involved in protein dimerization (Sunde 2002).

Another transcription factor with diverse functions is the zinc finger protein Yin Yang 1 (YY1) that is highly conserved from *Xenopus* to mammals (Shiro Iuchi 2005). YY1 binds DNA and can interact directly with Histone Acetyltransferases (HATs) and Histone Deacetyltransferases (HDACs). Histone Acetylases that interact with

YY1 directly regulate the transcriptional activities of YY1 (Shiro Iuchi 2005). In the mammalian dosage compensation system, the zinc finger transcription factor YY1 binds DNA and RNA through different motifs (Jeon and Lee 2011).

In *Drosophila* dosage compensation, we determined that the CLAMP C₂H₂ zinc finger protein is required for MSL complex recruitment and directly binds to MREs (Soruco et al. 2013). We therefore investigated whether analogously to the YY1 zinc finger protein, the zinc finger protein CLAMP could associate with RNAs in addition to its previously defined function as a DNA binding transcription factor. Specifically, we studied the ability of CLAMP to interact with the *roX* RNAs of MSL complex. The *roX* RNAs are required for targeting of the MSL complex to the X-chromosome, and the RNA helicase MLE that directly interacts with the *roX* RNAs *in vitro* incorporates the *roX* RNAs into the MSL complex (Maenner et al. 2013).

Recent studies revealed that CLAMP and MSL complex interact (Wang et al. 2013a). We therefore also investigated whether CLAMP could interact with MSL complex via protein-protein interactions. The MSL2-CXC domain was chosen as a first point of investigation because it was previously demonstrated to be the principal domain of MSL2 required for faithful targeting of MSL complex to CES *in vivo* (Fauth et al. 2010). However, the MSL2-CXC domain itself possesses no DNA sequence specific binding *in vitro* (Fauth et al. 2010). Therefore, we hypothesized that MSL2-CXC interacts with CLAMP during recruitment to the X-chromosome.

Results

RNA Immunoprecipitation

Drosophila male cells (SL2) were harvested and antibodies to CLAMP, MLE, and H3K9me3 were used for immunoprecipitation. MLE RNA immunoprecipitation (RIP) was used as a positive control because the MLE protein is a known RNA/DNA helicase and has been demonstrated to interact with the *roX1* and *roX2* RNAs (Maenner et al. 2013) (Ilik et al. 2013). MLE RIP showed high levels of *roX2* IP (Fig 1). CLAMP RIP also showed significant enrichment of *roX2* RNAs (Fig 1). This interaction was specific to the *roX2* RNA because PKA (Protein Kinase A) RIP with CLAMP antibody showed very low levels of enrichment compared to *roX2*. To ensure our assay was capturing specific *roX2* RNA interactions, we also checked H3K9me3 and a no antibody IP. Both these negative controls showed very little to no enrichment for *roX2* and PKA RNA, indicating that the high levels of *roX2* RNA observed in the CLAMP RNA IP was specific to CLAMP's *in vivo* interaction with *roX2* RNA (Fig 1). All IPs included a DNase treatment step, ensuring that the RNA interactions we captured in our assay were due to RNA-protein interactions and not mediated by DNA. In addition, RNase treatment of all samples showed no enrichment for any RNAs, further supporting that the interactions we observed via qPCR assessment of fold RNA enrichment were in fact due to RNA-protein interactions (data not shown). *Rox1* RNA was not assessed for CLAMP interaction because *rox1* is not expressed in *Drosophila* male (SL2) cells possibly due to *rox1* rearrangements.

Defining interactions between CLAMP and MSL complex components

Using male *Drosophila* cells, we tested whether the interaction we observed between CLAMP and *roX2* RNAs was a consequence of interactions between MLE and CLAMP. We performed co-immunoprecipitation with MLE and CLAMP. CLAMP antibody was used to immunoprecipitate material from the cells and a western blot was performed using MLE antibody. MLE antibody was also used to immunoprecipitate material from the cells and CLAMP antibody used to probe the western blot to determine if CLAMP was present. Co-immunoprecipitation was performed using four different conditions: crosslinking with 5% formaldehyde, native (non-crosslinked), +/- RNase treatment, and with CLAMP coupled antibody beads. MSL1, MSL2, GAF, and H3K9me3 IPs were also performed in tandem (Fig 2). MSL1 and MSL2 IPs appeared either overloaded or exhibited too strong signal, preventing any interaction from being detected (Fig 2). CLAMP appeared to be present in the GAF and H3K9me3 IPs, and MLE and CLAMP appeared to interact both in the presence and absence of RNase treatment (Fig 2C, 2D). When CLAMP antibody was coupled to beads or cells were crosslinked with formaldehyde and immunoprecipitation performed, no MLE and CLAMP interaction was detected (Fig 2A, 2B). Using native (non-crosslinking) conditions MLE was not observed in the CLAMP IPs and neither was CLAMP observed in the MLE IPs (Fig 2E). These results indicate that CLAMP and MLE may not have a direct protein-protein interaction *in vivo*. However, further studies will be needed to fully explore or detect the precise

conditions under which MLE and CLAMP may or may not interact as the findings presented here are inconclusive.

CLAMP and MSL complex super-shift EMSAs

The C-terminal 6 C₂H₂ zinc finger domain of CLAMP was expressed and fused to MBP (Maltose Binding Tag) using standard cloning procedures. This MBP-CLAMP was expressed and purified from *E.coli* using a maltose column. The CXC domain of MSL2 was expressed fused to a Glutathione S-transferase (GST) tag and purified from *E.coli*. Nuclear extraction of male (SL2) and female (KC) *Drosophila* cells was performed simultaneously. Because MSL complex is only expressed in males, the female cells served as a comparison for effects specific to MSL complex. Purified MBP-CLAMP was then incubated with a) nuclear extracts containing MSL complex: males, b) nuclear extracts without MSL complex: females, c) GST-MSL2-CXC in three separate samples. Biotin-labelled probe was added to each sample. The probe used was previously demonstrated to be directly bound by CLAMP *in vitro* using both Protein Binding Microarray experiments and Electrophoretic Mobility Shift Assays (EMSAs). In addition this probe sequence is an endogenous genomic sequence bound by CLAMP *in vivo*. MBP-CLAMP + MSL complex extract demonstrated a super-shift (Fig 3) relative to the normal shifting band observed when only MBP-CLAMP and DNA are bound without other proteins present. This supershift was not observed in extract that did not contain MSL complex or MSL2-CXC domain.

Discussion

CLAMP interacts with *roX2*

Our data provide the first evidence supporting a model in which CLAMP functions as a key initial regulator of sub-nuclear domain formation via dual specificity for non-coding RNA and DNA motifs. Using RNA immunoprecipitation, we determined that CLAMP has an *in vivo* association with the non-coding *roX* (*RNA on X*) of MSL complex. This association was specific to *roX2* as other RNAs (PKA) examined showed little enrichment after CLAMP antibody pull-down (Fig 1). Because this is an *in vivo* association, it is possible that this is not a direct interaction but mediated by another factor of the MSL complex, such as MLE, the previously characterized RNA/DNA helicase of MSL complex.

CLAMP co-immunoprecipitation with MSL complex components

We performed Co-immunoprecipitation of MLE with CLAMP. Using native and non-native immunoprecipitation, we did not observe a reproducible interaction between MLE and CLAMP, indicating that CLAMP's interaction with the *roX2* RNA may be direct or mediated by a yet unidentified factor (Fig 2).

Each zinc finger can recognize 3 bp of DNA. Therefore only 3 zinc fingers would theoretically be necessary to bind a 8bp sequence. We have observed a ~15bp motif *in vivo* at full-length CLAMP binding sites giving further support to the idea that not all seven zinc fingers of CLAMP are used for DNA binding. A seven-zinc finger motif

is predicted to be 21bp, not 15bp, if every zinc finger is binding the DNA motif. MSL complex components are required for stability of roX RNAs, and tandem stem loops of the roX RNAs are required for appropriate targeting of MSL complex (Ilik et al. 2013). Due to inability to ensure the accurate co-transcriptional folding and stability of the roX RNAs, gel-shift experiments could not be used to determine if there was a direct interaction between CLAMP and roX RNAs.

The CLAMP zinc fingers may form a complex with both MSL complex and DNA

Using MBP-CLAMP EMSA experiments +/- nuclear extracts containing MSL complex and +/- MS2-CXC domain, our data indicate that it may be possible to map the interaction between MSL complex and CLAMP using this approach. Nuclear extracts containing MSL complex were super-shifted in the presence of purified MBP-CLAMP 6-Zinc finger and a CLAMP DNA binding sequence (Fig 3). Such binding was not observed in nuclear extracts without MSL complex (Fig 3). These results suggest interactions between some CLAMP zinc fingers and MSL complex.

The nuclear extract containing MSL complex also contained endogenous CLAMP, therefore it is also possible that the observed MSL complex and CLAMP binding was due to MBP-CLAMP interaction with endogenous CLAMP. However, when we incubated cell extracts from female cells (KC) in which MSL complex is not expressed but full-length endogenous CLAMP is present, we did not observe a super-shifting with MBP-CLAMP (Fig 3). This result disfavors the notion that the

super-shifting we observe in the presence of MSL complex and MBP-CLAMP is mediated by CLAMP-CLAMP multimerization unless this multimerization only occurs when MSL complex is expressed. Perhaps MSL complex interaction with CLAMP's glutamine domain stabilizes the endogenous CLAMP thereby facilitating zinc fingers not bound to DNA or RNA to multimerize. This model is supported by the well-studied phenomenon that glutamine rich domains are unstructured and often aggregate (Atanesyan et al. 2012). Therefore endogenous CLAMP containing the glutamine domain could aggregate and form inclusion bodies in the absence of MSL complex with whom to anchor an interaction.

When purified recombinant MSL2-CXC was incubated with MBP-CLAMP 6-Zinc finger domain and CLAMP binding sequence, we did not observe an interaction (Fig 3). These results demonstrate that the ring finger of MSL2 (CXC) does not interact with the C-terminal 6-Zinc fingers of CLAMP *in vitro* under these conditions. Therefore, the interaction between MBP-CLAMP 6-Zinc finger and MSL complex nuclear extract may be mediated by another component of the MSL complex. Our RNA immunoprecipitation results support a model whereby CLAMP interacts with MSL complex *in vivo*.

Conclusion/Future Directions

Based on our findings presented here CLAMP appears to exhibit an *in vivo* interaction with *roX2* RNAs of MSL complex. In addition, CLAMP zinc fingers may

form a complex with both MSL complex and DNA. Further work is needed however to map these exact interactions and to determine the conditions under which they occur. More studies are underway to investigate which precise component(s) of MSL complex may interact with CLAMP and also which exact domain of CLAMP can mediate a direct interaction with MSL complex.

