

The CLAMP protein is essential in *Drosophila* in mediating dosage compensation through increased targeting to the X-chromosome and during development as a non-sex specific transcription factor.

By

Marcela M.L Soruco

B.S, Biology, San Diego State University, 2001

M.S. Biology, New York University, 2006

Thesis

Submitted in partial fulfillment of the requirements for the for the Degree of
Doctorate of Philosophy in the Department of Molecular and Cellular Biology and
Biochemistry at Brown University,

Providence, RI

May 2014

© Copyright 2014 by Marcela M.L. Soruco

NOTE: The year of copyright is the same as the year the degree is conferred.

This dissertation, by Marcela M.L. Soruco, 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_____ Advisor (Signature)_____

Advisor (Print)_____

Recommended to the Graduate Council

Date_____ Reader (Signature)_____

Reader (Print)_____

Date_____ Reader (Signature)_____

Reader (Print)_____

Date_____ Reader (Signature)_____

Reader (Print)_____

Date_____ Reader (Signature)_____

Reader (Print)_____

Approved by the Graduate Council

Date_____ Dean (Signature)_____

Dean (Print)_____

Marcela Soruco

Providence, RI

Email: Marcela_Soruco@brown.edu

Cell: 415.328.5988

Curriculum Vitae

Education

Brown University, Providence, RI 2009-present
Molecular Biology, Cellular Biology and Biochemistry Graduate Program

New York University, New York, NY 2006-2009
M.S Biology

San Diego State University, San Diego, CA 1996-2001
B.S. Biology, Minor Art Applied Design

Grants

National Institute of Health, F31 Fellowship, Brown University 2011-present

Current Research

Graduate Thesis Research 2010-present
MCB PhD Program, Brown University, Providence RI
Dr. Erica Larschan, Drosophila Dosage Compensation Laboratory
Modeling the mechanism involved in identifying the X-chromosome in *Drosophila* males resulting in the up-regulation of X-linked genes. Characterizing the synergistic interaction between the Male Specific Lethal (MSL) Complex and the novel CLAMP protein leading to the first step of X-identification.

Publications

Soruco, MM*, Chery*, J, Bishop, E, Siggers, T, Tolstorukov, M, Leydon, A, Sugden, A, Goebel, K, Feng, J, Xia, P, Vedenko, A, Bulyk, M, Park, PJ, Larschan, E. "The CLAMP protein links the MSL complex to the X chromosome during *Drosophila* dosage compensation." Genes Dev 2013

* These authors contributed equally to the work

Larschan, E, **Soruco, MM**, Lee, OK, Peng, S, Bishop, E, Chery, J, Goebel, K, Feng, J, Park, PJ, Kuroda, MI. "A genome-wide screen to identify essential regulators of dosage compensation." PLoS Genet 2012

Honors and Awards

Awarded Student Service Award Annual MCB Retreat 2013
MCB PhD Program, Brown University, Providence, RI

Awarded Student Talk Annual MCB Retreat 2013
MCB PhD Program, Brown University, Providence, RI

National Institute of Health, T32 Training Grant, Brown University 2009-2010
Masters Poster Presentation, Second Place 2009

NYU Master Poster Presentation, New York University, NY
Marin Education Fund Scholarship 1996-1998

Related Research Experience

- Rotation Research 2010
MCB PhD Program, Brown University, Providence RI
Dr. William Fairbrother, RNA Splicing Laboratory
Topic: "Identifying candidate proteins that enhance or suppress the binding of Polypyrimidine Tract Binding (PTB) protein involved in neuronal alternative splicing."
- Rotation Research 2010
MCB PhD Program, Brown University, Providence RI
Dr. Alison DeLong, Plant Ethylene Signaling Laboratory
Topic: "Gene expression analysis of the *rcn1* mutant resulting in increased ethylene production in *Arabidopsis*."
- Rotation Research 2009
MCB PhD Program, Brown University, Providence RI
Dr. Mark Johnson, Pollen Genetics Laboratory
Topic: "Gene expression validation of candidate Toll Interleukin Leucine Rich Repeat proteins induced in the male gamete during fertilization in *Arabidopsis*."
- Masters Thesis Research 2007-2009
Center for Genomics and Systems Biology
Biology Department, New York University, New York, NY
Dr. Gloria Coruzzi, Plant Molecular Biology Laboratory
Thesis: "A Genetic and Genomic Approach to Improve Nitrogen Efficiency."
Identified and tested putative genes involved in increased levels of nitrogen assimilation in *Arabidopsis*.

Teaching Experience

- Summer @ Brown, Brown University, Providence, RI 2013
Graduate Teaching Assistant for Pre-college Course
From Mad Cow to Alzheimer's: The Biological Basis for Disease Course
Assisted in laboratory section.
- Summer @ Brown, Brown University, Providence, RI 2012
Graduate Instructor for Pre-college Course
From Mad Cow to Alzheimer's: The Biological Basis for Disease Course
Developed and taught lecture and laboratory section.
- Brown University, Providence, RI 2011
Graduate Teaching Assistant for Undergraduate Course
Cellular Physiology and Biophysics Course
Assisted in lecture course.
- Summer @ Brown, Brown University, Providence, RI 2010-2011
Graduate Teaching Assistant for Pre-college Course
From Mad Cow to Alzheimer's: The Biological Basis for Disease Course
Assisted in laboratory section.

Professional Experience

- Laboratory Technician 2007-2009
Center for Genomics and Systems Biology, Dr. Gloria Coruzzi
Biology Department, New York University, New York, NY
Assisted Post Doctoral Fellows in research.

Laboratory Technician	2006
Undergraduate Teaching Laboratory	
Biology Department, New York University, New York, NY	
Prepared laboratory materials for requisite biology course.	
DNA Synthesis Specialist	2002-2005
Liquid Handling Group, Biosearch Technologies Inc., Novato, CA	
Developed oligonucleotide assay synthesis protocols and programmed robotics for liquid handling. Managed part-time employees.	

Conferences

FASEB Conference Chromatin and Epigenetics, Nassau Bahamas	
Oral Presentation	2013
Soruco, MM , Chery, J, Bishop, E, Siggers, T, Tolstorukov, M, Leydon, A, Sugden, A, Goebel, K, Feng, J, Xia, P, Vedenko, A, Bulyk, M, Park, PJ, Larschan, E	
"Synergistic interactions between MSL complex and the CLAMP protein regulate <i>Drosophila</i> dosage compensation"	
Drosophila Research Conference, Washington DC	
Oral Presentation	2013
Soruco, MM , Chery, J, Bishop, E, Siggers, T, Tolstorukov, M, Leydon, A, Sugden, A, Goebel, K, Feng, J, Xia, P, Vedenko, A, Bulyk, M, Park, PJ, Larschan, E	
"Synergistic interactions between MSL complex and the CLAMP protein regulate <i>Drosophila</i> dosage compensation"	
St. Jude Children's Research Hospital, National Graduate Student Symposium, Memphis, Tennessee	
Selected for Oral Presentation, declined	2013
Society for the Advancement of Chicanos and Native American National Conference, Seattle WA	
Oral Presentation	2012
Soruco, MM , Chery, J, Bishop, E, Siggers, T, Tolstorukov, M, Leydon, A, Sugden, A, Goebel, K, Feng, J, Xia, P, Vedenko, A, Bulyk, M, Park, PJ, Larschan, E	
"Synergistic interactions promote chromosome-specificity during <i>Drosophila</i> dosage compensation"	
Gordon Research Conference, Lucca, Italy	
Poster Presentation	2012
Soruco, MM , Chery, J, Bishop, E, Siggers, T, Tolstorukov, M, Leydon, A, Sugden, A, Goebel, K, Feng, J, Xia, P, Vedenko, A, Bulyk, M, Park, PJ, Larschan, E.	
"Synergistic interactions promote chromosome-specificity during <i>Drosophila</i> dosage compensation"	
Summer Symposium in Molecular Biology Conference, Penn State University	
Poster Presentation	2011
Soruco, M, Bishop, E, Dong, L, Park, PJ, Larschan, E.	
"A positive feedback mechanism contributes to X chromosome identification during <i>Drosophila</i> dosage compensation"	
Drosophila Research Conference, San Diego, CA	
Poster Presentation	2011
Soruco, M, Peng, S, Dong, L, Park, PJ, Larschan, E.	
"A positive feedback mechanism contributes to X chromosome identification during <i>Drosophila</i> dosage compensation"	

Departmental Service

Admissions Committee MCB PhD Program Brown University	2012-2013
Project ARISE: Advancing Rhode Island Science Education Providing high school teachers with molecular biology equipment	2012-2013
Recruitment Representative MCB PhD Program Brown University SACNAS National Conference, various locations	2011-2013

Additional Professional Training

Center for Environmental Research and Conservation Columbia University, New York, NY Certificate Conservation Biology	2006
University of Granada, Granada, Spain Certificate Spanish Language	2000

Professional Associations

Society for the Advancement of Chicanos and Native Americans in Science Treasurer, Brown University Chapter	2014-present
President, Brown University Chapter	2013-2014
Vice President, Brown University Chapter	2012-2013
Genetics Society of America	2011
American Association for the Advancement of Science	2011

References

Dr. Erica Larschan
Assistant Professor of Biology
Molecular Biology, Cell Biology, & Biochemistry
Brown University, Providence, RI
Erica_Larschan@brown.edu
(401) 863-1070

Dr. Mark Johnson
Associate Professor of Biology
Molecular Biology, Cell Biology, & Biochemistry
Brown University, Providence, RI
Mark_Johnson_1@brown.edu
(401) 863-3556

Dr. Robert Reenan
Professor of Biology
Molecular Biology, Cell Biology, & Biochemistry
Brown University, Providence, RI
Robert_Reenan@brown.edu
(401) 863-6353

Dr. Nicola Neretti
Assistant Professor of Biology
Molecular Biology, Cell Biology, & Biochemistry
Brown University, Providence, RI
Nicola_Neretti@brown.edu

(401) 863-6702

Dr. Alison DeLong
Associate Professor of Biology
Molecular Biology, Cell Biology, & Biochemistry
Brown University, Providence, RI
Alison_DeLong@brown.edu
(401) 863-3888

Dr. Judith Bender
Professor of Biology
Molecular Biology, Cell Biology, & Biochemistry
Judith_Bender@brown.edu
(401) 863-6238

Dr. William Fairbrother
Associate Professor of Biology
Molecular Biology, Cell Biology, & Biochemistry
Brown University, Providence, RI
William_Fairbrother@brown.edu
(401) 863-6215

Acknowledgements

I would like to begin by thanking my mentor, Dr. Erica Larschan. Her limitless enthusiasm and positive outlook permeated the workplace and made it a joyful place in which to conduct research. Erica is the type of mentor that fully supports her students through both good results and bad. Her infinite support created an environment where I always felt valued, appreciated and respected. This supportive setting gave me the intellectual freedom to be inquisitive with my research interests, which in turn fueled my passion for performing new experiments. In addition to being a considerate mentor, she is one of the most thoughtful and rigorous scientists. Under her comprehensive scientific guidance, I have received the finest training needed for future contributions to the greater scientific community. She is one of my most respected role models and I feel lucky to have been given the opportunity to work and learn in her lab.

Next, I would like to thank my committee members, Dr. Mark Johnson, Dr. Robert Reenan, and Dr. Nicola Neretti. To Dr. Johnson, I am thankful for the probing questions that helped me consider aspects of my project more thoroughly. I am thankful for the expansive knowledge of transcription and fly genetics from Dr. Reenan that helped me interpret important results. I would like to thank Dr. Neretti for his critical eye in dissecting my bioinformatics analysis in an effort to relay concepts in a clear and concise manner. With their guidance

and insight, I was able accomplish all that I had hoped during my PhD training at Brown University.

Additionally, I would like to thank faculty and support staff of the MCB program. I would like to thank Dr. Judith Bender for reading and critiquing numerous grant applications and manuscripts. I would like to thank Dr. Alison DeLong for her continued support of the SACNAS Chapter at Brown University. I would like to thank Dr. Tricia Serio for being a advocating for graduates student benefits. I would also like to thank Elaine Butler for her administrative support during the many milestone presentations and conference preparations throughout my time in the MCB program.

I would also like to thank the members of the Larschan lab, both current and past. They are both colleagues and friends and it was a pleasure to work along side them. I would like to give a special thanks to Jessica Feng, who began working in the lab as very focused undergraduate student. As her graduate student mentor, I saw first hand her dedication and attention to detail. Now serving as our lab manager, she has kept the lab running smoothly with her organizational skills that at times are pure magic.

Importantly, I would like to thank the many friends I have made at Brown University. Without their support, patient ears and laughter, I would not have

been able to manage my way through some difficult hurdles I know they are lifelong friends as well as amazing people who inspire me on a daily basis.

Last, but definitely not least, I would like to thank my family, who have encouraged my pursuit of a career in the sciences since my childhood. There are not enough words to describe how much their support has meant to me throughout the long journey of graduate school. To Antonio, Carmen, Monica, Alex, Peter, Carla and Victoria: thank you and I love you very much!!! I also want to include my very large extended family of aunts, uncles and cousins, which are too numerous to name. Thank you!

Table of Contents

Title page	i
Copyright page	ii
Signature page	iii
Curriculum Vitae	iv
Acknowledgements	ix
Table of Contents	xii
List of Tables	xx
List of Figures	xxi

Chapter 1: Introduction

A new player in X-identification: The CLAMP protein is a key factor in

<i>Drosophila</i> dosage compensation	1
Abstract	2
Introduction	3
Sex chromosome dosage compensation is an essential process	3
<i>Drosophila</i> as a model for X-identification	6
Protein components of the MSL complex	7
RNA components of the MSL complex	10
High affinity sites and MRE DNA elements	11
Mechanisms for spreading of the MSL complex and increasing transcription	

of X-linked genes	13
The CLAMP protein: a newly discovered regulator of <i>Drosophila</i> dosage compensation	14
Unique X-chromosome characteristics	15
The role of nuclear organization in <i>Drosophila</i> dosage compensation	16
Discussion	17
Acknowledgements	19
Figure 1: Dosage compensation patterns across species	20
Figure 2: The MSL complex and Multi-step targeting	21
Figure 3: The role of CLAMP in X-targeting	22
Figure 4: Unique X-characteristics and nuclear organization	23
References	24

Chapter 2

Identification of chromatin-associated regulators of MSL complex

targeting in <i>Drosophila</i> dosage compensation	37
Abstract	39
Introduction	41
Results	43
A genome-side RNAi screen identifies genes that regulate and	
MSL-dependent reporter in male SL2 cells	43
Selection of NSL complex , PAF complex, and CG1832 as candidate	

direct regulators of MSL targeting or function	48
CG1832 and NSL complex modulate MSL complex recruitment to endogenous CES loci <i>in vivo</i>	51
Genome-wide occupancy of NSL components and CG1832 support a direct role in MSL targeting	52
Discussion	54
Methods	58
Acknowledgements	64
Figure 1: An RNAi screening system in <i>Drosophila</i> SL2 cells to identify regulators of MSL complex recruitment and function	65
Figure 2: Candidate regulators of MSL complex identified from the genome-wide RNAi screen	67
Figure 3: NSL1 and CG 1832 contribute to MSL complex recruitment ...	69
Figure 4: NSL complex and CG1832 enrichment at CES loci	71
Table 1: Gene ontology categories that were enriched in RNAi screen ..	73
Figure S1: mRNA levels of MSL complex components and candidates after CG1832, Nsl1, and Paf1 RNAi treatments.	74
Figure S2: Western blots for CG1832 RNAi treatment and antibody validation	75
Figure S3: Validation of CG1832 antibody by ChIP at CES loci	76
Figure S4: Immunostaining of polytene chromosomes	77
Table S1: Initial candidates identified from a genome-wide RNAi screen to identify novel regulators of dosage compensation	79

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

that passed secondary screens	80
References	81

Chapter 3

The CLAMP protein links the MSL complex to the X-chromosome during

***Drosophila* dosage compensation**

Abstract

Introduction

Results

CLAMP directly recognizes MREs to tether the MSL complex to the

X-chromosome

CLAMP is required for MSL complex recruitment to the male

X-chromosome *in vivo*

CLAMP is enriched on the X-chromosome independent of the

MSL complex

Discussion

Acknowledgements

Methods

Figure 1: CLAMP localizes precisely at MREs at high levels when

co-localized with the MSL complex

Figure 2: CLAMP directly binds to MRE sequences

Figure 3: CES have three classes of CLAMP occupancy patterns	125
Figure 4: Three subclasses of CESs differ in their number of tandem MREs and two-dimensional clustering along the X chromosome	127
Supplemental Figure S1: Global comparisons of CLAMP and MSL complex occupancy at MREs	129
Supplemental 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	131
Supplemental Figure S3: CLAMP is required for roX gene transcription and MSL localization to the X	133
Supplemental Figure S4: Comparison of CLAMP binding characteristics in the presence and absence of MSL complex	136
References	138

Chapter 4

The non-sex specific role of CLAMP as a genome-wide transcription

factor	143
Abstract	144
Introduction	146
Results	147
Homozygous insertion line shows decreased CLAMP mRNA and protein levels	148

Homozygous insertion line shows additional dosage compensation defects	149
P-element insertion causes a decrease in viability of homozygous flies	150
Decreases in CLAMP expression causes severe motor defects in males and reduced flying in females	152
mRNA-seq in male and female cells shows that the loss of CLAMP shows positive and negative affects on gene expression	152
Discussion	155
Acknowledgements	158
Methods	159
Figure 1: Overview of CLAMP locus and P-element insertion	164
Figure 2: mRNA and protein levels are reduced in the homozygous insertion line	165
Figure 3: Reduced survival of homozygous insertion flies	167
Figure 4: Severe motor defects in surviving homozygous flies at 18 degrees	169
Figure 5: CLAMP RNAi reveals both positive and negative regulation genome-wide	170
Figure 6: CLAMP regulates genes involved in brain development and morphogenesis	171
Figure 7: Putative CLAMP enhancer driving GFP expression in the fly brain	172

References	173
------------------	-----

Chapter 5: Discussion

Co-evolution of the sex chromosomes and dosage compensation

components	178
Discussion	179
References	185

Appendix 1

Induction of MSL complex in female cells as a tool to identify initial sites of X-targeting during Drosophila dosage compensation

Introduction	191
Results	194
Sex Lethal RNAi stabilizes MSL complex and increases	
H4K16ac	194
CLAMP is enriched at high affinity sites after MSL complex	
induction	195
Sex lethal RNAi time course shows early induction of MSL2	
RNA	196
Discussion	197
Methods	199

Figure 1: Sex Lethal RNAi in females induces MSL2 expression with possible models of X-chromosome targeting	207
Figure 2: Sex Lethal RNAi after 6 days increases MSL2 and <i>roX</i> RNA levels and increases functional MSL complex at high affinity sites	208
Figure 3: CLAMP enrichment is increased after Sxl RNAi	209
Figure 4: Sex Lethal RNAi time course over 6 days shows early induction of MSL complex	210
References	211

List of Tables

Chapter 2

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

Table 1: Gene ontology categories that were enriched in RNAi screen ... 73

Table S1: Initial candidates identified from a genome-wide RNAi screen
to identify novel regulators of dosage compensation 79

Table S2: Final candidates identified from a genome-wide RNAi screen
that passed secondary screens 80

List of Figures

Chapter 1: Introduction

A new player in X-identification: The CLAMP protein is a key factor in

<i>Drosophila</i> dosage compensation	1
Figure 1: Dosage compensation patterns across species	20
Figure 2: The MSL complex and Multi-step targeting	21
Figure 3: The role of CLAMP in X-targeting	22
Figure 4: Unique X-characteristics and nuclear organization	23

Chapter 2

Identification of chromatin-associated regulators of MSL complex

targeting in *Drosophila* dosage compensation

37

Figure 1: An RNAi screening system in <i>Drosophila</i> SL2 cells to identify regulators of MSL complex recruitment and function	65
---	----

Figure 2: Candidate regulators of MSL complex identified from the genome-wide RNAi screen	67
--	----

Figure 3: NSL1 and CG 1832 contribute to MSL complex recruitment ...	69
--	----

Figure 4: NSL complex and CG1832 enrichment at CES loci	71
---	----

Figure S1: mRNA levels of MSL complex components and candidates after CG1832, Nsl1, and Paf1 RNAi treatments.	74
---	----

Figure S2: Western blots for CG1832 RNAi treatment and antibody	
---	--

validation	75
Figure S3: Validation of CG1832 antibody by ChIP at CES loci	76
Figure S4: Immunostaining of polytene chromosomes	77

Chapter 3

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

Figure 1: CLAMP localizes precisely at MREs at high levels when co-localized with the MSL complex	122
Figure 2: CLAMP directly binds to MRE sequences	124
Figure 3: CES have three classes of CLAMP occupancy patterns	125
Figure 4: Three subclasses of CESs differ in their number of tandem MREs and two-dimensional clustering along the X chromosome	127
Supplemental Figure S1: Global comparisons of CLAMP and MSL complex occupancy at MREs	129
Supplemental 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	131
Supplemental Figure S3: CLAMP is required for roX gene transcription and MSL localization to the X	133
Supplemental Figure S4: Comparison of CLAMP binding characteristics in the presence and absence of MSL complex	136

Chapter 4

The non-sex specific role of CLAMP as a genome-wide transcription

factor	143
Figure 1: Overview of CLAMP locus and P-element insertion	164
Figure 2: mRNA and protein levels are reduced in the homozygous insertion line	165
Figure 3: Reduced survival of homozygous insertion flies	167
Figure 4: Severe motor defects in surviving homozygous flies at 18 degrees	169
Figure 5: CLAMP RNAi reveals both positive and negative regulation genome-wide	170
Figure 6: CLAMP regulates genes involved in brain development and morphogenesis	171
Figure 7: Putative CLAMP enhancer driving GFP expression in the fly brain	172

Appendix 1

Induction of MSL complex in female cells as a tool to identify initial sites of X-targeting during Drosophila dosage compensation

190

Figure 1: Sex Lethal RNAi in females induces MSL2 expression with

possible models of X-chromosome targeting	207
Figure 2: Sex Lethal RNAi after 6 days increases MSL2 and <i>roX</i> RNA levels	
and increases functional MSL complex at high affinity sites	208
Figure 3: CLAMP enrichment is increased after Sxl RNAi	209
Figure 4: Sex Lethal RNAi time course over 6 days shows early induction	
of MSL complex	210

Chapter 1: Introduction

A new player in X-identification: The CLAMP protein is a key factor in *Drosophila* dosage compensation

Marcela M.L Soruco¹ and Erica Larschan¹

¹Department of Molecular Biology, Cellular Biology and Biochemistry, Brown University, Providence, RI, 02912, USA

This manuscript was prepared for submission to Chromosome Research as an invited review article. I wrote this manuscript and prepared figures. Erica Larschan provided comments and edits.

Abstract

In a process conserved from flies to mammals, the X chromosome is targeted for regulation to equalize transcription between males (XY) and females (XX). This process is termed dosage compensation and in *Drosophila*, the key regulator of dosage compensation is the Male Specific Lethal (MSL) complex. The MSL complex specifically identifies the X chromosome only in males and increases transcription of X-linked genes approximately two-fold. Targeting of the X chromosome is known to take place in multiple steps. First, there is a sequence dependent step involving 21 bp GA-rich *cis*-acting MSL recognition elements (MRE) that is enriched on the X chromosome. Importantly, these sequences are found within MSL complex high affinity sites. Two of these sites, the *roX* loci, encode the two long, non-coding RNA components of MSL complex. It is not known how the MRE sequences and non-coding RNAs are integrated to generate X-specificity because the MREs are not X-specific and the MSL components are not sufficient to target the MRE sequence. Recently, we identified a previously unstudied transcription factor, CLAMP, as the key DNA binding factor that directly recognizes MRE sequences. Though CLAMP is also found throughout the genome, it is enriched at MSL high affinity sites on the X and at the *roX* loci. CLAMP enrichment could further identify these sites as “seed sites” for MSL recruitment. Here we review the components of *Drosophila* dosage compensation that could facilitate the first steps of X-identification.

Introduction

Concurrent with the evolution of sex chromosomes, an imbalance in gene copy number developed between heterogametic and homogametic sexes (Charlesworth 1978). Therefore, species-specific mechanisms co-evolved to equalize this potentially lethal gene dose imbalance (Livernois et al. 2012). These mechanisms, collectively called dosage compensation, adjust the expression level of one or both targeted chromosome. These changes in expression simultaneously normalize transcriptional output of important and essential gene families across the genome (Disteche 2012). Though the mechanism of dosage compensations differs between various species, there are distinct similarities, which highlight the importance of this conserved process.

Sex chromosome dosage compensation is an essential process

In mammalian species in which males are XY and females are XX, the X chromosome is up-regulated in both the sexes in a process mediated by the H4K16ac chromatin mark (Deng et al. 2013). In an additional step, one X chromosome is randomly inactivated in females (Figure 1). Much research has focused on the mechanism of X-inactivation in mammals, where an increase in transcription of the *Xist* non-coding RNA identifies the future silenced X in a two-step model (Jonkers et al. 2008). Initial spreading of *Xist* from the X inactivation center (XIC) first targets gene rich regions followed by gene poor regions aided

by long interspersed nuclear element (LINE) retrotransposons (Chow et al. 2010; Simon et al. 2013). Initial transcription of *Xist* is regulated by an additional, but activating, non-coding RNA, JPX (Tian et al. 2010). *Xist* RNA is tethered to the X chromosome by the zinc finger protein YY1 (Jeon and Lee 2011) and the CTCF DNA binding proteins (Chao et al. 2002; Donohoe et al. 2007). Maintenance of silencing is accomplished by the Polycomb Repressive 2 Complex (PCR2), which deposits the histone modification H3K27me3, leading to the condensation and silencing of X-linked genes (Plath et al. 2003; Simon et al. 2013).

In *C. elegans*, the X chromosome is up-regulated in hermaphroditic females (XX) and males (X0) in a process not well understood (Deng et al. 2011). Then, in hermaphrodites, the expression of both X chromosomes is down-regulated by half (Figure 1). Repression of the X chromosome is initiated by homolog pairing and counting of the X to autosome (X:A) ratio (Nicoll et al. 1997). The Dosage Compensation Complex (DCC), which shares components with the condensin complex responsible for compaction and segregation of mitotic and meiotic chromosome, mediates repression of both X chromosomes by 50% (Meyer 2010; Wood et al. 2010).

Sequence specific binding of the DCC also takes place in a two-step model. First, sequence-specific targeting of sites containing Recruitment Element on X (*rex*) identifies high affinity sites on both X chromosomes. Then the DCC spreads to Dependent on X (*dox*) sites within gene promoters, resulting in X-

specific targeting at active genes (Jans et al. 2009; Pferdehirt et al. 2011). Recently, a decrease in RNA Pol II recruitment to the transcription start site of genes on the dosage compensated X chromosome was identified as the mechanism by which the DCC is able to reduce expression of X-linked genes (Kruesi et al. 2013).

In *Drosophila*, the single X chromosome in males (XY) is up-regulated approximately two-fold while no regulation takes place in females (XX) (Larschan et al. 2011) (Figure 1). The key regulator of dosage compensation in *Drosophila* is the Male Specific Lethal (MSL) complex, which is only present in males due to translational regulation of the core component MSL2 (Kelley et al. 1997) (Figure 2A). The MSL complex is thought to initially target two sites on the X chromosome, the RNA on the X (*roX*) loci, which encode the two non-coding RNA components of the complex, during co-transcriptional incorporation of the RNAs into the complex (Kelley et al. 1999; Bai et al. 2004) (Figure 2B).

Spreading of the MSL complex from the *roX* loci is facilitated by a series of evenly spaced high affinity sites. These sites are approximately 1.5 kb in length and are able to recruit very high levels of MSL complex (Alekseyenko et al. 2008; Sural et al. 2008). These sites contain smaller 21-bp MSL Recognition Element (MRE) sequence motifs necessary for MSL recruitment. MREs are found throughout the genome, but are 2-fold enriched on the X (Alekseyenko et al. 2008). The MSL complex then spreads to active genes in a sequence

independent step that involves recognizing the H3K36me3 histone modification that marks the gene bodies of active genes (Larschan et al. 2007; Sural et al. 2008) (Figure 2B). The MSL complex deposits the H4K16ac mark, which is likely to cause increased transcription elongation through elevated release of paused RNA Pol II into gene bodies (Gelbart et al. 2009; Larschan et al. 2011).

Interestingly, there is one feature of the dosage compensation process that is shared across species. The initial identification of the X chromosome is required for all subsequent steps of regulation (Lucchesi et al. 2005; Payer and Lee 2008) (Figure 1). Additionally, recent work has revealed that H4K16ac is a conserved step in activation of X chromosome transcription during dosage compensation in mammals and *Drosophila* (Gelbart et al. 2009; Deng et al. 2013). However, the mechanism by which H4K16ac is specifically targeted to the X chromosome remains poorly understood in any organism. Because *Drosophila* dosage compensation only involves targeting of H4K16ac and not additional silencing steps, it is an ideal model with which to study the X-identification step of dosage compensation.