Materials and Methods

RNA Immunoprecipitation

RNA Immunoprecipitation protocol was adapted from Keene, JD Nature Protocols 2006 with the following modifications (Keene et al. 2006). RNase, DNase-free filter-sterilized water was used for all buffer preparations.

I. Isolation of Nuclear and Cytoplasmic extracts from *Drosophila* cells

1ml of tissue cells was collected from T225 flasks of cells to calculate protein concentration. 2-5 mg of total protein was needed per RNA Immunoprecipitation. The concentration of protein was calculated in 1ml of *Drosophila* SL2 cells having a density of 5×10^6 cells. Protein was extracted from *Drosophila* SL2 cells using the following method to isolate nuclear and cytoplasmic extracts from *Drosophila* cells.

1. Cells were grown in T225 flasks. 1ml of cells was pelleted at 1500xg for 3 min at 4°C. Cells were rinsed with 1ml of cold PBS (1X) and re-pelleted at 1500xg for 3 min 4°C. Cells were pipetted up and down 20 times in 100 μ L of Extraction Buffer and pelleted at 1500xg for 3 min at 4°C.

Extraction / Lysis Buffer:

Sucrose (10%)
Tris-HCl pH 8 (10mM)
NaCl (10mM)
MgCl₂ (3mM)
DTT (2mM)
Triton X100 (0.2%)
PMSF (1mM)

2. The supernatant (cytoplasmic fraction) was transferred to a new tube and centrifuged at 1500xg for 3 min at 4°C to remove residual nuclei. The supernatant was transferred to a new tube and 20 uL of 6X loading dye added. The sample was heated at 95 degrees for 10 min. This fraction contained the cytoplasmic extract.
3. The nuclear extract from step 1 above was washed with 200 ul of TEN140 buffer.

TEN 140 Buffer:

Tris-HCl pH 8 (10mM)
EDTA pH 8 (1mM)
NaCl (0.14M)
DTT (2mM)
PMSF (1mM)

4. Nuclei were pelleted at 1500xg for 5min at 4°C. The supernatant was removed and the nuclei resuspended in 100uL of TEN 140 Buffer. 20ul of 6X loading dye was added to the nuclei and the sample boiled for 10 min at 95°C.
5. Protein extraction was calculated using Pierce BCA Protein Assay Kit (Catalog #23225) and additionally verified with Qbit Protein Assay Kit (Catalog #Q33221) to ensure a healthy concentration of protein was present at the time and cell density of harvesting. At least 2-5 mg of total protein was needed per RNA Immunoprecipitation.

II. Native RNA-ChIP-qPCR

The remaining cells in the T225 flasks were used for RNA Immunoprecipitation.

Day 1

1. The bottom of the flasks was scraped with cell media to remove cells that adhered to the bottom. Flasks were scraped as thoroughly as possible, i.e. the entire bottom surface was scraped.
2. Cells were poured into sterile 50 ml centrifuge tubes and pelleted by centrifugation for 10 min at 720 g at 4°C.
3. Supernatant was discarded and each pellet resuspended in 10 ml of ice-cold PBS (1X) then centrifuged for 10 min at 720 g at 4°C.
4. 1ml of ice-cold lysis buffer was added to each RIP sample.

Lysis Buffer:

1X Lysis Buffer (Active Motif Cat # 101212)

5 ul Protease Inhibitor Cocktail (PIC, Roche Cat # 116974980001)

5 ul Phenylmethanesulfonyl fluoride (PMSF, Sigma-Aldrich Product# P7626)

2 uL RNase Inhibitor (RNase Out 100 units/ml, Invitrogen Cat#10777-019)

5. Samples were incubated on ice for 30 minutes
6. Cells were transferred to ice-cold 2 ml size dounce homogenizer and dounced 10X to aid the release of nuclei.
7. Cells were centrifuged at 2400 g for 10 min.
8. Supernatant was removed and the pellet resuspended in 1mL polysome lysis buffer.

Polysome lysis buffer

100 mM KCl

5 mM MgCl₂

10 mM HEPES (pH 7.0) - keep refrigerated 4°C

0.5% NP40

Before use the following reagents were added to 1ml of buffer.

1 µL 1M DTT, 1 mM final concentration (Dithiothreitol, Fermentas Cat#R0861)

2 µL RNase inhibitor (RNase Out)

10 µL PIC (1:100)

2 µL 200mM VRCs (Vanadryl Ribonucleoside complexes, New England Biolabs Cat# S1402S)

9. Samples were passed through 3mL 27½ gauge needles 4 times and then incubated with 400 units DNase (RNase-free) per 1ml of sample on ice for 30 minutes.
10. Lysates were centrifuged for 2,000 g for 10 mins to clear larger particles. Supernatant was collected and transferred to clean microcentrifuge tube and stored on ice.
11. 200 ul of cleared lysate was removed as input and stored at 20°C.
12. To the remaining lysate 2 uL RNase Inhibitor and 2ul of antibody was added and samples rotated at 4°C overnight.

Day 2**Prepare beads**

Keep speeds around 1500 – 2000g when working with beads

13. Tubes were prepared with 50 µL protein A agarose (non-blocked, no antibody) beads per lysate reaction: 50ul beads * # lysate reactions
14. Beads were spun down at 2000 g, 2 min, 4°C and the supernatant discarded.

15. 850uL ice-cold NT2 buffer was added to 50ul beads.

NT2 Buffer

50 mM Tris-HCl (pH 7.4, JT Baker Cat#4103-01)

150 mM NaCl (Mallinckrodt, Cat# 7581)

1 mM MgCl₂ (EMD Bioscience, Cat# MX0045-1)

0.05% NP40 (Igepal Nonidet, Sigma Cat#I-30201)

16. Beads were spun with NT2 buffer at 2000 g 2 min 4°C. Supernatant was removed and discarded.

17. 50ul of beads was resuspended in 850 uL of ice-cold NT2 buffer. To each 50ul of beads, the following was added

2 uL RNase inhibitor (RNase Out)

2 uL VRC (200mM)

1 uL DTT (1M)

35 ul EDTA (0.5M) to yield final concentration of 20mM

18. Tubes were inverted 3 to 4 times to ensure beads were thoroughly resuspended in buffer.

19. Centrifuge 2 min 2000 g 4°C

20. Supernatant was removed to leave ~50 ul of beads volume at bottom of tube. Beads were stored on ice.

Immunoprecipitation

21. Samples were removed from 4°C overnight rotation. *If there is precipitate or debris floating in sample*, spin at 2,000 g 2 mins and discard pellet.

22. Prepared 50 uL protein-A sepharose beads from above was added, 50 uL beads per IP sample

23. Rotate 4°C for 3 hrs (3-5hrs)

24. Samples were spun 2000g 2min 4°C. Supernatant was removed with pipettor and discard.

25. 1ml ice-cold NT2 buffer was added.

26. Rotate 4°C 2 min.
27. Spin 2000 g 2 min 4°C. Remove supernatant with pipettor and discard.
28. Repeat steps 13 – 15 four times.
29. Resuspend beads in 100ul NT2 buffer.
30. Add 1.5 uL (30 ug) proteinase K (20 mg/mL = 20 ug/uL) to release RNP components.
31. Flick tube with finger four to five times.
32. Incubate for 30min at 55°C.
33. Add 4 uL (20ug) glycogen (5 mg/ml = 5ug/uL) to precipitation reaction to help make RNA pellet more visible for extraction.
(Glycogen, Roche Cat#10901393001)

III. Trizol Extraction

(adapted from Invitrogen Trizol Reagent: RNA Isolation Procedure Cat# 15596-026)

Homogenization.

1ml Trizol was added to each sample

Phase separation

1. Incubate the homogenized sample for 5 minutes at room temperature to permit complete dissociation of the nucleoprotein complex.
2. Add 0.2 mL of chloroform per 1 mL of TRIzol® Reagent used for homogenization. Cap the tube securely.
3. Shake tube vigorously by hand for 15 seconds. Or Vortex for 15 seconds.
4. Incubate for 2–3 minutes at room temperature.
5. Centrifuge the sample at 12,000 × g for 15 minutes at 4°C.

Note: The mixture separates into a lower red phenol-chloroform phase, an interphase, and a colorless upper aqueous phase. *RNA remains exclusively in the aqueous phase.*

The upper aqueous phase is ~50% of the total volume.

6. Remove the aqueous phase of the sample by angling the tube at 45° and pipetting the solution out. Avoid drawing any of the interphase or organic layer into the pipette when removing the aqueous phase.

7. Place the aqueous phase into a new tube and proceed to the RNA Isolation Procedure.

8. Save the interphase and organic phenol-chloroform phase if isolation of DNA or protein is desired. The organic phase can be stored at 4°C overnight

RNA precipitation

1. Add 0.5 mL of 100% isopropanol to the aqueous phase, per 1 mL of TRIzol® Reagent used for homogenization.

2. Incubate at room temperature for 10 minutes.

3. Centrifuge at $12,000 \times g$ for 10 minutes at 4°C.

Note: The RNA is often invisible prior to centrifugation, and forms a gel-like pellet on the side and bottom of the tube.

4. Proceed to RNA wash.

RNA wash

1. Remove the supernatant from the tube, leaving only the RNA pellet.

2. Wash the pellet, with 1 mL of 75% ethanol per 1 mL of TRIzol® Reagent used in the initial homogenization.

Note: The RNA can be stored in 75% ethanol at least 1 year at -20°C, or at least 1 week at 4°C.

3. Vortex the sample briefly, then centrifuge the tube at $7500 \times g$ for 5 minutes at 4°C. Discard the wash.

4. Air dry the RNA pellet for 5–10 minutes. Do not dry the pellet by vacuum centrifuge.

Note: Do not allow the RNA to dry completely, because the pellet can lose solubility. Partially dissolved RNA samples have an A260/280 ratio <1.6.

5. Proceed to RNA resuspension.

RNA resuspension

1. Resuspend the RNA pellet in 30ul RNase-free water or 0.5% SDS solution (20–50 µL) by passing the solution up and down several times through a pipette tip. Note: Do not dissolve the RNA in 0.5% SDS if it is to be used in subsequent enzymatic reactions.

2. Incubate in a water bath or heat block set at 55–60°C for 10–15 minutes.

3. Proceed to downstream application, or store at –70°C.

RNA concentration

1. Measure absorbance at 260 nm on Nandrop or Qbit

IV. cDNA Synthesis

RNA was reverse transcribed into cDNA for qPCR using Vilo Kit: Superscript II Reverse Transcription Kit (Invitrogen Cat#11754-050).

1. Use 1 ug of RNA/reaction (quantified on nanodrop)
2. RNA + water (RNAase free) = total volume of 14 ul
3. Add 4 ul vilo mix (radom hexamers + oligo dT)
2 ul SSIII
4. Incubate 25 C 10 minutes (annealing)
5. Incubate 42 C 60 minutes (extension)
6. Incubate 85 C 5 minutes (inactivate enzymes)
7. Dilute cDNA 1:20 before qPCR

cDNA concentrations were quantified by Qbit (Invitrogen Cat# Q32851).

qPCR was performed with appropriate primers.

Co-Immunoprecipitation

RNA Immunoprecipitation was performed as described in steps I, II, and III above.

Protein Co-Immunoprecipitation was analyzed on a Western with CLAMP and MLE antibodies and the WesternBreeze Chemiluminescent Kit (Novex Cat#WB7106).

Electrophoretic Mobility Shift Assay

EMSAs were performed using the Chemiluminescent Nucleic Acid Detection Module (Thermo Scientific Cat#89880)/LightShift Chemiluminescent EMSA Kit (Thermo Scientific Cat#20148). DNA binding conditions for CLAMP protein were previously optimized with 1 ug/uL Poly (dI•dC, Cat#20148E) and Elution Buffer: 5mM Tris-HCL, pH 8.5 (AE Macherey-Nagel Cat#740588.10).

DNA labeling

The CLAMP binding high affinity X sequence was end labeled with Biotin using Biotin 3'End Labeling Kit (Thermo Scientific Pierce Cat#89818) with the following adaptations. Each strand of DNA was Biotin labeled independently then annealed using a slow cool down cycle: 95°C 10 mins, 72°C 2 hours, 52°C 2 hours, 37°C 2 hours, 4°C using a PCR machine.

Single strand DNA digestion of annealed dsDNA was performed for 10 mins at 37°C with Mung Bean Nuclease 2,000 U/uL (New England Biolabs Cat#M0250S) followed by PCR Clean-up with Qiagen Kit: Qiaquick PCR Purification Kit Cat# 28104.

MSL complex Extraction

Whole cell extracts of cultured SL2 and KC cells were performed using the following method.

1. Cells grown in T25 flasks were scraped off the bottom and pipetted into 15 ml corning tubes.
2. Cells were spun at top speed for 10mins Room Temperature.
3. Media was aspirated and off and discarded.
4. 1ml of cold PBS (1X) was added to the pellet and pipetted up and down.
5. Cells were pelleted at 1500g for 3 min 4°C then PBS removed.
6. Pellets were resuspended in 0.5 ml cold extraction buffer and incubated on ice.

Extraction Buffer

200 mM KCl
100mM Tris pH 8.0
0.2mM EDTA
0.1% NP-40
10% Glycerol
1mM PMSF

7. Resuspended pellets were spun for 10 min at 14K RPM 4°C.
8. Supernatant was retained and used for gel-shift assays.

Figures

Figure 1: CLAMP exhibits *in vivo* interaction with *roX2*

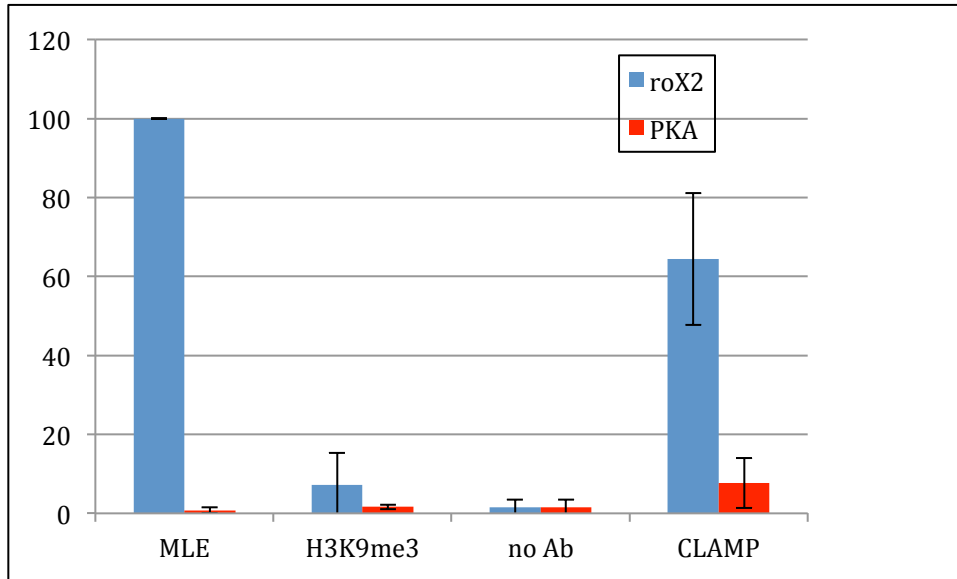


Figure 1: *roX2* RNA was immunoprecipitated from SL2 (male) cells using CLAMP antibody with MLE, the RNA/DNA helicase of MSL complex, antibody as a positive control. H3K9me3 and no Antibody IPs served as negative controls highlighting the specificity of CLAMP's interaction with *roX2* RNAs. *roX1* was not assayed because *roX1* is not expressed in male *Drosophila* cells possibly due to a rearrangement of the *roX1* gene in male cells.

Figure 2: Co-Immunoprecipitation of CLAMP and MLE

Figure 2a. Co-immunoprecipitation of CLAMP and MLE when cells are crosslinked with formaldehyde

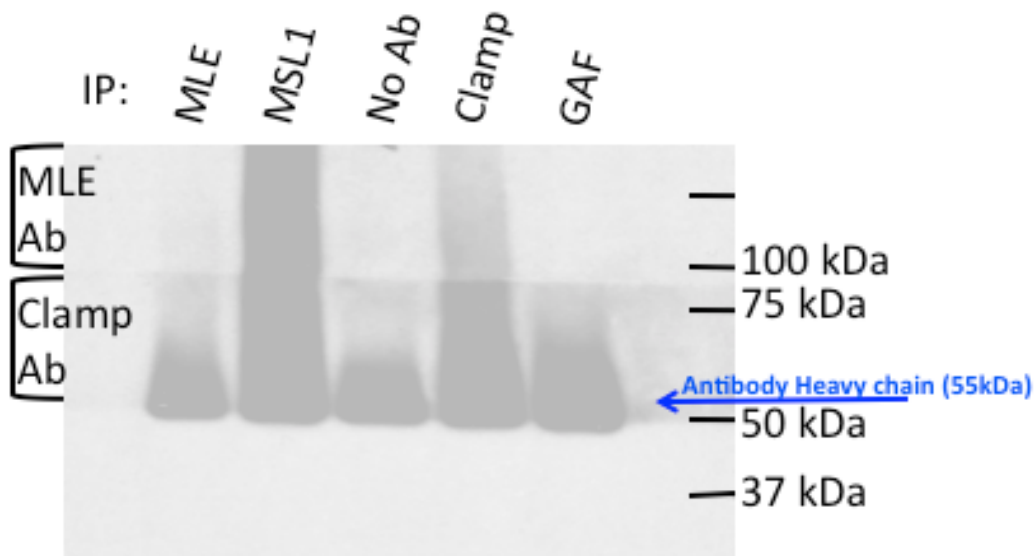


Figure 2a. Cells were crosslinked with formaldehyde and co-immunoprecipitation performed. MLE, MSL1, CLAMP, and GAF antibodies were used to immunoprecipitate and the Western blot probed with MLE and CLAMP antibodies. A no antibody sample was used as a negative control. The band around 50kDa common to all IPs corresponds to the antibody heavy chain (55kDa).

Figure 2: Co-Immunoprecipitation of CLAMP and MLE

Figure 2b. Co-Immunoprecipitation with CLAMP antibody coupled beads

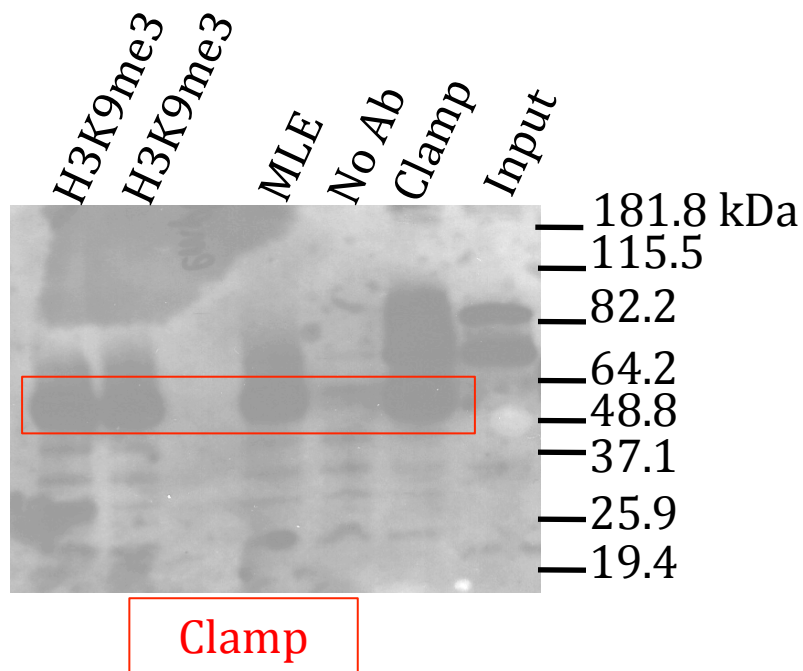


Figure 2b. CLAMP antibody was coupled to magnetic beads and used for immunoprecipitation from cells. CLAMP protein (~61.22 kDa) is detected around 64kDa ladder band. MLE (140kDa) appears to be absent.

Figure 2: Co-Immunoprecipitation of CLAMP and MLE

Figure 2c. Co-Immunoprecipitation without RNase treatment

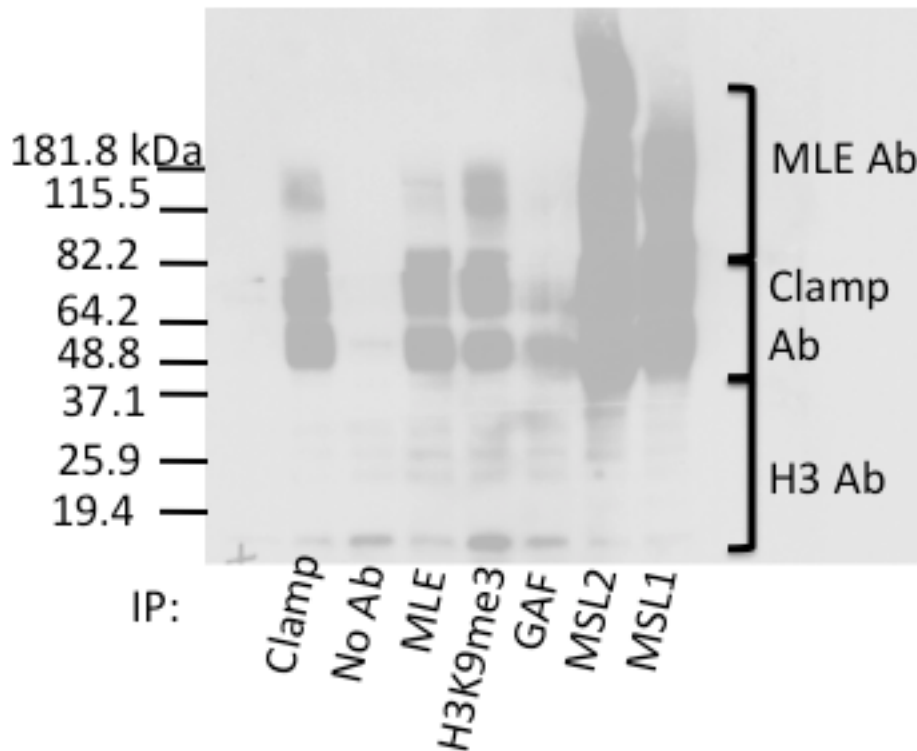


Figure 2c. Immunoprecipitation of cells using CLAMP, MLE, H3K9me3, GAF, MSL2, and MSL1 antibodies. H3 Antibody was used to probe the bottom third of the gel as a loading control. H3 is detected around 15-17kDa. CLAMP (~61.22kDa) appears to be present in all IPs except No Ab control. MSL2 and MSL1 IP signals were too strong to allow individual protein bands to be detected. Top third of gel was probed with MLE antibody. MLE (140kDa) maybe present in CLAMP and H3K9me3, but a lack of a more distinct MLE band in the MLE IPs make this result inconclusive.