***Drosophila* as a model for X-identification**

Until recently, there has remained a key question in the field: How does MSL complex specifically identify MRE sequences within high affinity sites on the X chromosome and not on autosomes? Though the MSL complex is able to

identify the MRE sequence elements *in vivo*, none of the MSL proteins have specificity for the MRE sequence *in vitro* (Fauth et al. 2010). Additionally, the MRE sequences are not X-specific as they are also found on autosomes (Alekseyenko et al. 2008). Furthermore, when the *roX* loci are moved to an autosome, the *roX* RNAs associated with the MSL complex can move *in trans* to the X chromosome. This indicates that the location of the *roX* loci on the X chromosome are not sufficient for X chromosome identification (Park et al. 2002). Therefore, the mechanism by which the MSL complex targets MRE sequences within high affinity sites on the X chromosome alone has remained elusive.

We recently identified a novel regulator of *Drosophila* dosage compensation that specifically identifies MRE sequences throughout the genome (Larschan et al. 2012; Soruco et al. 2013) (See Chapter 2). The Chromatin Linked Adapter of MSL Proteins (CLAMP) zinc finger protein directly links the MSL complex to MRE sequences, is enriched at MSL high affinity sites, including both the *roX* loci, and is enriched on the X overall (Larschan et al. 2012; Soruco et al. 2013) (See Chapter 3). This is the first time a factor has been able to link the function of the *roX* loci to key DNA elements required for MSL complex targeting. To understand this new interaction more clearly, the function of the MSL complex components and CLAMP are explained in more detail below.

Protein components of the MSL complex

The MSL complex is the ribonucleoprotein activator complex responsible for increasing transcript levels on the X chromosome. It is comprised of one of two redundant, long non-coding *roX* RNAs and at least five proteins: Male Specific Lethal 1 (MSL1), Male Specific Lethal 2 (MSL2), Male Specific Lethal 3 (MSL3), Maleless (MLE) and Males Absent On the First (MOF) (Figure 2A).

All of the protein components of MSL complex were identified in genetic screens for mutations that cause male-specific lethality (Belote and Lucchesi 1980; Hilfiker et al. 1997). However, all of the components except for MSL2 are also found in females (Kelley et al. 1997). MSL1 and MSL2 are the core structural components and nucleate complex assembly (Copps et al. 1998). MSL1 plays a scaffolding role associating with many other components such as MSL2, MSL3, MLE and MOF (Copps et al. 1998; Scott et al. 2000; Li et al. 2005; Kadlec et al. 2011). MSL1 can also form a homodimer with itself suggesting that the MSL complex may function as a heteroduplex potentially being comprised of two complete complex structures (Hallaceli et al. 2012).

MSL2 is the limiting component and is translationally repressed by the master regulator of sex-determination, Sex Lethal (Sxl). Functional Sxl protein is present only in females and prevents translation when bound to the 5' and 3' UTR of MSL2 RNA (Kelley et al. 1997; Beckmann et al. 2005). MSL2 is an E3 ubiquitin ligase which ubiquitylates H2B lysine 34 when associated with MSL1

(Wu et al. 2011). Additionally, MSL2 can ubiquitylate other MSL components when over-expressed (Villa et al. 2012).

MSL3 has a chromo-domain region with specificity for the H3K36me3 histone modification associated with actively transcribed genes, which is deposited by the RNA Pol II associated factor Set2 (Larschan et al. 2007; Kim et al. 2010). MOF is the major histone acetyl transferase (HAT) in *Drosophila* specific for H4K16ac leading to increased transcriptional activity (Smith et al. 2000; Smith et al. 2001). MOF is also a member of the Non-Specific Lethal (NSL) Complex, found at the promoters of many constitutively active genes on both the X and autosomes (Lam et al. 2012).

MLE is weakly associated with the MSL complex through interactions with the *roX* RNAs (Copps et al. 1998). MLE contains two RNA binding domains, which recognize specific secondary structures of the *roX* RNAs necessary for incorporation of the *roX* RNAs into the complex (Izzo et al. 2008; Maenner et al. 2013). In addition to functioning in the MSL complex, MLE affects splicing and RNA editing genome-wide (Reenan et al. 2000).

A more loosely associated factor is the serine/threonine kinase JIL-1 (Jin et al. 2000; Li et al. 2005). JIL-1 has specificity for H3S10 and maintains euchromatin boundaries to prevent heterochromatin formation (Wang et al. 2011). It typically co-localizes with H3K36me3 and H4K16ac marks associated

with MSL complex targeting (Regnard et al. 2011). However, the specific role of JIL-1 in *Drosophila* dosage compensation remains unknown.

RNA components of the MSL complex

The *roX* RNAs, *roX1* and *roX2*, are essential for complex targeting and function (Franke and Baker 1999; Deng and Meller 2006; Li et al. 2008). Though the *roX* RNAs are very different in sequence composition and length (4.4 kb and 1.4 kb respectively) they are functionally redundant (Meller and Rattner 2002). Both contain conserved minimal tandem repeat stem loop features necessary for MLE-mediated incorporation into the complex (Meller et al. 2000; Ilik et al. 2013). However, it is likely that only one RNA transcript, either *roX1* or *roX2*, is incorporated at a time into a functioning MSL complex (Ilik et al. 2013).

The RNAs are highly unstable unless incorporated into the complex and differ greatly in their transcriptional control. The *roX1* locus is transcribed early during development from an MSL-independent promoter, generating a short non-coding RNA (Meller et al. 2000; Lim and Kelley 2012). This likely generates the initial functional MSL complex early in development. Later, MSL mediated sustained transcription from active complex, either including a long *roX1* transcript or *roX2* transcript, is then responsible for global compensation (Lim and Kelley 2012). As an additional control preventing complex formation in

females, *roX1* RNA transcription is repressed in females by the NURF301 repressor complex (Bai et al. 2007).

The location of the *roX* loci on the X chromosome serve as an important element involved in the initial stages of X-targeting. It has been hypothesized that the *roX* loci are the first sites of MSL targeting because *roX* RNA are necessary for complex function (Deng and Meller 2006). However, this has yet to be demonstrated *in vivo*.

Importantly, the *roX* genes are proximal to MRE sequences that recruit MSL complex and act as enhancers to increase *roX* gene transcription (Park et al. 2003; Bai et al. 2004). In mutant flies lacking both *roX* loci on the X chromosome, but containing a *roX* locus inserted on an autosome, the MSL complex can move *in trans* to identify the X chromosome (Park et al. 2002; Oh et al. 2003). Additionally, increasing *roX* RNA transcription from an artificial autosomal locus results in MSL complex spreading *in cis* extending over a 5 MB region (Kelley et al. 2008; Park et al. 2010). However, spreading across the entire autosome does not occur. Therefore, there must be additional elements identifying the X chromosome, which allows MSL complex to spread along the length of the entire X chromosome.

High affinity sites and MRE DNA Elements

In the absence of the more peripheral MSL3, MOF and MLE components, the core MSL complex components, MSL1 and MSL2, can only identify a subset of the wild type sites on the X chromosome (Lyman et al. 1997; Gu et al. 1998). This finding lead to a multi-step model for MSL complex targeting the X chromosome, which is likely facilitated through the activity of one or more complex components. High-resolution characterization of MSL targeting, using genomic approaches and performed using MSL3 null mutants, led to the identification of 150-300 evenly spaced high affinity sites, called chromatin entry sites (CES) in the study, which retain very high levels of MSL complex compared to surrounding sequence. Importantly, from these high affinity sites, a 21 bp sequence was found 2-fold enriched on the X and 4-fold enriched at active genes on the X (Alekseyenko et al. 2008).

This sequence, termed MSL Recognition Element (MRE), has a highly conserved 8 bp GA-repeat core. Mutational analysis of both the 8-bp core sequence and additional flanking region of the MRE sequence showed that MREs are necessary for MSL recruitment and is sufficient for recruitment in the context of the high affinity sites. However, there are approximately 13,000 MRE sequences found throughout the genome. Therefore, in the context of the entire genome, where only 3.5% of MREs map to high affinity sites, MRE sequences alone are not sufficient for targeting MSL complex to the X chromosome alone (Alekseyenko et al. 2012).

In an alternative and supporting approach, low formaldehyde crosslinking conditions and individual RNAi treatments against MLE, MOF and MSL3 in male S2 cells, identified 130 high affinity sites (Straub et al. 2008). The high affinity sites from this study are located in non-coding gene regions and have almost complete overlap with the original 150 high affinity sites (Alekseyenko et al. 2008). In summary, these diverse approaches identified the same group of high affinity sites, which are likely to be key initial sites of X-identification.

Mechanisms for spreading of the MSL complex and increasing transcription of X-linked genes

Once the *roX* loci and high affinity sites are targeted through MRE-mediated activity, the MSL complex then spreads through MSL3 chromo-domain interaction with the H3K36me3 chromatin mark to approximately 1000 active X-linked genes (Larschan et al. 2007; Sural et al. 2008). Spreading of the MSL complex from high affinity sites is sequence independent because an active autosomal gene that does not contain any MRE sequences can be inserted on the X chromosome and is targeted by MSL complex (Gorchakov et al. 2009). The global H4K16ac mark deposited on the X chromosome by the MSL complex decompacts chromatin (Shogren-Knaak et al. 2006). In this way, it theorized that there is an increase in the release of paused RNA Pol II into gene bodies (Larschan et al. 2011). However, the mechanism by which transcription is up-regulated approximately two-fold remains unknown. It is possible that core

components of the MSL complex counteract high level of transcriptional activation that would otherwise be caused by H4K16ac elsewhere in the genome (Prestel et al. 2010; Sun et al. 2013).

The CLAMP protein: a newly discovered regulator of *Drosophila* dosage compensation

Despite many recent advances in the understanding of *Drosophila* dosage compensation, one question has remained elusive. None of the MSL complex components are able to specifically recognize the MRE sequence. While the *roX* RNAs play an important role in the initial targeting, the *roX* loci are not sufficient for this process because the RNAs can travel *in trans*. The MRE sequences are enriched on the X but are also found on autosomes. Consequently, the mechanism by which the MSL complex is able to initially distinguish the X from the autosomes was not known.

We recently used a new cell-based screening strategy that allowed us to identify factors that have both male-specific roles in dosage compensation and roles in transcription in both males and females (Larschan et al. 2012) (See Chapter 2). Analogous to the condensin complexes in *C. elegans*, we hypothesized that non-sex specific factors may have an additional role in dosage compensation. Such factors with non-sex specific roles would have been missed in previous screens, which only identified male-specific lethal factors. In this way,

we identified the zinc finger transcription factor, Chromatin-Linked Adaptor for MSL Proteins (CLAMP), that provides a physical link between the MSL complex and the X chromosome in males (Figure 3A). The CLAMP protein is able to directly recognize MRE sequences throughout the genome but is found enriched on the X compared to autosomes and is most highly enriched at the high affinity sites (Figure 3B). Furthermore, CLAMP is bound to MRE sequences at the *roX* loci and is required for *roX2* transcription (Soruco et al. 2013) (See Chapter 3).

The CLAMP protein directly interacts with MRE DNA sequences and is therefore able to identify MRE sequences on all chromosomes (Figure 3A). However, CLAMP is enriched on the X chromosome in the absence of MSL complex in females (Figure 3B) (See Chapter 3). Enrichment of CLAMP on the X chromosome in the absence of MSL complex expression could mark the high affinity sites as 'seed sites' for the very first step of MSL targeting (Figure 3B). Additionally, the CLAMP protein and MSL complex have a synergistic relationship, which results in increased levels of both factors when co-localized, and could therefore initiate the X-identification cascade (Soruco et al. 2013) (Figure 3B) (See Chapter 3). I am currently testing this model by inducing MSL complex in female cells (see Appendix 1).

Unique X chromosome characteristics

In addition to CLAMP being enrichment on the X, there are several additional features that distinguish the X chromosome and make it distinct from autosomes. The X chromosome has an increased GC content in the 10 Kb region surrounding high affinity sites and a decrease in GC content within 1Kb of functional MREs. However, this is not sufficient to predict functional MREs. There is an MSL-independent increase in H4K16ac noted in female Kc cells and further increased acetylation as the result of MOF function in male cells (Alekseyenko et al. 2012). Importantly, using the MODEncode histone modifications ChIP-seq library as a training data set, H3K36me3 and JIL-1 association were predictive marks for high affinity sites on the X chromosome in female cells. Additionally, significant nucleosome depletion at high affinity sites in female cells were also noted, which correlates with previous DNase hypersensitivity of high affinity sites seen in male cells (Oh et al. 2004; Alekseyenko et al. 2012). Together with the CLAMP protein and *roX* RNAs, these characteristics are likely to ensure X-identification. However, future work is needed to determine the relative contributions of these factors during the targeting process.

The role of nuclear organization in *Drosophila* dosage compensation

The targeting of MSL2 to the X chromosome results in the formation of a distinct nuclear compartment close to the nuclear envelope that is not present in females (Straub et al. 2005; Grimaud and Becker 2009). The MSL dependent nuclear domain of the X chromosome in males is mediated through association

with nuclear pore and exosome proteins, an association that is not seen in females (Mendjan et al. 2006). This important link could therefore couple transcription, RNA processing and nuclear export of X-transcribed genes in males during dosage compensation (Figure 4).

Discussion

With the discovery of the CLAMP protein as an important regulator of *Drosophila* dosage compensation, experiments expanding on the role of CLAMP during the early stages of X-identification are eagerly anticipated (see Appendix 1). Recently, an increased accumulation of GA-repeats on the X chromosome was noted when comparing *D. melanogaster* to *D. miranda*, which contains a newly evolved X chromosome (Alekseyenko et al. 2013). Therefore, evolution of X-specific GA-repeats could be the precise mechanism that has allowed for an increase in CLAMP targeting to the X chromosome.

However, there are other transcription factors that bind GA repeat sequences. For example, GAGA factor, also known as GAF in mammals or Trl in *Drosophila*, is notably decreased at high affinity sites in *Drosophila* male cells, whereas CLAMP is found enriched (Alekseyenko et al. 2012; Soruco et al. 2013). Therefore, it is possible there is an interaction between CLAMP and the MSL complex that is favored in the context of the X chromosome that is not favored between GAGA and MSL complex. This interaction could be mediated by

characteristics of the surrounding chromatin, additional proteins, or both.

To gain insight into the evolution of dosage compensation, it is interesting to consider the non-sex specific roles of this previously unstudied CLAMP protein. It is possible that the MSL complex co-opted the more ancient CLAMP protein because it recognized GA-rich sequences that are enriched on the X chromosome (Ellison and Bachtrog 2013). Positive selection may have resulted in increased GA-repeats on the X chromosome due to their interaction with the highly conserved CLAMP protein, or vice versa. Additionally, it is possible that there is a physical interaction between CLAMP and MSL complex. This would likely be mediated through the glutamine-rich domain of CLAMP, where glutamine-rich domains are known to be involved in protein-protein interactions (Wang et al. 2013; Gonzalez Nelson et al. 2014)(Boor and Larschan unpublished).

As a DNA binding protein, CLAMP is found at the transcription start site of many transcribed genes on all chromosomes, as well as the 3' end of genes where MSL is also typically localized (Alekseyenko et al. 2006; Soruco et al. 2013) (See Chapter 2 and Chapter 3). This spatial separation could also isolate the two distinct functions of the CLAMP protein: the male-specific dosage compensation function and the non-sex specific function. This idea is supported by a recent study using new sonication techniques, which revealed different MSL sub complexes or core components acting over varying gene regions (Straub et

al. 2013). This finding could identify which MSL complex component is more commonly co-localized with CLAMP.

Future experiments investigating the formation of an MSL-specific nuclear domain could reveal a possible link to targeting of specific sites along the X prior to active gene recognition. Additionally, a nuclear domain could maintain high concentrations of important transcriptional machinery within close proximity to the X chromosome, facilitating increases in transcription levels. It is also possible that the CLAMP protein, through its glutamine-rich domain and seen to aggregate *in vitro*, could mediate this process (Chery and Larschan, unpublished). Understanding the mechanism of X-targeting during dosage compensation could provide new insights into essential cellular mechanisms that allow transcription factors to find their targets and precisely regulate the genome.

Acknowledgements

We are grateful to Dr. Leila Rieder for insights and critical review of the manuscript.

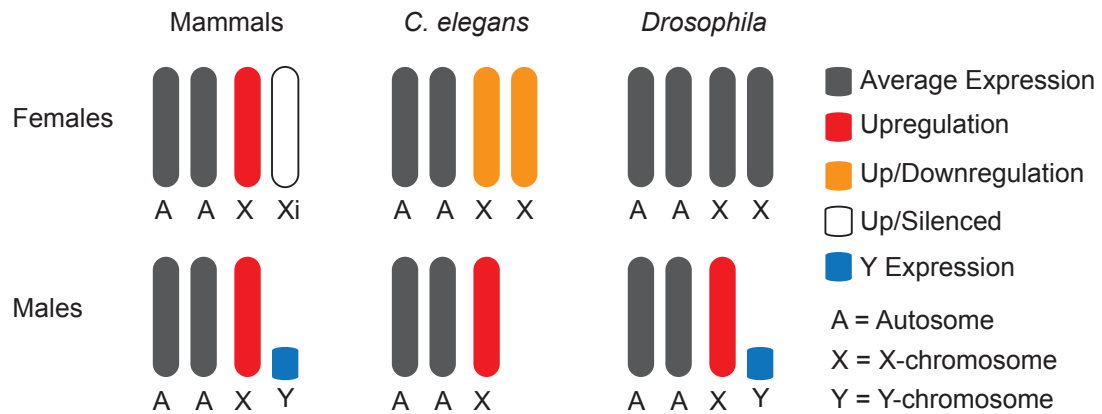


Figure 1: Dosage compensation patterns across species.

Increased transcription of the X chromosome is required for dosage compensation in mammals, *C. elegans* and *Drosophila*. The X chromosome is up-regulated in both sexes in mammals and *C.elegans*. A second step of regulation takes place in which the X chromosome is randomly inactivated in mammalian females. In *C.elegans*, both X chromosomes are down-regulated by half in hermaphroditic females. In *Drosophila*, the X chromosome is only up-regulated in males.

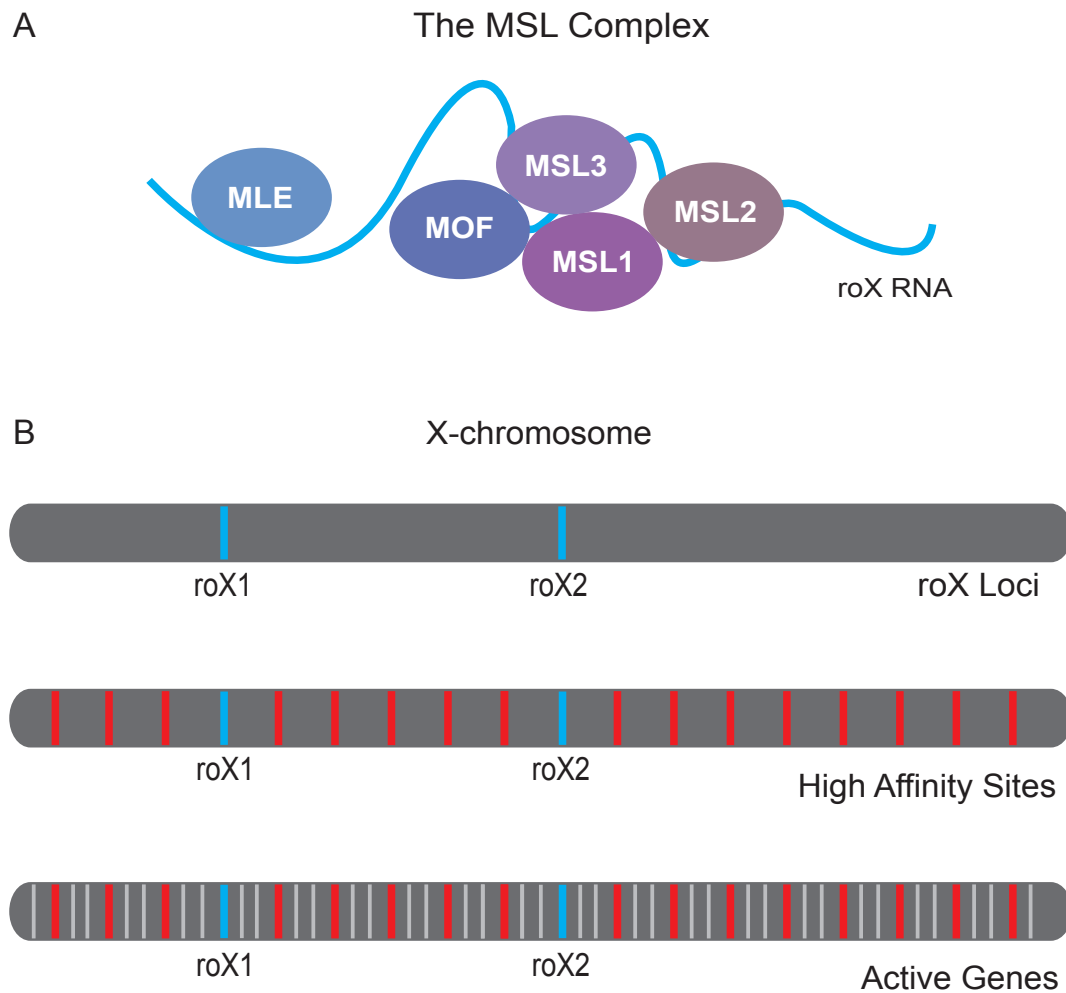


Figure 2: The MSL complex and multi-step targeting.

A) The Male Specific Lethal (MSL) complex components. B) The MSL Complex targets the X chromosome in multiple steps. First, it is thought that the *roX* loci (blue) are targeted during co-transcriptional incorporation of the RNAs into the complex. Second, a series of evenly spaced high affinity sites (red) containing MRE sequences are targeted. Finally, spreading from high affinity sites to the bodies of active genes (grey) results in an approximate two-fold increase in transcription of X-linked genes.

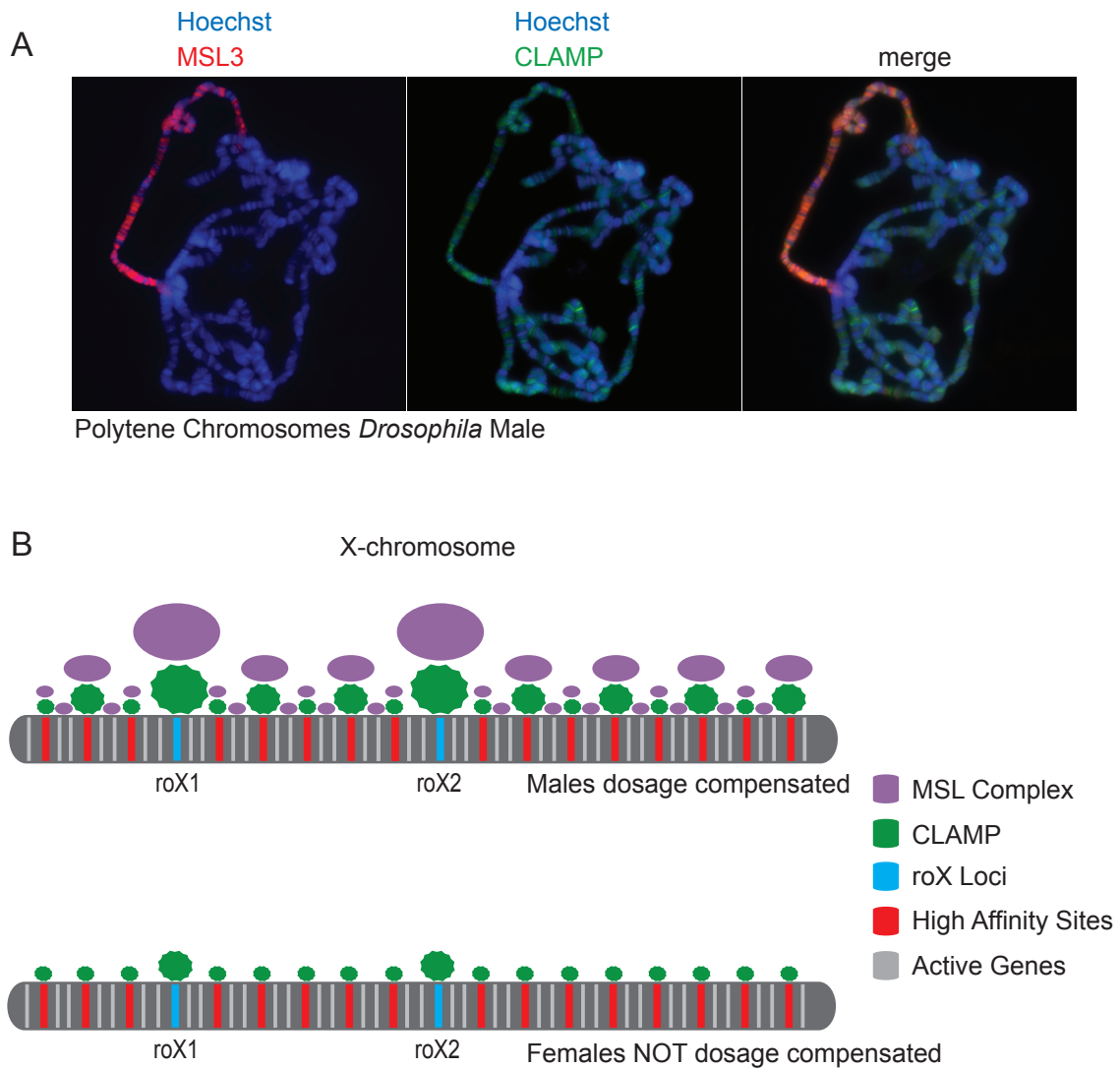


Figure 3: The role of CLAMP in X-targeting.

A) The zinc finger protein, CLAMP, is found throughout the genome. B) CLAMP is found enriched at high affinity sites, including the *roX* loci, in the absence of MSL complex in females. Then a synergistic interaction leads to increased levels of both factors when co-localized on the X chromosome.

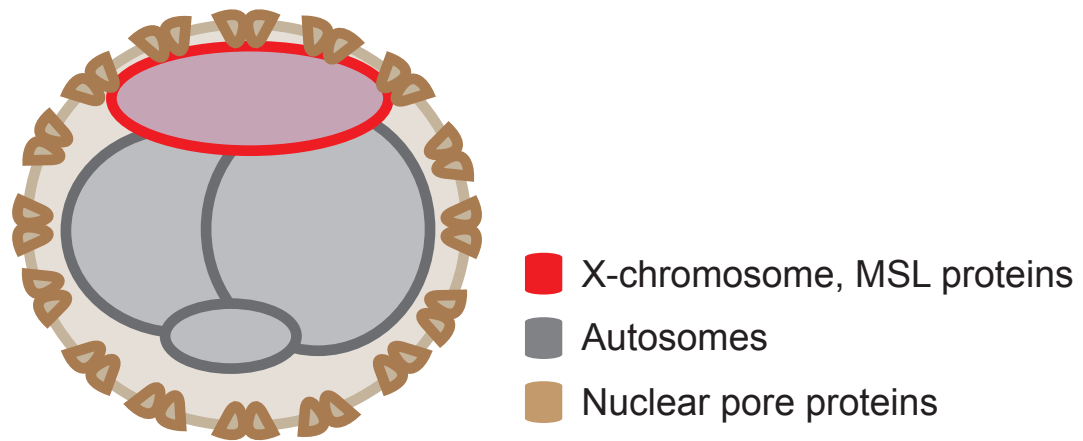


Figure 4: Unique X-characteristics and nuclear organization.

The X chromosome is maintained in a distinct nuclear space through association of MSL complex proteins associating with nuclear pore proteins.

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.
- 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.
- Beckmann K, Grskovic M, Gebauer F, Hentze MW. 2005. A dual inhibitory mechanism restricts msl-2 mRNA translation for dosage compensation in *Drosophila*. *Cell* **122**: 529-540.
- Belote JM, Lucchesi JC. 1980. Male-specific lethal mutations of *Drosophila melanogaster*. *Genetics* **96**: 165-186.
- Chao W, Huynh KD, Spencer RJ, Davidow LS, Lee JT. 2002. CTCF, a candidate trans-acting factor for X-inactivation choice. *Science* **295**: 345-347.
- Charlesworth B. 1978. Model for evolution of Y chromosomes and dosage compensation. *Proceedings of the National Academy of Sciences of the United States of America* **75**: 5618-5622.
- 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.
- 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. *The EMBO journal* **17**: 5409-5417.
- 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. *Developmental 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.
- Disteche CM. 2012. Dosage compensation of the sex chromosomes. *Annual review of genetics* **46**: 537-560.
- Donohoe ME, Zhang LF, Xu N, Shi Y, Lee JT. 2007. Identification of a Ctf cofactor, Yy1, for the X chromosome binary switch. *Molecular cell* **25**: 43-56.
- Ellison CE, Bachtrog D. 2013. Dosage compensation via transposable element mediated rewiring of a regulatory network. *Science* **342**: 846-850.
- 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.
- Franke A, Baker BS. 1999. The rox1 and rox2 RNAs are essential components of the compensasome, which mediates dosage compensation in *Drosophila*. *Molecular cell* **4**: 117-122.