Figure 2: Co-Immunoprecipitation of CLAMP and MLE

Figure 2d. Co-Immunoprecipitation with RNase treatment

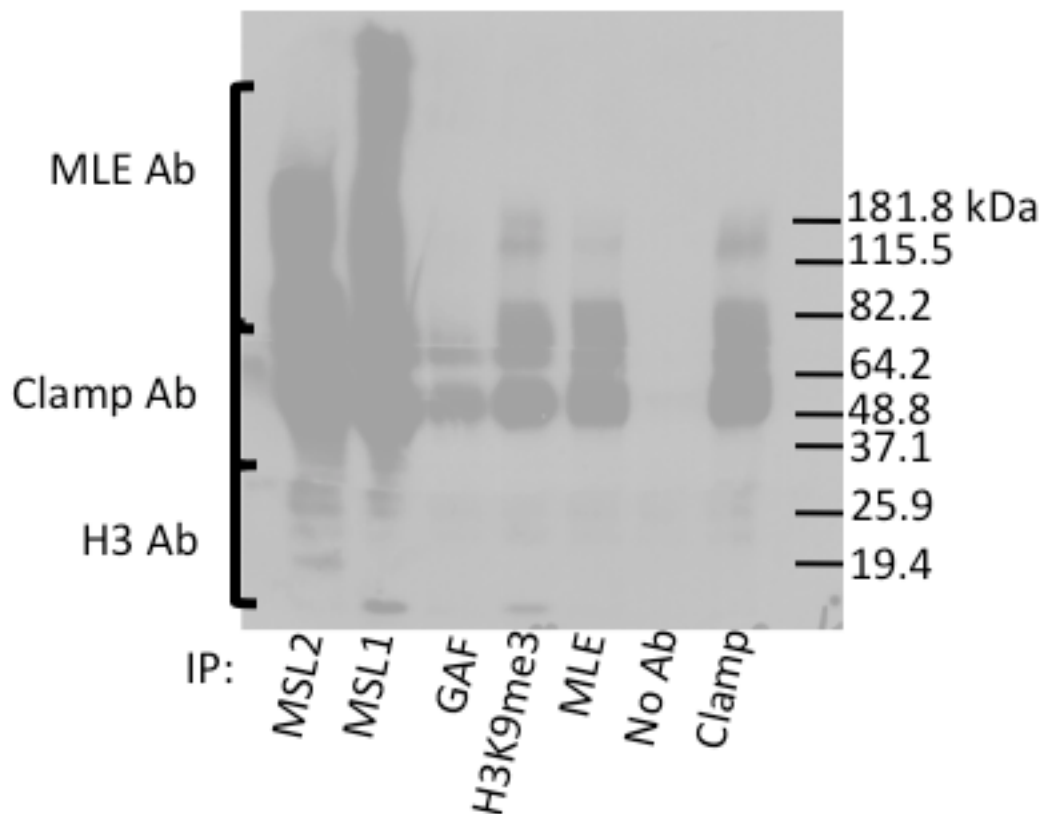


Figure 2d. Immunoprecipitation of cells using CLAMP, MLE, H3K9me3, GAF, MSL2, and MSL1 antibodies was performed with RNase digestion. H3 Antibody was used to probe bottom third of gel as a loading control. MSL2 and MSL1 IPs signal was too strong to allow individual protein bands to be detected. Top third of gel was probed with MLE antibody. MLE (140kDa) maybe present in CLAMP and H3K9me3, but a lack of a more distinct MLE band in MLE IPs make this result inconclusive.

Figure 2: Co-Immunoprecipitation of CLAMP and MLE

Figure 2e. CLAMP and MLE Co-Immunoprecipitation performed with Native conditions

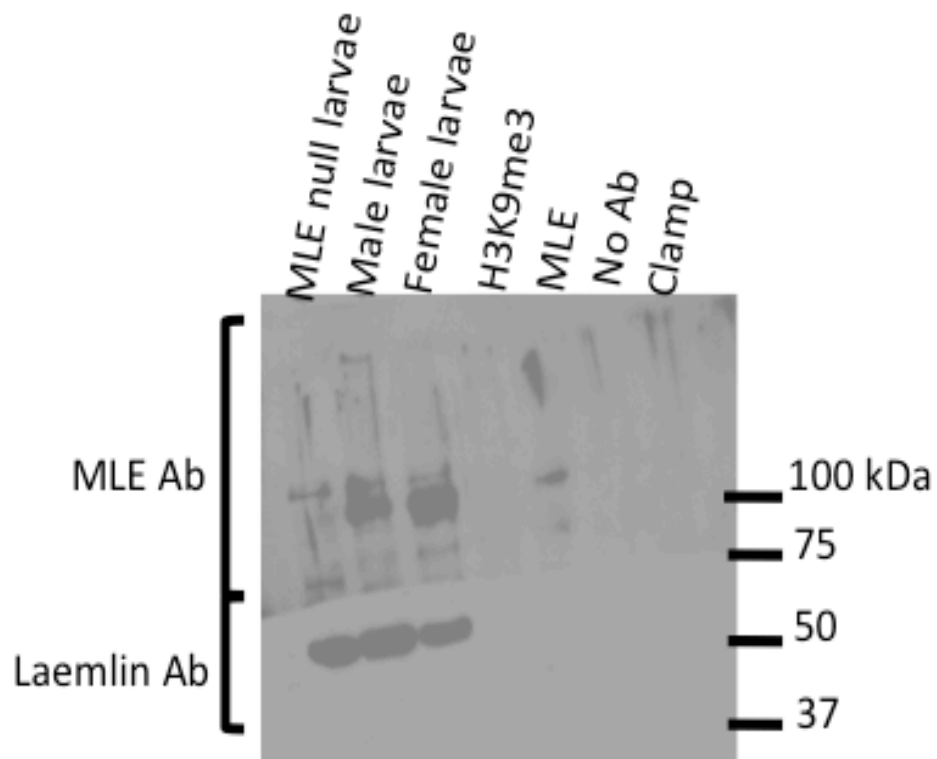


Figure 2e. Immunoprecipitation was performed using non-crosslinking conditions for MLE, Clamp, H3K9me3. Female and male wildtype larvae that normally express MLE were run on the Western as controls in addition to a MLE null larvae line. MLE (140 kDa) is detected in wt female, male larvae and MLE IPs but not CLAMP or H3K9me3 IPs. The top half of gel was probed with MLE antibody and bottom half probed with laemlin antibody.

Figure 3: Gel-shift with S2L (male) and KC (female) cell extracts

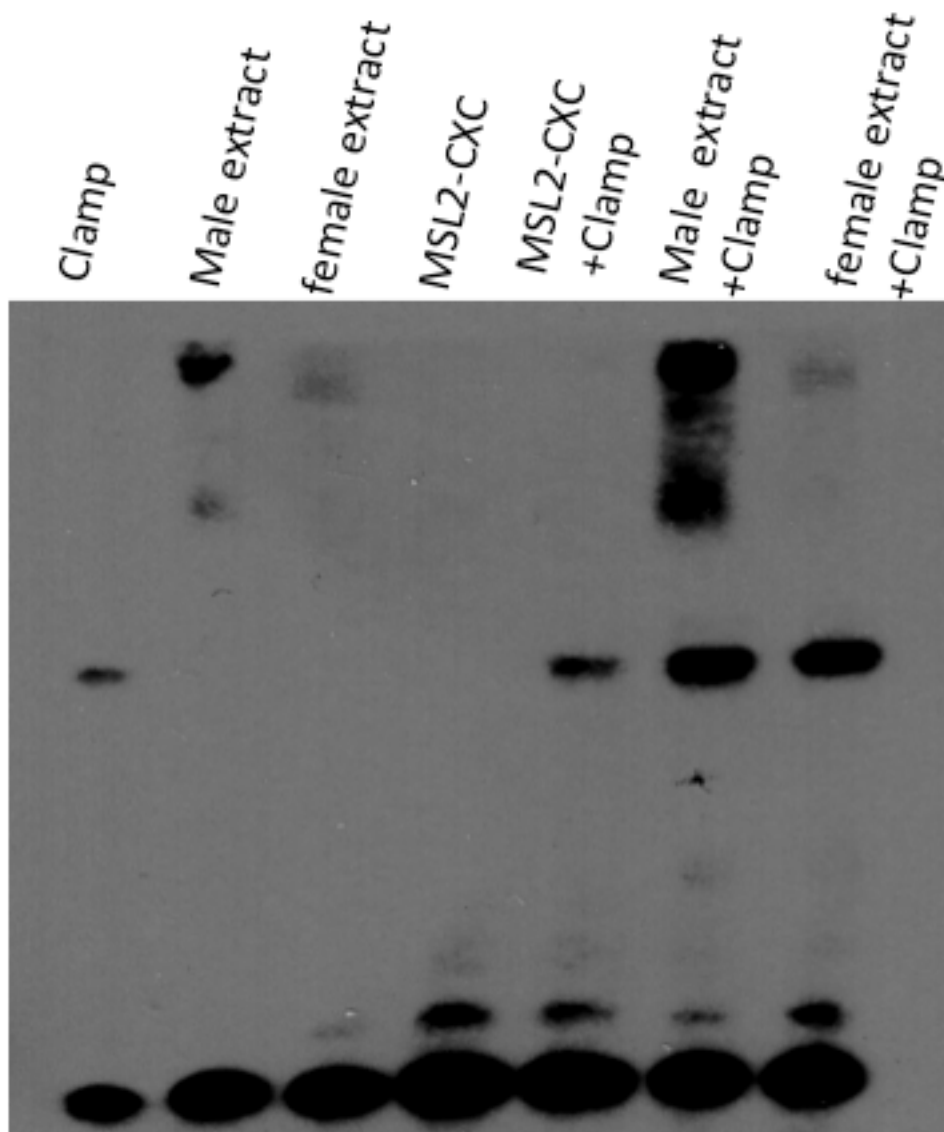


Figure 3. A Biotin-labeled oligonucleotide of a high affinity X-chromosome CLAMP binding sequence was incubated with MBP-CLAMP, GST-MSL2-CXC, SL2 extract, KC extract, and combinations of these. MBP-CLAMP + SL2 extract, which contains MSL complex, displays supershifted bands relative to MBP-CLAMP alone indicating MBP-CLAMP is interacting with a factor, likely MSL complex, within the SL2 (male) extract. Supershifted bands are absent in the MBP-CLAMP + KC extract (female). KC extract does not contain MSL complex, suggesting that the CLAMP interacting factor within SL2 extract is MSL complex. MBP-CLAMP + GFP-MSL2-CXC do not supershift the probe, indicating that Clamp is not interacting with CXC domain of MSL2. SL2 extract alone also supershifts the oligonucleotide probe, but not as strongly as SL2 extract +MBP-CLAMP. It is possible that endogenous CLAMP already associated with *in vivo* MSL complex is the supershifted band.