- 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.
- Gonzalez Nelson AC, Paul KR, Petri M, Flores N, Rogge RA, Cascarina SM, Ross ED. 2014. Increasing prion propensity by hydrophobic insertion. *PloS one* **9**: e89286.
- Gorchakov AA, Alekseyenko AA, Kharchenko P, Park PJ, Kuroda MI. 2009. Long-range spreading of dosage compensation in Drosophila captures transcribed autosomal genes inserted on X. *Genes & development* **23**: 2266-2271.
- Grimaud C, Becker PB. 2009. The dosage compensation complex shapes the conformation of the X chromosome in Drosophila. *Genes & development* **23**: 2490-2495.
- Gu W, Szauter P, Lucchesi JC. 1998. Targeting of MOF, a putative histone acetyltransferase, to the X chromosome of Drosophila melanogaster. *Developmental genetics* **22**: 56-64.
- 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.
- Hilfiker A, Hilfiker-Kleiner D, Pannuti A, Lucchesi JC. 1997. mof, a putative acetyltransferase 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.
- Izzo A, Regnard C, Morales V, Kremmer E, Becker PB. 2008. Structure-function analysis of the RNA helicase maleless. *Nucleic acids research* **36**: 950-962.
- 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. *The Journal of cell biology* **149**: 1005-1010.
- Jonkers I, Monkhorst K, Rentmeester E, Grootegoed JA, Grosveld F, Gribnau J. 2008. Xist RNA is confined to the nuclear territory of the silenced X chromosome throughout the cell cycle. *Molecular and cellular biology* **28**: 5583-5594.

- 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. *Nature structural & molecular biology* **18**: 142-149.
- 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.
- Kim D, Blus BJ, Chandra V, Huang P, Rastinejad F, Khorasanizadeh S. 2010. Corecognition of DNA and a methylated histone tail by the MSL3 chromodomain. *Nature structural & molecular biology* **17**: 1027-1029.
- Kruesi WS, Core LJ, Waters CT, Lis JT, Meyer BJ. 2013. Condensin controls recruitment of RNA polymerase II to achieve nematode X chromosome dosage compensation. *eLife* **2**: e00808.
- Lam KC, Muhlpfordt F, Vaquerizas JM, Raja SJ, Holz H, Luscombe NM, Manke T, Akhtar A. 2012. The NSL complex regulates housekeeping genes in Drosophila. *PLoS genetics* **8**: e1002736.

- 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. *Molecular 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 genetics* **8**: e1002830.
- 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 genetics* **8**: e1002564.
- Livernois AM, Graves JA, Waters PD. 2012. The origin and evolution of vertebrate sex chromosomes and dosage compensation. *Heredity* **108**: 50-58.

- Lucchesi JC, Kelly WG, Panning B. 2005. Chromatin remodeling in dosage compensation. *Annual review of genetics* **39**: 615-651.
- 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.
- Meller VH, Gordadze PR, Park Y, Chu X, Stuckenholtz C, Kelley RL, Kuroda MI. 2000. Ordered assembly of roX RNAs into MSL complexes on the dosage-compensated X chromosome in Drosophila. *Current biology : CB* **10**: 136-143.
- Meller VH, Rattner BP. 2002. The roX genes encode redundant male-specific lethal transcripts required for targeting of the MSL complex. *The EMBO journal* **21**: 1084-1091.
- 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. 2010. Targeting X chromosomes for repression. *Current opinion in genetics & development* **20**: 179-189.
- Nicoll M, Akerib CC, Meyer BJ. 1997. X chromosome-counting mechanisms that determine nematode sex. *Nature* **388**: 200-204.

- Oh H, Bone JR, Kuroda MI. 2004. Multiple classes of MSL binding sites target dosage compensation to the X chromosome of *Drosophila*. *Current biology : CB* **14**: 481-487.
- 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.
- 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. *Annual review of genetics* **42**: 733-772.
- 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.

- Plath K, Fang J, Mlynarczyk-Evans SK, Cao R, Worringer KA, Wang H, de la Cruz CC, Otte AP, Panning B, Zhang Y. 2003. Role of histone H3 lysine 27 methylation in X inactivation. *Science* **300**: 131-135.
- Prestel M, Feller C, Straub T, Mitlohner H, Becker PB. 2010. The activation potential of MOF is constrained for dosage compensation. *Molecular cell* **38**: 815-826.
- Reenan RA, Hanrahan CJ, Ganetzky B. 2000. The mle(napts) RNA helicase mutation in drosophila results in a splicing catastrophe of the para Na⁺ channel transcript in a region of RNA editing. *Neuron* **25**: 139-149.
- Regnard C, Straub T, Mitterweger A, Dahlsveen IK, Fabian V, Becker PB. 2011. Global analysis of the relationship between JIL-1 kinase and transcription. *PLoS genetics* **7**: e1001327.
- Scott MJ, Pan LL, Cleland SB, Knox AL, Heinrich J. 2000. MSL1 plays a central role in assembly of the MSL complex, essential for dosage compensation in Drosophila. *The EMBO journal* **19**: 144-155.
- 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.
- 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* **504**: 465-469.

- Smith ER, Allis CD, Lucchesi JC. 2001. Linking global histone acetylation to the transcription enhancement of X-chromosomal genes in *Drosophila* males. *The Journal of biological chemistry* **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.
- Straub T, Gilfillan GD, Maier VK, Becker PB. 2005. The *Drosophila* MSL complex activates the transcription of target genes. *Genes & development* **19**: 2284-2288.
- 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.
- Straub T, Zabel A, Gilfillan GD, Feller C, Becker PB. 2013. Different chromatin interfaces of the *Drosophila* dosage compensation complex revealed by high-shear ChIP-seq. *Genome research* **23**: 473-485.
- Sun L, Fernandez HR, Donohue RC, Li J, Cheng J, Birchler JA. 2013. Male-specific lethal complex in *Drosophila* counteracts histone acetylation and

- does not mediate dosage compensation. *Proceedings of the National Academy of Sciences of the United States of America* **110**: E808-817.
- 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.
- Tian D, Sun S, Lee JT. 2010. The long noncoding RNA, Jpx, is a molecular switch for X chromosome inactivation. *Cell* **143**: 390-403.
- 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.
- Wang C, Cai W, Li Y, Deng H, Bao X, Girton J, Johansen J, Johansen KM. 2011. The epigenetic H3S10 phosphorylation mark is required for counteracting heterochromatic spreading and gene silencing in *Drosophila melanogaster*. *Journal of cell science* **124**: 4309-4317.
- 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.
- Wood AJ, Severson AF, Meyer BJ. 2010. Condensin and cohesin complexity: the expanding repertoire of functions. *Nature reviews Genetics* **11**: 391-404.

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. *Molecular cell* **43**: 132-144.

Chapter 2

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

Erica Larschan¹, Marcela M.L. Soruco¹, Ok-Kyung Lee^{2,3,^}, Shouyong Peng^{2,#},
Eric Bishop⁴, Jessica Chery¹, Karen Goebel¹, Jessica Feng¹, Peter J. Park^{2,4},
and Mitzi I. Kuroda^{2,3,*}

¹Department of Molecular Biology, Cellular Biology and Biochemistry, Brown University, Providence, RI, 02912, USA

²Division of Genetics, Department of Medicine, Brigham and Women's Hospital, Boston, MA 02115, USA

³Department of Genetics, Harvard Medical School, Boston, MA, 02115, USA

⁴Center for Biomedical Informatics, Harvard Medical School, Boston, MA, 02115, USA

*Corresponding author: mkuroda@genetics.med.harvard.edu

Current affiliations:

[^]Korea Advanced Institute of Science and Technology, Department of Biological Sciences, 335 Gwahangno, Yuseong-gu, Daejeon 305-701, Korea

[#]Department of Medical Oncology, Dana-Farber Cancer Institute, Boston, MA 02115, USA

This manuscript was published in PLoS Genetics in July 2012. I performed the CLAMP ChIP-seq experiment and created figures. Erica Larschan wrote the manuscript. I edited and prepared final manuscript for submission.

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*.

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. 2011).

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 1980; 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 an MSL-dependent reporter in male SL2 cells

In order to identify genes that regulate MSL complex recruitment or function, 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). We do not wish to over-interpret any negative results; however, we did not identify the genes encoding HP1, SU(VAR)3-7, or ISWI. These each have a mutant phenotype that differentially affects the

polytene X chromosome (Deuring et al. 2000; Spierer et al. 2005), but 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 of 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 effect 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 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 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). Comparison of ChIP-seq and ChIP-chip profiles of CG1832, MBD-R2, and

Chromator show similar localization to 5' ends of active genes, with CG1832 more focused at CES (Fig. 4C). 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 for 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 regulating 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 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 and facilitates 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: 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.

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 Gallx 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) session GSE39271.

Generation of heatmaps

The average enrichment profiles of proteins around the MREs ($\pm$ 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 $\pm$ 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.

Acknowledgments

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 and reagents for performing the genome-wide RNAi screen. Next Generation sequencing was performed at the Brown University Genomics core facility.

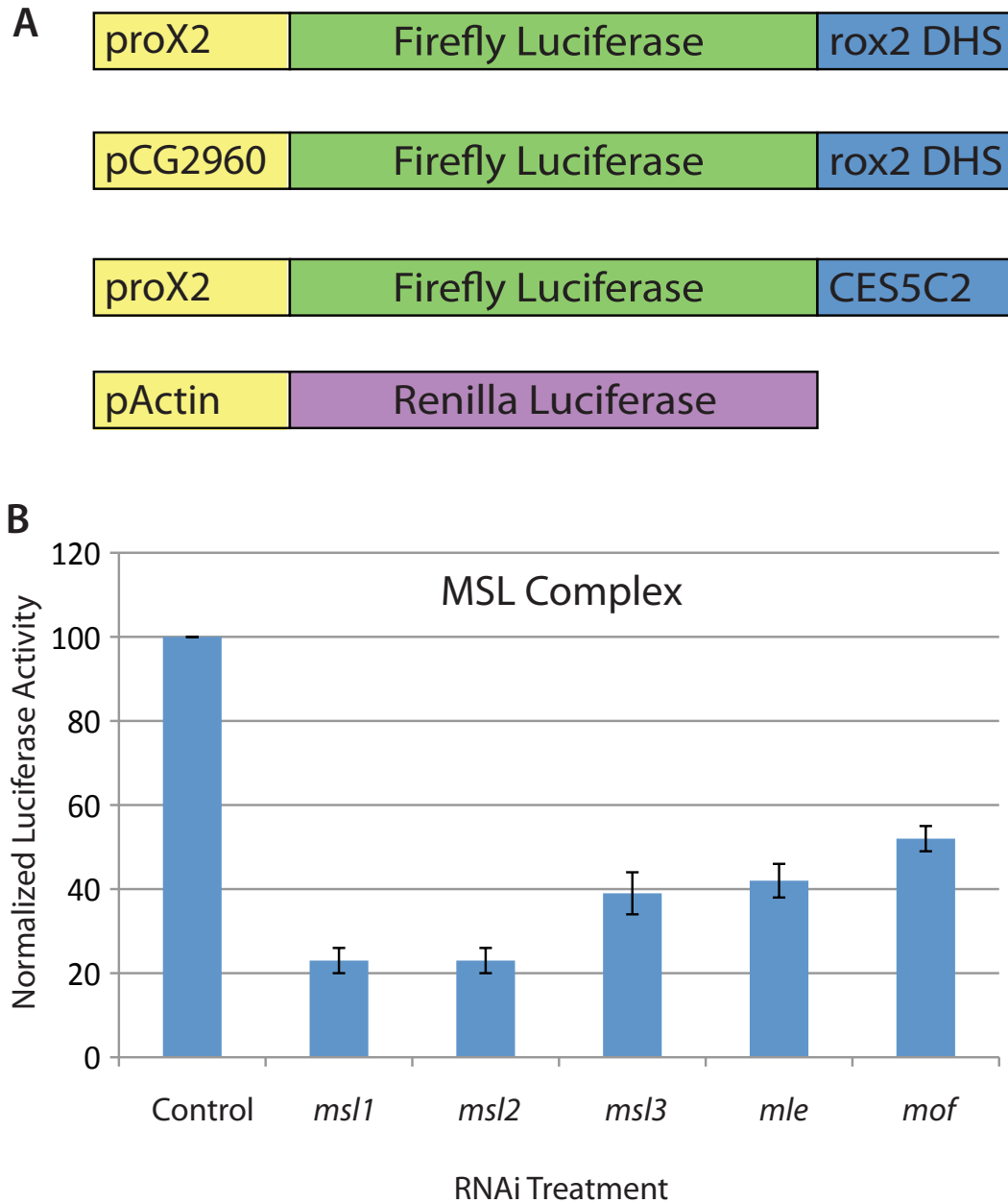


Figure 1: An RNAi screening system in *Drosophila* SL2 cells to identify regulators of MSL complex recruitment and function.

A) Schematic of the MSL-dependent reporter system for SL2 cell-based genome-wide RNAi screening. The 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 were altered. The alternative promoter is that of the CG2690 gene and the alternative CES 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.

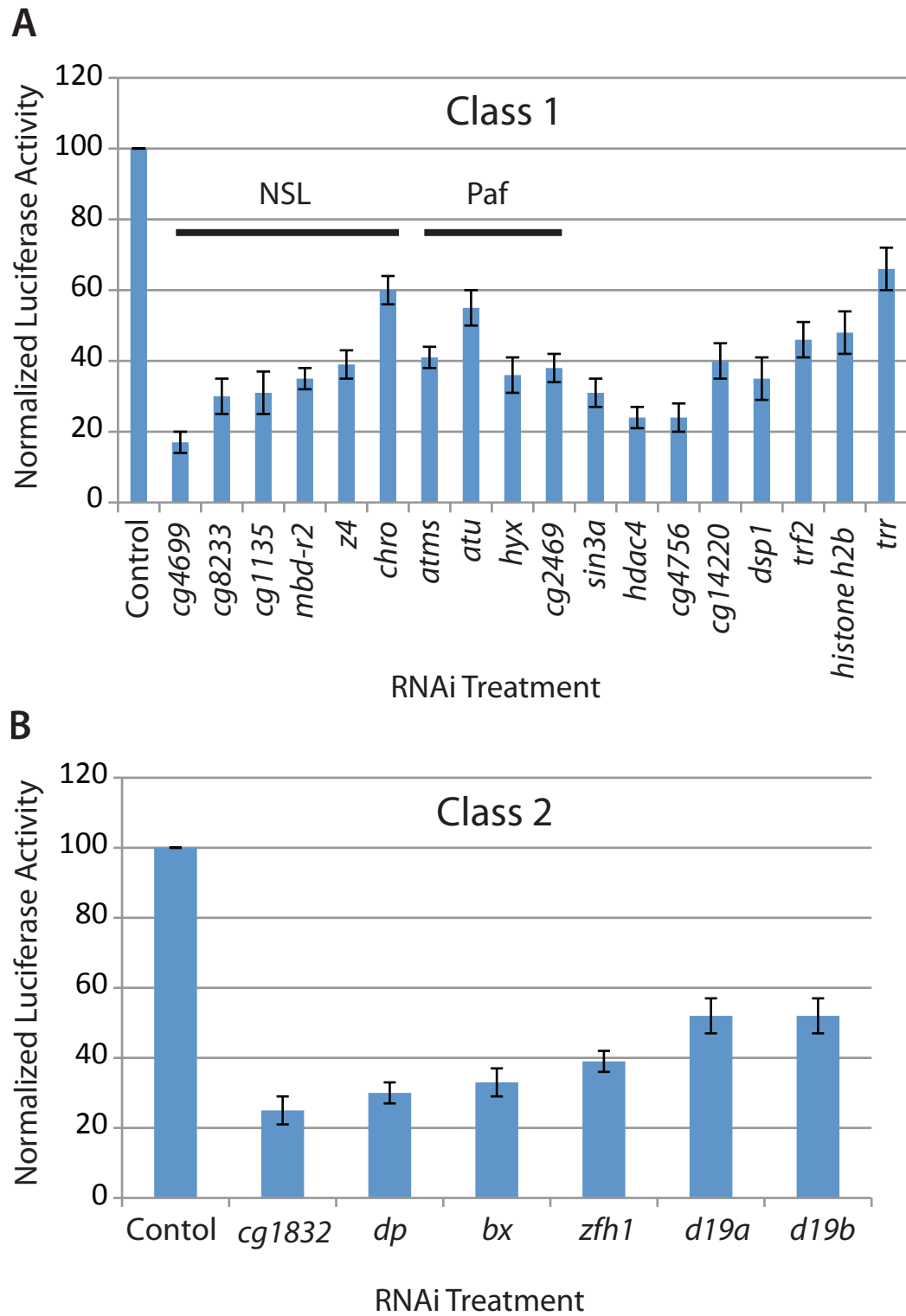


Figure 2: Candidate regulators of MSL complex identified from the genome-wide RNAi screen.

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.

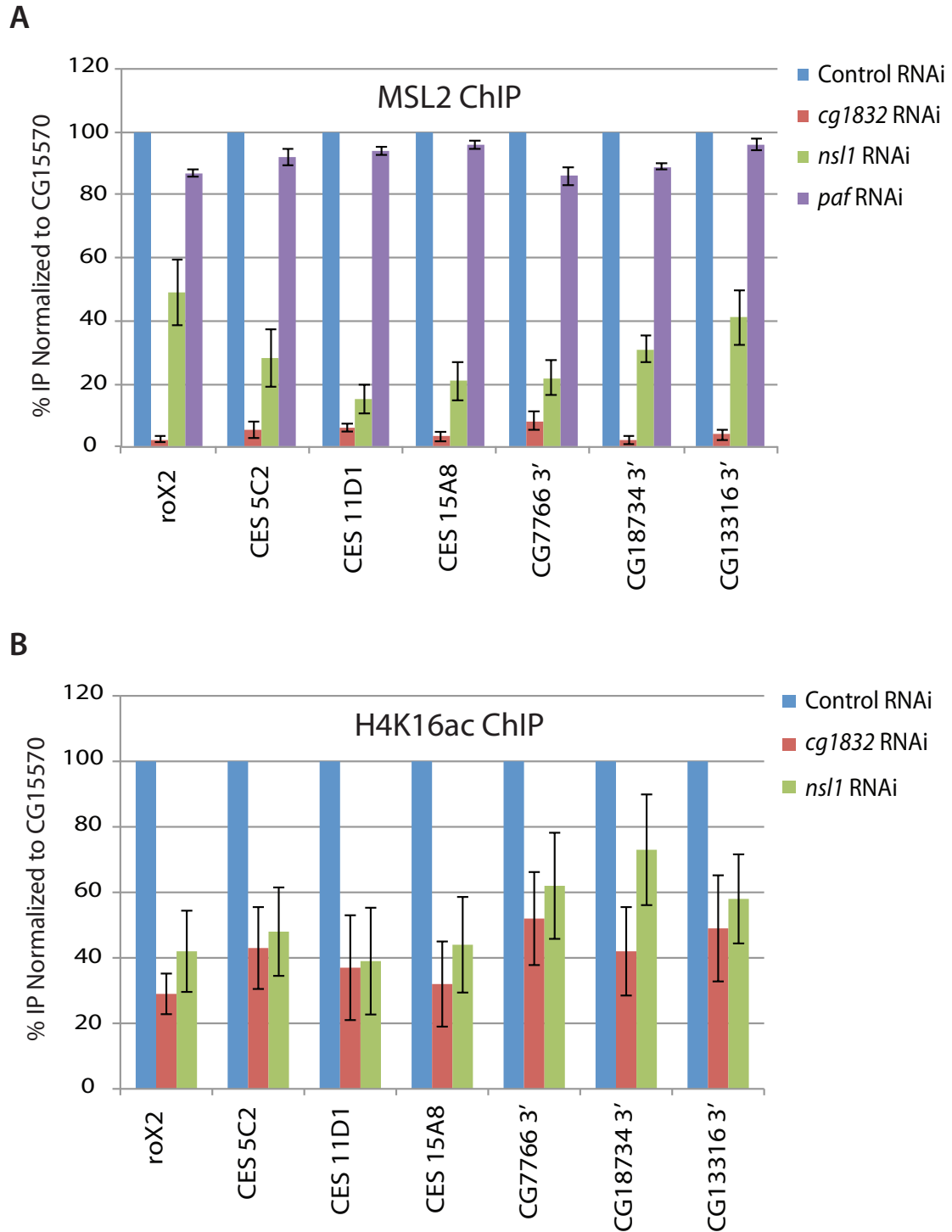


Figure 3: NSL1 and CG1832 contribute to MSL complex recruitment

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.

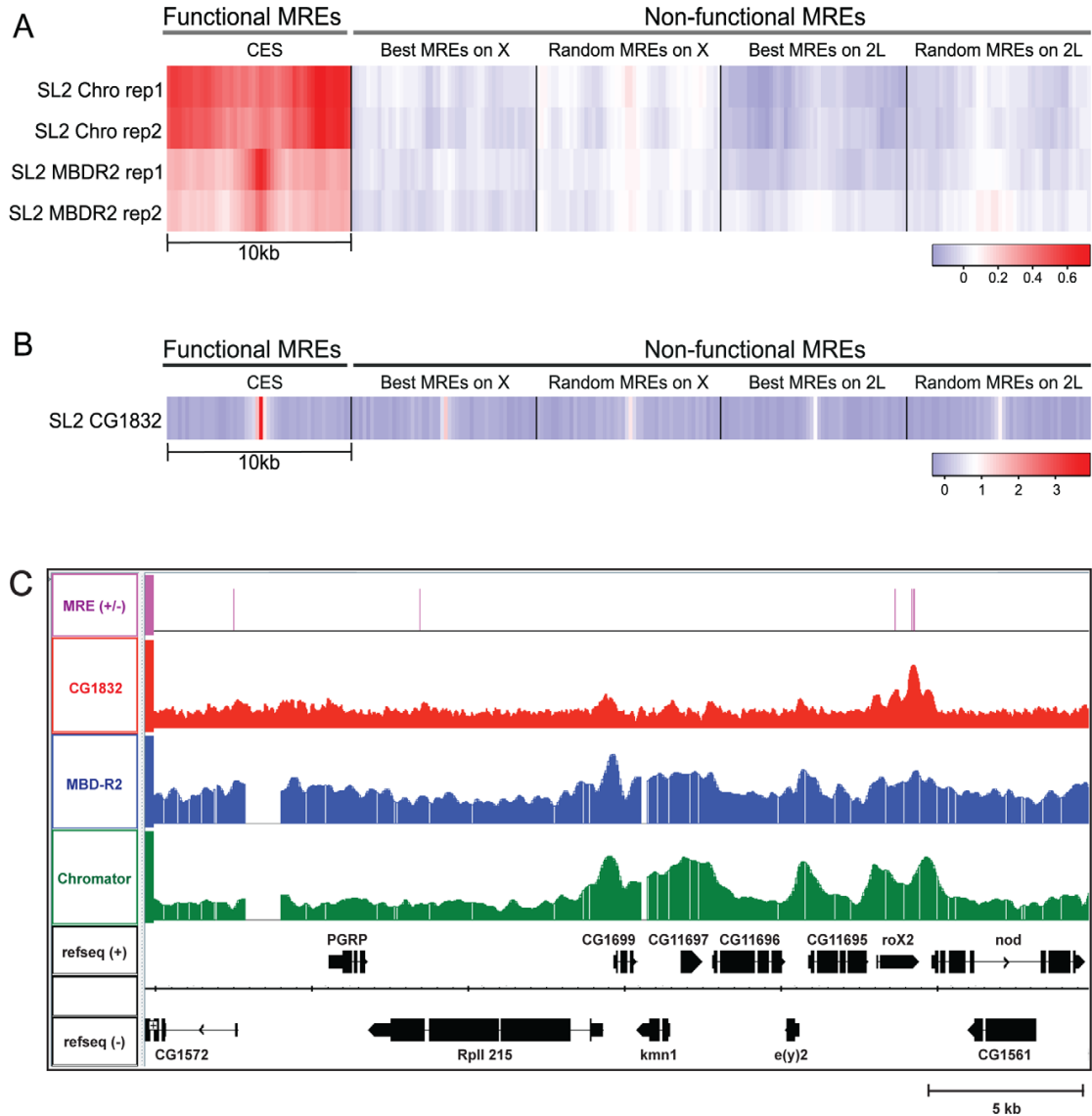


Figure 4: NSL complex and CG1832 enrichment at CES loci.

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 log₂ (ChIP/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 $\log_2$ (ChIP/input) scale is used for binned read count; the enrichment difference is enhanced in the ChIP-seq data owing to its greater dynamic range. C) Profiles of genome-wide localization data are shown for CG1832 (red), MBD-R2 (blue), and Chromator (green), in a 30 Kb region including the *roX2* locus.

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

Table 1. Gene ontology categories that were enriched in RNAi screen

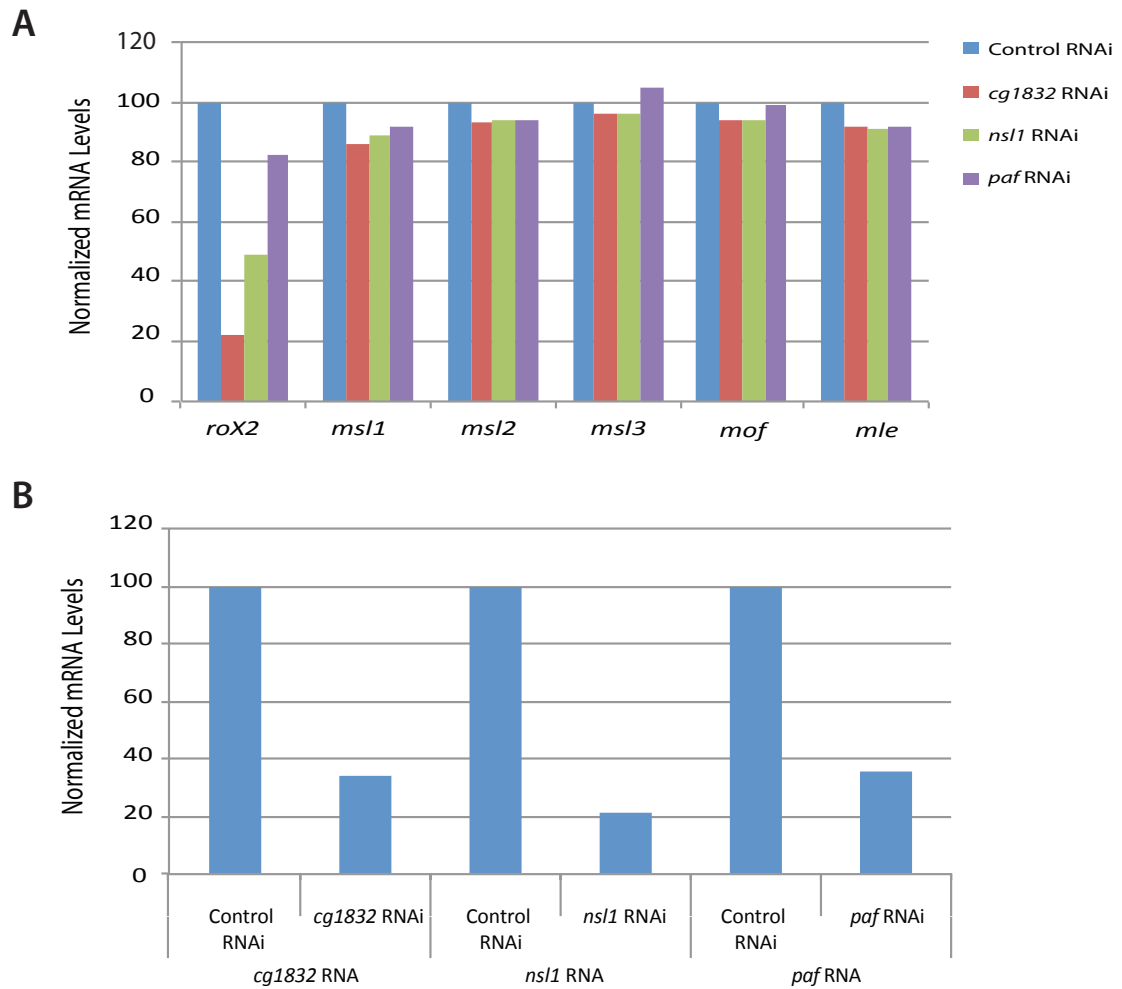


Fig. S1: mRNA levels of MSL complex components and candidates after CG1832, Nsl1, and Paf1 RNAi treatments.

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.

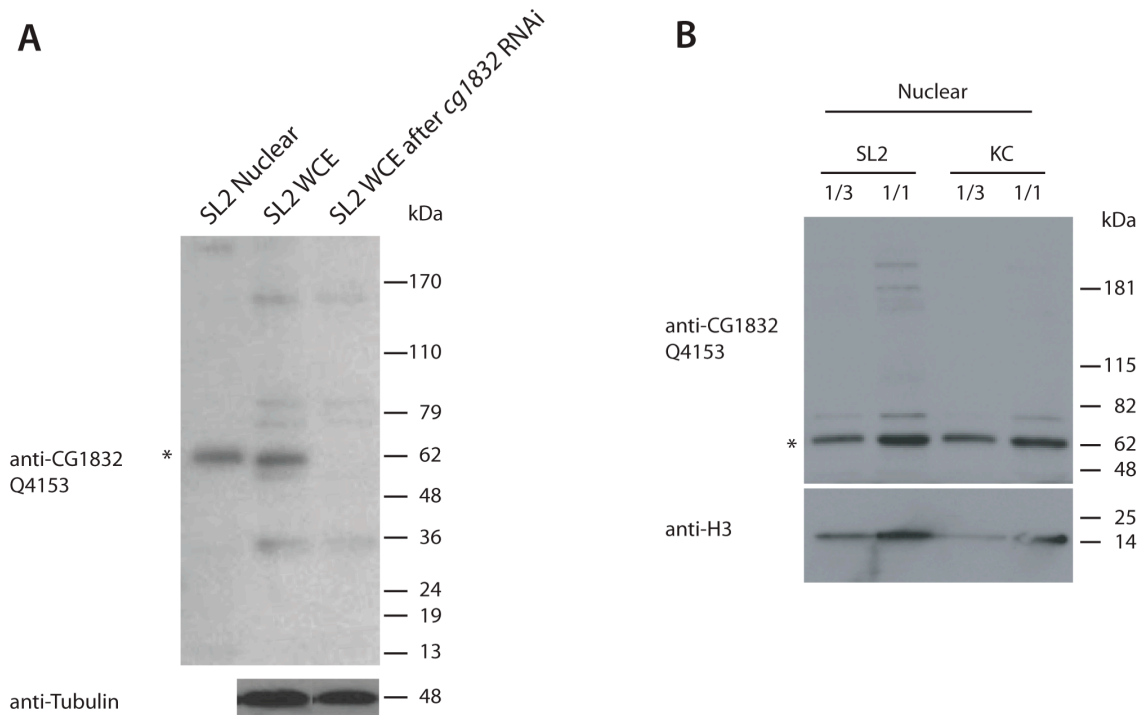


Figure S2: Western blots for CG1832 RNAi treatment and antibody validation.

A) Westerns on whole cell and nuclear extracts indicate that CG1832 is a nuclear protein and the CG1832 RNAi treatment reduces CG1832 protein levels. The arrow 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. Histone H3 was used as a loading control. Levels of CG1832 are similar in males and females.

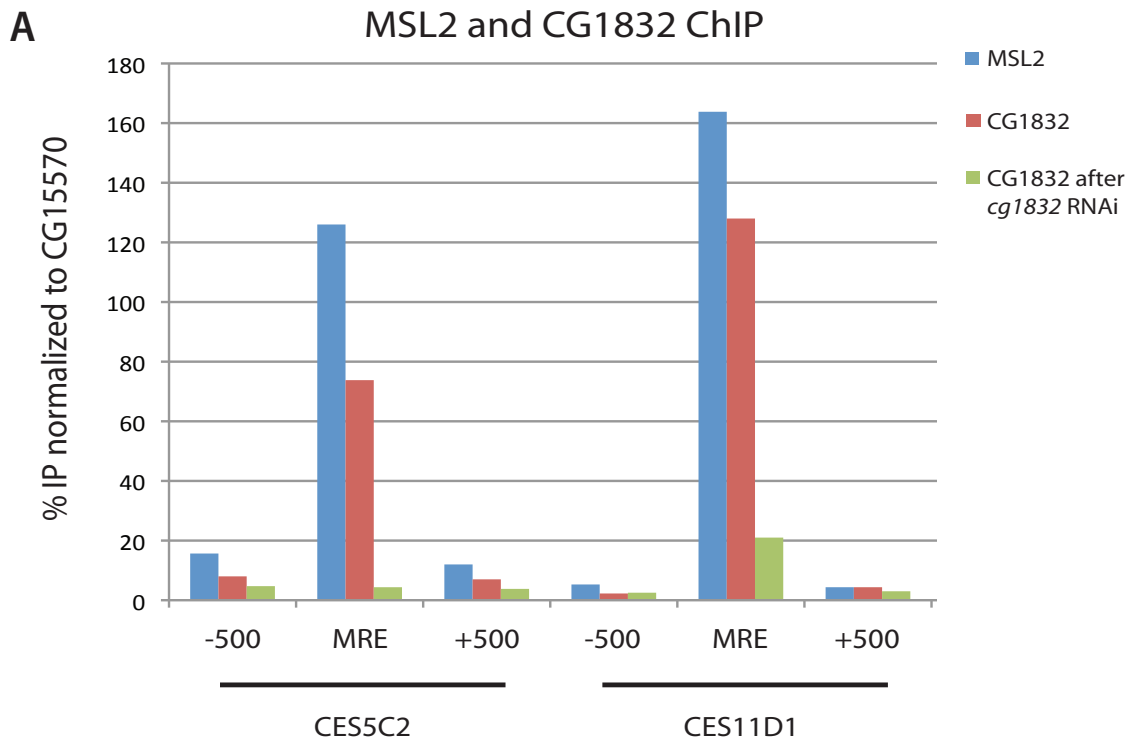
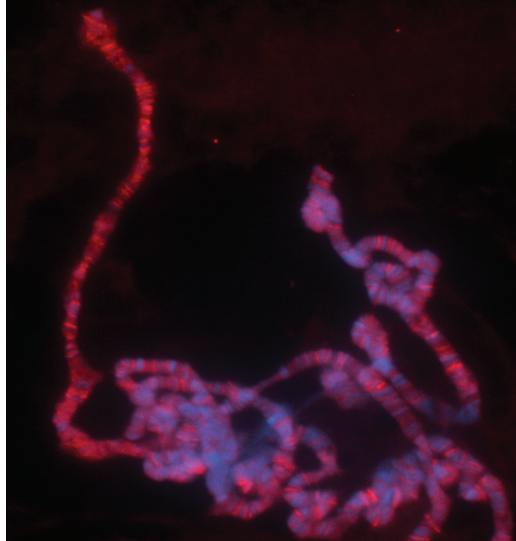


Figure S3: Validation of CG1832 antibody by ChIP at CES loci.

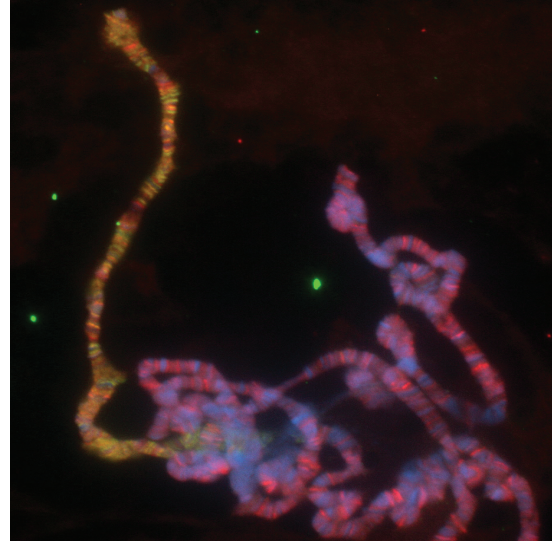
qPCR was performed on CG1832 ChIP samples before and after CG1832 RNAi treatment. Three independent CES loci were assayed and CG1832 RNAi strongly reduced the levels of CG1832 protein on chromatin at all three loci. An average of two independent experiments is shown.

A

Male
Hoechst (blue)
anti-CG1832 (red)



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



B

Female
Hoechst (blue)
anti-CG1832 (red)

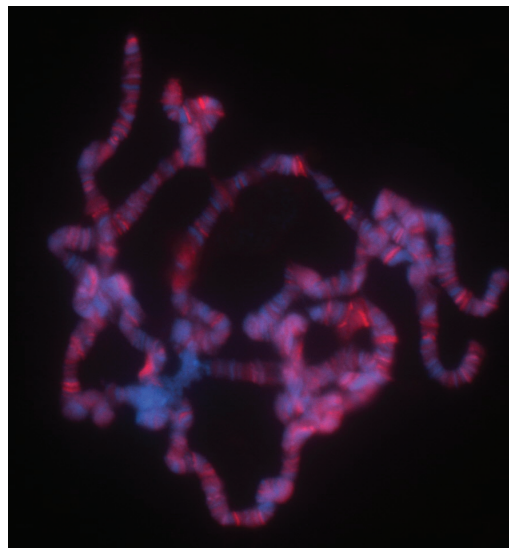


Figure S4. Immunostaining of polytene chromosomes.

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.

Link to table S1 (220 pages)

<https://www.dropbox.com/s/jdyqz30g5lbobc2/Table%20S1.pdf.zip>

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

Link to table S2 (5 pages)

<https://www.dropbox.com/s/0i8x2kad5c559ww/Table%20S2.pdf.zip>

Table S2: Final candidates identified from a genome-wide RNAi screen that passed secondary screens.

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. *Molecular and cellular biology* **26**: 250-260.
- 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, 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.
- 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.
- 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. 1980. Male-specific lethal mutations of *Drosophila melanogaster*. *Genetics* **96**: 165-186.
- 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.
- 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. *The EMBO journal* **17**: 5409-5417.
- 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. *Molecular cell* **5**: 355-365.
- 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.
- Furuhashi H, Nakajima M, Hirose S. 2006. DNA supercoiling factor contributes to dosage compensation in *Drosophila*. *Development* **133**: 4475-4483.
- 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 & development* **20**: 858-870.
- Hamada FN, Park PJ, Gordadze PR, Kuroda MI. 2005. Global regulation of X chromosomal genes by the MSL complex in *Drosophila melanogaster*. *Genes & development* **19**: 2289-2294.
- 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.
- 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.

- 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. *The Journal of cell biology* **149**: 1005-1010.
- 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. *Nature structural & molecular biology* **18**: 142-149.
- 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.
- 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, Tolstorukov MY, Park PJ. 2008. Design and analysis of ChIP-seq experiments for DNA-binding proteins. *Nat Biotechnol* **26**: 1351-1359.
- 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.
- 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.
- 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. *Molecular cell* **28**: 121-133.
- 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**.
- 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.
- 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.
- 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.

- Nybakken K, Perrimon N. 2002. Hedgehog signal transduction: recent findings. *Curr Opin Genet Dev* **12**: 503-511.
- 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.
- 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. *Molecular and cellular biology* **16**: 669-676.
- 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.

- 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.
- 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. *Molecular cell* **43**: 132-144.
- 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.

Chapter 3

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

Marcela M.L. Soruco^{1*}, Jessica Chery^{1*}, Eric P. Bishop^{2,7}, Trevor Siggers³, Michael Y. Tolstorukov^{2,3}, Alexander R. Leydon¹, Arthur U. Sugden¹, Karen Goebel¹, Jessica Feng¹, Peng Xia¹, Anastasia Vedenko³, Martha L. Bulyk^{3,4,5}, Peter J. Park^{2,3,6}, and Erica Larschan¹

¹Department of Molecular Biology, Cellular Biology and Biochemistry, Brown University, Providence, RI 02912

²Center for Biomedical Informatics, Harvard Medical School, Boston, MA 02115

³Division of Genetics, Department of Medicine, Brigham & Women's Hospital and Harvard Medical School, Boston, MA 02115

⁴Department of Pathology, Brigham & Women's Hospital and Harvard Medical School, Boston, MA, 02115.

⁵Harvard-MIT Division of Health Sciences and Technology, Harvard Medical School, Boston, MA 02115

⁶Children's Hospital Informatics Program, Boston, MA 02115

⁷Bioinformatics Graduate Program, Boston University, Boston, MA 02215

*These authors contributed equally to this work

This manuscript was published in *Genes and Development* in August 2013. I performed all experiments except for the protein binding microarrays in Figure 2 and ChIP-qPCR in Supplemental Figure S2D, which were performed by Jessica Chery. Bioinformatic analysis was done by Eric Bishop in the Park Lab for the ChIP-seq experiment and by Trevor Siggers in the Bulyk Lab for the protein binding microarrays. Erica Larschan wrote the initial manuscript. I created all figures and edited the final manuscript for submission.

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 the MSL complex. The discovery of CLAMP identifies a key factor required for the chromosome-specific targeting of dosage compensation, providing new insights into how subnuclear 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 Distèche 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 twofold to equal expression from the two female X chromosomes (Belote and Lucchesi 1980; 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, a process that remains poorly 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 1980; 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

cotranscriptionally incorporates them (Kelley et al. 1999, 2008). Second, the MSL complex recognizes 150–300 additional genomic loci distributed across the X chromosome, called “chromatin entry sites” (CESs). CESs 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-base-pair (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 the MSL complex and MREs (Alekseyenko et al. 2008; Straub et al. 2008; Fauth et al. 2010). Third, the MSL complex spreads from CESs to the bodies of active genes on the X chromosome by recognizing histone modifications associated with transcriptional activation, including trimethylation of H3K36 (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. 2011).

While both the *roX* noncoding RNAs and MREs contribute to the recognition of the male X chromosome, they are individually insufficient to generate X-chromosome specificity (Meller and Rattner 2002; 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) that would not have been recovered from the powerful MSL screens that identified all of the MSL complex components (Belote and Lucchesi 1980). Such non-sex-specific regulators provide a key opportunity to understand the first steps in X identification, including recognition of highly conserved MREs within CESs (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 CESs 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 CESs and that CG1832 RNAi strongly reduces MSL complex recruitment at five CESs tested in male tissue culture cells (Larschan et al. 2012). However, 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 flies? (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 named CG1832 CLAMP (chromatin-linked adaptor 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 CESs independent of the 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 CESs such as those at the *roX* loci. Our results support a mechanism whereby the MSL complex uses the X-enriched CLAMP protein at CESs followed by further enrichment from synergistic interactions between the two factors.

Results

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

To determine whether CLAMP and the MSL complex colocalize at CESs in vivo, we performed CLAMP chromatin immunoprecipitation (ChIP) sequencing (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: 10,009) and strong colocalization with MRE sequences (Fig. 1A,B; Supplemental Fig. S1). Moreover, CLAMP and

the MSL complex exhibit their highest occupancy levels when colocalized (Fig. 1A; Supplemental Fig. S2A, S2B). Nonoverlapping CLAMP occupancy sites occur at 59 ends of active genes on all chromosomes (Supplemental Fig. S2C), and nonoverlapping MSL-binding sites are localized over the bodies of active X-linked genes (Supplemental Fig. S2B). Furthermore, CLAMP sites occur precisely over MREs within CESs compared with the broader domains of MSL complex occupancy (Figs. 1A,B, 3A [below]). Together, these patterns are consistent with a role for CLAMP in recruiting the MSL complex to CESs, followed by spreading of the MSL complex to the bodies of active genes and additional non-sex-specific roles for CLAMP at 59 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 59 and 39 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). Each PBM is a dsDNA array that contains all possible 10-bp sequences embedded within

variable flanking sequences. After testing the complete seven-finger CLAMP construct and many additional combinations of zinc fingers, we were able to express soluble protein for the C-terminal six-finger and four-finger zinc finger regions of CLAMP (see the Materials and Methods). The in vitro PBM analysis for these proteins yielded statistically significant 8-bp motifs with high PBM enrichment (E-score = 0.497; 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; Supplemental Fig. S2D). Occupancy of both the MSL complex and CLAMP is greater on the X chromosome compared with the autosomal insertion (Supplemental 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.

To determine whether enrichment at CESs 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 CESs, GAF occupancy was lower at CESs compared with other GAF-binding sites throughout the genome (Supplemental Fig. S2E, S2F). These data suggest that CLAMP, but not GAF, directly recognizes MRE elements within CESs. In addition, CLAMP was recently shown to physically associate with the MSL complex on chromatin using an approach that combined ChIP with mass spectrometry (Wang et al. 2013). Therefore, CLAMP directly recognizes MREs and physically associates with the MSL complex, providing the first link between the MSL complex and the X chromosome.

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 five CESs 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 (<http://www.flyrnai.org>; see the Materials and Methods). Because CLAMP is highly expressed in all tissues, we expressed the RNAi construct ubiquitously (<http://www.flybase.org>; see the Materials and Methods). Although the CLAMP RNAi construct did not completely eliminate CLAMP mRNA levels (Supplemental Fig. S3A), immunostaining of polytene chromosomes for CLAMP (Supplemental Fig. S3A, S3C) and CLAMP

ChIP occupancy (Supplemental Fig. S3D, S3E) were strongly reduced. CLAMP RNAi causes a pupal lethal phenotype in both males and females (Larschan et al. 2012).

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 (Supplemental Fig. S3F). Our results showed that CLAMP RNAi strongly reduced the occupancy of the MSL complex in male larvae (Supplemental Fig. S3F). We measured H3 occupancy by ChIP to determine whether CLAMP RNAi globally disrupts chromatin and found only modest changes after CLAMP RNAi (Supplemental Fig. S3G). Also, we 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 (Supplemental 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 CESs (Fig. 1;

Supplemental Fig. 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 with autosomal controls (Supplemental Fig. S3B; Alekseyenko et al. 2008) but mediates at most a twofold 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 sequencing (mRNA-seq) experiments in male SL2 cells before and after CLAMP RNAi treatment. As expected for a non-sex-specific factor present on both the X chromosome and autosome (Supplemental Figs. 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 with autosomal genes ($P < 9.1 \times 10^{-5}$), consistent with a function in MSL complex targeting. Consistent with its localization throughout the genome and enrichment on the X chromosome (Supplemental Fig. S1), CLAMP RNAi affects gene expression on both the X chromosome and autosomes with a bias toward X-linked genes.

In contrast to average X-linked genes, the genes that encode the roX noncoding 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 quantitative RT-PCR (qRT-PCR) in male and female third

instar larvae (Supplemental Fig. S3A). In males, we found that roX2 levels were reduced fourfold 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 2002; Bai et al. 2004; Lim and Kelley 2012).

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

Identifying CLAMP as the protein that directly recognizes MREs 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 the MSL complex. Overall, we found that the percentage of MREs occupied by CLAMP was greater on the X chromosome when compared with autosomes in both male

cells (50% on X vs. 32% on autosomes) and female cells (62% on X vs. 38% on autosomes) (Supplemental Fig. S4A, S4B, S4C). 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 CESs. To define the total number of likely CESs for subsequent analysis, we determined how the largest class of previously characterized CESs (306 loci) (Sural et al. 2008) overlapped with CLAMP sites present in male cells and found 265 common sites (87% overlap). Next, 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 (Fig. 3A; see the Supplemental Material). We also plotted the average occupancy at each of these classes of CESs compared with the average CLAMP site on the X chromosome and the autosome (Fig. 3B). In this way, three distinct classes of CESs were defined (Fig. 3A, 3B). Group A: Largely MSL-dependent CESs exhibit a strongly decreased CLAMP occupancy in the absence of the MSL complex, and their occupancy in female cells is much below that of the average CLAMP site (177 sites). Group B: Partially MSL-dependent CESs, including both roX loci, exhibit higher than average levels of CLAMP occupancy in the absence of the MSL complex and further increase in their occupancy and breadth in the presence of the MSL complex (43 sites). Group C: Largely MSL-independent CESs have higher than average CLAMP occupancy in

both male and female cell lines (45 sites). ChIP-qPCR was used to validate these classes in both cell culture and larvae (Supplemental Fig. S4D).

Thus, group B and group C CESs, including both roX loci, have higher than average levels of MSL-independent CLAMP occupancy compared with group A sites, which have lower than average MSL-independent CLAMP occupancy. These data suggest that group B and group C CESs may function as beacons to promote X identification because they are occupied by CLAMP independent of the MSL complex. In addition, CLAMP associates with chromatin interdependently at group A and group B sites (Fig. 3; Supplemental Fig. S3B, S3F). 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 that distinguish each subclass of CESs, we compared the following properties: the number of tandem MREs within each CES, the two-dimensional clustering of CESs along the X chromosome, and the average nucleosome occupancy in the absence of the MSL complex (Fig. 4; Supplemental Fig. 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 the MSL complex using publicly available female cell data sets (Supplemental Fig. S4E; Henikoff et al. 2009; <http://www.modencode.org>). Under high-salt extraction conditions, which are likely to provide the most accurate nucleosome occupancy profiles over CESs, the decreased nucleosome density at group B and group C loci in female cells resembles that previously reported for male cells (Supplemental Fig. S4E; Alekseyenko et al. 2012).

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 (Supplemental Fig. S4E) are features that distinguish subclasses of CESs and were not previously identified by analysis of the average properties of all CESs (Alekseyenko et al. 2008, 2012). By defining three distinct subclasses of CESs, we provide a new tool for examining how the MSL complex identifies the X chromosome.

Discussion

Our identification of CLAMP as the previously unknown link that tethers the MSL complex to the X chromosome 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 group C CESs, 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 (Kelley et al. 1997; Meller and Rattner 2002) and is likely to facilitate cotranscriptional MSL complex assembly (Lim and Kelley 2012). CLAMP is enriched at the CESs adjacent to roX loci independent of the MSL complex and tethers the MSL complex to the X chromosome (Figs. 1, 2, 3). (3) The MSL complex catalyzes acetylation of H4K16 (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 (Fig. 3A; Supplemental Fig. S3B, S3C, S3F) are likely to contribute to the high occupancy of both factors at the roX loci, followed by spreading to additional CESs (Fig. 3A; Supplemental Figs. S3F, S4). Through this proposed mechanism, we suggest that CLAMP functions together with roX RNAs to target the MSL complex to the X chromosome. 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.

Methods

Flies and crosses

The CLAMP RNAi fly line was obtained from the TRiP project (www.flyrnai.org). 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

A protocol was adapted from Larschan et al., 2007.

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 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 μl 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

Standard cell culture approaches were used to maintain *Drosophila* SL2 and Kc cells and RNAi treatment conditions are described at www.flyrnai.org. 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 μ g 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

A protocol was adapted from Gelbart et al., 2009.

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 for ChIP-seq and mRNA-seq

A protocol was adapted from Alekseyenko et al., 2008.

Data will be available at NCBI GEO Short Read Archive (SRA). GEO release

number GSE39271.

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 μ M 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. 3xA, 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. 4xB, 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) and PBM**

A protocol was adapted from Berger et al., 2006.

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

A protocol was adapted from Kelley et al., 1999.

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.

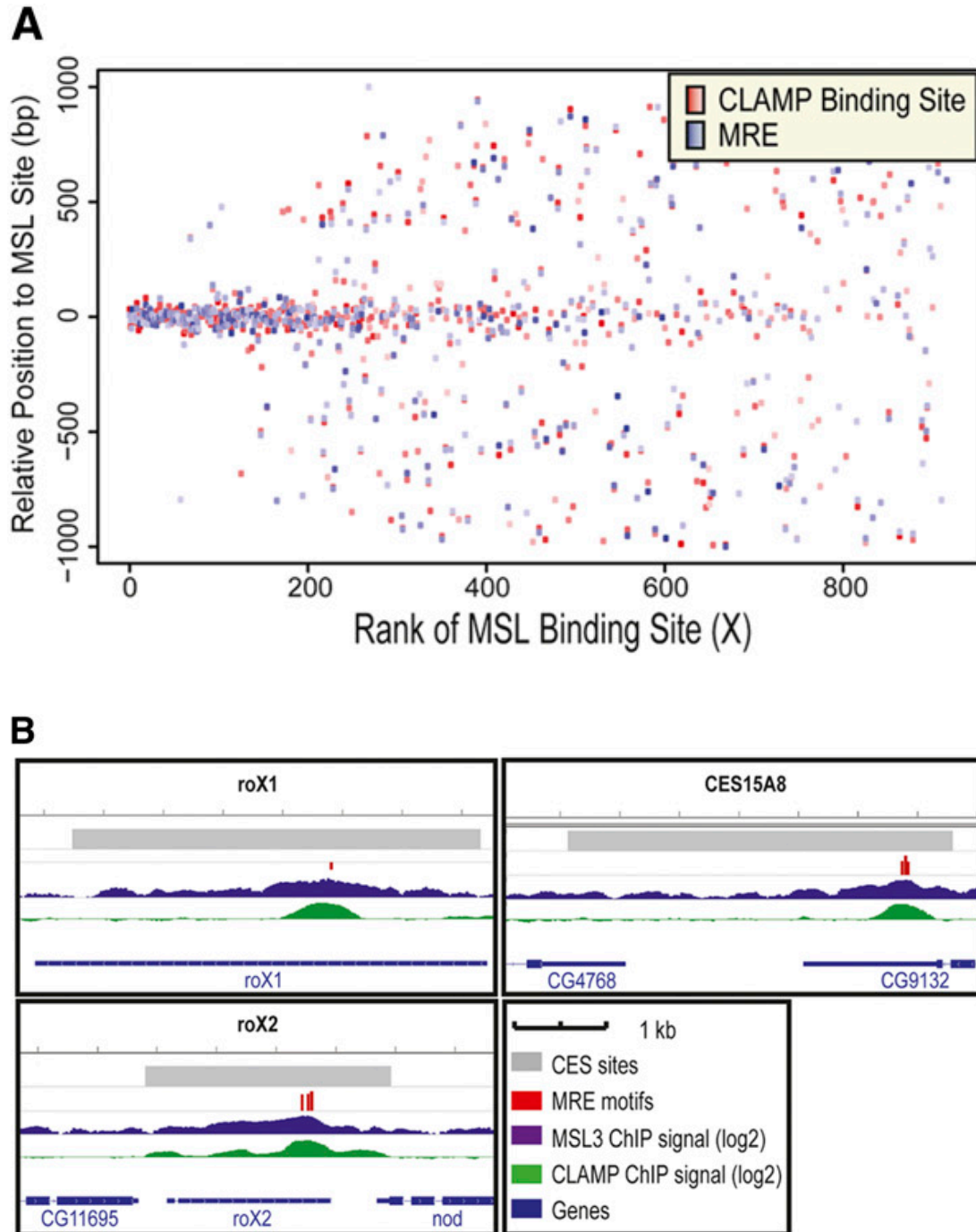


Figure 1. CLAMP localizes precisely at MREs at high levels when colocalized with the MSL complex.

(A) Occupancy relationships calculated from input-normalized CLAMP ChIP-seq from male SL2 cells (Larschan et al. 2012) and previously defined CESs

(Alekseyenko et al. 2008) (two biological replicates). MSL complex-binding sites 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 intensity of the red color is proportional to CLAMP occupancy (see the Materials and Methods). The intensity of the blue color is proportional to similarity of the individual MRE sequence to the consensus sequence. (B) Examples of CLAMP and MSL complex occupancy at three CES are shown from SL2 cells. Input- normalized read counts are shown on a log scale for CLAMP and the MSL complex.

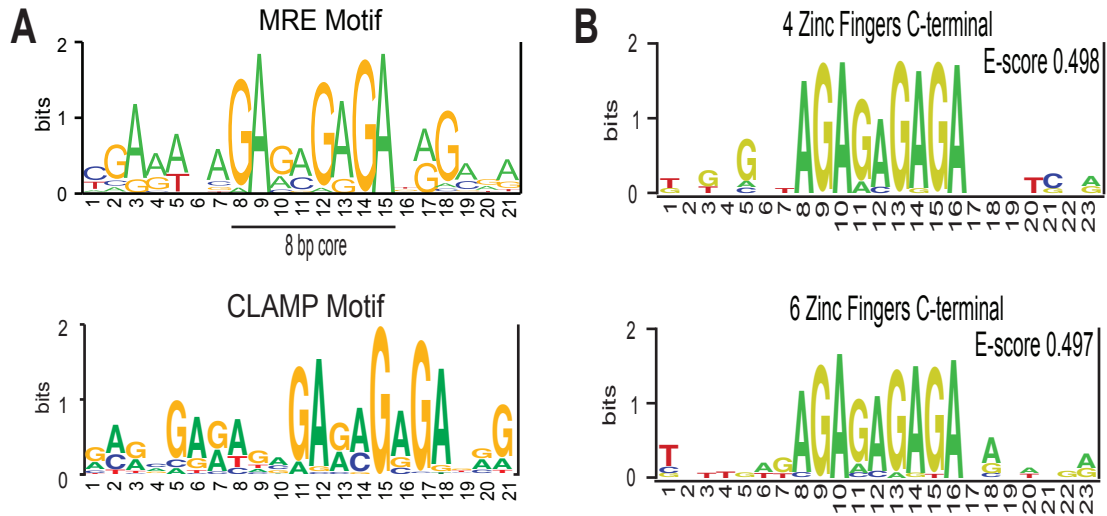


Figure 2. CLAMP directly binds to MRE sequences.