References

- Adelman K, Wei W, Ardehali MB, Werner J, Zhu B, Reinberg D, Lis JT. 2006. Drosophila Paf1 modulates chromatin structure at actively transcribed genes. *Mol Cell Biol* **26**: 250-260.
- Adkins NL, Hagerman TA, Georgel P. 2006. GAGA protein: a multi-faceted transcription factor. *Biochem Cell Biol* **84**: 559-567.
- Akhtar A, Becker PB. 2000. Activation of transcription through histone H4 acetylation by MOF, an acetyltransferase essential for dosage compensation in Drosophila. *Mol Cell* **5**: 367-375.
- Akhtar A, Zink D, Becker PB. 2000. Chromodomains are protein-RNA interaction modules. *Nature* **407**: 405-409.
- Alekseyenko AA, Ellison CE, Gorchakov AA, Zhou Q, Kaiser VB, Toda N, Walton Z, Peng S, Park PJ, Bachtrog D et al. 2013. Conservation and de novo acquisition of dosage compensation on newly evolved sex chromosomes in Drosophila. *Genes & development* **27**: 853-858.
- Alekseyenko AA HJ, Peng S, Gelbart M, Tolstorukov M, Plachetka A, Kharchenko PV, Jung YL, Gorchakov AA, Larschan E, Gu T, Minoda A, Riddle NC, Schwartz YB, Elgin SCR, Karpen GH, Pirrotta V, Kuroda MI**, Park PJ**. 2012. Sequence-specific targeting of dosage compensation in Drosophila favors an active chromatin context. *In press PLoS Genetics*.
- Alekseyenko AA, Ho JW, Peng S, Gelbart M, Tolstorukov MY, Plachetka A, Kharchenko PV, Jung YL, Gorchakov AA, Larschan E et al. 2012. Sequence-specific targeting of dosage compensation in Drosophila favors an active chromatin context. *PLoS genetics* **8**: e1002646.
- Alekseyenko AA, Larschan E, Lai WR, Park PJ, Kuroda MI. 2006. High-resolution ChIP-chip analysis reveals that the Drosophila MSL complex selectively identifies active genes on the male X chromosome. *Genes & development* **20**: 848-857.
- Alekseyenko AA, Peng S, Larschan E, Gorchakov AA, Lee OK, Kharchenko P, McGrath SD, Wang CI, Mardis ER, Park PJ et al. 2008. A sequence motif within chromatin entry sites directs MSL establishment on the Drosophila X chromosome. *Cell* **134**: 599-609.
- Amrein H, Axel R. 1997. Genes expressed in neurons of adult male Drosophila. *Cell* **88**: 459-469.
- Atanesyan L, Gunther V, Dichtl B, Georgiev O, Schaffner W. 2012. Polyglutamine tracts as modulators of transcriptional activation from yeast to mammals. *Biological chemistry* **393**: 63-70.
- Bai X, Alekseyenko AA, Kuroda MI. 2004. Sequence-specific targeting of MSL complex regulates transcription of the roX RNA genes. *The EMBO journal* **23**: 2853-2861.
- Bai X, Larschan E, Kwon SY, Badenhorst P, Kuroda MI. 2007. Regional control of chromatin organization by noncoding roX RNAs and the NURF remodeling complex in Drosophila melanogaster. *Genetics* **176**: 1491-1499.

- Bailey TL, Boden M, Buske FA, Frith M, Grant CE, Clementi L, Ren J, Li WW, Noble WS. 2009. MEME SUITE: tools for motif discovery and searching. *Nucleic Acids Res* **37**: W202-208.
- Bashaw GJ, Baker BS. 1997. The regulation of the *Drosophila* msl-2 gene reveals a function for Sex-lethal in translational control. *Cell* **89**: 789-798.
- Batista PJ, Chang HY. 2013a. Cytotopic localization by long noncoding RNAs. *Current opinion in cell biology* **25**: 195-199.
- . 2013b. Long Noncoding RNAs: Cellular Address Codes in Development and Disease. *Cell* **152**: 1298-1307.
- Beatriz Vicoso DB. 2009. Progress and prospects toward our understanding of the evolution of dosage compensation.
- Bell O, Conrad T, Kind J, Wirbelauer C, Akhtar A, Schubeler D. 2008. Transcription-coupled methylation of histone H3 at lysine 36 regulates dosage compensation by enhancing recruitment of the MSL complex in *Drosophila melanogaster*. *Molecular and cellular biology* **28**: 3401-3409.
- Belote JM, Lucchesi JC. 1980a. Control of X chromosome transcription by the maleless gene in *Drosophila*. *Nature* **285**: 573-575.
- . 1980b. Male-specific lethal mutations of *Drosophila melanogaster*. *Genetics* **96**: 165-186.
- Berger MF, Bulyk ML. 2009. Universal protein-binding microarrays for the comprehensive characterization of the DNA-binding specificities of transcription factors. *Nature protocols* **4**: 393-411.
- Berger MF, Philippakis AA, Qureshi AM, He FS, Estep PW, 3rd, Bulyk ML. 2006. Compact, universal DNA microarrays to comprehensively determine transcription-factor binding site specificities. *Nat Biotechnol* **24**: 1429-1435.
- Bryson TD, Weber CM, Henikoff S. 2010. Baculovirus-encoded protein expression for epigenomic profiling in *Drosophila* cells. *Fly (Austin)* **4**: 258-265.
- Buscaino A, Legube G, Akhtar A. 2006. X-chromosome targeting and dosage compensation are mediated by distinct domains in MSL-3. *EMBO reports* **7**: 531-538.
- Carrozza MJ, Li B, Florens L, Suganuma T, Swanson SK, Lee KK, Shia WJ, Anderson S, Yates J, Washburn MP et al. 2005. Histone H3 methylation by Set2 directs deacetylation of coding regions by Rpd3S to suppress spurious intragenic transcription. *Cell* **123**: 581-592.
- Chow JC, Ciaudo C, Fazzari MJ, Mise N, Servant N, Glass JL, Attreed M, Avner P, Wutz A, Barillot E et al. 2010. LINE-1 activity in facultative heterochromatin formation during X chromosome inactivation. *Cell* **141**: 956-969.
- Chu C, Qu K, Zhong FL, Artandi SE, Chang HY. 2011. Genomic maps of long noncoding RNA occupancy reveal principles of RNA-chromatin interactions. *Molecular cell* **44**: 667-678.
- Chuang PT, Albertson DG, Meyer BJ. 1994. DPY-27: a chromosome condensation protein homolog that regulates *C. elegans* dosage compensation through association with the X chromosome. *Cell* **79**: 459-474.

- Colgan J, Ashali H, Manley JL. 1995. A direct interaction between a glutamine-rich activator and the N terminus of TFIIB can mediate transcriptional activation in vivo. *Mol Cell Biol* **15**: 2311-2320.
- Conrad T, Cavalli FM, Holz H, Hallaceli E, Kind J, Ilik I, Vaquerizas JM, Luscombe NM, Akhtar A. 2012. The MOF chromobarrel domain controls genome-wide H4K16 acetylation and spreading of the MSL complex. *Dev Cell* **22**: 610-624.
- Copps K, Richman R, Lyman LM, Chang KA, Rampersad-Ammons J, Kuroda MI. 1998. Complex formation by the Drosophila MSL proteins: role of the MSL2 RING finger in protein complex assembly. *Embo J* **17**: 5409-5417.
- Csankovszki G, McDonel P, Meyer BJ. 2004. Recruitment and spreading of the C. elegans dosage compensation complex along X chromosomes. *Science* **303**: 1182-1185.
- Dahlsveen IK, Gilfillan GD, Shelest VI, Lamm R, Becker PB. 2006. Targeting determinants of dosage compensation in Drosophila. *PLoS genetics* **2**: e5.
- Davis TL, Meyer BJ. 1997. SDC-3 coordinates the assembly of a dosage compensation complex on the nematode X chromosome. *Development* **124**: 1019-1031.
- Deng X, Berletch JB, Ma W, Nguyen DK, Hiatt JB, Noble WS, Shendure J, Disteche CM. 2013. Mammalian X upregulation is associated with enhanced transcription initiation, RNA half-life, and MOF-mediated H4K16 acetylation. *Dev Cell* **25**: 55-68.
- Deng X BJ, Ma W, Nguyen DK, Hiatt JB, Noble WS, Shendure J, Disteche CM. 2013. Mammalian X Upregulation Is Associated with Enhanced Transcription Initiation, RNA Half-Life, and MOF-Mediated H4K16 Acetylation. *Developmental Cell* **25**: 1-14.
- Deng X, Hiatt JB, Nguyen DK, Ercan S, Sturgill D, Hillier LW, Schlesinger F, Davis CA, Reinke VJ, Gingeras TR et al. 2011. Evidence for compensatory upregulation of expressed X-linked genes in mammals, Caenorhabditis elegans and Drosophila melanogaster. *Nature genetics* **43**: 1179-1185.
- Deng X, Meller VH. 2006. roX RNAs are required for increased expression of X-linked genes in Drosophila melanogaster males. *Genetics* **174**: 1859-1866.
- Derrien T, Johnson R, Bussotti G, Tanzer A, Djebali S, Tilgner H, Guernec G, Martin D, Merkel A, Knowles DG et al. 2012. The GENCODE v7 catalog of human long noncoding RNAs: analysis of their gene structure, evolution, and expression. *Genome research* **22**: 1775-1789.
- Deuring R, Fanti L, Armstrong JA, Sarte M, Papoulas O, Prestel M, Daubresse G, Verardo M, Moseley SL, Berloco M et al. 2000. The ISWI chromatin-remodeling protein is required for gene expression and the maintenance of higher order chromatin structure in vivo. *Mol Cell* **5**: 355-365.
- Dickinson LA, Joh T, Kohwi Y, Kohwi-Shigematsu T. 1992. A tissue-specific MAR/SAR DNA-binding protein with unusual binding site recognition. *Cell* **70**: 631-645.
- Disteche CM. 2012. Dosage compensation of the sex chromosomes. *Annu Rev Genet* **46**: 537-560.
- Dundr M, Misteli T. 2010. Biogenesis of nuclear bodies. *Cold Spring Harbor perspectives in biology* **2**: a000711.

- Dunlap D, Yokoyama R, Ling H, Sun HY, McGill K, Cugusi S, Lucchesi JC. 2012. Distinct contributions of MSL complex subunits to the transcriptional enhancement responsible for dosage compensation in *Drosophila*. *Nucleic acids research* **40**: 11281-11291.
- Ellison CE, Bachtrog D. 2013. Dosage compensation via transposable element mediated rewiring of a regulatory network. *Science* **342**: 846-850.
- Engreitz JM, Pandya-Jones A, McDonel P, Shishkin A, Sirokman K, Surka C, Kadri S, Xing J, Goren A, Lander ES et al. 2013. The Xist lncRNA exploits three-dimensional genome architecture to spread across the X chromosome. *Science* **341**: 1237973.
- Enuameh MS, Asriyan Y, Richards A, Christensen RG, Hall VL, Kazemian M, Zhu C, Pham H, Cheng Q, Blatti C et al. 2013. Global analysis of *Drosophila* Cys(2)-His(2) zinc finger proteins reveals a multitude of novel recognition motifs and binding determinants. *Genome research* **23**: 928-940.
- Fauth T, Muller-Planitz F, Konig C, Straub T, Becker PB. 2010. The DNA binding CXC domain of MSL2 is required for faithful targeting the Dosage Compensation Complex to the X chromosome. *Nucleic acids research* **38**: 3209-3221.
- Fisher WW, Li JJ, Hammonds AS, Brown JB, Pfeiffer BD, Weiszmman R, MacArthur S, Thomas S, Stamatoyannopoulos JA, Eisen MB et al. 2012. DNA regions bound at low occupancy by transcription factors do not drive patterned reporter gene expression in *Drosophila*. *Proceedings of the National Academy of Sciences of the United States of America* **109**: 21330-21335.
- Furuhashi H, Nakajima M, Hirose S. 2006. DNA supercoiling factor contributes to dosage compensation in *Drosophila*. *Development* **133**: 4475-4483.
- Gallach M, Arnau V, Aldecoa R, Marin I. 2010. A sequence motif enriched in regions bound by the *Drosophila* dosage compensation complex. *BMC genomics* **11**: 169.
- Gamsjaeger R, Liew CK, Loughlin FE, Crossley M, Mackay JP. 2007. Sticky fingers: zinc-fingers as protein-recognition motifs. *Trends in biochemical sciences* **32**: 63-70.
- Garriga-Canut M, Agustin-Pavon C, Herrmann F, Sanchez A, Dierssen M, Fillat C, Isalan M. 2012. Synthetic zinc finger repressors reduce mutant huntingtin expression in the brain of R6/2 mice. *Proceedings of the National Academy of Sciences of the United States of America* **109**: E3136-3145.
- Gelbart ME, Kuroda MI. 2009. *Drosophila* dosage compensation: a complex voyage to the X chromosome. *Development* **136**: 1399-1410.
- Gelbart ME, Larschan E, Peng S, Park PJ, Kuroda MI. 2009. *Drosophila* MSL complex globally acetylates H4K16 on the male X chromosome for dosage compensation. *Nature structural & molecular biology* **16**: 825-832.
- Gilfillan GD, Straub T, de Wit E, Greil F, Lamm R, van Steensel B, Becker PB. 2006. Chromosome-wide gene-specific targeting of the *Drosophila* dosage compensation complex. *Genes Dev* **20**: 858-870.
- Girirajan S, Campbell CD, Eichler EE. 2011. Human copy number variation and complex genetic disease. *Annu Rev Genet* **45**: 203-226.

- Gontan C, Achame EM, Demmers J, Barakat TS, Rentmeester E, van IW, Grootegeod JA, Gribnau J. 2012. RNF12 initiates X-chromosome inactivation by targeting REX1 for degradation. *Nature* **485**: 386-390.
- Gu W, Szauter P, Lucchesi JC. 1998a. Targeting of MOF, a putative histone acetyl transferase, to the X chromosome of *Drosophila melanogaster*. *Developmental genetics* **22**: 56-64.
- Gu WG, Szauter P, Lucchesi JC. 1998b. Targeting of MOF, a putative histone acetyl transferase, to the X chromosome of *Drosophila melanogaster*. *Dev Genet* **22**: 56-64.
- Guertin MJ, Lis JT. 2012. Mechanisms by which transcription factors gain access to target sequence elements in chromatin. *Current opinion in genetics & development*.
- Hallacli E, Lipp M, Georgiev P, Spielman C, Cusack S, Akhtar A, Kadlec J. 2012. Msl1-mediated dimerization of the dosage compensation complex is essential for male X-chromosome regulation in *Drosophila*. *Molecular cell* **48**: 587-600.
- Hamada FN, Park PJ, Gordadze PR, Kuroda MI. 2005. Global regulation of X chromosomal genes by the MSL complex in *Drosophila melanogaster*. *Genes Dev* **19**: 2289-2294.
- Hattori M, Fujiyama A, Taylor TD, Watanabe H, Yada T, Park HS, Toyoda A, Ishii K, Totoki Y, Choi DK et al. 2000. The DNA sequence of human chromosome 21. *Nature* **405**: 311-319.
- Heard E, Turner J. 2011. Function of the sex chromosomes in mammalian fertility. *Cold Spring Harbor perspectives in biology* **3**: a002675.
- Henikoff S, Henikoff JG, Sakai A, Loeb GB, Ahmad K. 2009. Genome-wide profiling of salt fractions maps physical properties of chromatin. *Genome Res* **19**: 460-469.
- Henry KW, Wyce A, Lo WS, Duggan LJ, Emre NC, Kao CF, Pillus L, Shilatifard A, Osley MA, Berger SL. 2003. Transcriptional activation via sequential histone H2B ubiquitylation and deubiquitylation, mediated by SAGA-associated Ubp8. *Genes Dev* **17**: 2648-2663.
- Herpin A, Scharf M. 2011. Sex determination: switch and suppress. *Current biology : CB* **21**: R656-659.
- Hilfiker A, Hilfiker-Kleiner D, Pannuti A, Lucchesi JC. 1997. mof, a putative acetyl transferase gene related to the Tip60 and MOZ human genes and to the SAS genes of yeast, is required for dosage compensation in *Drosophila*. *The EMBO journal* **16**: 2054-2060.
- Hofer U, Henley JE, Exline CM, Mulhern O, Lopez E, Cannon PM. 2013. Pre-clinical modeling of CCR5 knockout in human hematopoietic stem cells by zinc finger nucleases using humanized mice. *The Journal of infectious diseases* **208 Suppl 2**: S160-164.
- Hsu DR, Meyer BJ. 1994. The dpy-30 gene encodes an essential component of the *Caenorhabditis elegans* dosage compensation machinery. *Genetics* **137**: 999-1018.
- Huppert JL. 2010. Structure, location and interactions of G-quadruplexes. *The FEBS journal* **277**: 3452-3458.

- Ilik IA, Quinn JJ, Georgiev P, Tavares-Cadete F, Maticzka D, Toscano S, Wan Y, Spitale RC, Luscombe N, Backofen R et al. 2013. Tandem stem-loops in roX RNAs act together to mediate X chromosome dosage compensation in *Drosophila*. *Molecular cell* **51**: 156-173.
- Jacobs PA, Baikie AG, Court Brown WM, Strong JA. 1959. The somatic chromosomes in mongolism. *Lancet* **1**: 710.
- James E, Darnell J. 2002. TRANSCRIPTION FACTORS AS TARGETS FOR CANCER THERAPY. *Nature Reviews Cancer* **2**.
- Jans J, Gladden JM, Ralston EJ, Pickle CS, Michel AH, Pferdehirt RR, Eisen MB, Meyer BJ. 2009. A condensin-like dosage compensation complex acts at a distance to control expression throughout the genome. *Genes & development* **23**: 602-618.
- Jeon Y, Lee JT. 2011. YY1 tethers Xist RNA to the inactive X nucleation center. *Cell* **146**: 119-133.
- Jin Y, Wang Y, Johansen J, Johansen KM. 2000. JIL-1, a chromosomal kinase implicated in regulation of chromatin structure, associates with the male specific lethal (MSL) dosage compensation complex. *J Cell Biol* **149**: 1005-1010.
- Johnson NA, Lachance J. 2012. The genetics of sex chromosomes: evolution and implications for hybrid incompatibility. *Annals of the New York Academy of Sciences* **1256**: E1-22.
- Kadlec J, Hallacli E, Lipp M, Holz H, Sanchez-Weatherby J, Cusack S, Akhtar A. 2011. Structural basis for MOF and MSL3 recruitment into the dosage compensation complex by MSL1. *Nat Struct Mol Biol* **18**: 142-149.
- Keene JD, Komisarow JM, Friedersdorf MB. 2006. RIP-Chip: the isolation and identification of mRNAs, microRNAs and protein components of ribonucleoprotein complexes from cell extracts. *Nature protocols* **1**: 302-307.
- Kelley RL, Lee OK, Shim YK. 2008. Transcription rate of noncoding roX1 RNA controls local spreading of the *Drosophila* MSL chromatin remodeling complex. *Mechanisms of development* **125**: 1009-1019.
- Kelley RL, Meller VH, Gordadze PR, Roman G, Davis RL, Kuroda MI. 1999. Epigenetic spreading of the *Drosophila* dosage compensation complex from roX RNA genes into flanking chromatin. *Cell* **98**: 513-522.
- Kelley RL, Wang J, Bell L, Kuroda MI. 1997. Sex lethal controls dosage compensation in *Drosophila* by a non-splicing mechanism. *Nature* **387**: 195-199.
- Keogh MC, Kurdistani SK, Morris SA, Ahn SH, Podolny V, Collins SR, Schuldiner M, Chin K, Punna T, Thompson NJ et al. 2005. Cotranscriptional set2 methylation of histone H3 lysine 36 recruits a repressive Rpd3 complex. *Cell* **123**: 593-605.
- Kharchenko PV, Alekseyenko AA, Schwartz YB, Minoda A, Riddle NC, Ernst J, Sabo PJ, Larschan E, Gorchakov AA, Gu T et al. 2011. Comprehensive analysis of the chromatin landscape in *Drosophila melanogaster*. *Nature* **471**: 480-485.
- Kharchenko PV, Tolstorukov MY, Park PJ. 2008. Design and analysis of ChIP-seq experiments for DNA-binding proteins. *Nat Biotechnol* **26**: 1351-1359.