(A, top) PWM for the original MRE motif (Alekseyenko et al. 2008). (Bottom) PWMs 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 at the 39 ends of active genes on the X chromosome. (B) PBM analysis resulted in direct in vitro binding specificities of the C-terminal CLAMP four (top) and six (bottom) zinc finger fragments.

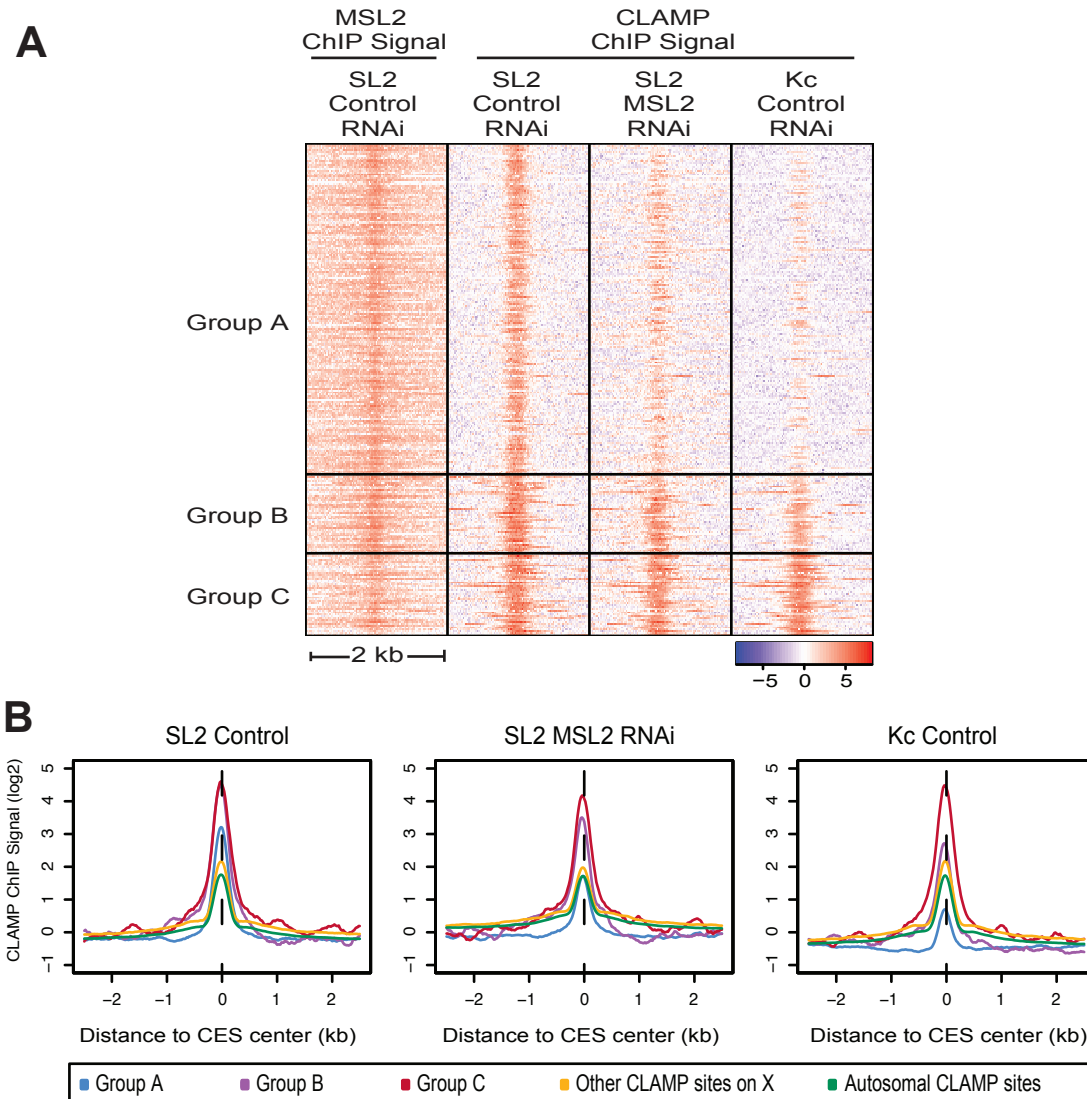


Figure 3. CESs have three classes of CLAMP occupancy patterns.

(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 CESs (265 total CESs). (1) Group A sites are highly MSL-dependent. (2) Group B sites are partially MSL-dependent. (3) Group C sites are independent of the MSL complex. (B) Average profiles of CLAMP occupancy (RPKM) normalized to input are centered at group A, group B,

or group C CES peaks and are shown compared with average occupancy at all CLAMP sites on the X or autosomes (log2 scale).

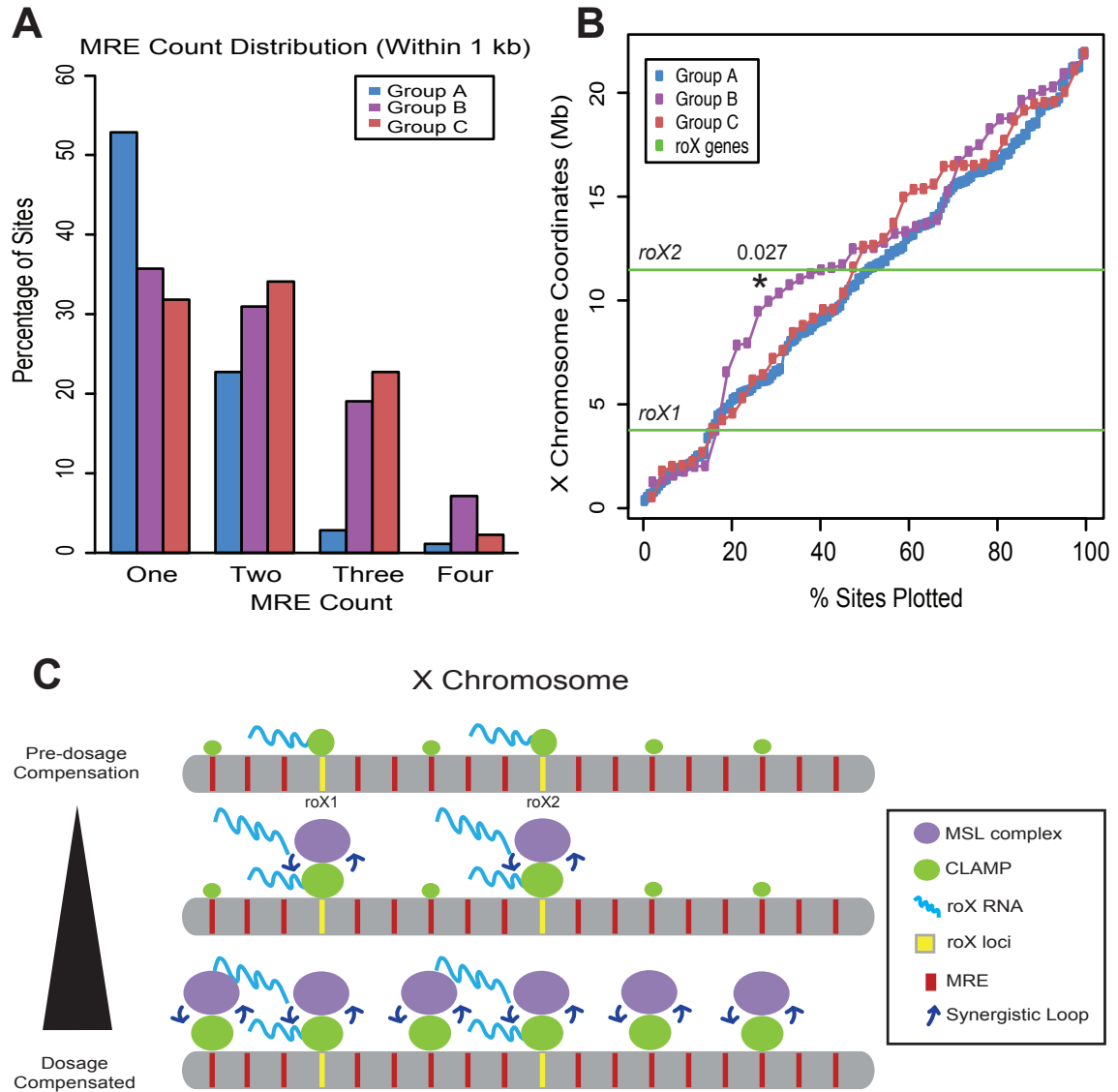
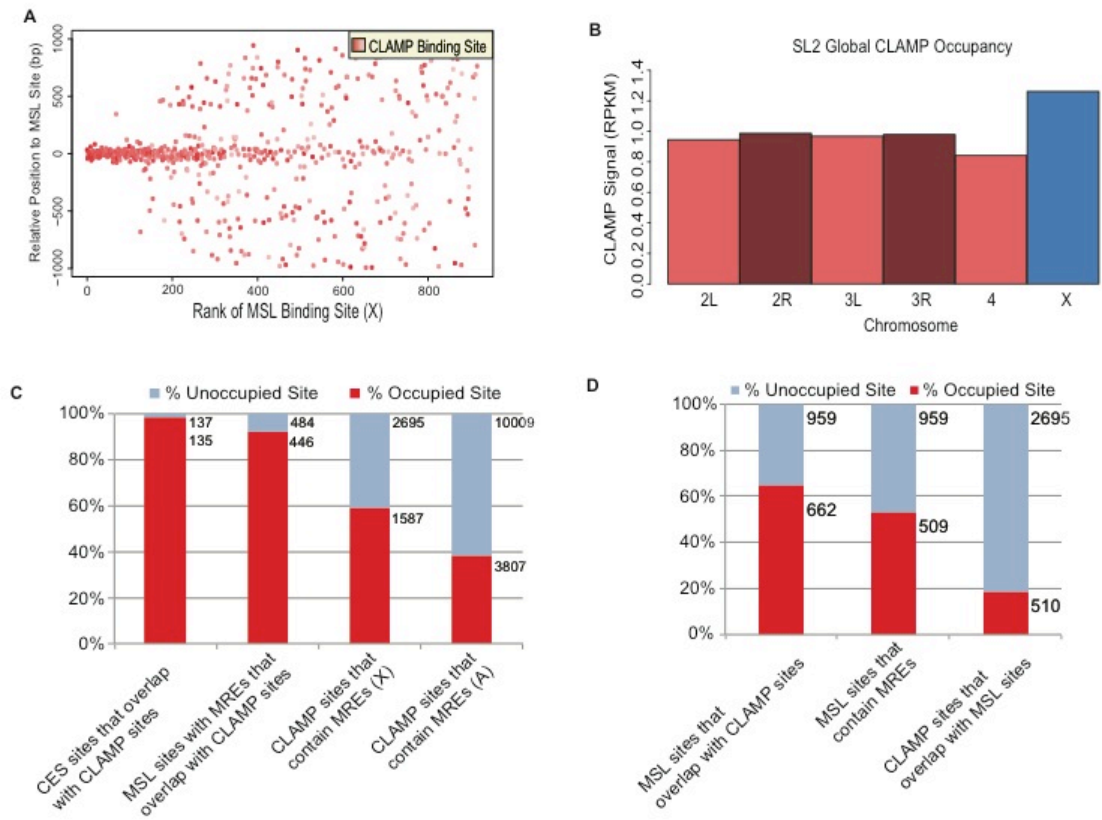


Figure 4. Three subclasses of CESSs differ in their number of tandem MREs and two-dimensional clustering along the X chromosome.

(A) Analysis of the number of tandem MREs within each subclass of CESSs. (B) The locations of CESSs from each subclass of sites are 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

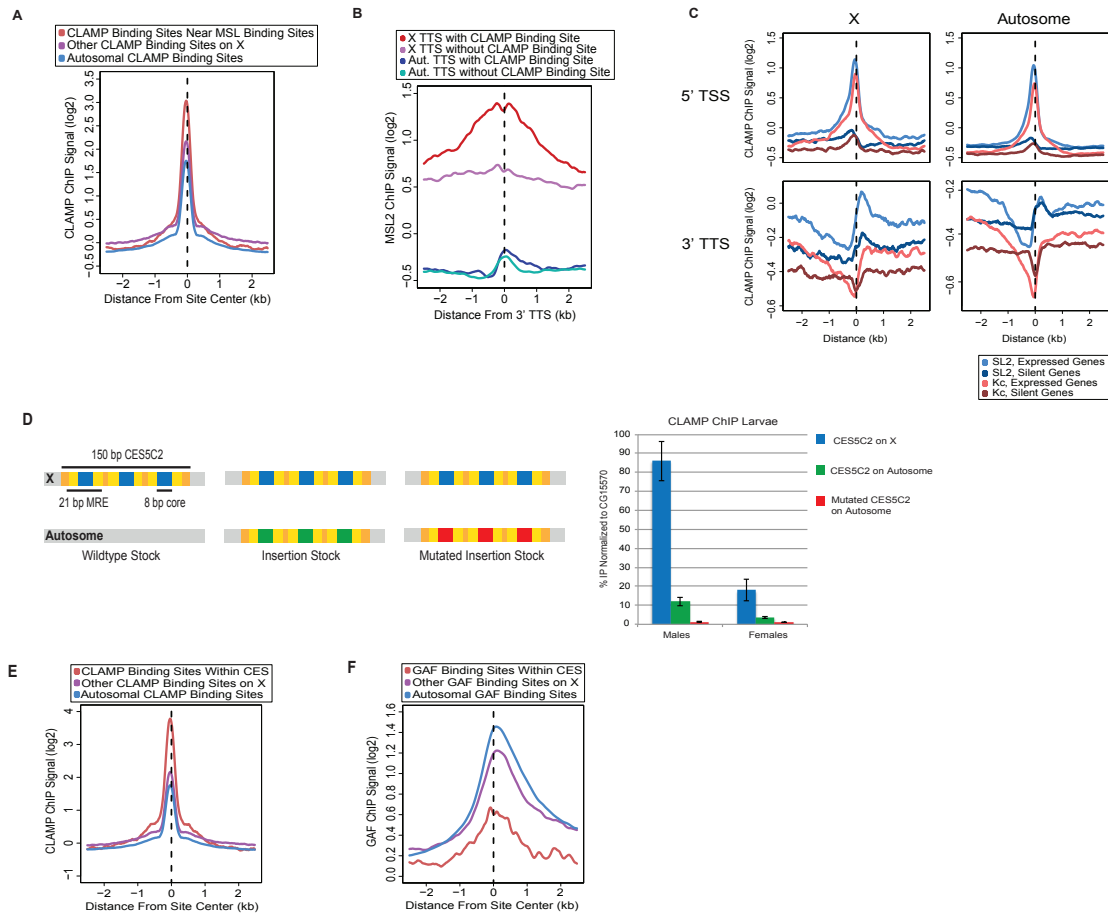
the percentage of sites plotted. The positions of the roX loci are shown as green lines. (C) Model for CLAMP and roX RNAs functioning together during X identification: (1) CLAMP binds to MREs within group B and group C independent of the MSL complex, and roX RNAs are transcribed independent of MSL complex recruitment. (2) The MSL complex is likely to be cotranscriptionally assembled at roX loci. CLAMP tethers the MSL complex to CESs, including roX loci. The MSL complex catalyzes H4K16ac, thereby opening chromatin and increasing the accessibility of MRE sequences for CLAMP recognition. Next, increased levels of CLAMP associate with CESs and thereby recruit additional MSL complexes via synergistic interactions. (3) The MSL complex spreads to additional CESs, potentially stabilized by CLAMP, thereby initiating additional synergistic interactions.



Supplemental 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.

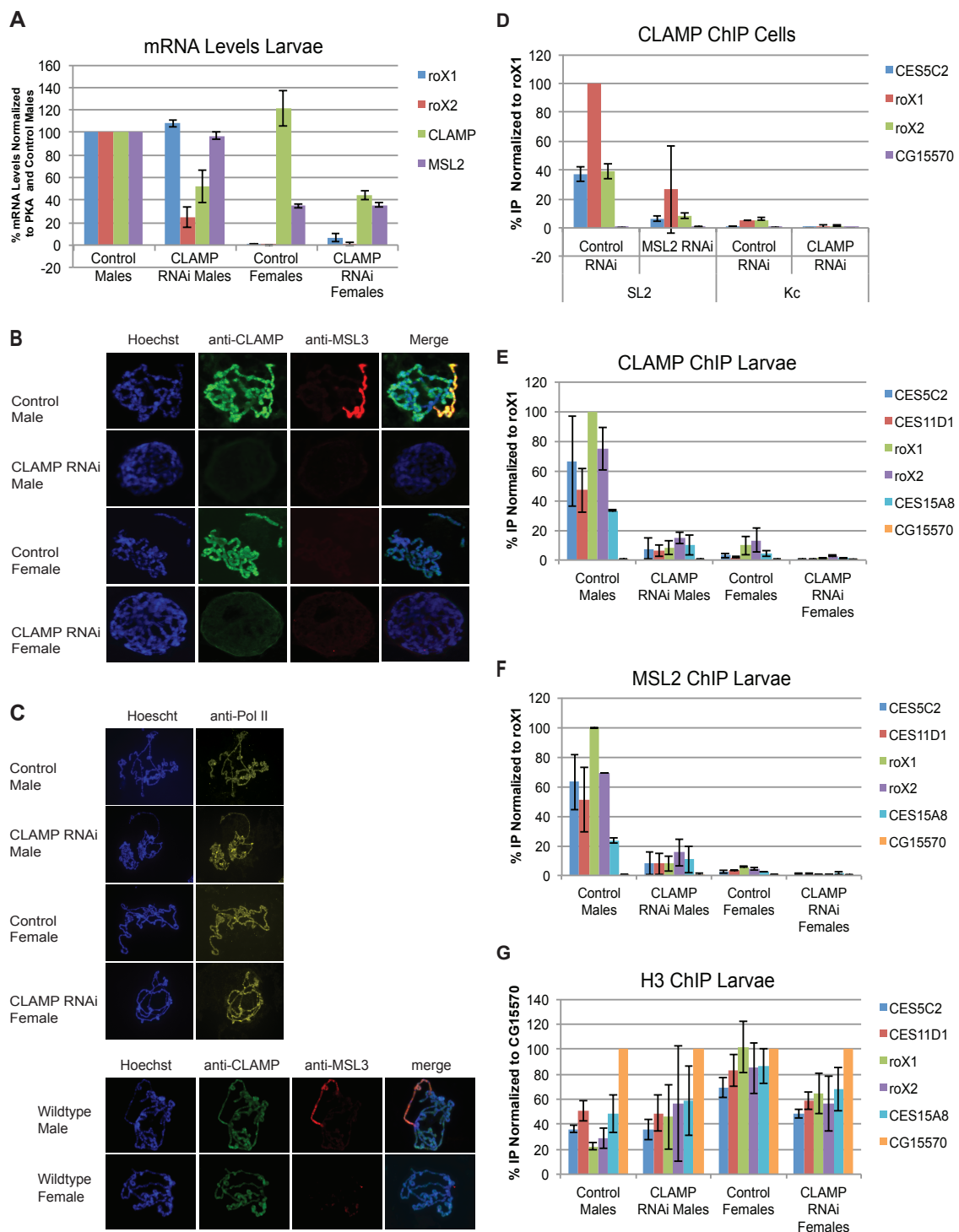
Each red dot represents the position of a CLAMP binding site that is within 1 kb of an MSL complex binding site. The closer the red dot is located to the zero position on the *y*-axis, the closer the CLAMP site is to the center of the MSL complex binding site. The intensity of the red color is proportional 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.



Supplemental 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

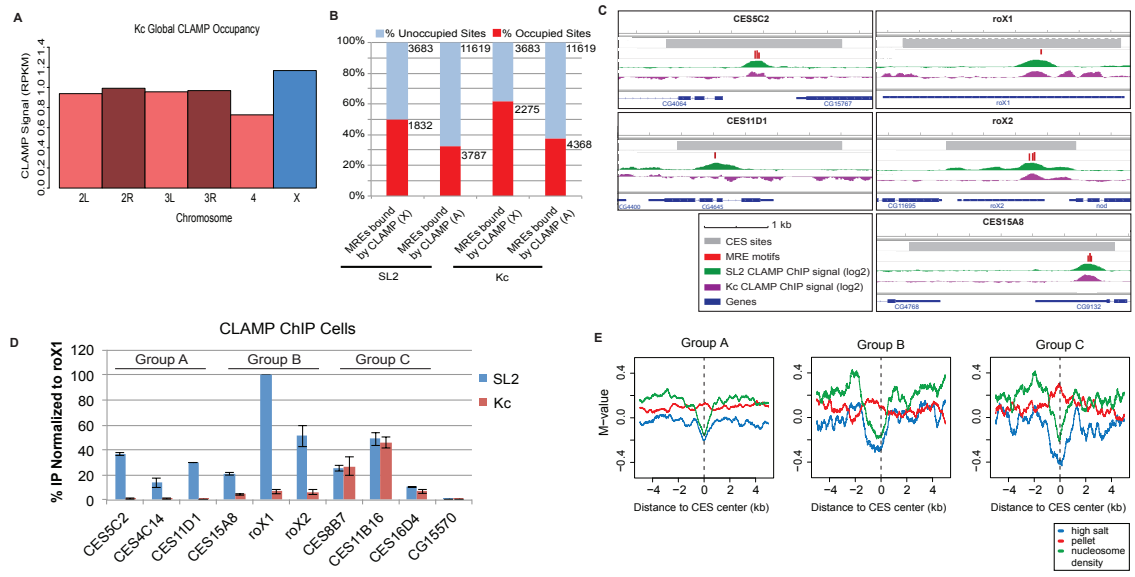
are centered around the transcription start site (TSS) or transcription termination site (TTS). CLAMP is highly enriched at TSS regions of active genes on all chromosomes in both SL2 and Kc cells. Enrichment near the TTS is specific to the male X chromosome and corresponds with the location of CES near 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)).



Supplemental 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 to 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. Normalized was conducted as described for **D)** except the CG15570 gene was used for normalization instead of *roX1* because it exhibits the highest H3 occupancy levels.



Supplemental 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.

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*.

Chapter 4

The non-sex specific role of CLAMP as a genome-wide transcription factor

Marcela M.L. Soruco¹, Leila Rieder¹, Jason Dobson¹, Catherine McMannus¹,
Tara Fresquez¹, Jessica Feng¹, Jacob Bliss¹, and Erica Larschan¹

¹Brown University, Providence, RI 02906, USA

This manuscript is in preparation for submission. I performed all the experiments except for the insertion immunostaining experiment in Figure 2, which was performed by Leila Reider. GFP imaging in Figure 7 was generated by the Fly Light group at the HHMI Janelia Farm Research Campus. Jacob Bliss maintained and counted flies. Jessica Feng maintained cell cultures. Jason Dobson processes raw mRNA-seq data. I wrote the text and generated all figures. Erica Larschan provided edits and comments.

Abstract

Dosage compensation evolved independently in various species containing heterogametic sex chromosomes. The protein complexes involved in dosage compensation also evolved independently and therefore differ across species. Many of the complexes responsible for the change in X-linked gene expression across species share components with other non-sex specific complexes. In *Drosophila*, the Male Specific Lethal (MSL) Complex is the co-activator complex responsible for increasing transcription of X-linked genes in males. Recently, we identified the CLAMP zinc finger protein as a key factor necessary for MSL complex targeting to the X chromosome. CLAMP works through direct tethering of the complex to GA-rich *cis*-acting sequences. However, CLAMP is not an X-specific factor and is found at GA-rich sequences throughout the genome. Therefore, it is critical to understand how the non-sex specific and male-specific functions of CLAMP are integrated.

Here we investigate the non-sex specific role of CLAMP as a genome-wide transcriptional regulator that occupies the promoters of active genes. CLAMP protein is ubiquitously expressed in embryos and larvae and CLAMP mRNA is most highly expressed in the ovaries of female flies and neurons of adult flies. Using a fly line containing a P-element insertion in the 5' UTR of the CLAMP gene, which results in reduced mRNA and protein levels, we see a reduction in survival of homozygous flies with males being the most severely

affected. Interestingly, we also observe a distinct defect in flying and movement of both male and female homozygous mutant flies that could be linked to reduced CLAMP expression in neurons. Additionally, to identify genes that could be responsible for the observed defect, we performed mRNA-seq in male and female cells. This revealed CLAMP regulating genes involved in brain development and morphogenesis. This is the first time that CLAMP function has been linked to important developmental processes and could aid in our understanding of the evolution of dosage compensation.

Introduction

Dosage compensation is an essential mechanism responsible for equalizing X-linked gene expression between the sexes in heterogametic species. In mammals, the X chromosome is up-regulated in males (XX) and females (XY) and then one X chromosome is randomly inactivated in females (Jonkers et al. 2008; Deng et al. 2013). In *C. elegans*, the X chromosome is up-regulated in both males (XX) and hermaphroditic females (X0); then transcripts produced from both X chromosomes are down-regulated two-fold in males (Meyer 2010; Deng et al. 2011). In *Drosophila*, the X chromosome is up-regulated approximately two-fold only in males (XX) while no regulation takes place in females (XY) (Kelley et al. 1995).

Diverse dosage compensation mechanisms have evolved independently in the different species (Livernois et al. 2012). As expected, the protein complexes which are responsible for these changes in gene expression are also not shared among species. In mammals, the Polycomb repressive complex (PRC) is responsible for X-inactivation in females (Plath et al. 2003; Simon et al. 2013). In *C. elegans*, the dosage compensation complex (DCC), which shares components with the condensin complex, down-regulates X-linked gene expression in hermaphrodites (Meyer 2010; Wood et al. 2010). In *Drosophila*, the male specific lethal (MSL) complex, which shares a component with the essential

non-specific lethal (NSL) complex, up-regulates the X chromosome in males (Belote and Lucchesi 1980; Kelley et al. 1997; Lam et al. 2012).

We recently identified a previously unstudied zinc finger protein using a cell-based screening approach to identify non-sex specific factors that are also involved in dosage compensation in *Drosophila* (Larschan et al. 2012). The Chromatin Linked Adapter for MSL Proteins (CLAMP) protein was found to be an important regulator of *Drosophila* dosage compensation. CLAMP directly recognizes important *cis*-acting sequences enriched on the X chromosome and is required for MSL localization to the X chromosome. Additionally, CLAMP is found targeting GA-rich sequences throughout the genome and is localized to the 5' end of transcribed genes (Soruco et al. 2013). Here we investigate the role of CLAMP as a genome-wide transcription factor. We first characterized a fly line containing a P-element insertion in the 5' UTR of the CLAMP gene to identify developmental defects due to the loss of CLAMP expression. Additionally, we aimed to identify genes that could be directly regulated by the CLAMP protein using mRNA-seq in cell culture before and after CLAMP RNAi.

Results

CLAMP RNA is expressed at most developmental time points in many tissues type but it is found very highly enriched in the early embryo, ovaries in adult females and neurons of adult flies (www.flybase.org)(Graveley et al. 2011).

Ovary expression is likely due to increased expression necessary for maternal deposition of CLAMP mRNA in the developing egg (www.flybase.org). To investigate the non-sex specific role of CLAMP, we first characterized defects in fly development that result from the loss of CLAMP expression.

The CLAMP locus is located in heterochromatin flanking the centromere on the arm of chromosome 2L (Fig 1A). Due to the highly repetitive nature of this region, there are currently no available null or deletion mutants. Our recent attempts at generating a null mutant using imprecise excision and homologous recombination for this difficult region have been unsuccessful (data not shown). To investigate the non-sex specific role of CLAMP, we characterized an available P-element insertion line containing an 11.5 kb transposable element artificially inserted in the 5' UTR of the CLAMP gene to observe developmental defects due to the loss of CLAMP expression (Fig 1B).

Homozygous insertion line shows decreased CLAMP mRNA and protein levels

To determine whether the insertion line shows a decrease in CLAMP levels, we quantified mRNA and protein levels in third instar larvae raised at 18 degrees and compared levels to a control line (yw) raised at the same temperature. To assess mRNA levels, we performed qRT-PCR on cDNA from heterozygous and homozygous insertion larvae. In the insertion line, we see a

70% and 80% decrease in CLAMP mRNA in male and female homozygous larvae, respectively, compared to control larvae (Fig 2A, right).

To assess protein levels and localization, we performed immunostaining on polytene chromosomes from the salivary glands of heterozygous and homozygous males and compared levels to control larvae. Here we see CLAMP levels similar to control larvae for heterozygous males. However, in homozygous males and females, we see a dramatic decrease in CLAMP localization (Fig 2B). Together, these data indicate that the P-element insertion is a haplo-sufficient mutation and when homozygous is disrupting CLAMP expression.

Homozygous insertion males show additional dosage compensation defects

Previously, we observed an absence of MSL3 localization to the X chromosome in larvae expressing a CLAMP RNAi construct driven by the Ga4-UAS system (See Chapter 3) (Soruco et al. 2013). To see if MSL3 localization is altered in the insertion line, we stained for MSL3 on polytene chromosomes in heterozygous and homozygous larvae. Here we see a marked decrease in MSL3 localization to the X chromosome in heterozygous insertion larvae and a dramatic decrease in homozygous larvae compared to control larvae (Fig 2B). The reduction of MSL3 localization due to reduced levels of CLAMP protein

supports the requirement of CLAMP in recruiting MSL complex to the X chromosome (See Chapter 3) (Soruco et al. 2013).

Additionally, in both heterozygous and homozygous insertion males we see a modest increase in *roX* mRNA levels (Fig. 2A, left). We had previously observed a decrease in *roX2* levels and a modest increase in *roX1* levels after CLAMP RNAi treatment in both male and female larvae (See Chapter 3)(Soruco et al. 2013). Therefore, if CLAMP is regulating *roX* transcription, it is possible that the residual CLAMP protein levels differ between the RNAi treatment and the insertion line resulting in differences in *roX* mRNA levels.

P-element insertion causes a decrease in viability of homozygous flies

We previously illustrated that a reduction of CLAMP mRNA and protein levels using a CLAMP RNAi line results in pupal lethality in both males and females flies suggesting that CLAMP is an essential protein (See Chapter 3) (Soruco et al. 2013). To quantify defects in survivorship for the insertion line, we counted both heterozygous and homozygous larvae and adult flies over a 16 day time course. We compared homozygous to heterozygous animals where the expected Mendelian ratio for homozygous animal would be 50% for a mutant lines raised over a balancer chromosome.

Overall, we detect a less severe phenotype in the P-element insertion line compared to the RNAi line. For male larvae containing the insertion, we see a slight reduction in homozygous male larvae comparing 11 homozygous to 25 heterozygous larvae or 44% for homozygous male larvae compared to a 50% expected for the Mendelian ratio (Fig. 3A). Interestingly, the homozygous female larvae counts are higher than expected with 22 homozygous to 20 heterozygous larvae or 110% of homozygous female larvae compared to 50% for expected (Fig 3A). This nearly 2-fold increase in homozygous female larvae counts could potentially be caused by changes in morphogenesis due to changes in transcription.

We then counted the adult flies comparing homozygous to heterozygous flies. For male flies, we see a dramatic reduction in homozygous flies with 4 homozygous to 32 heterozygous insertion flies or 12.5% homozygous compared to 50% for expected (Fig 3B). For female flies, we see a modest reduction in homozygous flies with 9 homozygous to 21 heterozygous insertion flies or 42.8% homozygous compared to 50% expected (Fig. 3B). Together, these results indicate that CLAMP is an important protein for both males and females with male survival being very dependent on normal CLAMP levels. This dramatic sensitivity seen in males is potentially due to the additional requirement of CLAMP during dosage compensation.

Decreases in CLAMP expression causes severe motor defects in males and reduced flying in females

In addition to the decrease in the number of male flies surviving to adulthood (Fig 3B), there is a severe motor or neuronal defect in both the males and females of the surviving homozygous flies containing the P-element insertion. In these surviving flies, we see a visible defect in crawling and flying in the females and a severe defect in the males where they are unable to fly and stumble while walking (Fig 4) (Link to movies <https://www.dropbox.com/sh/8oqppast5imuyp8/l6gJ0x3AkR>). This defect is most likely due to decreased mRNA and protein levels in the developing fly (Fig 2). However, further neurological characterizations and behavior assays are needed to understand the specific role CLAMP is playing in adult development.

mRNA-seq in male and female cells shows that the loss of CLAMP results in both positive and negative changes in gene expression

We previously performed ChIP-seq in male and female cells and found CLAMP enriched at the 3' end of active genes on the X chromosome in male cells (See Chapter 3) (Soruco et al. 2013). This 3' localization is similar to MSL Complex localization, which is responsible for increased transcription of X-linked genes during dosage compensation (Alekseyenko et al. 2006). We also found CLAMP enriched at the 5' end of active genes on all chromosomes similar to

other ubiquitous transcription factors (Alekseyenko et al. 2013; Soruco et al. 2013). Therefore, to identify changes in gene expression due to the loss of CLAMP expression we performed paired-end mRNA-seq in male and female cells before and after CLAMP RNAi. Assaying changes in gene regulation can be difficult in whole animals due to the differences in the varying transcriptional profiles of the many cell lineages. Importantly, CLAMP expression is reduced after the CLAMP RNAi in both male and female cells (See Chapter 3) (Soruco et al. 2013). Using this technique, we performed two biological replicates and selected all genes with an RPKM greater than 3 and that have a greater than 2-fold change in gene expression after RNAi treatment. We further refined these gene lists by selecting genes that had a greater than 0.2 RPKM in either the control or CLAMP RNAi sample, identifying top 50% of genes in each each list. This ensured that sufficiently expressed genes were selected. Using these criteria, we identified 422 genes in male cells and 329 genes in female cells that were differentially expressed after CLAMP RNAi treatment.

We then identified 323 genes in which expression is decreased after CLAMP RNAi and 428 genes that increased in expression after RNAi treatment. We then categorized these genes as being either positively or negatively regulated by CLAMP, respectively. In male cells, we identified 156 genes that are positively regulated by CLAMP and 266 genes that are negatively regulated by CLAMP. In female cells, we identified 167 genes that are positively regulated and 162 genes that are negatively regulated by CLAMP (Fig 5A). We then

arranged these genes base on their chromosome location where we see genes in both groups located on all chromosomes. This is expected considering that CLAMP is localized to every chromosome in both males and female larvae (Fig. 2).

Next, we compared genes located on each chromosome to a random assignment of genes genome-wide (intermine.modencode.org). In male cells, we detected a large difference in the number of genes regulated on the X chromosome and 2L compared to a random assignment that is not detected in female cells. The large difference in the number of genes regulated on the X chromosome after CLAMP RNAi could be due to a dual role CLAMP is playing when localized to both the 5' and 3' ends of active genes on the X chromosome (See Chapter 3) (Soruco et al. 2013). We also detected a large difference in the number of regulated genes on chromosome 2R and 3L in both male and female cells compared to a random assignment. Further analysis is needed identify if any additional transcription factors which could be contributing to this affect.

We next performed Gene Ontology (GO) Term analysis using the ModMine software (intermine.modencode.org). Using this technique, we observe an over-representation of genes involved in metabolic processes and defense response both positively and negatively regulated in male and female cells (data not shown). Additionally, we look at genes that overlap in the two cell types where we observe 24 positively regulated and 19 negatively regulated genes in

both male and female cells (Fig 6). Importantly, CLAMP regulates genes involved in morphogenesis and neurological development in both males and females (Fig. 6, red). Therefore, it is likely that CLAMP is required for appropriate neurological function due to regulation of genes affected in both sexes. Importantly, this is consistent with the defect observed in the homozygous insertion flies (Fig 4). CLAMP positively regulates genes involved in olfactory behavior and circadian rhythms (Fig. 6 left, blue). Additionally, the CLAMP gene was detected as being positively regulated due to the decrease in expression after RNAi treatment (Fig. 6 left, green). Detection of this decrease in CLAMP expression is an important validation of our filtering criteria. Finally, CLAMP negatively regulates genes involved in motor behavior, visual behavior and reproduction (Fig. 6 right, purple).

Discussion

Using an available P-element insertion fly line and mRNA-seq in cells, we have uncovered a new potential role for the CLAMP protein during development. Because a complete null mutant is not yet available despite our extensive efforts, we characterized the phenotypic changes caused by an 11.5 kb P-element insertion of the 5' UTR of the CLAMP gene. We determined that decreased CLAMP protein levels causes severe viability defects in adult male flies as well as results in neuromuscular defects in male and female flies. Furthermore, CLAMP regulates genes involved in neurological development, which could potentially link changes in gene expression with the observed phenotype in the

P-element insertion line. It is possible that this is a direct effect of CLAMP binding to promoters of these genes or an indirect effect of CLAMP regulating another transcription factor.

CLAMP was recently found associated with the Myb transcription complex in a mass spectrometry experiment (Wang et al. 2013). The Myb transcription complex can act as both a repressor and activator complex depending on the components that are colocalized (Lewis et al. 2004). In neurons, the Myb-MuvB complex works to suppress specific olfactory genes to ensure proper differentiation and CO₂ sensing (Sim et al. 2012). It is possible that either CLAMP is directly regulating target genes identified by our mRNA-seq or regulation occurs via recruitment of a larger transcription complex. Further work will include additional replicates of mRNA-seq analysis and comparison with ChIP-seq data to determine whether target genes are in close proximity to a CLAMP peak. We will also correlate our mRNA-seq and ChIP-seq with existing data that identifies genes targeted by the Myb transcription complex.

Recently, the Fly Light group at the HHMI Janelia Farm Research Campus developed lines for genes that are specifically expressed in neurons which express GFP protein from 1 - 2 kb promoter regions. From this data set, there is a line containing the 1 kb region upstream relative to the CLAMP locus which is a potential enhancer region (Fig 7A) (Pfeiffer et al. 2008). In this line, the putative CLAMP enhancer region (R39B02) drives broad expression in the larval central

nervous system and limited expression in the pupal ventral nerve cord. In the adult fly brain, there are just a few GFP expressing neurons specific to the periesophageal neuropils center responsible for mechanosensory function of antennal segments (Fig 7B) (Virtual Fly Brain) (Milyaev et al. 2012). In the future, this line could be a valuable tool in assessing the specific region of brain in which CLAMP is required because these neurons can be specifically stimulated using genetic tools. We will perform CLAMP immunostaining in the brain to determine how this driver line correlates with CLAMP protein expression in the brain.

Finally, we have been diligent in our attempts to generate a true null allele for CLAMP. We are currently using TALEN (transcription activator-like effector nuclease) and CRISPR (clustered regularly interspaced short palindromic repeats) technologies available through the Drosophila RNAi Screening Center (DRSC) to edit the genome by deleting the first intron of the CLAMP locus (Cermak et al. 2011; Briggs et al. 2012). We believe this method will succeed where others have failed due to the sequence specificity engineered into these technologies. We are also currently designing a CLAMP rescue line using the P(acman) recombineering system, which allows for site-specific integration of large fragments of DNA into the genome (Venken et al. 2009). We have designed a construct to contain the entire CLAMP locus (~3 kb) and a 6 kb upstream region, 9 kb in total, to be artificially targeted to a previously characterized region on chromosome 3L. By crossing this rescue line to the mutant and insertion flies we would hope to rescue any defects in development,

morphology or behavior and confirm the essential role of CLAMP in fly development.

Acknowledgements

We are grateful to the Bloomington stock center for providing the insertion line and *Drosophila* RNAi Screening Center (DRSC) for advice and reagents for generating the TALEN and CRISPR constructs. Next Generation sequencing was performed at the Brown University Genomics core facility.

Methods

Flies and Larval counts

Fly stocks were provided by the Bloomington stock center and raised on yeast, sugar, cornmeal-based media at 18 °C for multiple generations. The CLAMP insertion line is from Bloomington #14352 (y(1); P{SUPor-P}Clamp(KG00473)/SM6a; ry(506)). The first wave of third instar larvae was scored as 1st wave and all subsequent days were scored as larvae continued to emerge.

Larval and cell mRNA sample preparation

10 sexed third instar larvae were collected, flash frozen in liquid nitrogen and homogenized using a Tissue Lyser (Invitrogen) and stainless steel beads. Total RNA was extracted using the Qiagen RNeasy Plus Kit at room temperature. 1 mL of cells was centrifuged at 600 g for 2 minutes and the media was removed. The cells were disrupted by adding 350 μ l Buffer RLT with 2-mercaptoethanol (10 μ l/1ml). Lysate was transferred to a QIAshredder spin column in a 2 ml collection tube and centrifuge for 2 min at maximum speed. The homogenized lysate was transferred to a gDNA Eliminator spin column and centrifuge for 30 s at $\geq 8000 \times g$ ($\geq 10,000$ rpm). To the flow through, 1 volume (usually 350 μ l) of 70% ethanol was added and mixed by pipetting. The entire volume (700 μ l) was transfer to an

RNeasy spin column and centrifuged for 15 s at $\geq 8000 \times g$ ($\geq 10,000$ rpm). The flow-through was discarded. 700 μ l Buffer RW1 was added to the RNeasy spin column and centrifuged for 15 s at $\geq 8000 \times g$ ($\geq 10,000$ rpm) to wash the spin column membrane. Discard the flow-through and reuse collection tube. The flow-through was discarded. 500 μ l Buffer RPE with ethanol was added to the RNeasy spin column and centrifuge for 15 s at $\geq 8000 \times g$ ($\geq 10,000$ rpm) to wash the spin column membrane. The flow-through was discarded. 500 μ l Buffer RPE with ethanol was added to the RNeasy spin column and centrifuge for 2 min at $\geq 8000 \times g$ ($\geq 10,000$ rpm) to wash the spin column membrane. The RNeasy spin column was placed onto a new 2 ml collection tube and centrifuged at full speed for 1 min to eliminate any residual ethanol. The RNeasy spin column was placed onto a new 1.5 ml collection tube. 30 μ l RNase-free water directly to the spin column membrane and centrifuged for 1 min at $\geq 8000 \times g$ ($\geq 10,000$ rpm) to elute the RNA. Using the eluate from previous step was pipetted onto the RNeasy spin column and centrifuged for 1 min at $\geq 8000 \times g$ ($\geq 10,000$ rpm) to increase yield and concentration of RNA.

cDNA synthesis and Quantitative real time RT-PCR

Synthesis of cDNA was conducted as described in Alekseyenko et al. 2006 using 1 mg 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 StepOnePlus

real time PCR machine.. cDNA was diluted prior to qPCR to ensure values were in linear range. Values were normalized to PKA, a highly expressed gene. All primers were tested to assure that they have 100% efficiency and good linearity prior to use with an ABI StepOnePlus real time PCR machine. The average relative quantitation (ABI User Bulletin #2) of three technical replicate reactions was determined for all samples. Platinum SYBR master mix w/ROX (Invitrogen) was used. Primer sequences are as follows. roX1 , F- ATGCGAGCGAGACAATGATACT, R-GACTTGCAGTCCGCCCTATG, roX2, F- CGTTACTCTTGCTTGATTTTGC, R-AGCTCGGATGGCCATCGAAA, MSL2 , F- GCCCTGTCCGTATGAATG, R-CAACATGGGTAAACAACC, CLAMP, F- CTGACCACGGCAATAGAA, R-TCCCTCATTTGGCACTTC, PKA, F- CAATCAGCAGATTCTCCGGCT, R-AGCCGCACTCGCGCTTCTAC

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.

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., submitted). Primer sequences are available on request and at www.flyrnai.org.

RNAi treatment of SL2 and Kc cells

RNAi treatment was conducted in T225 flasks as described at www.flyrnai.org. SL2 and Kc 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, mRNA preparation using the RNAeasy Plus kit (Qiagen) was performed (as described above). SL2 and Kc cells were maintained in Gibco Schneider's medium with 10% heat inactivated FBS (JRH).

mRNA seq and analysis

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

GO Term Analysis

Gene ontology analysis was performed using gene lists with a greater than 2-fold change in expression after CLAMP RNAi. Gene lists were then imported into the online modMine list analysis tool developed by the ModEncode consortium. The GO terms were reported with no correction for genes with a less than 0.5 p-value.

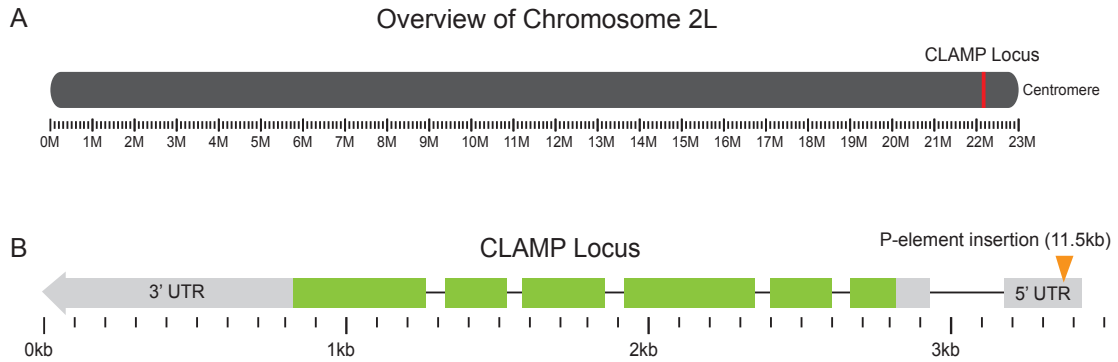


Figure 1: Overview of CLAMP locus and P-element insertion.

A) The CLAMP locus is located in heterochromatin surrounding the centromere on chromosome 2L. B) The CLAMP insertion line consists of an 11.5 kb transposable element, P-element, inserted in the 5' UTR of the CLAMP gene.

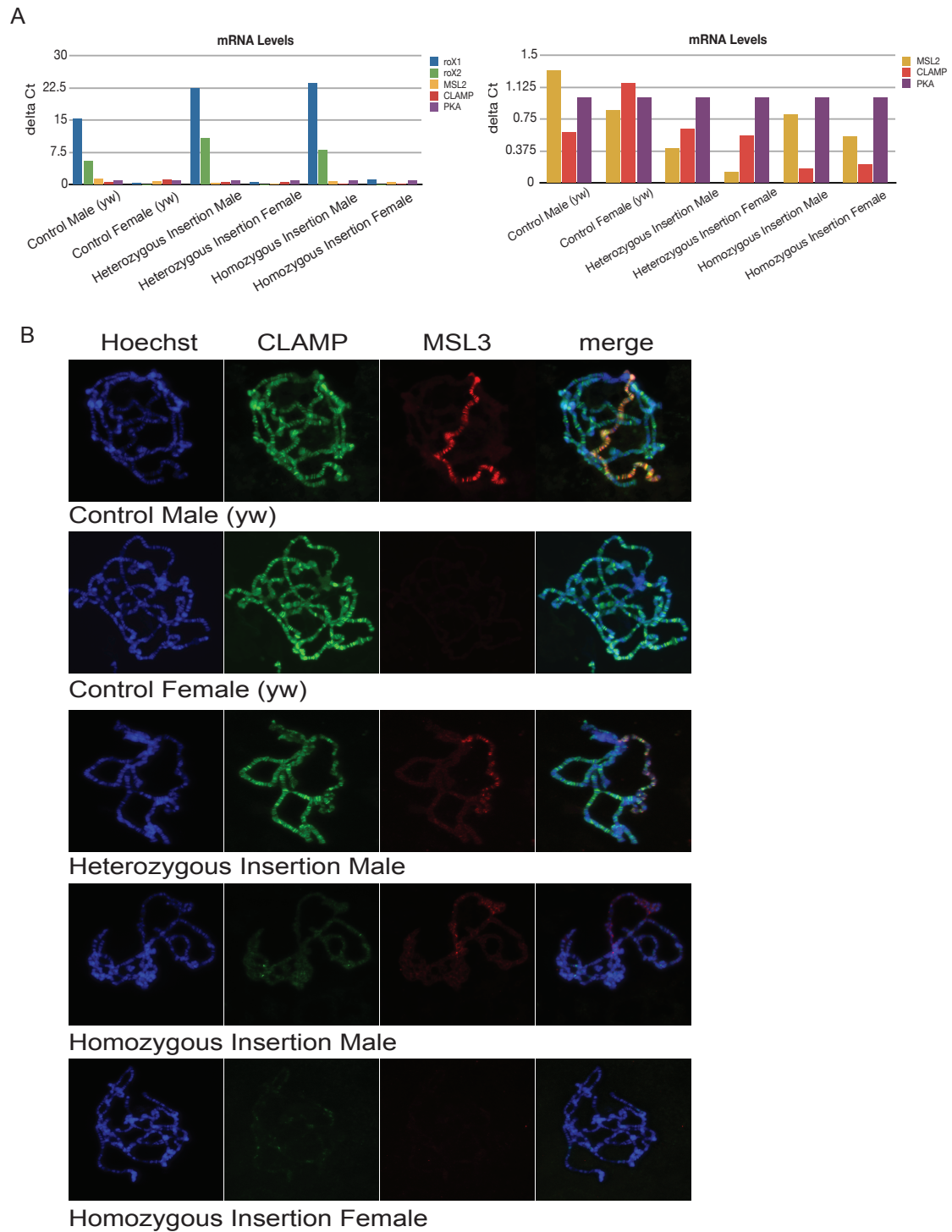


Figure 2: mRNA and protein levels are reduced in the homozygous insertion line.

A) mRNA analysis of cDNA from 3rd instar larvae raised at 18 degrees show a

70% and 80% reduction in CLAMP mRNA in homozygous insertion males and females, respectively, compared to control larvae, *yw* (right). Heterozygous and homozygous insertion males show a modest increase in *roX* mRNA levels (left)

B) Immunostaining of polytene chromosomes from male 3rd instar larvae raised at 18 degrees for heterozygous males show near wild type CLAMP levels. Homozygous male and female larvae show a dramatic reduction of CLAMP protein levels and localization. Additionally, there is a marked reduction of MSL3 localization on the X chromosome in heterozygous males and a near complete loss of MSL3 in homozygous males.

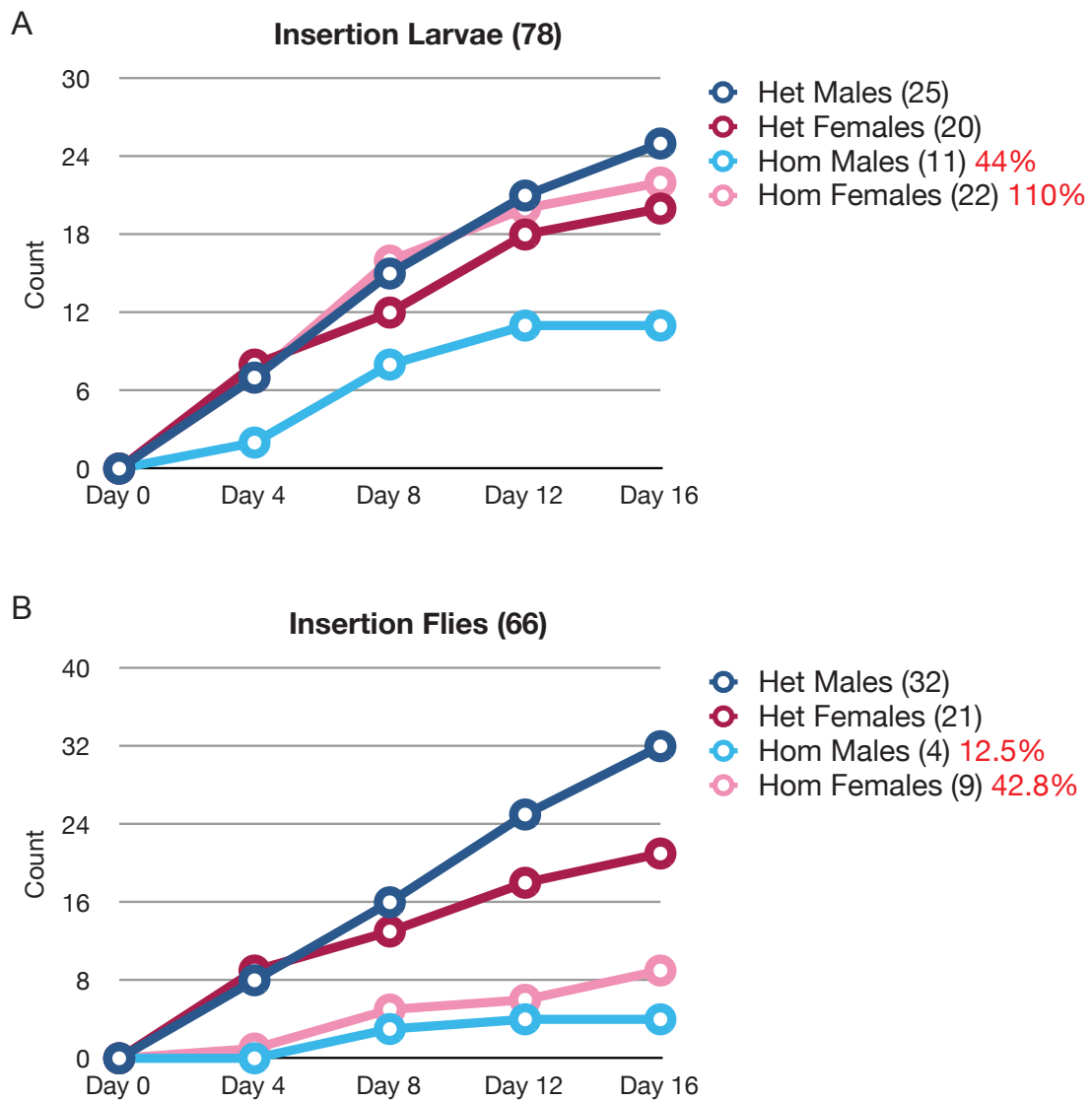
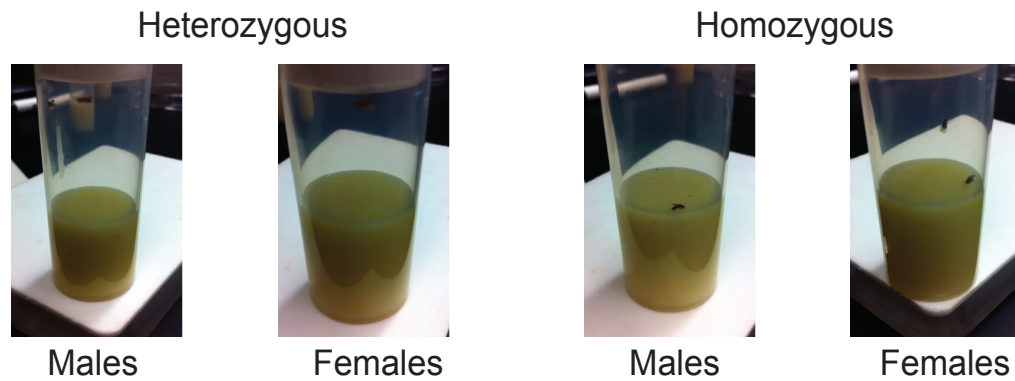


Figure 3: Reduced survival of homozygous insertion flies.

A) Heterozygous and homozygous larvae counts over a 16 day time course reveals a slight reduction in homozygous male larvae (44%) when compared to the expected Mendelian ratio of 50% for homozygous insertion larvae maintained over a balancer chromosome. Homozygous female larvae are higher than expected (110%). B) Adult fly counts over a 16 day time course shows a dramatic reduction in homozygous male flies (12.5%) surviving while

homozygous females shows a slight reduction (42.8%) compared to the expected Mendelian ratio of 50%.



Link to movies: <https://www.dropbox.com/sh/8oqppast5imuyp8/l6gJ0x3AkR>

Figure 4: Severe motor defects observed in surviving homozygous flies.

Homozygous males and females show an obvious defect in movement and flying with males being most severely affected.

Link to movies <https://www.dropbox.com/sh/8oqppast5imuyp8/l6gJ0x3AkR>



Figure 5. CLAMP RNAi reveals both positive and negative regulation genome-wide

A) mRNA-seq analysis before and after CLAMP RNAi revealed 422 genes in male cells and 329 genes in females cells with greater than 2-fold changes in expression further filtered for the top 50% of expressed genes. 323 genes are positively (decreased after RNAi) and 428 are negatively (increased after RNAi) regulated by CLAMP in both cell types. Genes from both these groups are found on all chromosomes in both male and female cells. B) Comparison of the difference in regulated genes on each chromosome to a random assignment of genes genome-wide reveals a marked difference for X and 2L genes in males cells that is not seen in female cells. Additionally, there is a marked difference for 2R and 3L genes in both males and females compared to a random assignment.