- Kind J, Vaquerizas JM, Gebhardt P, Gentzel M, Luscombe NM, Bertone P, Akhtar A. 2008. Genome-wide analysis reveals MOF as a key regulator of dosage compensation and gene expression in *Drosophila*. *Cell* **133**: 813-828.
- Klug A. 2010. The discovery of zinc fingers and their development for practical applications in gene regulation and genome manipulation. *Q Rev Biophys* **43**: 1-21.
- Klug A, Schwabe JW. 1995. Protein motifs 5. Zinc fingers. *FASEB journal : official publication of the Federation of American Societies for Experimental Biology* **9**: 597-604.
- Ko-Hsuan Hung YW, Jin Crystal Zhao. 2013. Regulation of Mammalian Gene Dosage by Long Non-coding RNAs. *Biomolecules* **3(1)**: 124-142.
- Kulkarni MM, Booker M, Silver SJ, Friedman A, Hong P, Perrimon N, Mathey-Prevot B. 2006. Evidence of off-target effects associated with long dsRNAs in *Drosophila melanogaster* cell-based assays. *Nat Methods* **3**: 833-838.
- Lai F, Orom UA, Cesaroni M, Beringer M, Taatjes DJ, Blobel GA, Shiekhattar R. 2013. Activating RNAs associate with Mediator to enhance chromatin architecture and transcription. *Nature* **494**: 497-501.
- Langmead B, Trapnell C, Pop M, Salzberg SL. 2009. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol* **10**: R25.
- Larschan E, Alekseyenko AA, Gortchakov AA, Peng S, Li B, Yang P, Workman JL, Park PJ, Kuroda MI. 2007. MSL complex is attracted to genes marked by H3K36 trimethylation using a sequence-independent mechanism. *Mol Cell* **28**: 121-133.
- Larschan E, Alekseyenko AA, Lai WR, Park PJ, Kuroda MI. 2006. MSL complex associates with clusters of actively transcribed genes along the *Drosophila* male X chromosome. *Cold Spring Harbor symposia on quantitative biology* **71**: 385-394.
- Larschan E, Bishop E, Kharchenko P, Park P, Kuroda M. 2011a. X chromosome dosage compensation via enhanced transcription elongation in *Drosophila* males. *Nature* **In press**.
- Larschan E, Bishop EP, Kharchenko PV, Core LJ, Lis JT, Park PJ, Kuroda MI. 2011b. X chromosome dosage compensation via enhanced transcriptional elongation in *Drosophila*. *Nature* **471**: 115-118.
- Larschan E, Soruco MM, Lee OK, Peng S, Bishop E, Chery J, Goebel K, Feng J, Park PJ, Kuroda MI. 2012. Identification of chromatin-associated regulators of MSL complex targeting in *Drosophila* dosage compensation. *PLoS genetics* **8**: e1002830.
- Lee C, Li X, Hechmer A, Eisen M, Biggin MD, Venters BJ, Jiang C, Li J, Pugh BF, Gilmour DS. 2008. NELF and GAGA factor are linked to promoter-proximal pausing at many genes in *Drosophila*. *Mol Cell Biol* **28**: 3290-3300.
- Lee JT. 2012. Epigenetic regulation by long noncoding RNAs. *Science* **338**: 1435-1439.
- Lee JT, Bartolomei MS. 2013. X-Inactivation, Imprinting, and Long Noncoding RNAs in Health and Disease. *Cell* **152**: 1308-1323.

- Lejeune J, Gautier M, Turpin R. 1959. [Study of somatic chromosomes from 9 mongoloid children]. *Comptes rendus hebdomadaires des seances de l'Academie des sciences* **248**: 1721-1722.
- Leppig KA, Distech CM. 2001. Ring X and other structural X chromosome abnormalities: X inactivation and phenotype. *Seminars in reproductive medicine* **19**: 147-157.
- Li F, Parry DA, Scott MJ. 2005. The amino-terminal region of Drosophila MSL1 contains basic, glycine-rich, and leucine zipper-like motifs that promote X chromosome binding, self-association, and MSL2 binding, respectively. *Molecular and cellular biology* **25**: 8913-8924.
- Li F, Schiemann AH, Scott MJ. 2008. Incorporation of the noncoding roX RNAs alters the chromatin-binding specificity of the Drosophila MSL1/MSL2 complex. *Molecular and cellular biology* **28**: 1252-1264.
- Lim CK, Kelley RL. 2012. Autoregulation of the Drosophila Noncoding roX1 RNA Gene. *PLoS Genet* **8**: e1002564.
- Lipeng Wu BMZ, Yanming Wang, Benjamin A. Garcia, Yali Dou. 2011. The RING Finger Protein MSL2 in the MOF Complex Is an E3 Ubiquitin Ligase for H2B K34 and Is Involved in Crosstalk with H3 K4 and K79 Methylation. *Molecular cell* **43**: 132-144.
- Little M, Holmes G, Walsh P. 1999. WT1: what has the last decade told us? *BioEssays : news and reviews in molecular, cellular and developmental biology* **21**: 191-202.
- Lours C, Bardot O, Godt D, Laski FA, Couderc JL. 2003. The Drosophila melanogaster BTB proteins bric a brac bind DNA through a composite DNA binding domain containing a pipsqueak and an AT-Hook motif. *Nucleic acids research* **31**: 5389-5398.
- Lucchesi JC. 1998. Dosage compensation in flies and worms: the ups and downs of X-chromosome regulation. *Current opinion in genetics & development* **8**: 179-184.
- Lucchesi JC, Kelly WG, Panning B. 2005. Chromatin remodeling in dosage compensation. *Annu Rev Genet* **39**: 615-651.
- Lucchesi JMBJC. 1980. Control of X Chromosome transcription by the maleless gene in *Drosophila*. *Nature* **285**: 573-575.
- Lusk RW, Eisen MB. 2013. Spatial promoter recognition signatures may enhance transcription factor specificity in yeast. *PloS one* **8**: e53778.
- Lyman LM, Copps K, Rastelli L, Kelley RL, Kuroda MI. 1997. Drosophila male-specific lethal-2 protein: Structure/function analysis and dependence on MSL-1 for chromosome association. *Genetics* **147**: 1743-1753.
- Maenner S, Muller M, Frohlich J, Langer D, Becker PB. 2013. ATP-dependent roX RNA remodeling by the helicase maleless enables specific association of MSL proteins. *Molecular cell* **51**: 174-184.
- Mank JE. 2009. The W, X, Y and Z of sex-chromosome dosage compensation. *Trends in genetics : TIG* **25**: 226-233.

- McDonel P, Jans J, Peterson BK, Meyer BJ. 2006. Clustered DNA motifs mark X chromosomes for repression by a dosage compensation complex. *Nature* **444**: 614-618.
- Meller VH. 2003. Initiation of dosage compensation in *Drosophila* embryos depends on expression of the roX RNAs. *Mechanisms of development* **120**: 759-767.
- Meller VH, Kuroda MI. 2002. Sex and the single chromosome. *Advances in genetics* **46**: 1-24.
- Meller VH, Rattner BP. 2002a. The roX genes encode redundant male-specific lethal transcripts required for targeting of the MSL complex. *Embo J* **21**: 1084-1091.
- . 2002b. The roX genes encode redundant male-specific lethal transcripts required for targeting of the MSL complex. *The EMBO journal* **21**: 1084-1091.
- Meller VH, Wu KH, Roman G, Kuroda MI, Davis RL. 1997. roX1 RNA paints the X chromosome of male *Drosophila* and is regulated by the dosage compensation system. *Cell* **88**: 445-457.
- Mendjan S, Taipale M, Kind J, Holz H, Gebhardt P, Schelder M, Vermeulen M, Buscaino A, Duncan K, Mueller J et al. 2006. Nuclear pore components are involved in the transcriptional regulation of dosage compensation in *Drosophila*. *Molecular cell* **21**: 811-823.
- Meyer BJ. 1997. Sex Determination and X Chromosome Dosage Compensation.
- . 2010. Targeting X chromosomes for repression. *Current opinion in genetics & development* **20**: 179-189.
- Mito Y, Henikoff JG, Henikoff S. 2007. Histone replacement marks the boundaries of cis-regulatory domains. *Science* **315**: 1408-1411.
- Morra R, Smith ER, Yokoyama R, Lucchesi JC. 2008. The MLE subunit of the *Drosophila* MSL complex uses its ATPase activity for dosage compensation and its helicase activity for targeting. *Molecular and cellular biology* **28**: 958-966.
- Nagaike T, Logan C, Hotta I, Rozenblatt-Rosen O, Meyerson M, Manley JL. 2011. Transcriptional activators enhance polyadenylation of mRNA precursors. *Mol Cell* **41**: 409-418.
- Nguyen DK, Disteche CM. 2006. Dosage compensation of the active X chromosome in mammals. *Nat Genet* **38**: 47-53.
- Nikiforov MA, Chandriani S, Park J, Kotenko I, Matheos D, Johnsson A, McMahon SB, Cole MD. 2002. TRRAP-dependent and TRRAP-independent transcriptional activation by Myc family oncoproteins. *Mol Cell Biol* **22**: 5054-5063.
- Nora EP, Lajoie BR, Schulz EG, Giorgetti L, Okamoto I, Servant N, Piolot T, van Berkum NL, Meisig J, Sedat J et al. 2012. Spatial partitioning of the regulatory landscape of the X-inactivation centre. *Nature* **485**: 381-385.
- Nybakken K, Perrimon N. 2002. Hedgehog signal transduction: recent findings. *Curr Opin Genet Dev* **12**: 503-511.
- Oh H, Park Y, Kuroda MI. 2003. Local spreading of MSL complexes from roX genes on the *Drosophila* X chromosome. *Genes & development* **17**: 1334-1339.
- Onodera O, Roses AD, Tsuji S, Vance JM, Strittmatter WJ, Burke JR. 1996. Toxicity of expanded polyglutamine-domain proteins in *Escherichia coli*. *FEBS letters* **399**: 135-139.