Positively Regulated by CLAMP
in Male and Female Cells (24 genes)

Gene	Symbol	Name	Function
CG42457	CG42457		olfactory behavior; peripheral nervous system development; dormancy process; regulation of hemocyte proliferation; cellular process; synaptic transmission.
CG12369	Lac	Lachesin	cell adhesion involved in heart morphogenesis; axonogenesis; homophilic cell adhesion; morphogenesis of embryonic epithelium; regulation of tube size, open tracheal system; maintenance of epithelial integrity, open tracheal system; open tracheal system development; septate junction assembly.
CG3318	Dat	Dopamine N acetyltransferase	catecholamine metabolic process; sleep; developmental pigmentation; dopamine catabolic process; regulation of circadian sleep/wake cycle, sleep; serotonin catabolic process; octopamine catabolic process.
CG3131	Duox	Dual oxidase	reactive oxygen species metabolic process; response to oxidative stress; oxidation-reduction process; wing disc development; mucosal immune response; negative regulation of apoptotic process; adult chitin-based cuticle development; adult chitin-containing cuticle pigmentation; defense response to bacterium.
CG18557	CG18557		proteolysis.
CG10383	CG10383		determination of adult lifespan; GPI anchor metabolic process; intracellular protein transport.
CG1832	CG1832	CLAMP	unknown
CG12142	Tsp42Eg	Tetraspanin 42Eg	unknown
CG12133	CG12133		proteolysis.
CG18371	CG18371		unknown
CG14688	CG14688		fatty acid alpha-oxidation.
CG4115	CG4115		unknown
CG15695	CG15695		metabolic process
CG1976	RhoGAP100F	RhoGAP100F	Rho protein signal transduction; cytoskeletal matrix organization at active zone; axon extension involved in axon guidance; maintenance of presynaptic active zone structure; R7 cell development; signal transduction.
CG31127	Wskc	Wskc	protein phosphorylation
CG34054	CG34054		unknown
CR33662	snoRNA:U3.9B	snoRNA:U3.9B	rRNA modification guide activity
CR34620	snoRNA:Psi28S-3405b	snoRNA:Psi28S-3405b	unknown
CR34597	snoRNA:Psi28S-1192a	snoRNA:Psi28S-1192a	unknown
CR33753	snoRNA:Psi28S-3327c	snoRNA:Psi28S-3327c	unknown
CR33652	snoRNA:Psi28S-1837a	snoRNA:Psi28S-1837a	unknown
CR42652	7SLRNA:CR42652	Signal recognition particle 7SL RNA CR42652	unknown
CG42872	MtnE	Metallothionein E	response to mercury ion; detoxification of cadmium ion; response to silver ion; response to copper ion; response to zinc ion
CG43738	cpo	couch potato	olfactory behavior; peripheral nervous system development; dormancy process; regulation of hemocyte proliferation; cellular process; synaptic transmission.

Negatively Regulated CLAMP
in Male and Female Cells (19 genes)

Gene	Symbol	Name	Function
CG1759	cad	caudal	single-organism developmental process; biological regulation; genital disc development; multicellular organismal development; regionalization; system development; digestive tract development; positive regulation of cellular biosynthetic process; positive regulation of cell proliferation; segment specification; anterior/posterior axis specification.
CG12763	Dpt	Dipterin	innate immune response; response to bacterium; defense response to Gram-negative bacterium; positive regulation of innate immune response; antibacterial humoral response; defense response to bacterium.
CG1462	Aph-4	Alkaline phosphatase 4	nervous system development; multicellular organism reproduction; metabolic process; epithelial fluid transport.
CG32540	CCKLR-17D3	CCK-like receptor at 17D3	adult locomotory behavior; G-protein coupled receptor signaling pathway.
CG6870	CG6870		unknown
CG10794	DptB	Dipterin B	defense response to Gram-positive bacterium; antibacterial humoral response; humoral immune response.
CG4476	CG4476		visual behavior; neurotransmitter transport.
CG3288	CG3288		unknown
CG30324	CG30324		unknown
CG31272	CG31272		transport; transmembrane transport; lipid metabolic process.
CR34561	snoRNA:Or-aca1	snoRNA:Or-aca1	unknown
CR34643	snoRNA:Psi28S-3327b	snoRNA:Psi28S-3327b	unknown
CR34622	snoRNA:Psi28S-3305b	snoRNA:Psi28S-3305b	unknown
CR34660	snoRNA:Me28S-U1230		unknown
CR33661	snoRNA:Psi28S-1180	snoRNA:Psi28S-1180	unknown
CR33709	snoRNA:Psi28S-3186	snoRNA:Psi28S-3186	unknown
CG42603	Slp26Ad	Seminal fluid protein 26Ad	multicellular organism reproduction
CG43065	bru-2	bruno-2	negative regulation of translation.
	dro5	-	-

Figure 6. CLAMP regulates genes involved in brain development and morphogenesis.

Overlap of genes in both cells types reveals CLAMP both positively and negatively regulates genes involved in brain development and morphogenesis (red). CLAMP positively regulates genes involved in olfactory behavior and circadian rhythms (blue) (left). CLAMP negatively regulates genes involved in motor and visual behavior and reproduction (purple) (right). CLAMP gene identified in positively regulated group due to reduction of mRNA from RNAi treatment (green).

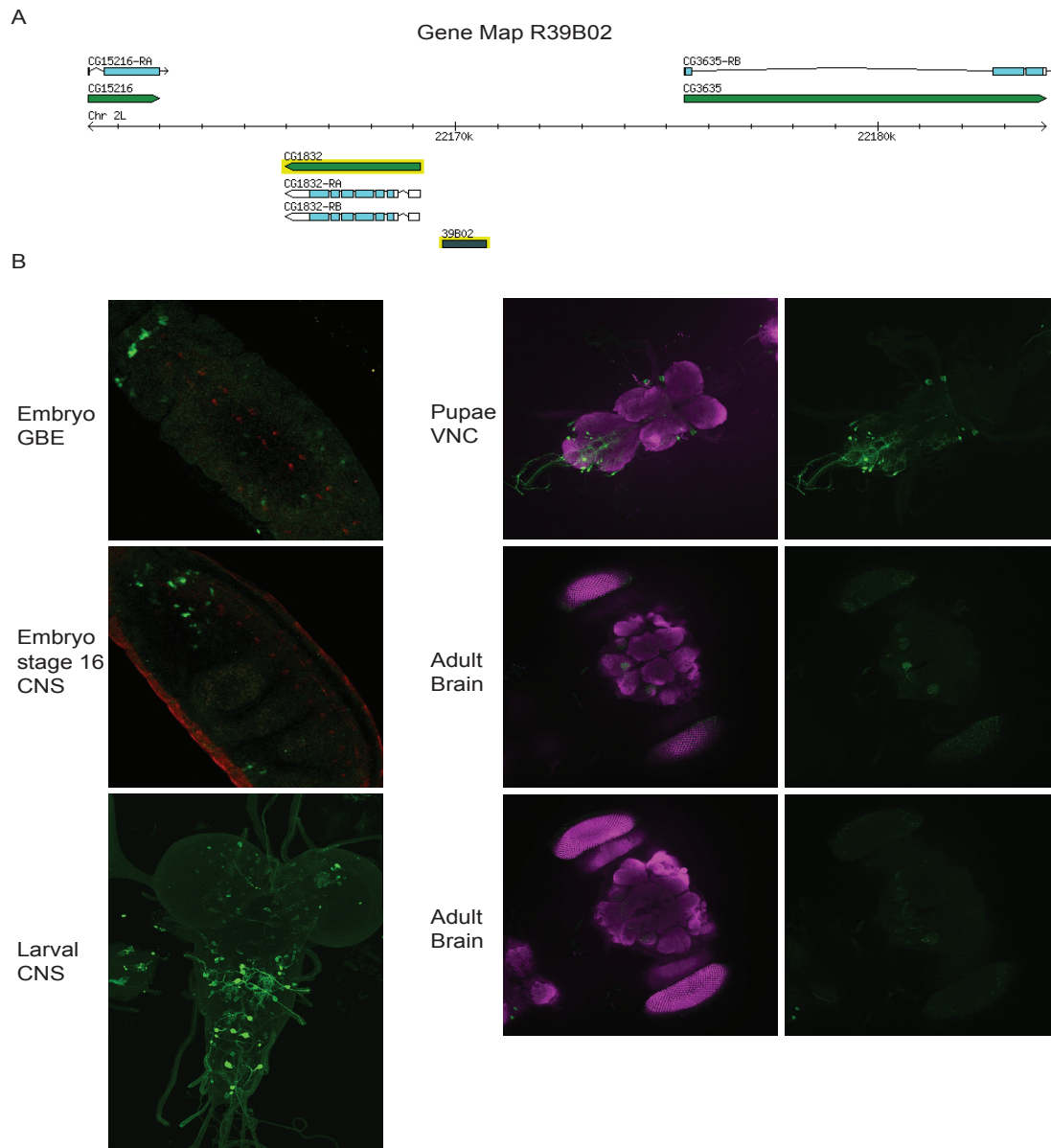


Figure 7: Putative CLAMP enhancer driving GFP expression in the fly brain.

A) Gene map of putative enhancer region R39B02 (grey with yellow highlight) located upstream relative to the CLAMP locus (green with yellow highlight) B) GFP expression (green) of R39B02 region in developing fly brain from the Fly Light group at HHMI Janelia Farm Research Campus reveals neuronal localization.

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, 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.
- Belote JM, Lucchesi JC. 1980. Male-specific lethal mutations of *Drosophila melanogaster*. *Genetics* **96**: 165-186.
- Briggs AW, Rios X, Chari R, Yang L, Zhang F, Mali P, Church GM. 2012. Iterative capped assembly: rapid and scalable synthesis of repeat-module DNA such as TAL effectors from individual monomers. *Nucleic acids research* **40**: e117.
- Cermak T, Doyle EL, Christian M, Wang L, Zhang Y, Schmidt C, Baller JA, Somia NV, Bogdanove AJ, Voytas DF. 2011. Efficient design and assembly of custom TALEN and other TAL effector-based constructs for DNA targeting. *Nucleic acids research* **39**: e82.
- Deng X, Berletch JB, Ma W, Nguyen DK, Hiatt JB, Noble WS, Shendure J, Distech CM. 2013. Mammalian X upregulation is associated with

- enhanced transcription initiation, RNA half-life, and MOF-mediated H4K16 acetylation. *Developmental 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.
- Graveley BR, Brooks AN, Carlson JW, Duff MO, Landolin JM, Yang L, Artieri CG, van Baren MJ, Boley N, Booth BW et al. 2011. The developmental transcriptome of *Drosophila melanogaster*. *Nature* **471**: 473-479.
- Jonkers I, Monkhorst K, Rentmeester E, Grootegoed JA, Grosveld F, Gribnau J. 2008. Xist RNA is confined to the nuclear territory of the silenced X chromosome throughout the cell cycle. *Molecular and cellular biology* **28**: 5583-5594.
- Kelley RL, Solovyeva I, Lyman LM, Richman R, Solovyev V, Kuroda MI. 1995. Expression of msl-2 causes assembly of dosage compensation regulators on the X chromosomes and female lethality in *Drosophila*. *Cell* **81**: 867-877.
- 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.

- Lam KC, Muhlpfordt F, Vaquerizas JM, Raja SJ, Holz H, Luscombe NM, Manke T, Akhtar A. 2012. The NSL complex regulates housekeeping genes in *Drosophila*. *PLoS genetics* **8**: e1002736.
- 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.
- Lewis PW, Beall EL, Fleischer TC, Georlette D, Link AJ, Botchan MR. 2004. Identification of a *Drosophila* Myb-E2F2/RBF transcriptional repressor complex. *Genes & development* **18**: 2929-2940.
- Livernois AM, Graves JA, Waters PD. 2012. The origin and evolution of vertebrate sex chromosomes and dosage compensation. *Heredity* **108**: 50-58.
- Meyer BJ. 2010. Targeting X chromosomes for repression. *Current opinion in genetics & development* **20**: 179-189.
- Milyaev N, Osumi-Sutherland D, Reeve S, Burton N, Baldock RA, Armstrong JD. 2012. The Virtual Fly Brain browser and query interface. *Bioinformatics* **28**: 411-415.
- Pfeiffer BD, Jenett A, Hammonds AS, Ngo TT, Misra S, Murphy C, Scully A, Carlson JW, Wan KH, Lavery TR et al. 2008. Tools for neuroanatomy and neurogenetics in *Drosophila*. *Proceedings of the National Academy of Sciences of the United States of America* **105**: 9715-9720.

- Plath K, Fang J, Mlynarczyk-Evans SK, Cao R, Worringer KA, Wang H, de la Cruz CC, Otte AP, Panning B, Zhang Y. 2003. Role of histone H3 lysine 27 methylation in X inactivation. *Science* **300**: 131-135.
- Sim CK, Perry S, Tharadra SK, Lipsick JS, Ray A. 2012. Epigenetic regulation of olfactory receptor gene expression by the Myb-MuvB/dREAM complex. *Genes & development* **26**: 2483-2498.
- 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* **504**: 465-469.
- 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.
- Venken KJ, Carlson JW, Schulze KL, Pan H, He Y, Spokony R, Wan KH, Koriabine M, de Jong PJ, White KP et al. 2009. Versatile P[acman] BAC libraries for transgenesis studies in Drosophila melanogaster. *Nature methods* **6**: 431-434.
- 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.

Wood AJ, Severson AF, Meyer BJ. 2010. Condensin and cohesin complexity: the expanding repertoire of functions. *Nature reviews Genetics* **11**: 391-404.

Chapter 5: Discussion

Co-evolution of the sex chromosomes and dosage compensation components

Marcela M.L. Soruco¹, Jason Dobson¹, Nadia Singh² and Erica Larschan¹

¹Department of Molecular Biology, Cellular Biology and Biochemistry, Brown University, Providence, RI 02912, USA

²Genetics Department, North Carolina State University, Raleigh, NC 27695

This manuscript is in preparation for submission. I wrote the text and Erica Larschan provided comments and edits.

Discussion

Sex chromosomes are thought to have evolved slowly as a result of translocations between autosomes in regions surrounding genes involved in sex determination (Charlesworth 1978). This process evolved independently in many lineages, creating a wide variety of sex chromosome (Livernois et al. 2012). In the XY system, this created heterogametic sex chromosomes where the X contained many essential and non-sex specific genes and the Y became degenerate through lack of recombination and contained very few genes, most of which are specific to spermatogenesis (Charlesworth and Charlesworth 2000; Charlesworth et al. 2005).

In both mammals and *Drosophila*, females are XX and males are XY. In 1967, Susumu Ohno first proposed a mechanism to correct for this gene dose imbalance stating that all genes on the X chromosome would need to be increased in expression two-fold to compensate for expression from the degenerative Y homologue (Ohno 1967). In the absence of such a mechanism, expression of the X-chromosome would be insufficient, likely causing lethality. The mechanisms to correct for the imbalance of gene dose due to heterogametic sex chromosomes are collectively called dosage compensation.

Through the effort of many research teams, the increase in X-linked transcription has proven to be necessary in many species (Gupta et al. 2006;

Nguyen and Distèche 2006; Lin et al. 2007). In some species, a secondary compensatory mechanism also became necessary to equalize expression of the X-chromosome to autosomes (Deng et al. 2011; Kharchenko et al. 2011; Lin et al. 2011). For example, in mammals and *C. elegans*, the X-chromosome is up-regulated in both sexes, then as an additional step one X-chromosome is randomly inactivated in mammalian females and both X-chromosomes are down-regulated by half in *C. elegans* hermaphroditic females (Jonkers et al. 2008; Meyer 2010). In *Drosophila*, the X-chromosome is only up-regulated in a single step mechanism for dosage compensation and only in males.

As the different sex chromosomes and dosage compensation mechanisms evolved independently in varying lineages, the protein complexes responsible for compensation also evolved differently across species. For example, in mammals, the Polycomb repressive complex (PRC) is responsible for X-chromosome silencing in females (Plath et al. 2003; Simon et al. 2013). In *C. elegans*, the dosage compensation complex (DCC), which shares components with the condensin complex, down-regulates X-expression in hermaphrodites (Meyer 2010; Wood et al. 2010). In *Drosophila*, the male specific lethal (MSL) complex, which shares a component with the non-specific lethal (NSL) complex, up-regulates the X-chromosome only in males (Belote and Lucchesi 1980; Kelley et al. 1997; Lam et al. 2012). One possible explanation for how these mechanisms and complexes evolved independently among different species is that evolution co-opted functions of existing protein to also function in

dosage compensation. One example would be the MOF protein, which deposits H4K16ac on the X-chromosome and is a shared feature between *Drosophila* and mammals (Gelbart et al. 2009; Deng et al. 2013).

Neo-functionalization of existing proteins that interact with transcriptional machinery would have been under selective pressure to be co-opted to work in dosage compensation. For example, certain DNA binding proteins would have been prime candidates for dosage compensation cofactors if the sequence motifs they target accumulated on the X-chromosome and these co-factors act in transcription regulation. Our work has revealed that is the case in *Drosophila* and the MSL complex (Larschan et al. 2012; Soruco et al. 2013).

We recently discovered a previously unstudied, non-sex specific protein, CLAMP, which is involved in dosage compensation in *Drosophila* (See Chapter 2) (Larschan et al. 2012). CLAMP targets GA-rich MSL recognition elements (MREs) that are enriched on the X-chromosome (Alekseyenko et al. 2008; Soruco et al. 2013). CLAMP is also required for MSL localization at MSL high affinity sites on the X, two of which are the *roX* loci that encode the long non-coding RNA components of the complex (Soruco et al. 2013). However, CLAMP is also found targeting GA-rich sequences throughout the genome and is therefore not an X-specific factor (See Chapter 3) (Soruco et al. 2013).

It was recently reported that the X-chromosome in *Drosophila* has accumulated GA-repeats when comparing *D. miranda*, which has a newly evolved X-chromosome, and *D. melanogaster*, which has a much older X-chromosome (Jason Dobson, data not shown) (Alekseyenko et al. 2013). Interestingly, the CLAMP protein is in the top 15% of the most conserved proteins in the *Drosophila* lineage when comparing rates of non-synonymous to synonymous changes in protein coding genes across three species *melanogaster*, *yakuba* and *simulans* (Nadia Singh, data not shown). Since CLAMP targets GA-rich sequences, this provides a link between evolution and accumulation of GA-repeats on the X-chromosome, which are highly conserved, and dosage compensation. Additionally, if we look at conservation for protein components of the MSL complex, these proteins are evolving very rapidly (Nadia Sing data not shown) (Rodriguez et al. 2007). Therefore, there must be some element, which is preserving the interaction between the GA-rich MRE sequences and the MSL complex components. The CLAMP protein has a highly unstructured glutamine rich domain that is flexible and could potentially accommodate structural changes in the MSL proteins. However, there are still a number of questions that need to be answered. For example, does the complexity of sequences required for CLAMP and MSL targeting differ on a newly evolving X vs. a fully evolved X-chromosome? To answer this question, a more detailed analysis of CLAMP binding to GA-repeats in various *Drosophila* species is required.

It is not only important to look at evolution and accumulation of GA-repeats in the *Drosophila* lineage, but also where these sequences are accumulating in relation to genes. MSL high affinity sites are often found close to the 3' end of genes where MREs and CLAMP are also localized (Alekseyenko et al. 2008). CLAMP is also found at promoter regions of active genes on all chromosomes (See Chapter 3) (Soruco et al. 2013). This spatial separation of CLAMP localization could have evolved to separate the two specific functions of CLAMP as a non-sex specific transcription regulator and a dosage compensation cofactor. This could be tested further using a reporter gene assay where regions of the X-chromosome that CLAMP alone or MSL complex and CLAMP together are known to target are positioned to the 5' and 3' end of the reporter. Monitoring of transcriptional changes would identify which sequences act as more efficient recruiters and activators.

Characterizing sequence requirements for CLAMP and MSL targeting will also be important to further model spreading of the MSL complex from nucleation sites of the *roX* loci to high affinity sites along the X chromosome. Currently, it is thought that the *roX* loci are targeted first during incorporation of RNAs into the complex followed by a series of 150-300 high affinity sites (Kelley et al. 1999; Bai et al. 2004)(Alekseyenko et al. 2008). However, until now, visualizing these targeting events in real time has not been possible. Therefore, we are eagerly awaiting data from our MSL induction system in female cells to model this targeting in real time (See Appendix 1). We hope to be able to distinguish

between continuous and dis-continuous spreading. Additionally, we have new evidence that H3K36me3 is an important feature in identifying MSL high affinity sites in the absence of MSL complex (Jason Dobson, data not shown). Therefore, we would like to test this requirement in our MSL induction system.

In *Drosophila* and *C. elegans*, most active X-linked genes are compensated with lowly transcribed genes not requiring full compensation (Disteche 2012). In *Drosophila*, this may be due to X-linked genes evolving faster (Meisel et al. 2012). In mammals, there are still a number of genes that escape X-inactivation making further exceptions to the need for full compensation of all genes (Berletch et al. 2011). Interestingly in mammals, accumulation of GATA-repeats on the X-chromosome has been found enriched at genes that escape X-inactivation (McNeil et al. 2006). Therefore, accumulation of GA-repeats could work across species to recruit activating factors. However, if the GATA binding protein in mammals is a true homologue to the CLAMP protein is still not known.

Finally, understanding the role the newly identified CLAMP protein is playing as an essential transcription factor responsible for regulating developmental processes genome-wide will contribute to our understanding of the evolution of dosage compensation in *Drosophila* (See Chapter 4).

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, 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. *The EMBO journal* **23**: 2853-2861.
- Belote JM, Lucchesi JC. 1980. Male-specific lethal mutations of *Drosophila melanogaster*. *Genetics* **96**: 165-186.
- Berletch JB, Yang F, Xu J, Carrel L, Disteche CM. 2011. Genes that escape from X inactivation. *Human genetics* **130**: 237-245.
- Charlesworth B. 1978. Model for evolution of Y chromosomes and dosage compensation. *Proceedings of the National Academy of Sciences of the United States of America* **75**: 5618-5622.
- Charlesworth B, Charlesworth D. 2000. The degeneration of Y chromosomes. *Philosophical transactions of the Royal Society of London Series B, Biological sciences* **355**: 1563-1572.

- Charlesworth D, Charlesworth B, Marais G. 2005. Steps in the evolution of heteromorphic sex chromosomes. *Heredity* **95**: 118-128.
- 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. *Developmental 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.
- Disteche CM. 2012. Dosage compensation of the sex chromosomes. *Annual review of genetics* **46**: 537-560.
- 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.
- Gupta V, Parisi M, Sturgill D, Nuttall R, Doctolero M, Dudko OK, Malley JD, Eastman PS, Oliver B. 2006. Global analysis of X-chromosome dosage compensation. *Journal of biology* **5**: 3.
- Jonkers I, Monkhorst K, Rentmeester E, Grootegoed JA, Grosveld F, Gribnau J. 2008. Xist RNA is confined to the nuclear territory of the silenced X chromosome throughout the cell cycle. *Molecular and cellular biology* **28**: 5583-5594.

- 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.
- Kharchenko PV, Xi R, Park PJ. 2011. Evidence for dosage compensation between the X chromosome and autosomes in mammals. *Nature genetics* **43**: 1167-1169; author reply 1171-1162.
- Lam KC, Muhlpfordt F, Vaquerizas JM, Raja SJ, Holz H, Luscombe NM, Manke T, Akhtar A. 2012. The NSL complex regulates housekeeping genes in Drosophila. *PLoS genetics* **8**: e1002736.
- 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.
- Lin H, Gupta V, Vermilyea MD, Falciani F, Lee JT, O'Neill LP, Turner BM. 2007. Dosage compensation in the mouse balances up-regulation and silencing of X-linked genes. *PLoS biology* **5**: e326.
- Lin H, Halsall JA, Antczak P, O'Neill LP, Falciani F, Turner BM. 2011. Relative overexpression of X-linked genes in mouse embryonic stem cells is consistent with Ohno's hypothesis. *Nature genetics* **43**: 1169-1170; author reply 1171-1162.

- Livernois AM, Graves JA, Waters PD. 2012. The origin and evolution of vertebrate sex chromosomes and dosage compensation. *Heredity* **108**: 50-58.
- McNeil JA, Smith KP, Hall LL, Lawrence JB. 2006. Word frequency analysis reveals enrichment of dinucleotide repeats on the human X chromosome and [GATA]_n in the X escape region. *Genome research* **16**: 477-484.
- Meisel RP, Malone JH, Clark AG. 2012. Faster-X evolution of gene expression in *Drosophila*. *PLoS genetics* **8**: e1003013.
- Meyer BJ. 2010. Targeting X chromosomes for repression. *Current opinion in genetics & development* **20**: 179-189.
- Nguyen DK, Disteche CM. 2006. Dosage compensation of the active X chromosome in mammals. *Nature genetics* **38**: 47-53.
- Ohno S. 1967. *Sex chromosomes and sex-linked genes*. Springer-Verlag, Berlin, New York, ect.
- Plath K, Fang J, Mlynarczyk-Evans SK, Cao R, Worringer KA, Wang H, de la Cruz CC, Otte AP, Panning B, Zhang Y. 2003. Role of histone H3 lysine 27 methylation in X inactivation. *Science* **300**: 131-135.
- Rodriguez MA, Vermaak D, Bayes JJ, Malik HS. 2007. Species-specific positive selection of the male-specific lethal complex that participates in dosage compensation in *Drosophila*. *Proceedings of the National Academy of Sciences of the United States of America* **104**: 15412-15417.
- 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* **504**: 465-469.

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.

Wood AJ, Severson AF, Meyer BJ. 2010. Condensin and cohesin complexity: the expanding repertoire of functions. *Nature reviews Genetics* **11**: 391-404.

Appendix 1

Induction of MSL Complex in female cells as a tool to identify initial sites of X-targeting during *Drosophila* dosage compensation

Marcela M.L Soruco¹, Jessica Feng¹, Erica Larschan¹

¹Brown University, Providence RI

This manuscript is in preparation for submission. I performed all the experiments except for the protein western blot in Fig. 3C which was done by Erica Larschan. Jessica Feng prepared and maintained the cell cultures for the mRNA and ChIP analysis. I prepared all the Figures and wrote the text. Erica Larschan edited and provided comments.

Introduction

In *Drosophila*, dosage compensation is necessary to equalize transcripts of X-linked genes between males (XX) and females (XY). The Male Specific Lethal (MSL) Complex is expressed only in males and specifically identifies the X-chromosome from the autosomes for up-regulation. Targeting of the MSL complex to the X-chromosome is known to take place in multiple steps. First, it is thought that the *roX* loci are targeted during cotranscriptional incorporation of these essential RNA components into the complex (Kelley et al. 1999; Bai et al. 2004). Next, a subset of high affinity sites, also known as Chromatin Entry Sites (CES), are targeted in a sequence-dependent step (Alekseyenko et al. 2008). These high affinity sites contain 21 bp GA-rich *cis-acting* sequences termed MSL recognition elements (MREs) (Lyman et al. 1997; Gu et al. 1998; Alekseyenko et al. 2008). Finally, the MSL complex spreads in a sequence independent step from these sites to the bodies of active genes through the recognition of the H3K36me3 mark associated with active transcription. The MSL complex globally deposits H4K16ac on the X-chromosome through activity of the maleless absent on first (MOF) component (Gelbart et al. 2009; Gorchakov et al. 2009). H4K16ac modification enables increased levels of RNA Pol II over the bodies of active genes resulting in an approximately two-fold increase in X-linked gene transcription (Larschan et al. 2011).

Identification of the MRE sequences involved in dosage compensation was a critical step in identifying elements, which work to initiate the early stages of X-targeting. However, the MRE sequences are not X-specific. They are two fold enriched on the X overall and 4-fold enriched over the bodies of active genes on the X, but MREs are also found on autosomes (Alekseyenko et al. 2008). Additionally, when the *roX* loci are artificially placed at an autosomal location, the RNAs and complex move *in trans* to target the X (Park et al. 2002; Oh et al. 2003), oh 2003). Therefore, defining the chromatin context in which the MRE sequences are located on the X-chromosome is essential in understanding how MSL complex is able to accurately target the X-chromosome.

Recently, we identified a novel zinc finger protein, Chromatin Linked Adapter for MSL Proteins (CLAMP), which is essential in targeting MSL complex to high affinity sites (Larschan et al. 2012). Importantly, CLAMP is able to directly target MRE sequences. However, CLAMP is also found targeting MREs throughout the genome and is therefore not an X-specific factor. Our earlier work was able to detect three classes of high affinity sites within 256 total sites when comparing levels of CLAMP enrichment in female cells in the absence of MSL complex to male cells where MSL complex is expressed. In this study we identified approximately 100 sites in group A, 100 sites in group B sites and 50 sites in group. In the group A or 'MSL dependent' sites, there was very low levels of CLAMP in the absence of MSL complex in female cells and showed a marked increase in CLAMP enrichment the presence of MSL complex in male cells. In

the group B or 'semi MSL dependent' sites there was modest enrichment of CLAMP in the absence of MSL complex and then showed a modest increase of CLAMP enrichment in the presence of MSL complex. In the group C or 'MSL independent' sites, there was very high levels of CLAMP enrichment in the presence and absence of MSL complex in both male and female cells (Soruco et al. 2013). From this data, we hypothesized that variations in CLAMP enrichment at these classes of sites could potentially identify a subset of sites for targeting by MSL complex very early in development before MSL complex is expressed.

Here we use RNAi treatment against the master sex regulator, Sex Lethal (Sxl), to induce MSL complex formation in female cells (Alekseyenko et al. 2012). Using this same system, we performed a 6-day time course experiment in which we hoped to capture the early stages of X-identification, which until now has not been captured due to the narrow window of time between fertilization and MSL complex expression in late cell cycle 14 (Lott et al. 2011). Performing chromatin immunoprecipitation (ChIP) for MSL2, H4K16 we hope to distinguish between a continuous or discontinuous targeting model. In a continuous model, the MSL complex would scan the length of X-chromosome and target adjacent sites. In a discontinuous model, a sub-nuclear domain would facilitate targeting of sites (Mendjan et al. 2006; Vaquerizas et al. 2010). Additionally, performing ChIP for CLAMP, we hope to observe if enrichment of the newly identified CLAMP protein plays a role in identifying these sites as "seed sites" for MSL recruitment early in the X-identification process.

Results

Sex Lethal RNAi stabilizes MSL Complex and increases H4K16ac

The MSL complex is comprised of a minimum of five proteins and one of two redundant long non-coding RNAs (Belote and Lucchesi 1980; Hilfiker et al. 1997; Franke and Baker 1999; Deng and Meller 2006; Li et al. 2008; Ilik et al. 2013). In females, one of two essential core structural proteins, MSL2, is translationally repressed by a female specific isoform of the master sex regulator protein, Sex Lethal (Sxl) (Fig. 1A) (Kelley et al. 1997). Binding of Sxl to the 5' and 3' UTR of MSL2 RNA in females prevents MSL2 protein expression (Kelley et al. 1997; Smith et al. 2000; Smith et al. 2001; Beckmann et al. 2005). Removing repression of MSL2 in female cells by performing Sxl RNAi induces MSL2 protein expression, stabilizes MSL complex and leads to MOF driven H4K16ac deposition on the X-chromosome (Alekseyenko et al. 2012). This global increase in X-specific acetylation results in an increase in X-linked gene expression (Alekseyenko et al. 2012).

To test this new system, we initially performed an end point experiment in which we assayed MSL2 and H4K16ac by ChIP-qPCR six days after Sxl RNAi in female cells, similar to Alekseyenko, Ho et al 2012. First, to assess MSL complex induction, we measured RNA levels for MSL2 and the *roX* genes by performing

real-time qPCR. These were normalized to PKA, which is ubiquitously expressed. Testing *roX* expression is an accurate reflection of functioning complex because MSL complex is a strong inducer of *roX* transcription in male cells. Using this approach, we observed a stabilization of MSL2 RNA and induction of *roX* RNA levels as expected if functional MSL complex is present (Fig. 2A).

Second, we performed chromatin immunoprecipitation followed by qPCR (ChIP-qPCR) for MSL2 and H4K16ac at seven loci on the X-chromosome, which include five high affinity sites, one random MRE site where MSL does not have a strong peak and one control locus of a untranscribed gene (CG15570), to assess localization and function of MSL complex. Importantly, MSL2 is enriched at all of the five high affinity sites, with as much as a 5-fold increase in enrichment at the 11D1 site (Fig. 2B). As expected we see a negligible enrichment of MSL2 at the random MRE and no enrichment at the untranscribed genes (Fig. 2B). This localization correlates with the induction of functional MSL complex because we also observe a specific increase in H4K16ac at the five high affinity sites that does not occur at the two control loci (Fig. 2C). In the future, we could assay expression levels of nearby genes for increases due to MSL complex activity.

CLAMP is enriched at high affinity sites after MSL complex induction

Next, we performed CLAMP ChIP-qPCR at the same seven loci to assay for changes in CLAMP levels. Here we also see a 2-fold to 3-fold increase in

CLAMP enrichment to the five high affinity sites after Sxl RNAi, which is absent at the two control loci. These data support a synergistic interaction between MSL complex and the CLAMP protein where co-localization increases the enrichment of both proteins. We hypothesize that increased CLAMP occupancy in the presence of MSL complex could be due to increased chromatin accessibility from H4K16ac deposition from functional MSL complex (Gelbart et al. 2009). In the future, we will test this model on acetylated nucleosome arrays. These data are consistent with our previous data that revealed a CLAMP-MSL synergistic interaction in both flies and cell lines (Soruco et al. 2013) (Fig. 3).

Sex Lethal RNAi time course shows early induction of MSL2 RNA

Using this same Sxl RNAi system, we performed a six-day time course experiment with a corresponding control RNAi (GFP) at each day, in an effort to capture the initial sites of MSL complex targeting to high affinity sites. We first assayed mRNA levels of the *roX* loci, MSL2, Sxl, and CLAMP where we normalized expression relative to PKA. Here we see a 25% reduction of Sxl RNA over the six-day experiment (Fig. 4A). This corresponds to a modest increase in MSL2 RNA as early as day 1 with the highest increase at day 5 (Fig. 4B). Additionally, there is a modest induction in *roX* RNA levels as early as day 2 (Fig. 4B). Finally, stabilization of MSL2 protein levels as early as day 2 is detected which is slightly delayed compared to increases in MSL2 mRNA levels as expected (Fig 4C). We are in the process of performing ChIP for MSL2,

H4K16ac and CLAMP to model this induction system in real time. We also plan on generating Illumina sequencing libraries from this experiment, which will provide high-resolution targeting data for targeting of these three proteins over the entire X-chromosome.

Discussion

Sex Lethal RNAi in female cells proves to be a valuable tool to observe MSL complex dynamically targeting the X-chromosome. In the near future, we will assess MSL localization by performing ChIP-qPCR for MSL2 and H4K16ac. In addition, we will perform ChIP-qPCR for CLAMP recruitment to see if dynamic re-localization of CLAMP takes place as MSL complex changes chromatin accessibility through deposition of H4K14ac.

Next generation sequencing of ChIP for MSL2, H4K16ac and CLAMP will provide a global view of MSL targeting the X-chromosome and chromatin remodeling through deposition of H4K16ac. Analyzing ChIP-seq data for the CLAMP protein will allow us to test the hypothesis of whether the varying levels CLAMP enrichment at the three classes of high affinity sites before MSL complex expression plays a role in identifying these sites as “seed sites” for early MSL identification.

Finally, we hope these data sets will allow us to answer some outstanding questions in the field. These being: Are the *roX* loci the first sites targeted? Does H3K36me3 levels affect MSL targeting? Does MSL complex target the X-chromosome in a continuous or discontinuous model?

Methods

Kc cell culture and RNAi treatment

dsRNA to target the Sex lethal (Sxl) gene (CG43770) was generated as described in detail at www.flyrnai.org. The DRSC 21490 and DRSC 28896 constructs were synthesized using the T7 megascript kit (Ambion) with additional RNA polymerase II 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). GFP dsRNA used as a control construct was 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 was conducted in T75 flasks as described at www.flyrnai.org. Cells were diluted to 3.0×10^6 cells/mL in Gibco Schneider's medium and 3 mL of cells were added to each flask with 67.5 µg of dsRNA in 1.5 mL of nuclease free water. Cells were maintained in 15 mL of Gibco Schneider's medium with 10% heat inactivated FBS (JRH). Six days after dsRNA treatment, mRNA preparation using the RNAeasy Plus kit (Qiagen) and chromatin preparation and were performed as described below.

mRNA sample preparation

Total RNA was extracted using the Qiagen RNeasy Plus Kit at room temperature. 1 mL of cells was centrifuged at 600 g for 2 minutes and the media was removed. The cells were disrupted by adding 350 μ l Buffer RLT with 2-mercaptoethanol (10 μ l/1ml). Lysate was transferred to a QIAshredder spin column in a 2 ml collection tube and centrifuge for 2 min at maximum speed. The homogenized lysate was transferred to a gDNA Eliminator spin column and centrifuge for 30 s at $\geq 8000 \times g$ ($\geq 10,000$ rpm). To the flow through, 1 volume (usually 350 μ l) of 70% ethanol was added and mixed by pipetting. The entire volume (700 μ l) was transfer to an RNeasy spin column and centrifuged for 15 s at $\geq 8000 \times g$ ($\geq 10,000$ rpm). The flow-through was discarded. 700 μ l Buffer RW1 was added to the RNeasy spin column and centrifuged for 15 s at $\geq 8000 \times g$ ($\geq 10,000$ rpm) to wash the spin column membrane. Discard the flow-through and reuse collection tube. The flow-through was discarded. 500 μ l Buffer RPE with ethanol was added to the RNeasy spin column and centrifuge for 15 s at $\geq 8000 \times g$ ($\geq 10,000$ rpm) to wash the spin column membrane. The flow-through was discarded. 500 μ l Buffer RPE with ethanol was added to the RNeasy spin column and centrifuge for 2 min at $\geq 8000 \times g$ ($\geq 10,000$ rpm) to wash the spin column membrane. The RNeasy spin column was placed onto a new 2 ml collection tube and centrifuged at full speed for 1 min to eliminate any residual ethanol. The RNeasy spin column was placed onto a new 1.5 ml collection tube. 30 μ l RNase-free water directly to the spin column membrane and centrifuged for 1 min at $\geq 8000 \times g$ ($\geq 10,000$ rpm) to elute the RNA. Using the eluate from previous step

was pipetted onto the RNeasy spin column and centrifuged for 1 min at $\geq 8000 \times g$ ($\geq 10,000$ rpm) to increase yield and concentration of RNA.

cDNA synthesis and Quantitative real time RT-PCR

Synthesis of cDNA was conducted as described in Alekseyenko et al. 2006 using 1 mg 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 StepOnePlus real time PCR machine. cDNA was diluted 1:20 prior to qPCR to ensure values were in linear range. Values were normalized to PKA, a highly expressed gene. Briefly, all primers were tested to assure that they have 100% efficiency and good linearity prior to use with an ABI StepOnePlus real time PCR machine. The average relative quantitation (ABI User Bulletin #2) of three technical replicate reactions was determined for all samples. Platinum SYBR master mix w/ROX (Invitrogen) was used. Normalization was conducted using primers within either PKA internal control genes. Primer sequences uses roX1, F-ATGCGAGCGAGACAATGATACT, R-GACTTGCAGTCCGCCCTATG, roX2, F-CGTTACTCTTGCTTGATTTTGC, R-AGCTCGGATGGCCATCGAAA, MSL2, F-GCCCTGTCCGTATGAATG, R-CAACATGGGTAAACAACC, Sxl, F-CGCCGAAATGGCCTCCTGGG, R-TGCGTGACAAGCTCACAGGTCTG, CLAMP, F-CTGACCACGGCAATAGAA, R-TCCCTCATTTGGCACTTC, PKA, F-CAATCAGCAGATTCTCCGGCT, R-AGCCGCACTCGCGCTTCTAC.

Chromatin Preparation

The following protocol is modified from Larschan et al, 2007

The cells were scraped from bottom of T75 flask and two 0.5 mL aliquots were removed for RNA extraction and protein gels. To crosslink the cells, 0.354 mL of 36.9% formaldehyde was added to 13 mL cells to a final concentration of 1%. The cells were then agitated on a shaker at 110 rpm for 10 min at room temp. To quench crosslinking, 1.662 mL of 2.5 M glycine was added to a final concentration of 125 mM. The cells were agitated at 110 rpm for 5 min at room temp. The cell were then transferred to a 15 mL Falcon tube on ice and centrifuged at 4°C at 1500 g for 5 min. The supernatant was removed and the pellet was suspended in 15mL PBS-EDTA followed by a centrifugation at 4°C 1500 g for 5 min. The supernatant was removed and the pellet was suspended in 10 mL ChIP Wash A buffer (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). Solution was rotated at 4°C for 10 min followed by a centrifugation at 4°C 1500 g for 5 min. The supernatant was removed and the pellet was suspended in 10 mL ChIP Wash B buffer (10 mM HEPES pH 7.6, 100 mM NaCl, 1 mM EDTA pH 8.0, 0.5 mM EGTA pH 8.0, 0.01% Triton X-100, 1x PI, 0.2 mM PMSF). Solution was rotated at 4°C for 5 min. To count cells, 100 ul of cells was removed and 900 ul trypan blue was added (1:10 dilution). Note all cells will be dead and stained blue. The cell

solution were continued to rotate at 4°C while the cell dilution was counted using a hemacytometer. The cells were centrifuged at 4°C 1500 g for 5 min and all supernate was removed. The cells were diluted to a target cell count of 2.5×10^8 cells/mL in lysis buffer (0.1% SDS, 10 mM EDTA pH 8.0, 50 mM Tris-HCL pH 8.0). The cells were lyse by rotating at 4°C for 10 min. 300 ul aliquots of the solution was transferred to Protein Low Bind Safelock tubes for Bioruptor (Diagenode) sonication at 3 rounds of 5 minutes each round. Note: Each round is programmed to alternate 30 seconds “on” then 30 seconds “off”. Total time for 3 rounds of 5 min each round is 15 min but total time for sonication is 7.5 min. Add ice after each round. Make sure the water level does not go above line on sonicator. Water line should match sample volume in eppendorf tubes. Sonicate in batches of 6 tubes. The tubes were pooled into a 15 mL falcon tube and spin at 4°C 13,000 rpm for 10 min. The supernatant was transferred to a Falcon tube and 9 volumes of ChIP Dilution Buffer was added. Chromatin preparation was filtered through a PolyPrep column (Biorad). 400 ul was removed for Input DNA into Protein Low Bind tubes and the remainder was frozen in liquid N₂. To reverse the crosslinks on the input samples, the following reagents were added: 1) 200 mL 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 mL 5M NaCl; 4) 1 mL 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 mL of RNase A followed by incubation at 37 °C for 30 minutes; 2) 20 mL 1M Tris-

HCl pH 6.8; 3) 10 mL 0.5M EDTA pH 8.0; and 4) 3 mL 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 mL 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 Lysis buffer 20% SDS buffer to a final concentration of 40 ug/ 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., submitted) and the anti-H4K16ac was obtained from Abcam. 60 mL 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 µl 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

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 the CG15570 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. Primers used: 5C2, F-ATCAATGTTTCGATGTAGA, R-CTTCCAACCTATCTCGCTCT, 4C14, F-ATCGACAGAGATGGCAACTGTCAC, R-GTTGTCAGCTCCATAAGCAATCCCA, 11D1, F-GTGGAAACGGACAGCGTAA, R-CACATCAGCGACAAGAGGC, roX1, F-ATGCGAGCGAGACAATGATACT, R-GACTTGCAGTCCGCCCTATG, roX2, F-ACGGTGCTGGCTTAGAGAGA, R-AGTTCTGGTCACCCTGGAAA, Random MRE on X, F-CCCCATTTACATTTGCCCC, R-AGGTTTCCTGTCCCCTCTCT, CG15570, F-TACTCCTGTCCACCTGCTTT, R-CATTGCCAGTTTCCGTTTCC

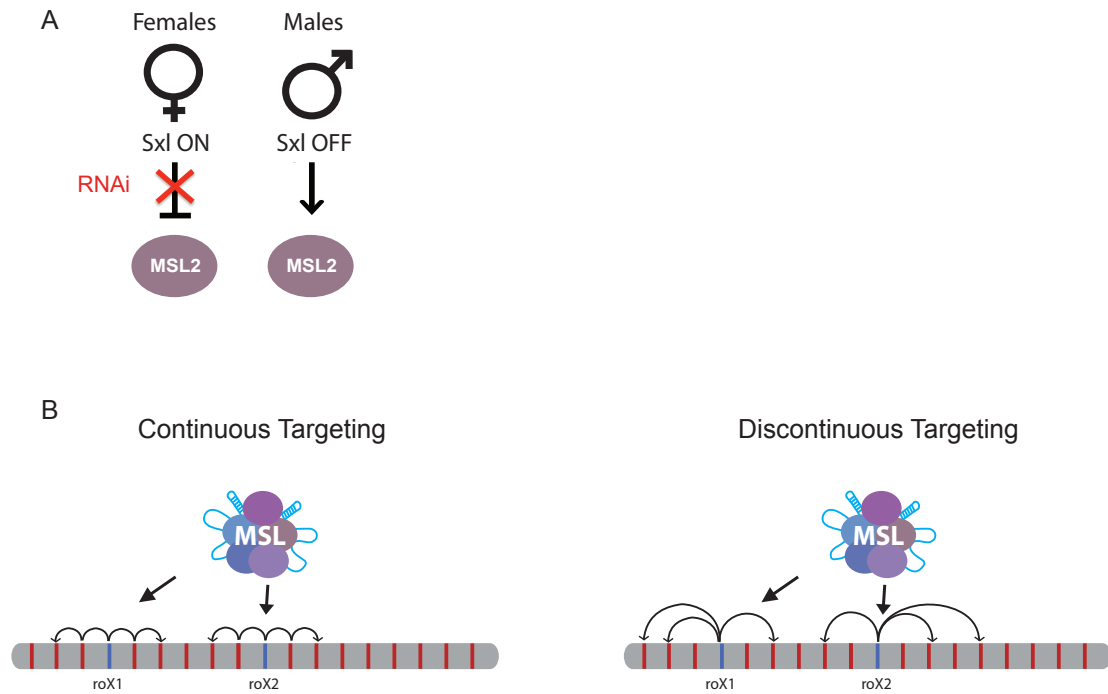


Fig. 1: Sex Lethal RNAi in females induces MSL2 expression with possible models of X-chromosome targeting

A) Repression of MSL2 takes place in females where Sex Lethal (Sxl) is expressed. B) Two possible models for MSL targeting the X-chromosome, continuous (left) and discontinuous (right).

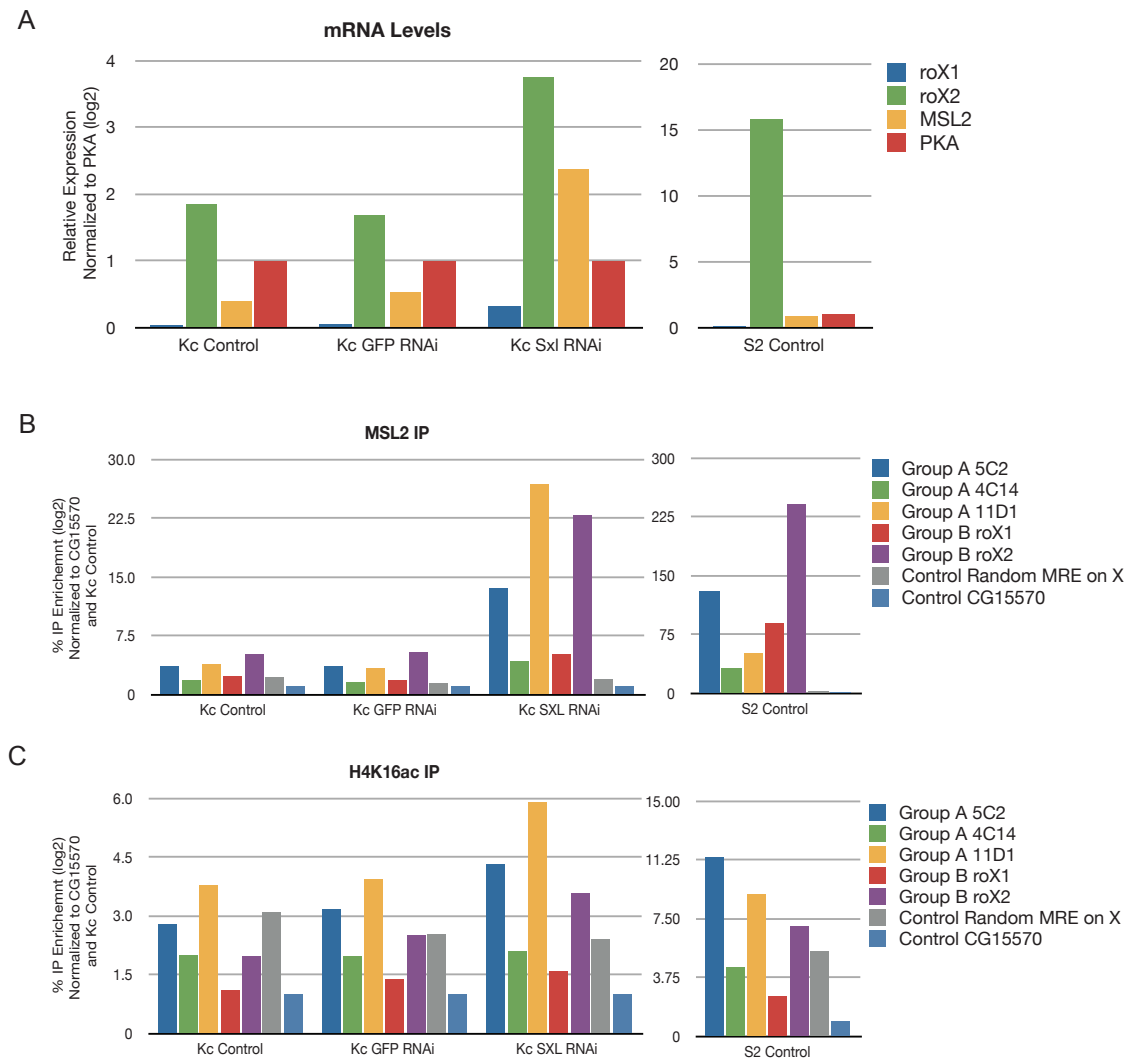


Fig. 2: Sex Lethal RNAi after 6 days increases MSL2 and *roX* RNA levels and increases functional MSL complex to high affinity sites.

A) *Sxl* RNAi increases MSL2 mRNA levels after 6-days. This leads to an increase in *roX* transcription. B) MSL2 is recruited is increase at five high affinity sites. C) MSL complex is functional at high affinity sites because there is a corresponding increase in H4K16ac deposited through MSL complex activity.

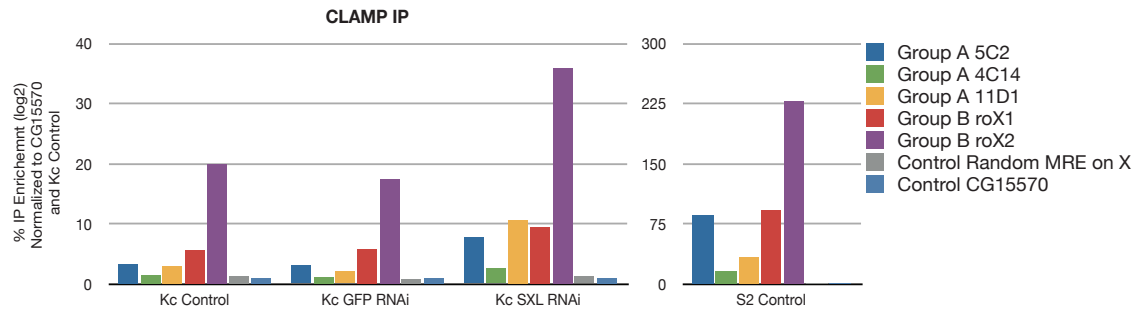


Fig. 3: CLAMP enrichment is increased after Sxl RNAi

CLAMP enrichment is increased after 6-days of Sxl RNAi likely due to decondensation of chromatin through H4K16ac activity of MSL complex. This supports a synergistic model where co-localization of MSL complex and CLAMP increases enrichment of both proteins.

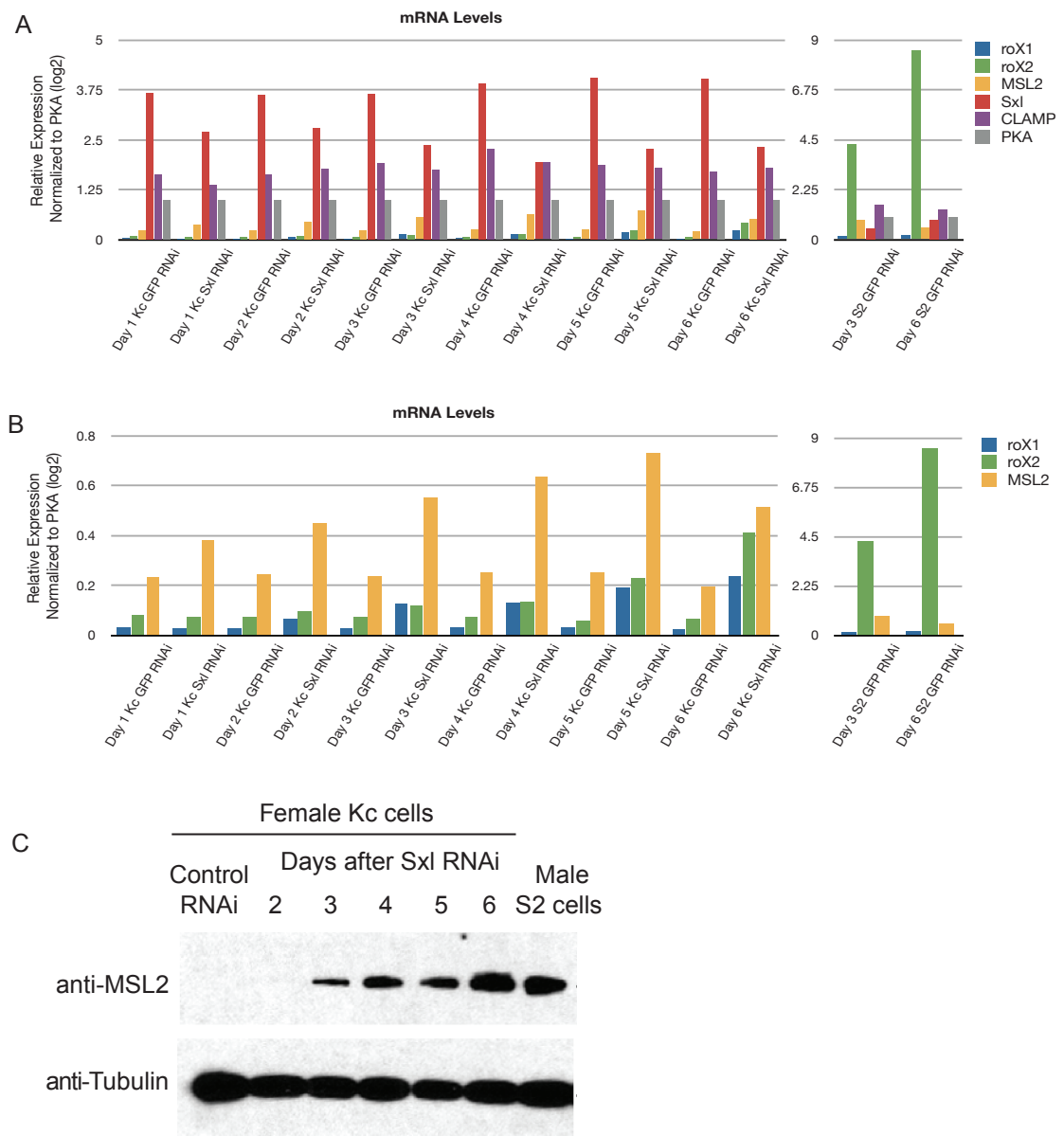


Fig. 4: Sex Lethal RNAi time courses over 6 days shows early induction of MSL complex

A) Sxl RNAi shows an average 25% reduction in Sxl RNA over six days. B) Sxl RNAi increases MSL2 mRNA as early as day 1, where *roX* transcription increases as early as day 2. C) MSL2 protein is induced as early as day three with day 6 being most similar to male cells.

References

- 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, 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. *The EMBO journal* **23**: 2853-2861.
- Beckmann K, Grskovic M, Gebauer F, Hentze MW. 2005. A dual inhibitory mechanism restricts msl-2 mRNA translation for dosage compensation in *Drosophila*. *Cell* **122**: 529-540.
- Belote JM, Lucchesi JC. 1980. Male-specific lethal mutations of *Drosophila melanogaster*. *Genetics* **96**: 165-186.
- Deng X, Meller VH. 2006. roX RNAs are required for increased expression of X-linked genes in *Drosophila melanogaster* males. *Genetics* **174**: 1859-1866.
- Franke A, Baker BS. 1999. The rox1 and rox2 RNAs are essential components of the compensasome, which mediates dosage compensation in *Drosophila*. *Molecular cell* **4**: 117-122.

- 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.
- Gorchakov AA, Alekseyenko AA, Kharchenko P, Park PJ, Kuroda MI. 2009. Long-range spreading of dosage compensation in Drosophila captures transcribed autosomal genes inserted on X. *Genes & development* **23**: 2266-2271.
- Gu W, Szauter P, Lucchesi JC. 1998. Targeting of MOF, a putative histone acetyltransferase, 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 acetyltransferase 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.
- 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, 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 genetics* **8**: e1002830.
- 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.
- Lott SE, Villalta JE, Schroth GP, Luo S, Tonkin LA, Eisen MB. 2011. Noncanonical compensation of zygotic X transcription in early *Drosophila melanogaster* development revealed through single-embryo RNA-seq. *PLoS biology* **9**: e1000590.
- 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.
- 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.
- 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.
- 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.
- Smith ER, Allis CD, Lucchesi JC. 2001. Linking global histone acetylation to the transcription enhancement of X-chromosomal genes in *Drosophila* males. *The Journal of biological chemistry* **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.
- Vaquerizas JM, Suyama R, Kind J, Miura K, Luscombe NM, Akhtar A. 2010. Nuclear pore proteins nup153 and megator define transcriptionally active regions in the *Drosophila* genome. *PLoS genetics* **6**: e1000846.