- Orom UA, Derrien T, Beringer M, Gumireddy K, Gardini A, Bussotti G, Lai F, Zytnicki M, Notredame C, Huang Q et al. 2010. Long noncoding RNAs with enhancer-like function in human cells. *Cell* **143**: 46-58.
- Palmer MJ, Richman R, Richter L, Kuroda MI. 1994. Sex-Specific Regulation of the Male-Specific Lethal-1 Dosage Compensation Gene in *Drosophila*. *Gene Dev* **8**: 698-706.
- Paris M, Kaplan T, Li XY, Villalta JE, Lott SE, Eisen MB. 2013. Extensive divergence of transcription factor binding in *Drosophila* embryos with highly conserved gene expression. *PLoS genetics* **9**: e1003748.
- Park SW, Oh H, Lin YR, Park Y. 2010. MSL cis-spreading from roX gene up-regulates the neighboring genes. *Biochemical and biophysical research communications* **399**: 227-231.
- Park Y, Kelley RL, Oh H, Kuroda MI, Meller VH. 2002. Extent of chromatin spreading determined by roX RNA recruitment of MSL proteins. *Science* **298**: 1620-1623.
- Park Y, Mengus G, Bai X, Kageyama Y, Meller VH, Becker PB, Kuroda MI. 2003. Sequence-specific targeting of *Drosophila* roX genes by the MSL dosage compensation complex. *Molecular cell* **11**: 977-986.
- Payer B, Lee JT. 2008. X chromosome dosage compensation: how mammals keep the balance. *Annu Rev Genet* **42**: 733-772.
- Pelassa I, Cora D, Cesano F, Monje FJ, Montarolo PG, Fiumara F. 2014. Association of polyalanine and polyglutamine coiled coils mediates expansion disease-related protein aggregation and dysfunction. *Hum Mol Genet*.
- Pferdehirt RR, Kruesi WS, Meyer BJ. 2011. An MLL/COMPASS subunit functions in the *C. elegans* dosage compensation complex to target X chromosomes for transcriptional regulation of gene expression. *Genes & development* **25**: 499-515.
- Pinter SF, Sadreyev RI, Yildirim E, Jeon Y, Ohsumi TK, Borowsky M, Lee JT. 2012. Spreading of X chromosome inactivation via a hierarchy of defined Polycomb stations. *Genome research* **22**: 1864-1876.
- Prestel M, Feller C, Becker PB. 2010a. Dosage compensation and the global re-balancing of aneuploid genomes. *Genome biology* **11**: 216.
- Prestel M, Feller C, Straub T, Mitlohner H, Becker PB. 2010b. The activation potential of MOF is constrained for dosage compensation. *Molecular cell* **38**: 815-826.
- Raja SJ, Charapitsa I, Conrad T, Vaquerizas JM, Gebhardt P, Holz H, Kadlec J, Fraterman S, Luscombe NM, Akhtar A. 2010. The nonspecific lethal complex is a transcriptional regulator in *Drosophila*. *Mol Cell* **38**: 827-841.
- Ramadan N, Flockhart I, Booker M, Perrimon N, Mathey-Prevot B. 2007. Design and implementation of high-throughput RNAi screens in cultured *Drosophila* cells. *Nat Protoc* **2**: 2245-2264.
- Rinn JL, Chang HY. 2012. Genome regulation by long noncoding RNAs. *Annual review of biochemistry* **81**: 145-166.
- Robasky K, Bulyk ML. 2011. UniPROBE, update 2011: expanded content and search tools in the online database of protein-binding microarray data on protein-DNA interactions. *Nucleic Acids Res* **39**: D124-128.

- Salz HK, Erickson JW. 2010. Sex determination in *Drosophila*: The view from the top. *Fly (Austin)* **4**: 60-70.
- Schulz EG, Heard E. 2013. Role and control of X chromosome dosage in mammalian development. *Current opinion in genetics & development*.
- Sekido R, Lovell-Badge R. 2009. Sex determination and SRY: down to a wink and a nudge? *Trends in genetics : TIG* **25**: 19-29.
- Shi X, Finkelstein A, Wolf AJ, Wade PA, Burton ZF, Jaehning JA. 1996. Paf1p, an RNA polymerase II-associated factor in *Saccharomyces cerevisiae*, may have both positive and negative roles in transcription. *Mol Cell Biol* **16**: 669-676.
- Shiro Iuchi NK. 2005. *Zinc Finger Proteins: From Atomic Contact to Cellular Function*. Landes Bioscience / Eureka.com, Kluwer Academic / Plenum Publishers 233 Spring Street New York, New York U.S.A. 100013.
- Shogren-Knaak M, Ishii H, Sun JM, Pazin MJ, Davie JR, Peterson CL. 2006. Histone H4-K16 acetylation controls chromatin structure and protein interactions. *Science* **311**: 844-847.
- Shogren-Knaak M, Peterson CL. 2006. Switching on chromatin: mechanistic role of histone H4-K16 acetylation. *Cell Cycle* **5**: 1361-1365.
- Simon MD, Pinter SF, Fang R, Sarma K, Rutenberg-Schoenberg M, Bowman SK, Kesner BA, Maier VK, Kingston RE, Lee JT. 2013. High-resolution Xist binding maps reveal two-step spreading during X-chromosome inactivation. *Nature*.
- Smith ER, Allis CD, Lucchesi JC. 2001. Linking global histone acetylation to the transcription enhancement of X-chromosomal genes in *Drosophila* males. *J Biol Chem* **276**: 31483-31486.
- Smith ER, Pannuti A, Gu W, Steurnagel A, Cook RG, Allis CD, Lucchesi JC. 2000. The *drosophila* MSL complex acetylates histone H4 at lysine 16, a chromatin modification linked to dosage compensation. *Molecular and cellular biology* **20**: 312-318.
- Soruco MM, Chery J, Bishop EP, Siggers T, Tolstorukov MY, Leydon AR, Sugden AU, Goebel K, Feng J, Xia P et al. 2013. The CLAMP protein links the MSL complex to the X chromosome during *Drosophila* dosage compensation. *Genes & development* **27**: 1551-1556.
- Spierer A, Seum C, Delattre M, Spierer P. 2005. Loss of the modifiers of variegation Su(var)3-7 or HP1 impacts male X polytene chromosome morphology and dosage compensation. *J Cell Sci* **118**: 5047-5057.
- Straub T, Grimaud C, Gilfillan GD, Mitterweger A, Becker PB. 2008. The chromosomal high-affinity binding sites for the *Drosophila* dosage compensation complex. *PLoS genetics* **4**: e1000302.
- Sunde JMMaM. 2002. Zinc Fingers--Folds for Many Occasions. *IUBMB Life* **54**: 351-355.
- Sural TH, Peng S, Li B, Workman JL, Park PJ, Kuroda MI. 2008. The MSL3 chromodomain directs a key targeting step for dosage compensation of the *Drosophila melanogaster* X chromosome. *Nature structural & molecular biology* **15**: 1318-1325.

- Tanny JC, Erdjument-Bromage H, Tempst P, Allis CD. 2007. Ubiquitylation of histone H2B controls RNA polymerase II transcription elongation independently of histone H3 methylation. *Genes Dev* **21**: 835-847.
- Thomas CJ, Kotova E, Andrade M, Adolf-Bryfogle J, Glaser R, Regnard C, Tulin AV. 2014. Kinase-Mediated Changes in Nucleosome Conformation Trigger Chromatin Decondensation via Poly(ADP-Ribosyl)ation. *Molecular cell* **53**: 831-842.
- Thomas Cremer MC. 2010. Chromosome Territories. *Cold Spring Harbor perspectives in biology*.
- Thomas S, Li XY, Sabo PJ, Sandstrom R, Thurman RE, Canfield TK, Giste E, Fisher W, Hammonds A, Celniker SE et al. 2011. Dynamic reprogramming of chromatin accessibility during Drosophila embryo development. *Genome biology* **12**: R43.
- Vaquerizas JM, Kummerfeld SK, Teichmann SA, Luscombe NM. 2009. A census of human transcription factors: function, expression and evolution. *Nature reviews Genetics* **10**: 252-263.
- Vaquero A, Blanch M, Espinas ML, Bernues J. 2008a. Activation properties of GAGA transcription factor. *Bba-Gene Regul Mech* **1779**: 312-317.
- . 2008b. Activation properties of GAGA transcription factor. *Biochimica et biophysica acta* **1779**: 312-317.
- Villa R, Forne I, Muller M, Imhof A, Straub T, Becker PB. 2012. MSL2 combines sensor and effector functions in homeostatic control of the Drosophila dosage compensation machinery. *Molecular cell* **48**: 647-654.
- Voss TC, Hager GL. 2014. Dynamic regulation of transcriptional states by chromatin and transcription factors. *Nature reviews Genetics* **15**: 69-81.
- Wang CI, Alekseyenko AA, LeRoy G, Elia AE, Gorchakov AA, Britton LM, Elledge SJ, Kharchenko PV, Garcia BA, Kuroda MI. 2013a. Chromatin proteins captured by ChIP-mass spectrometry are linked to dosage compensation in Drosophila. *Nature structural & molecular biology* **20**: 202-209.
- . 2013b. Chromatin proteins captured by ChIP-mass spectrometry are linked to dosage compensation in Drosophila. *Nat Struct Mol Biol*.
- Wells MB, Snyder MJ, Custer LM, Csankovszki G. 2012. Caenorhabditis elegans dosage compensation regulates histone H4 chromatin state on X chromosomes. *Molecular and cellular biology* **32**: 1710-1719.
- Worringer KA, Panning B. 2007. Zinc finger protein Zn72D promotes productive splicing of the maleless transcript. *Mol Cell Biol* **27**: 8760-8769.
- Wu L, Zee BM, Wang Y, Garcia BA, Dou Y. 2011. The RING Finger Protein MSL2 in the MOF Complex Is an E3 Ubiquitin Ligase for H2B K34 and Is Involved in Crosstalk with H3 K4 and K79 Methylation. *Mol Cell* **43**: 132-144.
- Xiao-Yong Li ST, Peter J Sabo, Michael B Eisen, John A Stamatoyannopoulos and Mark D Biggin. 2011. The role of chromatin accessibility in directing the widespread, overlapping patterns of Drosophila transcription factor binding. *Genome biology* **12**.

- Yonker SA, Meyer BJ. 2003. Recruitment of *C. elegans* dosage compensation proteins for gene-specific versus chromosome-wide repression. *Development* **130**: 6519-6532.
- Zhao J, Sun BK, Erwin JA, Song JJ, Lee JT. 2008. Polycomb proteins targeted by a short repeat RNA to the mouse X chromosome. *Science* **322**: 750-756.
- Zhou K, Kuo WH, Fillingham J, Greenblatt JF. 2009. Control of transcriptional elongation and cotranscriptional histone modification by the yeast BUR kinase substrate Spt5. *Proc Natl Acad Sci U S A* **106**: 6956-6961.
- Zhou T, Yang L, Lu Y, Dror I, Dantas Machado AC, Ghane T, Di Felice R, Rohs R. 2013. DNASHape: a method for the high-throughput prediction of DNA structural features on a genomic scale. *Nucleic acids research* **41**: W56-62.